frontiers | Research Topics

The digitalization of neurology

Edited by
Daniel B. Hier, Michael Carrithers,
Jorge Mario Rodríguez-Fernández and Benjamin Kummer

Published in
Frontiers in Digital Health

ISSN 1664-8714
ISBN 978-2-8325-3768-8
DOI 10.3389/978-2-8325-3768-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

The digitalization of neurology

Topic editors

Daniel B. Hier — Missouri University of Science and Technology, United States
Michael Carrithers — University of Illinois, United States
Jorge Mario Rodríguez-Fernández — University of Texas Medical Branch at Galveston, United States
Benjamin Kummer — Icahn School of Medicine at Mount Sinai, United States

Citation

Hier, D. B., Carrithers, M., Rodríguez-Fernández, J. M., Kummer, B., eds. (2023). *The digitalization of neurology*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-3768-8

Table of contents

frontiers | Frontiers in Digital Health

TYPE Editorial
PUBLISHED 09 October 2023
DOI 10.3389/fdgth.2023.1291110

OPEN ACCESS

EDITED AND REVIEWED BY
Uwe Aickelin,
The University of Melbourne, Australia

*CORRESPONDENCE
Daniel B. Hier
dhier@uic.edu

RECEIVED 08 September 2023
ACCEPTED 27 September 2023
PUBLISHED 09 October 2023

CITATION
Hier DB, Carrithers MD, Rodríguez-Fernández JM and Kummer B (2023) Editorial: The digitalization of neurology.
Front. Digit. Health 5:1291110.
doi: 10.3389/fdgth.2023.1291110

Editorial: The digitalization of neurology

Daniel B. Hier[1,2]*, Michael D. Carrithers[1], Jorge M. Rodríguez-Fernández[3] and Benjamin Kummer[4]

[1]Department of Neurology and Rehabilitation, University of Illinois at Chicago, Chicago, IL, United States, [2]Kummer Institute for Artificial Intelligence and Autonomous Systems, Missouri University of Science and Technology, Rolla, MO, United States, [3]Department of Neurology, University of Texas Medical Branch, Galveston, TX, United States, [4]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

KEYWORDS

neurology, digitalization, precision medicine, precision neurology, teleneurology, computation

Editorial on the Research Topic
The digitalization of neurology

Introduction

Modern neurology originated in the nineteenth century with the writings of Charcot, Wernicke, Gowers, Hughlings Jackson, Cajal, Broca, and others (1). For much of the history of neurology, observations were qualitative, not quantitative. Case reports and small case series dominated disease descriptions. Neuroimaging was characterized by plain radiographs that were not suitable for computer analysis. EEG and EMG waveforms were analog and not digital. Medical records were on paper and were not suitable for computer analysis. Patients were examined in person. Before modern neuroimaging, a definitive diagnosis was often impossible without a biopsy or autopsy.

> *We define the digitalization of neurology as the transition to computable observations, treatments, diagnoses, and outcomes.*

Neurology digitalization began slowly 50 years ago and has accelerated in the last 20 years. Progress has occurred on multiple fronts, including *Precision Neurology, Big Data in Neurology, Computable Neurology*, and *Remote Neurology*.

Precision neurology

> *Precision neurology is the application of precision medicine principles to the field of neurology.*

Precision neurology involves tailoring neurological care and treatment to individual patients according to their unique genetic, molecular, and clinical profiles. In this Research Topic, *Differential DNA methylation associated with multiple sclerosis and*

disease modifying treatments in an underrepresented minority population by Bingen et al. found that MS is associated with differential DNA methylation in genes that regulate immune cell differentiation, host defense against gastrointestinal pathogens, and susceptibility to acute myeloid leukemia. They also found an epigenetic signature associated with dimethyl fumarate treatment in genes that regulate cytokine signaling, axon guidance, and adherens junctions.

In another example of precision neurology, *Designing evidence-based support aids for social media access for individuals with moderate-severe traumatic brain injury: A preliminary acceptability study* by Zhao et al. examines the feasibility of modifying access to social networks for people affected by brain injury.

Another precision neurology study by Howlett-Prieto et al. is *Subtypes of relapsing-remitting multiple sclerosis identified by network analysis*. They used network analysis of multiple sclerosis phenotypes to identify the first subtypes of relapsing and remitting multiple sclerosis that could differ in response to treatment or outcome.

Big data in neurology

> *Big data in neurology is the creation of large data sets that have both depth (many patients) and breadth (many variables that include data from imaging, proteomics, clinical observations, genomics, etc.)*

In a big data approach to epilepsy, *INTUITION: a data platform to integrate human epilepsy clinical care and support for discovery* by Maharathi et al. describes the merging of pathological, clinical, radiographic, pharmacological and electroencephalographic data captured in the surgical treatment of epilepsy.

Another article from this Research Topic *Parkinson's disease population-wide registries in the United States: Current and future opportunities* examines opportunities and barriers to the creation of large Parkinson's disease registries in the United States (Wu and Wilson). The California Parkinson's Disease Registry has already collected information on 93,928 unique Parkinson's disease patients.

Another article from this Research Topic is entitled *Workflow for health-related and brain data lifecycle* and it examines best practices for curating and maintaining data related to brain health (Brůha et al.).

Computable neurology

> *Computable neurology converts neurological observations into machine-readable codes that can be entered into machine learning and deep learning applications.*

Traditional observations in neurology have been qualitative rather than quantitative. Computation with qualitative observations has been difficult. Initially, electroencephalographic and electromyographic waveforms were analog. The digitalization of these waveforms has made them computable. Similarly, analog images on radiographic films have been made computable by digitalization.

The Hecker et al. article *Voice Analysis for Neurological Disorder Recognition–A Systematic Review and Perspective on Emerging Trends* illustrates how voice features can be digitally encoded to enhance the diagnosis and treatment of neurological disorders.

Medical records on paper have been converted to electronic health records. However, unstructured patient data in electronic health records must be normalized using ontologies and natural language processing methods to create computable concepts suitable for machine learning and deep learning applications. Several articles examine methods for extracting computable concepts from electronic health records, including methods for annotating neurological concepts. Azizi et al.'s *Enhanced neurologic concept recognition using a named entity recognition model based on transformers* explores the use of neural networks to extract neurological concepts from unstructured text. *Inter-rater agreement for the annotation of neurologic signs and symptoms in electronic health records* by Oommen et al. evaluates how well different raters perform in identifying neurological concepts in free text from electronic health records. After phenotypes have been extracted from electronic health records, *The visualization of Orphadata neurology phenotypes* explores the visualization of these neurological phenotypes with heat maps and word clouds (Hier et al.).

Although extracting concepts from unstructured text in electronic health records shows great promise for precision neurology and big data, documentation has burdened physicians and other providers. It contributes to physician burnout (2). An article in this Research Topic entitled *It's time to change our documentation philosophy: writing better neurology notes without the burnout* describes strategies that can reduce documentation burden and take advantage of recent changes in documentation regulations from the Center for Medicare and Medicaid Services (CMS) (Rodríguez-Fernández et al.).

Remote neurology

> *Remote neurology is using technology to provide neurological services at a distance.*

Teleneurology has been used for acute and non-acute neurological consultations, including stroke (3). Another example of remote neurology is monitoring neurological patients by actimetry or telemetry (4, 5). In this Research Topic, Ward et al.'s *Implementation and impact of a point of care electroencephalography platform in a community hospital: a cohort study* demonstrates the feasibility of providing emergency EEG services in a community hospital when neurologists and technicians are not available via a point-of-care EEG device.

The twelve articles in this Research Topic highlight the accelerated pace of digitalization in neurology and illustrate the varied uses of neurological observations once they have been

made computable. We believe that further advances in neurology will depend on the increasing digitalization of neurology and that the abundant availability of observations in a computable form will support the implementation of advanced methods from machine learning and artificial intelligence.

Author contributions

DH: Writing – original draft. MC: Writing – review & editing. JR: Writing – review & editing. BK: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article.

MDC acknowledges research funding from the Department of Veterans Affairs (BLR&D Merit Award BX000467) and prior support from Biogen.

Conflict of interest

BK has no relationship with the Kummer Institute at Missouri University of Science and Technology.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

1. Haymaker W, Schiller F. *The founders of neurology: One hundred and forty-six biographical sketches by eighty-eight authors*. Springfield IL: Charles C Thomas Publishers (1970).

2. Fred HL, Scheid MS. Physician burnout: causes, consequences, and (?) cures. *Tex Heart Inst J*. (2018) 45:198. doi: 10.14503/THIJ-18-6842

3. Rubin MN, Wellik KE, Channer DD, Demaerschalk BM. Systematic review of teleneurology: methodology. *Front Neurol*. (2012) 3:156. doi: 10.3389/fneur.2012.00156

4. Merkelbach S, Schulz H, Kölmel H, Gora G, Klingelhöfer J, Dachsel R, et al. Fatigue, sleepiness, and physical activity in patients with multiple sclerosis. *J Neurol*. (2011) 258:74–9. doi: 10.1007/s00415-010-5684-3

5. Tatum WO, Mani J, Jin K, Halford JJ, Gloss D, Fahoum F, et al. Minimum standards for inpatient long-term video-eeg monitoring: a clinical practice guideline of the international league against epilepsy and international federation of clinical neurophysiology. *Clin Neurophysiol*. (2022) 134:111–28. doi: 10.1016/j.clinph.2021.07.016

frontiers | Frontiers in Digital Health

SYSTEMATIC REVIEW
published: 07 July 2022
doi: 10.3389/fdgth.2022.842301

Voice Analysis for Neurological Disorder Recognition–A Systematic Review and Perspective on Emerging Trends

Pascal Hecker[1,2]*, *Nico Steckhan*[1], *Florian Eyben*[2], *Björn W. Schuller*[2,3,4] *and Bert Arnrich*[1]

[1] *Digital Health – Connected Healthcare, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany,* [2] *audEERING GmbH, Gilching, Germany,* [3] *EIHW – Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Augsburg, Germany,* [4] *GLAM – Group on Language, Audio, & Music, Imperial College London, London, United Kingdom*

OPEN ACCESS

Edited by:
Max Little,
University of Birmingham,
United Kingdom

Reviewed by:
Ian Cleland,
Ulster University, United Kingdom
Dhiraj Kumar,
Jamia Millia Islamia, India

***Correspondence:**
Pascal Hecker
Pascal.Hecker@hpi.de
orcid.org/0000-0001-6604-1671

Specialty section:
This article was submitted to Health Technology Innovation, a section of the journal Frontiers in Digital Health

Received: *23 December 2021*
Accepted: *25 May 2022*
Published: *07 July 2022*

Citation:
Hecker P, Steckhan N, Eyben F, Schuller BW and Arnrich B (2022) Voice Analysis for Neurological Disorder Recognition–A Systematic Review and Perspective on Emerging Trends. Front. Digit. Health 4:842301. doi: 10.3389/fdgth.2022.842301

Quantifying neurological disorders from voice is a rapidly growing field of research and holds promise for unobtrusive and large-scale disorder monitoring. The data recording setup and data analysis pipelines are both crucial aspects to effectively obtain relevant information from participants. Therefore, we performed a systematic review to provide a high-level overview of practices across various neurological disorders and highlight emerging trends. PRISMA-based literature searches were conducted through PubMed, Web of Science, and IEEE Xplore to identify publications in which original (i.e., newly recorded) datasets were collected. Disorders of interest were psychiatric as well as neurodegenerative disorders, such as bipolar disorder, depression, and stress, as well as amyotrophic lateral sclerosis amyotrophic lateral sclerosis, Alzheimer's, and Parkinson's disease, and speech impairments (aphasia, dysarthria, and dysphonia). Of the 43 retrieved studies, Parkinson's disease is represented most prominently with 19 discovered datasets. Free speech and read speech tasks are most commonly used across disorders. Besides popular feature extraction toolkits, many studies utilise custom-built feature sets. Correlations of acoustic features with psychiatric and neurodegenerative disorders are presented. In terms of analysis, statistical analysis for significance of individual features is commonly used, as well as predictive modeling approaches, especially with support vector machines and a small number of artificial neural networks. An emerging trend and recommendation for future studies is to collect data in everyday life to facilitate longitudinal data collection and to capture the behavior of participants more naturally. Another emerging trend is to record additional modalities to voice, which can potentially increase analytical performance.

Keywords: neurological disorders, voice, speech, everyday life, multiple modalities, machine learning, disorder recognition

1. INTRODUCTION

1.1. Neurological Disorders and Speech

The burden of neurological disorders on the healthcare system is heavy (1). Neurological disorders manifest themselves with various symptoms at different disease stages. Recognition and diagnosis of most neurological disorders still rely on clinical examinations, mostly upon the manifestation of prominent symptoms. With modern machine learning approaches, researchers have attempted to quantify neurological disorders through various modalities from unobtrusive sensors to gain a longitudinal and holistic picture of the individual patient and course of disease (2). Speech, in particular, is a promising modality, since its production is shown to be very susceptible to slight perturbations caused by those disorders (3). Furthermore, voice recordings are unobtrusive and readily available through the widespread usage of smartphones and other smart devices (4).

To record voice data in a clinical setting, the principle approach is to access a patient cohort and compare it with a representative healthy control cohort. An experimental protocol is developed, which includes a medical assessment to quantify the disorder as well as the recording of voice samples according to clearly defined speech elicitation tasks. The medical assessment provides a 'ground truth' description of the disease status, and the voice recordings are then used to indirectly infer that disease status.

Existing studies have regarded a multitude of neurological disorders, which were reported to have a measurable impact on voice. Those can be loosely grouped, for the scope of this review, into psychiatric disorders, neurodegenerative disorders and speech impairments. Psychiatric disorders encompass depression (3), anxiety, obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD) (5), schizophrenia (6), and, to a certain extent, stress (7). Neurodegenerative disorders include disorders leading to cognitive decline, such as Alzheimer's disease (AD) and mild cognitive impairment (MCI) (8, 9), as well as a broader range of disorders that do not primarily affect cognition, such as amyotrophic lateral sclerosis (ALS) (10), multiple sclerosis (MS) (11), and Parkinson's disease (PD) (12). Lastly, there are several disorders, which affect speech production itself, such as aphasia, dysarthria, and dysphonia. Aphasia is the inability to comprehend or formulate language, dysarthria emerges when muscle coordination for speech production is impaired and dysphonia is when voice is hoarse due to problems with the larynx.

1.2. Data Processing Pipeline

1.2.1. Speech Tasks

The human voice can be produced in various ways, such as reading text, singing or laughing. Recommendations for the technical details on how data for the acoustic assessment of voice production in a clinical setup should be recorded, are provided by Patel et al. (13). These guidelines are compiled by an expert panel from the American Speech-Language-Hearing Association (ASHA), and we strongly recommend consulting these suggestions before setting up novel data collection efforts.

In research settings, participants are asked to produce specific vocalisations, which elicit distinct information for comparable analyses. Those speech tasks, which provide the basis for voice-based disorder quantification, can be grouped into certain categories for the scope of this review. Participants can be asked to produce a sustained phonation of a phone, typically the vowel /a/. Diadochokinesis is the ability to produce antagonistic movements in quick succession, these are typically rapid syllable successions in the case of speech tasks, such as pa-ta-ka. Read speech categorizes tasks, in which written material is provided to be read out aloud. Those materials can be customized for a specific research question or standardised text passages, for example 'the north wind and the sun,' which is constructed to contain every phone in the English language. Free speech encompasses tasks, which do at most provide an initial association point, but then require the participant to associate or behave freely. Examples are clinical interviews between a physician and a patient or a 'picture description task,' in which the patient is asked to describe a picture in their own words.

1.2.2. Feature Extraction

With the obtained data at hand, data analysis is performed next. The typical data analysis pipeline consists of preprocessing the collected data and then applying analytical methods to obtain quantitative insights. The very first step here is to enhance the quality of the raw audio signal by applying, amongst others, denoising and dereverberation. For data preprocessing, audio recordings are often filtered for segments containing speech through voice activity detection (VAD). If multiple speakers are present in one recording, speaker diarisation can be applied to try to separate voice segments, for example, from the patient and a doctor in a clinical interview setting. To perform linguistic analysis, recent advances in automated speech recognition (ASR) enable automatic transcription of the content. With transcriptions, analysis can include, for example, aspects of the semantic structure of the recorded speech [e.g., as done by Tóth et al. (14)].

To make the raw audio signal accessible for automated analysis, statistical derivatives of the signal, namely, features, are extracted. To quantify voice, several features stem from the acoustic aspects of the speech signal that account for the structure of the vocal production system. Prominent and commonly used acoustic feature sets in the community are the expert-knowledge driven Geneva Minimalistic Acoustic Parameter Set [GeMAPS, Eyben et al. (15)] on one hand and the large-scale, general-purpose driven Computational Paralinguistics Challenge [ComParE, Weninger et al. (16)] feature set. Further, there are features, which are tailored for disease-specific vocal dynamics [e.g., (8) on AD]. Low et al. (17) provide a comprehensive overview of the commonly used acoustic features derived from speech in neurological disorder quantification. They regard the GeMAPS features and provide a glossary on the regarded features [based on Cummins et al. (3) and Horwitz et al. (18)], to which we refer the interested reader.

Recent additions to those 'traditional' acoustic features were introduced at ComParE 2018 and 2019 (19, 20), and are based on representations of the audio signal found through deep

neuronal networks (see 1.2.3 Analysis), as well as a high-level summary of speech segments through the Bag-of-Audio-Words (BoAWs) approach. There are out-of-the-box toolkits to extract features, most prominently Praat (21), openSMILE (22), and VOICEBOX. BoAWs can be extracted using the openXBOW framework (23), and learnt representations of the speech signal can be extracted with the DeepSpectrum (24) and AuDeep (25, 26) toolkits. Nonetheless, it is not uncommon to write custom code to perform feature extraction.

1.2.3. Analysis

After preprocessing and feature extraction, data analysis is performed. There are two general approaches for data analysis: statistical analysis and predictive modeling.

For statistical analysis, extracted features are tested with various statistical means to find significant correlations of individual features for the tested conditions, which then express changes in vocal characteristics. The sum of those identified correlating features can ideally serve as general and reliable indicators for different disorders, and are occasionally referred to as 'vocal biomarkers.'

In predictive modelling, on the other hand, machine learning approaches are used to try and build statistical models, which can discriminate between different categories or a general scale, relevant for the regarded disease at hand. Common machine learning models employed for categorical classification are, support vector machines (SVM), the k-nearest neighbors algorithm (k-NN), decision trees (DT), random forests (RF), Gaussian Mixture Models (GMMs), and Hidden Markov Models (HMM). If values from continuous scales are to be predicted, regression models, such as linear regression, logistic regression, support vector regression, and regression trees can be utilised.

If sufficient data is available, artificial neural networks (ANN) can be employed as well, which promise a high performance on large data sets. For ANNs, organizational architectures of neuronal networks inspired by the dynamics in the human brain, are constructed for specific tasks in specific domains. In the field of disease recognition from voice, convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are commonly used, see Cummins et al. (27) for a review of recent developments and examples in the field. Foremost, CNNs learn feature representations of input spectrograms of the audio signal or directly from the raw audio waveform. They either contain architectural elements to perform a classification decision right within the network architecture or other predictive modeling approaches are employed based on those feature representations.

With ComParE 2018 and 2019 (19, 20), learnt deep representations are used as additional baseline feature sets. With the DeepSpectrum toolkit, CNNs pre-trained for image recognition tasks, are used to extract abstract representations of spectrograms from the raw audio signal. auDeep first uses spectrograms from the input audio signal to train encoder-decoder networks without providing class labels (sequence-to-sequence autoencoder), specific to the data at hand. The outputs of the trained encoder can then be used to output features in the form of abstract representations based on the spectrograms of the input signal.

1.3. Related Work

Previous reviews in the field have summarized the state of voice analysis for individual disorders and a few reviews outlined the state of research across several neurological disorders. One prominent systematic review was performed by Low et al. (17), in which they regarded a variety of psychiatric disorders (depression, PTSD, OCD, bulimia, anorexia, schizophrenia, hypomania, and anxiety). Therein, they synthesised which acoustic features are prominently changed in voice in each disorder. They further provided an overview of recent developments and guidelines for data collection. Another review was performed by Voleti et al. (28), which regarded neurological thought disorders (such as AD, schizophrenia, etc.) and created a taxonomy for speech and language features used.

However, the scope of the review of Low et al. (17) was limited to psychiatric disorders and Voleti et al. (28) did not perform a systematic literature search. In this context, a comprehensive review that provides a broad overview of the field of neurological disorder recognition from voice is needed. Therefore, we extended to the scope of Low et al. (17) by also including the neurodegenerative disorders ALS, AD, MCI, MS, and PD. Further, we adopted a reproducible, systematic approach by querying bibliographic databases.

1.4. Scope of the Review

The aim of this review is to provide a general overview of the field of neurological disorder recognition from voice. The main contribution is to survey how voice data is commonly collected across psychiatric and neurodegenerative disorders, how data is frequently analysed, and to highlight emerging trends. The novel insights from this review will be helpful when setting up future data collection efforts.

We do this by searching for publications on original datasets. From these retrieved publications, we extract information on the study setup, the speech tasks utilised, the analysis methods used, and particularities in the voice recording setup (to uncover emerging trends). Furthermore, we provide an overview of significantly correlating acoustic features in common psychiatric and neurodegenerative disorders by extending the work of Low et al. (17). **Figure 1** presents an overview of these outlined topics addressed within this review.

2. MATERIALS AND METHODS

This systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (29).

2.1. Literature Screening

2.1.1. Information Sources/Identification

The following electronic databases were searched for relevant articles: PubMed, Web of Science (Web of Science Core Collection, version 5.35), and IEEE Xplore. Those databases were queried in August 2020 with the following search term:

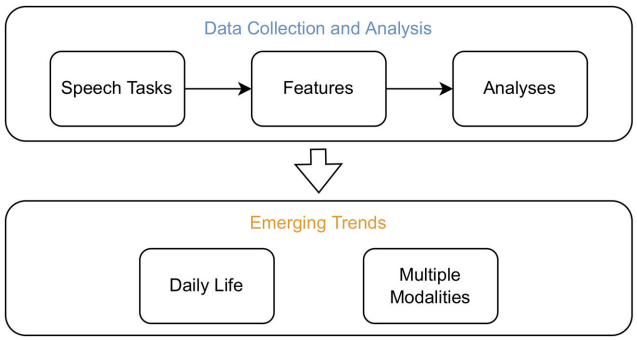


FIGURE 1 | Overview of the topics addressed within this review. Studies which recorded original datasets to assess neurological disorders from voice were screened. From those, information was extracted to assess which speech tasks are commonly employed, which methods are frequently used to extract features, and which analysis methods are prevalent. In addition, significantly correlating acoustic features in common psychiatric and neurodegenerative disorders were summarised. The emerging trends to record data in daily life and from multiple modalities were uncovered by screening for particularities in the recording setup of the retained studies.

((speech OR voice) AND (dataset OR 'data set') AND <disorder specification>). In place of <disorder specification>, a search term for each regarded disorder was inserted:

- ('mental health' OR psychiatry OR psychiatric OR 'affective disorder' OR 'psychological disorder' OR 'mental illness')
- (Anxiety)
- (Depress∗)
- (Stress)
- ('Acute stress reaction')
- ('Obsessive-compulsive disorder' OR OCD)
- ('Post-traumatic stress disorder' OR PTSD)
- (Schizophrenia)
- (Hypomania)
- (Bulimia)
- (Anorexia)
- (Alzheimer∗)
- (Dementia)
- ('Cognitive impairment∗')
- ('Multiple sclerosis')
- (Parkinson∗)
- (Aphasia).

The disorders to be regarded were primarily based on work from other reviews on individual and multiple disorders. The aim was to cover psychiatric as well as prominent neurodegenerative disorders, stress as well as speech impairments such as aphasia. No restriction on the date of publication was imposed.

Google Scholar is an ambivalent source for systematic literature reviews. On one hand, it covers a broad range of publications, especially those in conference proceedings, but on the other hand, it is crawler-based instead of bibliographic and more focused on exploitative instead of systematic search behavior and does not allow bulk downloads of the returned results (30). Therefore, we decided not to use Google Scholar for the systematic search here but can recommend it as well as explicit dataset search engines such as Google Dataset Search to the interested reader to explore individual disorders and aspects of the field.

2.1.2. Screening

Only articles published in English language were considered. After duplicate removal, the first author (P.H.) screened the title and abstract of all records. The focus was to include studies, which report a newly recorded ('original') dataset, and whose research was primarily based on voice and speech. Emphasis was put on studies, which regraded acoustic features (omitting purely linguistic analyses to keep the scope manageable). Studies had to focus on the above-mentioned disorders and include recording voice data from patients. The exclusion criteria for screening were: (a) publications that used existing datasets (i.e., did not record data themselves), (b) publications that were not focused on the above-mentioned neurological and psychiatric disorders, studies involving children, publications, which focused on qualitative or quantitative interview analyzes as well as literature reviews. 203 duplicates were removed with the 'check for duplicates' function in the reference manager Mendeley Desktop (version 1.19.6, Elsevier, Amsterdam, Netherlands); the other bibliography organization of this literature review was done in Zotero (version 5.0.90, Corporation for Digital Scholarship, Vienna, Virginia, USA).

2.2. Data Extraction

Data extraction was performed by P.H. with assistance of N.S. Our approach was to extract a wealth of information to assess common practices and to identify emerging trends in the field later on. Data to be extracted consisted of information on (a) the study setup (number of patients and patient assessment), (b) the voice recording setup (additional modalities, recording conditions: in everyday life or laboratory), (c) the speech tasks (elicitation protocols) used in the study (elicitation material used, if applicable: performance comparison), and (d) analysis methods employed (features extracted, analysis methods used: statistical and predictive modeling and validation schemes).

In published studies, the focus is often put on analysis and it is not clearly stated in the title or abstract, whether original data was recorded or an existing dataset was used. The search term ('dataset' or 'data set') in this systematic review was introduced to search for original datasets However, some original studies might not have been covered. Therefore, we conducted an additional systematic search for literature reviews, which are focused on acoustic analysis of individual disorders and synthesized their identified features.

2.3. Acoustic Features

The aspect of which acoustic features are found to correlate with which neurological disorder was addressed prominently by Low et al. for psychiatric disorders (17). In the broader scope of this review, we aimed to extend their synthesis to also incorporate acoustic features of the neurodegenerative disorders addressed in this review.

Several recent reviews summarized significantly correlating acoustic features in individual neurodegenerative disorders, and we systematically screened an electronic database to retrieve those. We queried Web of Science and used their 'refine' function to retain only review articles published from 2015 on. The search terms to retrieve reviews were:

- TS=((ALS) AND (speech OR voice) AND (analysis))
- TS=((Alzheimer∗) AND (speech OR voice) AND (analysis))
- TS=((Multiple sclerosis OR MS) AND (speech OR voice) AND (analysis))
- TS=((Parkinson∗) AND (speech OR voice) AND (analysis))
- TS=((stress) AND (speech OR voice) AND (analysis)).

Title and abstract were screened and full-text articles were retrieved for the matching candidates. Reviews that provided syntheses in which publications were explicitly listed that found correlating acoustic features with the respective disorder, were retained.

With the publicly available source code[1] and permission provided by Low et al. (17), we extended their synthesis of Figure 3 by adding data of the studies listed in the found reviews. Studies identifying a significant positive correlation received a score of 1, studies finding a significant negative correlation received a score of −1 and non-significant or contradictory studies were scored with 0. Only the most comprehensive review on each disorder (clearly stating the studies found with correlating acoustic features) was used so to cover a comparable number of studies. Reviews used to extract data for extending the figure were the following (9–12). The code to extend the figure of Low et al. (17), and to plot all figures from this review, can be found at GitHub[2]. The aspects of stress and speech impairments were omitted from that overview to fully focus on neurodegenerative disorders.

Furthermore, stress and speech impairments were found to be very heterogeneous. Different manifestations of stress were described by Van Puyvelde et al. (7) for physical, delirious, emotional, and cognitive load and they presented an own model for Voice and Effort (MoVE) to characterize those interactions with voice. Speech impairments such as aphasia, dysarthria, and dysphonia amongst others, stem from general dysfunctions of the speech production systems, and for example, dysarthria can be the consequence of stroke as well as MS.

3. RESULTS

The PRISMA flow diagram is depicted in **Figure 2** and shows the study selection process.

The search terms described in 2.1.1 were used to retrieve 1,492 publications and ultimately, 43 studies were included.

[1]https://github.com/danielmlow/review/tree/389fc387a91f2d38004775ba7c94a970e3d1ae02

[2]https://github.com/Pascal-H/speech_analysis_for_neurological_disease_recognition

After obtaining the final included studies, we noticed that the disorders described in those studies fell into slightly different categories than searched for in the search terms. The categories that started to emerge after data extraction were the following: the neurodegenerative disorders ALS, AD, and PD, the psychiatric disorders bipolar disorder, depression and, to some extend, stress as well as the group of speech impairments, such as aphasia, dysarthria, and dysphonia. Our results and the discussion are therefore based on those categories.

Table 1 presents the number of studies found for each disorder and summary statistics on the number of participants (patients and controls) for all studies of each disorder. Most studies describing original datasets were included for PD followed by stress. PD also has on average most patients included, while for datasets on stress, usually no patients but only healthy participants are recruited.

3.1. Speech Tasks

Figure 3 is a synthesis of the included studies and provides an overview of the proportion of how often each speech task was recorded for each disorder. To provide an overview of the proportion of speech tasks represented in general, dependent on disorder, **Figure 3B** is an inverse view on the data of **Figure 3A**. Here, it is noticeable that speech tasks eliciting free speech (FS) are used most frequently in the included studies. Furthermore, that speech task category was used in all disorders analysed, except for ALS.

Since studies could employ multiple speech tasks, the number of speech tasks may differ from the number of original datasets (**Table 1**). Roughly half of the speech tasks described were utilised in datasets collected from PD patients, stress was represented second often.

In comparison to the other speech tasks regarded in this review, free speech and read speech tasks are less strictly defined. Nevertheless, several typical setups could be identified. Common setups for free speech tasks include (clinical) interview situations (31–37), acted interactions (38–40), picture description (41–44), letting participants talk about a specific question or topic (45–47), or even smartphone conversations (48, 49), as well as specific memory and association tasks suitable for quantifying AD (44). Read speech includes standardised (36, 42, 47, 50) and custom (51–60) sentences or text passages, such as 'the north wind and the sun' (46, 61), 'the rainbow passage' (62, 63), and other passages (64, 65) as well as disease specific tasks, such as constructed sentences with emotionally evoking words for depression quantification (31, 66). Especially in PD, utilising sustained phonation of the vowel /a/ appear to be popular [e.g., (60, 67–71)]. The most specific speech task used was diadochokinesis (DD), which was only used in datasets concerned with PD [e.g., (67)].

Data underlying **Figures 3A,B**, resulting from data extraction, are included in **Supplementary Tables S1**, **S2**.

3.2. Feature Extraction

Figure 4A presents a synthesis of the feature extraction toolkits used. PRAAT, OPENSMILE, and VOICEBOX emerged

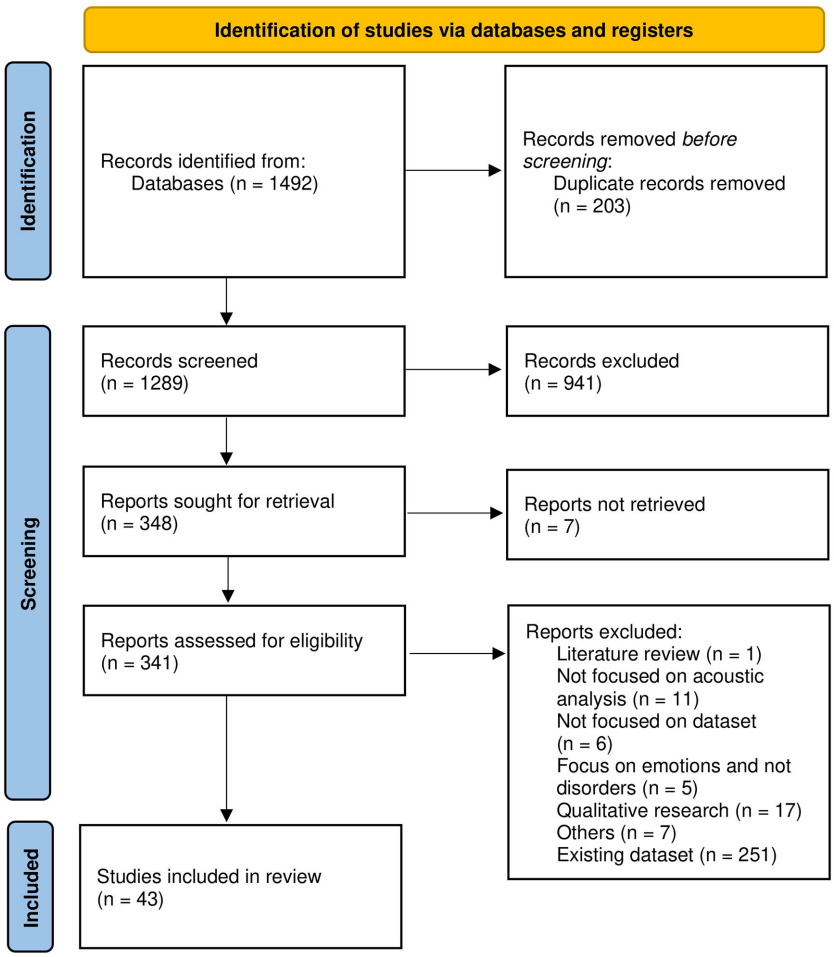


FIGURE 2 | PRISMA flow chart for study selection.

TABLE 1 | Overview of the included studies reporting on original (newly recorded) datasets from neurological disorders to provide a survey over emerging trends in the field.

Disorder	# studies	Patients	Controls
		Median (range)	Median (range)
Parkinson's	20	36 (3–1,513)	20 (8–64)
Stress	6	-	44 (4–60)
Depression	5	92 (12–224)	61 (12–397)
Speech impairments	4	12 (8–21)	13 (8–21)
Alzheimer's	3	82 (71–214)	93 (82–268)
ALS	3	13 (11–25)	12 (11–13)
Bipolar	2	31 (10–51)	9 (9)

as commonly used out-of-the-box toolkits for feature extraction. Roughly half of the included studies used custom code or did not specify the toolkit used.

3.3. Analysis

3.3.1. Statistical Analysis

Figure 4B aggregates broad categories of analysis methods. Statistical analyses, where individual features are tested for significance, are relatively frequently used.

Figure 5 is an extended version of the synthesis created by Low et al. (17). Acoustic feature categorisation is based on Eyben et al. (15). Each cell represents a summary of the studies with statistical tests performed for the respective feature. The more studies were found for a respective feature, the larger the cell. The found correlation of each study determines the shading: if a feature correlates positively with the disorder, the cell is shaded red. In case of a negative correlation, the cell is shaded blue and if non-significant findings are presented, the shading is gray. The final shading of a cell is determined by accounting for all correlations for all reported studies: the more intense, the more unanimous the findings across all studies and the less intense, the less unanimous are the aggregated studies. For each of the added neurodegenerative conditions, a review was systematically

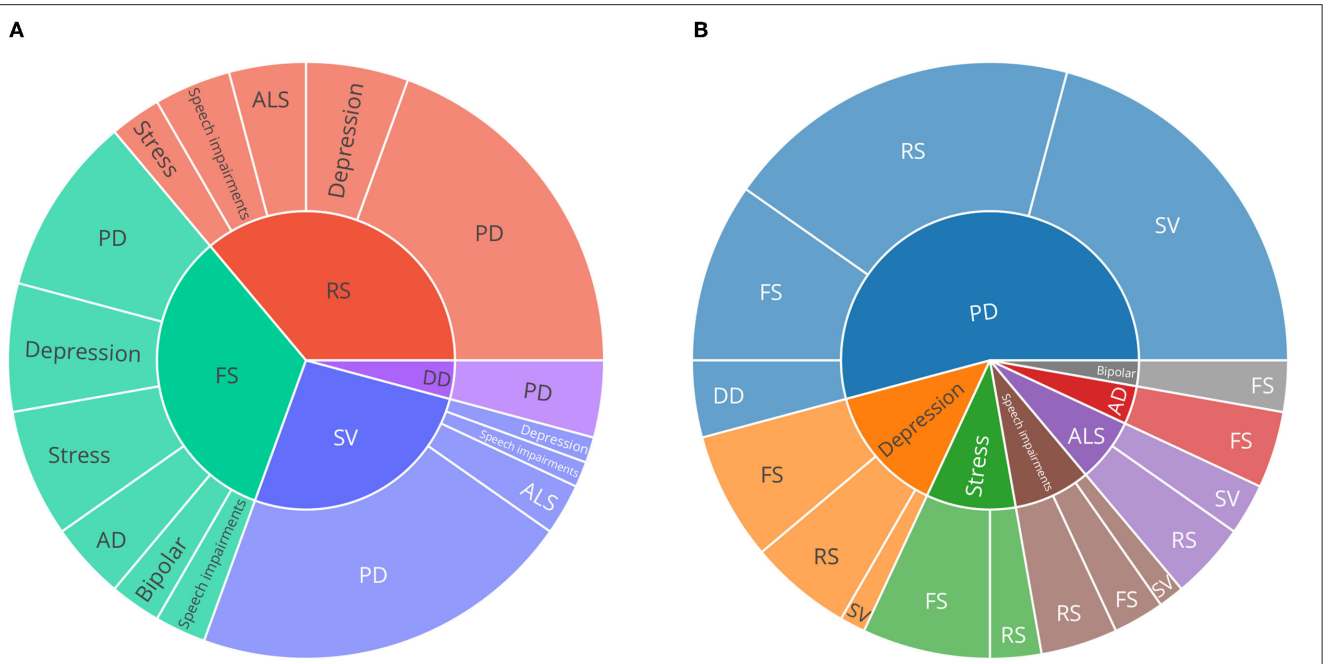


FIGURE 3 | Sunburst charts describing the proportion of speech tasks and neurological disorders. **(A)** speech task categories on the inner circle, **(B)** disorders on the inner circle. Only publications describing original (newly recorded) datasets were considered to provide an overview of emerging trends in the field. PD, Parkinson's disease; AD, Alzheimer's disease; ALS, amyotrophic lateral sclerosis; SV, sustained vowels; DD, diadochokinesis; RS, read speech; FS, free speech.

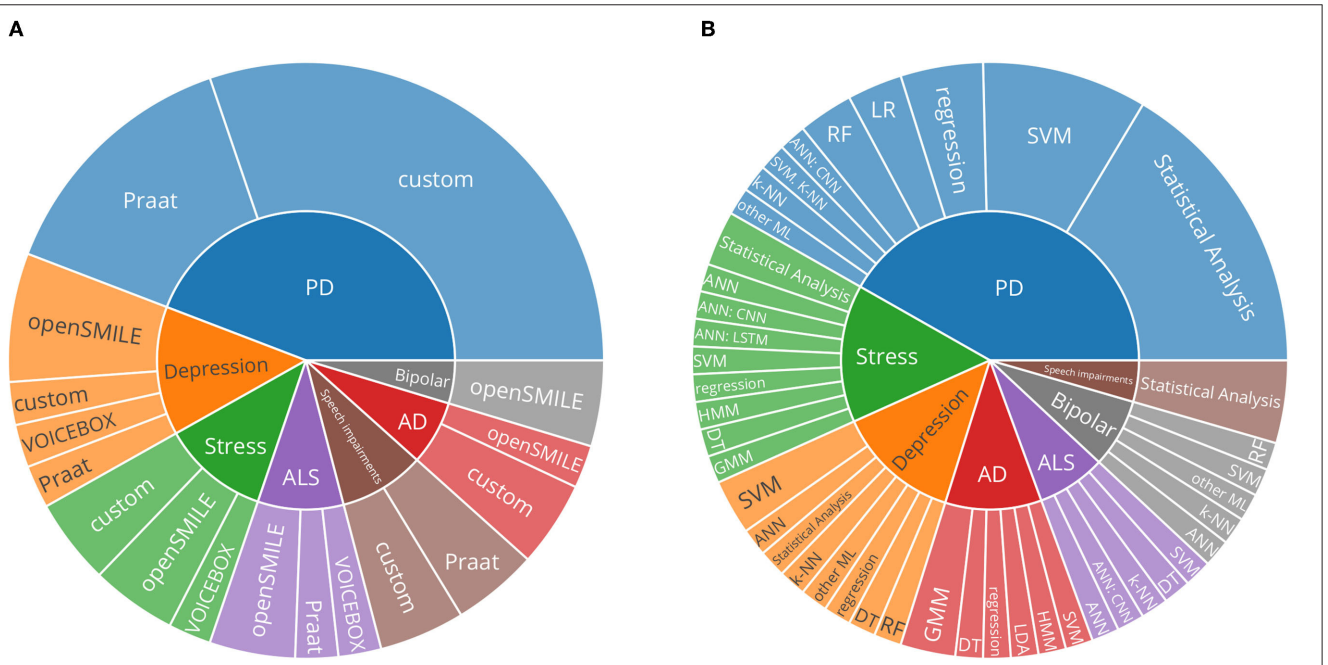


FIGURE 4 | (A) Sunburst chart describing the proportion of feature extraction toolkits used and neurological disorders. 'Custom' describes studies which did not mention the toolkit used or which utilised custom methods to extract features. **(B)** Sunburst chart describing the proportion of predictive modeling approaches used and neurological disorders. Categorical classification: SVM, support vector machines; k-NN, k-nearest neighbors algorithm; DT, decision trees; RF, random forests; GMM, Gaussian Mixture Models; HMM, Hidden Markov Models; Regression models: LR, linear regression; regression: other regression methods. ANNs, Artificial neural networks; CNN, convolutional neural networks; LSTM, long short-term memory networks; PD, Parkinson's disease; AD, Alzheimer's disease; ALS, amyotrophic lateral sclerosis.

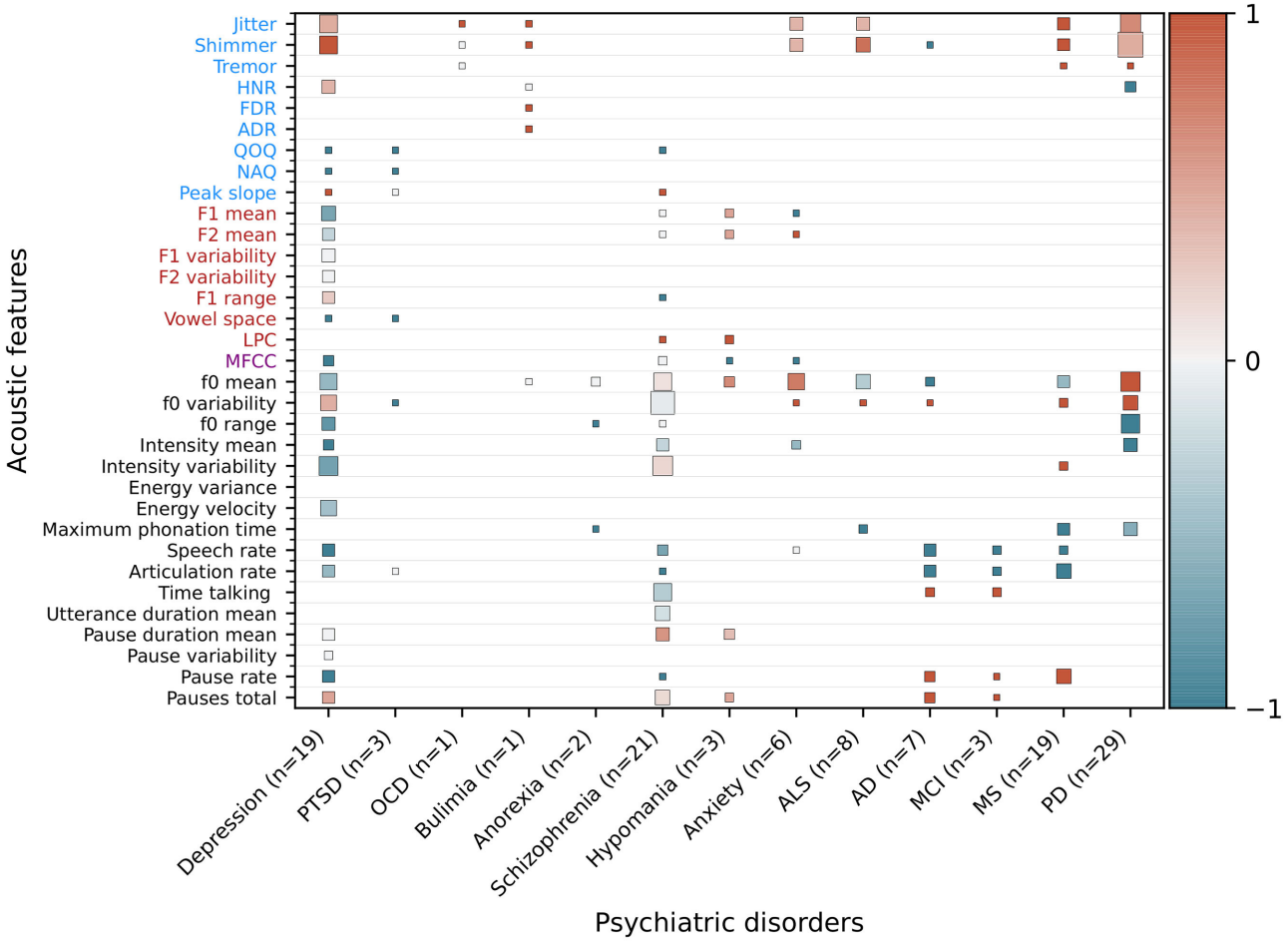


FIGURE 5 | Extended heatmap based on Figure 3 from Low et al. (17). In addition to psychiatric disorders, significantly correlating features from neurodegenerative disorders were extracted and added based on recent reviews of the respective disorders. Features that are significantly higher in a psychiatric population than healthy controls or that correlate positively with the severity of a disorder receive a score of 1 (red), features that are lower or correlate negatively receive a score of –1 (blue), and non-significant or contradicting findings receive a score of 0 (gray). The mean is computed for features with multiple results. The cell size is weighted by the number of studies. Features not studied in a disorder are blank. Additionally, the number of studies (n), of which the correlating features are extracted, is given for each disorder. OCD, obsessive-compulsive disorder; PTSD, post-traumatic stress disorder; ALS, amyotrophic lateral sclerosis; ALS, amyotrophic lateral sclerosis; AD, Alzheimer's disease; MCI, mild cognitive impairment; MS, multiple sclerosis; PD, Parkinson's disease.

identified, which synthesized several studies which reported correlations of acoustic features with the respective condition. The review used to extract studies for ALS was (10), the one for AD and MCI was (9), the one for MS was (11), and the one for PD was (12).

3.3.2. Predictive Modeling

The predictive modeling approaches pursued by the retrieved studies are presented in **Figure 4B**. Classical (non-neural-network-based) approaches are in the majority. Of those approaches, support vector machines followed by regression approaches, are most prominent. General artificial neural networks (ANNs) and convolutional neural networks (CNNs) are most widespread in the included studies. Neural networks can consume the raw audio signal in various ways. The introduced learnt representations with DeepSpectrum and AuDeep were used in Baird et al. (61), and AuDeep achieved the best results. Often, features based on the mel-frequency cepstral coefficients (MFCCs) are used as input to the studies that employ deep learning approaches. MFCCs, simplified, aim to represent a spectrum based on how speech is perceived by human hearing. Mendiratta et al. (46), Khorram et al. (48), and An et al. (56) use MFCCs to represent the speech signal for their deep learning approaches. In addition, Khorram et al. (48), Baird et al. (61), An et al. (56), and Prince et al. (72) provide hand-crafted feature sets to the neural network, for example, Khorram et al. (48) used eGeMAPS features as input for a ANN.

Table 2 shows datasets, in which further modalities in addition to audio were recorded. Only included datasets for ALS, PD and stress recorded multiple modalities.

TABLE 2 | Included studies with original datasets, in which multiple modalities were recorded.

	Year	Condition	Additional modalities
Garcia-Gancedo et al. (65)	2019	ALS	Physical activity, heart rate variability (HRV)
An et al. (56)	2018	ALS	Articulatory movement data
Wang et al. (53)	2016	ALS	Articulatory movement data
Prince et al. (72)	2019	Parkinson's	Sensor data: Finger tapping, walking, memory task
Barnish et al. (36)	2017	Parkinson's	Video, Respiratory Sinus Arrhythmia (RSA) and Heart Rate (HR)
Gratch et al. (32)	2014	Depression	Videos
Baird et al. (61)	2019	Stress	Biosignals: Blood volume pulse (BVP), Skin conductance (SC)
Lefter et al. (38)	2014	Stress	

TABLE 3 | Included studies with original datasets, in which data was collected outside a traditional laboratory setup: in everyday life.

	Year	# Subjects	Condition	Recording condition
Khorram et al. (48)	2018	60	Bipolar disorder	Conversations during daily smartphone usage
Maxhuni et al. (49)	2016	10	Bipolar disorder	Smartphone recorded constantly in the background
Zhang et al. (45)	2020	222	Depression	Web forms
Prince et al. (72)	2015	1,513	Parkinson's	User smartphones
Dubey et al. (55)	2015		Parkinson's	Smartwatch in group session for vocal exercises
Palacios-Alonso et al. (40)	2019	32	Stress	Smartphone
Garcia-Gancedo et al. (65)	2019	25	ALS	Home monitoring and clinical site visits for sensor data recording; audio only collected at clinical site

Table 3 presents datasets, in which data was collected outside a controlled laboratory setup ('in everyday life'). Recordings here were most prominently done via user smartphones or web forms.

4. DISCUSSION

In this review, we systematically screened for publications, in which voice data for various neurological disorders were recorded. Syntheses of included studies provide a high-level overview of different disorders and insights into emerging trends in the field. Previous work was extended to provide an overview of which features are correlated with changes in voice in psychiatric and neurodegenerative disorders.

The respective subsections in the discussion aim to provide valuable guidance when performing such data collection. We cover the aspects of which speech tasks are frequently used, which confounders might be encountered, which feature extraction toolkits are available, which analysis methods are common, and which validation procedures should be employed.

4.1. Neurological Conditions and Speech

As presented in **Table 1**, systematic literature screening returned the most original datasets for PD. Research done in this domain was one of the earliest approaches in the whole field of speech analysis for disease recognition (73) and therefore, the high aggregation of datasets could be reasonable. ALS and bipolar disorder, on the other hand, appear to be relatively under-explored research areas in terms of datasets published.

4.1.1. Speech Tasks

When regarding the numbers of speech tasks used for different disorders as presented in **Figure 3**, it appears that the free speech task category is most commonly used in existing datasets, closely followed by read speech with only one dataset less. Both task categories show broad heterogeneity and can be divided into individual subcategories. In essence, however, free speech tasks aim at capturing 'naturally flowing' speech, in which especially hesitations and pauses can be valuable disease indications, for example, when regarding AD or MCI (9). A very standardised approach used across multiple disorders appears to be the picture description task, utilised in PD (41), stress (42), and AD (43, 44). Recently, Slegers and Filiou (74) reviewed several studies that employ picture description tasks to describe their potential in clinical practice to assess AD. Similarly, Mueller et al. (75) assessed how picture description tasks can be used in diagnosing AD and even potentially already in MCI. Speech tasks prompting read speech cover a wide range of the participant's

language (in contrast to e.g., the task of the constrained sustained pronunciation of vowels), while still having a fixed body of text that is consistent for all participants.

A few publications ced the performances of different speech tasks used in the same dataset. This can provide valuable insights into which tasks appear to cover the best information on a disorder status in an actual recording setup. However, only 6 of the included publications provide those analyses, therefore, unfortunately, these reports can be only regarded as anecdotal. Sakar et al. (57) and Karan et al. (51) each report in their analysis on PD that performance on sustained phonations of vowels performed better than read speech. Interestingly, (59) recorded Czech speaking participants with and without PD and regarded a neutral and a word-stress-modified reading passage and found that the passage with word stress modifications performed better. Further, they achieved their best performance with a free speech task, in which participants had to recite a poem from memory. Alghowinem et al. (31), Liu et al. (37), and Zhang et al. (45) reported that tasks using free speech performed better than sustained vowels and read speech for depression, respectively. A recent study assessed differences in performance of various speech tasks eliciting connected speech in patients with early AD and MCI. That study, therefore, offers some practical consideration for which particular free speech task might be best suitable for these conditions (76). Analysing the performance of speech tasks is valuable for the community, since choosing the best performing speech task can reduce time effort and burden imposed on the patient in a clinical as well as in an everyday-life setup.

4.1.2. Confounders

In their review (section 4.2), Low et al. (17) portray several relevant confounding factors, which should be considered and avoided during data collection. Regarding rather symptoms and problems and not only disorder rating scales promises to provide a more fine-grained view of a patient and account for disorders, in which more heterogeneous symptoms are present (77). A central aspect that needs to be controlled for in voice analysis, are confounding factors that influence voice production. Commonly assessed factors are, for example, age, sex, and native language, less common are comorbidities, race, education, height, weight, and dialect. Especially medication is not frequently reported but plays a crucial role since its side effects might influence speech production.

4.2. Data Processing Pipeline

4.2.1. Feature Extraction Toolkits

Regarding the toolkits used for feature extraction, as portrayed in **Figure 4A**, of all studies actually extracting features, almost half used custom methods. In particular, in the field of PD, datasets are described, which validate and explore the impact of Lee Silverman Voice Treatment (LSVT) (78) to mitigate voice-based impairments due to PD. Success in that treatment routine is measured in increased vocal intensity [e.g., (63, 79)], and therefore in those studies, features are very specific and focused only on that outcome. As pointed out by Low et al. (17), standardising feature extraction yields the benefit of better comparability across studies, but specific approaches in which anatomically informed and manually constructed features can reflect an aspect of a disorder, which might not be covered by standardised feature sets, can be valuable as well. Within the scope of this review, relevant feature extraction toolkits were presented. Studies using custom methods are hard to quantify systematically since the performance obtained on one dataset might not transfer well to another dataset. Further, it is worth emphasising that, since studies included in this review are limited to original datasets, the actual usage in all analytical studies might vary.

4.2.2. Features Correlating With Neurodegenerative Conditions

We extended the figure of the synthesis of significantly correlating features for neurological disorders in Low et al. (17) by adding the neurodegenerative conditions ALS, AD and MCI, MS, and PD (**Figure 5**). Findings regarding the disorder-related features are summarized as the following:

Amyotrophic lateral sclerosis: Chiaramonte and Bonfiglio (10) conducted a meta-analysis and found that jitter and shimmer correlate positively, and maximum phonation time (MPT) correlates negatively, significantly with progression of bulbar ALS. The predominantly initial spinal type of ALS, characterised by muscle weakening, usually transitions to show some bulbar involvement at a later stage, at which speech impairments are surfacing. No significant correlations between F0 mean and F0 variability were observed in the meta-analysis.

Alzheimer's disease and mild cognitive impairment: Martínez-Nicolás et al. (9) systematically reviewed altered acoustic features in patients with AD and MCI. Decreased speech and articulation ratio, as well as an increased number of pauses, are characteristic for the early stages of AD. Fewer studies are concerned with MCI, but increased pause duration and longer speech and phonation time are reported. Language impairments are already present in the prodromal (pre-symptomatic) stage and the challenge of the field is to distinguish cognitive impairments due to age from the onset of AD.

Multiple sclerosis: Noffs et al. (11) systematically screened for studies describing speech impairments in MS and found, for acoustic analyses, that a slowing in tongue movement causes a lower speech and articulation rate. Further, glottal inefficiency causes increased jitter and shimmer, and intensity variability and symptoms are expected to worsen upon disease progression.

Parkinson's disorder: Chiaramonte and Bonfiglio (12) conducted a meta-analysis and concluded that jitter, shimmer and F0 variability are significantly increased in patients with PD. Increased F0 variability is likely to be caused by increased rigidity in laryngeal and respiratory muscles and the associated inability to keep the laryngeal muscles in a fixed position.

4.2.3. Analysis Methods

As depicted in **Figure 4B**, statistical analyses, where individual features are tested for significance, are described along with datasets for PD, speech impairments, stress, and depression. Lee

Silverman Voice Treatment (LSVT) is usually assessed in such manner (63, 79), and studies describing novel ways in collecting datasets [e.g., (45)] rely on such statistical descriptions.

From the 'traditional' predictive modeling approaches, support vector machines (SVMs) are most frequently used, which is in line with the baseline of the Interspeech ComParE challenge (16). Regression approaches are suitable to map disorder assessment scales (e.g., UPDRS for PD) but can potentially struggle with small sample sizes and unbalanced class distributions.

Approaches using neural networks are gaining popularity in recent years and are discussed in the following review (27). The recent ComParE 2018 and 2019 (19, 20) introduced features from deep representations as baseline methods in the domain of computational paralinguistics. This approach was pursued by Baird et al. (61) in the retrieved studies. In the other studies utilising neural networks, various network architectures are used. The way in which raw audio signals are processed and fed into neural networks depends strongly on the employed network architecture.

The overall goal of predictive modeling approaches is to create models that learn to generalise and therefore could classify voice samples of speakers, who were not present in the original dataset. To evaluate how well suggested predictive modeling approaches would perform at that task, the dataset should be split up into train, validation and test partitions. The train partition serves to adjust and fine-tune parameters of the model and those adjustments are then tested on the validation partition. The best performing model is then evaluated on the test partition, a hold-out part of the dataset (or ideally even a completely independent dataset with samples from the same disorder). This hold-out part should provide a sound judgement on how the model performs on data that it did not encounter during training/validation. Speakers have to be separated through all partitions since otherwise, the model can learn to identify a user and not learn the underlying information about the disease itself.

For imbalanced class distributions, which can be common in datasets with neurological disorders, the unweighted average recall (UAR) is the metric used in ComParE and should be used for comparing results across different predictive modeling approaches. Low et al. (17) provide some further, helpful advice for evaluation and validation of modeling approaches. Foremost, they advocate for using nested bootstrapping for a more robust performance estimation on small ($<$ 100 patients) datasets. Ideally, the train, validation and test partitions would each represent the whole subject population of the dataset, but since this is unlikely for smaller subject numbers, nested bootstrapping provides a means to describe the mean or median estimate over a multitude of evaluation runs.

4.3. Emerging Trends

Some of the studies included in this review used a non-conventional clinical data recording setup. Those approaches can be categorized in a) data collection performed 'in everyday life' and b) data collected from multiple modalities. Both categories are introduced further in the following section to provide an overview of these emerging trends.

4.3.1. Everyday-Life Data Collection

Traditionally, medical datasets for analysing the impact of a disorder on voice were recorded in controlled recording conditions with relatively small sample sizes, since access to patients is a big obstacle to overcome and only possible through clinical institutions. Predictive modeling approaches and results from statistical analyses should be as general and flexible as possible, and also work on novel participants, who were not part of the initially recorded data. This requirement led to efforts in recent years to collect large-scale datasets. In those datasets, participants are often recruited not only at a clinic, but through interest groups and networks for disorders (80). Data collection itself is then being done remotely, in an offsite setup, through mobile devices such as smartphones (45) and smartwatches (55). These efforts are very promising to push the field toward a real-world use case, in which enough data can be collected to extrapolate models to work sufficiently well when confronted with completely novel data.

4.3.1.1. Example Studies

In most clinical datasets, participants are only screened once since there is an increased effort to track and re-invite participants. Systems with which participants can provide several samples over a given observation time (49), are a big advantage and opportunity of large-scale data collection efforts. This can provide valuable insights in researching longitudinal disease courses [e.g., (48)], but recording sessions have to be designed differently than clinical sessions to put particular emphasis on adherence, therefore reducing user burden, and to motivate the user to record multiple times.

The overview in **Figure 3** presenting which speech tasks are most commonly used in existing datasets, can provide some considerations on which speech tasks can be prioritized when user time is a considerable factor. Therefore, a legit approach could be to design a minimalist, user-friendly recording protocol, set up a small, clinical pre-study to validate that the relevant indications for the disease to be assessed are covered, and then use that minimalist protocol in a large scale data collection effort. According to our systematic screening, it depends on the disorder, but free speech and read speech tasks are most commonly used and could therefore make up a minimalist protocol.

After literature screening for this review, a publication was released, which showcases the highlighted points for everyday-life data collection (81). The authors managed to gather voice samples via a web app of over 6,650 participants, of which roughly 10 % reported to be depressed. They are piloting an extensive survey with 17 speech tasks, which on one hand seems to impair adherence (of 6,650, only 1,382 participants completed at least two of the total four survey versions), but on the other hand, can provide valuable insights into which speech tasks indeed carry most relevant information. This goes to show that a careful balance between user burden and the information to be collected is to be considered.

4.3.1.2. Practical Considerations

The effort to bridge the gap between research and a real-world use case, however, is very high in the healthcare setup, since stakes are exceedingly more grave than in other fields. For example, providing an unsuitable product recommendation in an e-commerce setup is intuitively less detrimental than mislabelling a potential patient in a healthcare setup, where diagnosis or therapy decisions might be impacted. Therefore, even in large-scale data collection efforts, representing a whole population of potential later users is still a challenge, but a big step toward the right direction. Before generalising to everyday-life use cases, rigorous validation of experimental results is required, including quantification of changes in speech with time or treatment, as emphasised by Robin et al. (82).

Other challenges in large-scale data collections are non-standardised recording conditions. In controlled, clinical setups, high-quality microphones and even recording booths are used [e.g., (31)], but when collecting the data remotely from the user, microphone types might vary along with the variety of different smartphones on the market [e.g., (45)]. A few studies reported experiences and ideas to combat those issues [e.g., (83)]. Additionally, knowing beforehand which features are expected to be affected by the disorder to be studied can help when trying to adjust the data analysis pipeline respectively [e.g., (17)].

Obtaining reliable ground truth labels is another relevant aspect when participants are not recorded in a controlled clinical setup. Usually, participants are asked to self-annotate their data. To ensure a sufficient quality for these labels, it has to be ensured that participants can properly understand the applied labels themselves, and that the labeling process should be made as straightforward and effortless as possible (84).

A further consideration for large-scale data collection efforts is recruitment and user adherence. In clinical setups, cohorts are usually available through patients who are regularly treated in the clinic itself. If those patients are usually belonging to a rather elderly cohort (e.g., PD), specific considerations are required to ensure that smart devices to be used for large-scale data collection can be intuitively used and do not cause user frustration (72). To obtain data from a larger number of patients, the available cohort at a clinic might not be sufficient. Interest groups and networks for particular disorders can be a viable source to recruit patients (80), and healthy participants can be reached through online marketing or platforms such as Amazon Mechanical Turk [e.g., as done in R'mani Haulcy et al. (85)].

Another consideration and challenge for large-scale data collection is the identification of unique users. Machine learning systems in the voice analysis domain can easily overfit when no clear speaker separation is done. Since in anonymous data collection efforts [e.g., Zhang et al. (45)], it cannot be ruled out that the same speaker donates multiple samples, evaluation of the system's performance might be biased. Recruiting a clear set of speakers can be a solution, or using a setup in which the user has to register with a unique ID [e.g., via email address, Hecker et al. (86)].

4.3.1.3. Data Privacy

A major and not negligible caveat in data collection approaches in everyday life is that the collected voice data might contain identity revealing aspects, and therefore, potential misuse could bear severe consequences. Especially in longitudinal data collection efforts, the longer the data collection effort continues, the more information from a patient is being collected, and the likelier a potential breach could be.

In a commercial setting, the technology of voice assistants seems promising at first glance to be utilised to quantify the status of disorders from voice. Voice assistants like Amazon Echo and Google Home are widespread and people interact readily with them through 'free speech' prompts. Recently, some research has been done to find ways in which health-related processing of voice assistant queries can be implemented in a privacy-preserving way [e.g., (87, 88)]. However, privacy considerations on medical (voice) data collected in everyday life are a magnitude higher in the medical context than in private usage scenarios, and therefore, the technology is not yet widely used for medical voice collection yet (89). The majority of data collection efforts in everyday life identified within this review nevertheless focuses on dedicated implementations: custom apps on the smartphone (48, 49, 55, 72) and web sites (40, 45, 65). That way, data is not being processed or residing on the third party system of a voice assistant.

4.3.2. Multiple Modalities

Another trend is the collection of data from multiple modalities. Predictive modeling approaches can gain performance when using more than a single modality, and this approach is known for some time already (90). In PD for example, gait is prominently affected besides voice (91). In affect-related disorders, such as major depression and bipolar disorder, video as an additional modality can carry complementary information on expressed emotion. The prominent Audio/Visual Emotion Challenge and Workshop (AVEC) addressed this aspect: featured sub-challenges in which audio and video data or features from clinical interviews (92) and interviews with virtual agents (93, 94) from the Distress Analysis Interview Corpus [DAIC, (32)] are provided as well as data on bipolar disorder (95). In addition, setups in which data is collected from the smartphone's camera as additional video input within a commercial setup are nowadays easily conceivable (96). The number of smart devices with sensors is constantly growing and therefore this topic has also been increasingly reflected in more recent dataset publications in this review (40, 48, 65, 72).

4.3.2.1. Example Studies

The datasets we identified, which used multiple modalities, were recorded from voice data from patients with PD, stress, and ALS. Interestingly, apart from the traditional pairing of voice and video [as in Gratch et al. (32) for depression], some other modalities in combination with speech emerged. For PD, researchers used sensor data to additionally assess the motoric capabilities of the patients through a commonly used finger tapping task, a walking task, and a memory task (72). In another dataset, video, respiratory sinus arrhythmia, and heart rate data (36) were combined. Since PD affects motor coordination, assessing those modalities can yield some benefit, especially since (72) was done in a remote care setup.

Similarly, ALS affects muscle coordination and the studies using additional modalities to voice recorded physical activity and heart rate variability (65) as well as articulatory movement data (53, 64).

For stress, datasets were retrieved, which recorded biosignals such as blood volume pulse and skin conductance (61), as well as video data (38). Video data is frequently used to assist in the quantification of the expression of affect and therefore might also yield valuable additional information in a setting to elicit stress. Biosignals, such as skin conductance and blood volume pulse, are traditionally used to predict stress, and the attempt to infer them from the audio signal could pave the way to detect stress unobtrusively by voice only.

4.3.2.2. Adapting the Data Processing Pipeline

When recording and analysing data collected from multiple modalities, however, the complexity of the recording setup and analysis pipeline is increased, since the different modalities need to be fused at some point in the analysis pipeline. If features are fused before predictive modeling algorithms are employed, the approach is termed 'early fusion,' if multiple models for the respective modalities are created and their outputs are fused, it is termed 'late fusion.'

In practice, increased complexity when conducting a study to record and analyse data as well as the need to still fully understand the effect of disorders on the voice modality are likely the reasons for focused datasets. But in line with the emerging trend toward everyday-life data collection, multimodal approaches could gain further popularity. When utilising participants' smartphones for data collection, their sensors already provide intrinsic additional modalities such as video, location, movement, and even device usage data. On the other hand, relying only on the voice modality could in practice lead to applications in settings where only that modality is available, for example when assessing phone calls (48).

4.4. Future Work

Based on the systematic screening of various original datasets from voice recordings of neurological disorders, we further highlight the following emerging trends. Future data collection endeavors will benefit prominently from collecting data in an everyday-life setup. Recording data in a clinical setup is a good means to explore specific nuances and aspects (e.g., symptoms) of a disorder further while recording data in everyday life enables insights into longitudinal disorder manifestation. Recording further modalities apart from audio can boost the performance of predictive modeling approaches. More research should be done on multi-modal data processing to balance the benefit of additional information and the cost of increased complexity.

5. CONCLUSION

To summarize, a variety of speech tasks are used in clinical practice, and usually, multiple tasks are recorded within one study to ensure that the relevant, distinct information for comparable analyses are covered. When regarding the common analysis methods utilised, we observe that custom feature extraction methods are quite prominent. However, established feature extraction toolkits within the research community yield the benefit of better comparability of the analysed features across different studies. Recently, learnt representations from deep learning toolkits are finding their way into the research community and offer an addition to the standard acoustic features.

The main contribution of this review is to provide a general overview of the field of neurological disorder recognition from voice. We emphasise how data collection efforts are undertaken, which trends emerge in the field, and aim to provide the readers with valuable practical insights. Lastly, we extend the overview of significantly correlating features for psychiatric disorders from Low et al. (17) and added prominent neurodegenerative disorders. This overview is particularly helpful when planning a data collection approach for a respective disorder to see which manifestations in voice are to be expected and to see with which speech task these could be captured.

DATA AVAILABILITY STATEMENT

The source code for generating the figures, including the raw tables of extracted data from the literature search as well as the extension of the figure of Low et al. (17) can be found on GitHub: https://github.com/Pascal-H/voice_analysis_for_neurological_disorder_recognition.

AUTHOR CONTRIBUTIONS

PH, FE, BS, and BA: conceptualisation. PH and NS: methodology (data identification and screening). PH, FE, and BA: analysis (syntheses). PH: writing – original draft preparation. PH, NS, FE, BS, and BA: writing – review and editing. All authors contributed to the article and approved the submitted version

FUNDING

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 491466077.

ACKNOWLEDGMENTS

We would like to thank Arpita M. Kappattanavar for her valuable input on the data extraction methodology. Further, we express our gratitude to Lin Zhou for proofreading the manuscript, and Maximilian Schmitt for his input on deep learning approaches.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2022.842301/full#supplementary-material

REFERENCES

1. Feigin VL, Nichols E, Alam T, Bannick MS, Beghi E, Blake N, et al. Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the global burden of disease study 2016. *Lancet Neurol*. (2019) 18:459–80. doi: 10.1016/S1474-4422(18)30499-X
2. Dorsey E, Omberg L, Waddell E, Adams JL, Adams R, Ali MR, et al. Deep phenotyping of Parkinson's disease. *J Parkinsons Dis*. (2020) 10:855–73. doi: 10.3233/JPD-202006
3. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Commun*. (2015) 71:10–49. doi: 10.1016/j.specom.2015.03.004
4. Lu H, Frauendorfer D, Rabbi M, Mast MS, Chittaranjan GT, Campbell AT, et al. Stresssense: detecting stress in unconstrained acoustic environments using smartphones. In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. Pittsburgh, PA (2012). p. 351–60.
5. Bourla A, Mouchabac S, El Hage W, Ferreri F. e-PTSD: an overview on how new technologies can improve prediction and assessment of posttraumatic stress disorder (PTSD). *Eur J Psychotraumatol*. (2018) 9:1424448. doi: 10.1080/20008198.2018.1424448
6. Parola A, Simonsen A, Bliksted V, Fusaroli R. Voice patterns in schizophrenia: a systematic review and Bayesian meta-analysis. *Schizophrenia Res*. (2020) 216:24–40. doi: 10.1016/j.schres.2019.11.031
7. Van Puyvelde M, Neyt X, McGlone F, Pattyn N. Voice stress analysis: a new framework for voice and effort in human performance. *Front Psychol*. (2018) 9:1994. doi: 10.3389/fpsyg.2018.01994
8. Pulido MLB, Hernández JBA, Ballester MÁF, González CMT, Mekyska J, Smékal Z. Alzheimer's disease and automatic speech analysis: a review. *Expert systems with applications*. (2020) 150:113213. doi: 10.1016/j.eswa.2020.113213
9. Martínez-Nicolás I, Llorente TE, Martínez-Sánchez F, Meilán JJG. Ten years of research on automatic voice and speech analysis of people with alzheimer's disease and mild cognitive impairment: a systematic review article. *Front Psychol*. (2021) 12:645. doi: 10.3389/fpsyg.2021.620251
10. Chiaramonte R, Bonfiglio M. Acoustic analysis of voice in bulbar amyotrophic lateral sclerosis: a systematic review and meta-analysis of studies. *Logopedics Phoniatr Vocol*. (2020) 45:151–63. doi: 10.1080/14015439.2019.1687748
11. Noffs G, Perera T, Kolbe SC, Shanahan CJ, Boonstra FM, Evans A, et al. What speech can tell us: a systematic review of dysarthria characteristics in Multiple Sclerosis. *Autoimmunity Rev*. (2018) 17:1202–9. doi: 10.1016/j.autrev.2018.06.010
12. Chiaramonte R, Bonfiglio M. Acoustic analysis of voice in Parkinson's disease: a systematic review of voice disability and meta-analysis of studies. *Revista de Neurologia*. (2020) 70:393–405. doi: 10.33588/rn.7011.2019414
13. Patel RR, Awan SN, Barkmeier-Kraemer J, Courey M, Deliyski D, Eadie T, et al. Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function. *Am J Speech Lang Pathol*. (2018) 27:887–905. doi: 10.1044/2018_AJSLP-17-0009
14. Tóth L, Hoffmann I, Gosztolya G, Vincze V, Szatlóczki G, Bánréti Z, et al.. A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Curr Alzheimer Res*. (2018) 15:130–8. doi: 10.2174/1567205014666171121114930
15. Eyben F, Scherer KR, Schuller BW, Sundberg J, André E, Busso C, et al. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Trans Affect Comput*. (2015) 7:190–202. doi: 10.1109/TAFFC.2015.2457417
16. Weninger F, Eyben F, Schuller BW, Mortillaro M, Scherer KR. On the acoustics of emotion in audio: what speech, music, and sound have in common. *Front Psychol*. (2013) 4:292. doi: 10.3389/fpsyg.2013.00292
17. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: a systematic review. *Laryngosc Investigat Otolaryngol*. (2020) 5:96–116. doi: 10.31219/osf.io/5pwze
18. Horwitz R, Quatieri TF, Helfer BS, Yu B, Williamson JR, Mundt J. On the relative importance of vocal source, system, and prosody in human depression. In: *2013 IEEE International Conference on Body Sensor Networks*. Cambridge, MA: IEEE (2013). p. 1–6.
19. Schuller B, Steidl S, Batliner A, Marschik PB, Baumeister H, Dong F, et al. The INTERSPEECH 2018. Computational paralinguistics challenge: atypical & self-assessed affect, crying & heart beats. In: *Proceedings of Interspeech 2018*. Hyderabad (2018). p. 122–6.
20. Schuller BW, Batliner A, Bergler C, Pokorny FB, Krajewski J, Cychosz M, et al. The INTERSPEECH 2019. Computational paralinguistics challenge: styrian dialects, continuous sleepiness, baby sounds & orca activity. In: *Proceedings of Interspeech 2019*. (2019). p. 2378–82. doi: 10.21437/Interspeech.2019-1122
21. Boersma P, Van Heuven V. Speak and unSpeak with PRAAT. *Glot International*. (2001) 5:341–347.
22. Eyben F, Weninger F, Gross F, Schuller B. Recent developments in opensmile, the munich open-source multimedia feature extractor. In: *Proceedings of the 21st ACM International Conference on Multimedia*. (2013). p. 835–8.
23. Schmitt M, Schuller B. Openxbow: introducing the passau open-source crossmodal bag-of-words toolkit. *J Mach Learn Res*. (2017) 18:3370–4. doi: 10.48550/arXiv.1605.06778
24. Amiriparian S, Gerczuk M, Ottl S, Cummins N, Freitag M, Pugachevskiy S, et al. Snore sound classification using image-based deep spectrum features. In: *Proceedings of Interspeech 2017*. (2017). p. 3512–6. doi: 10.21437/Interspeech.2017-434
25. Amiriparian S, Freitag M, Cummins N, Schuller B. Sequence to Sequence autoencoders for unsupervised representation learning from audio. In: *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop (DCASE2017)*. Tampere: Tampere University of Technology, Laboratory of Signal Processing (2017). p. 17–21.
26. Freitag M, Amiriparian S, Pugachevskiy S, Cummins N, Schuller B. audeep: unsupervised learning of representations from audio with deep recurrent neural networks. *J Mach Learn Res*. (2017) 18:6340–4. doi: 10.48550/arXiv.1712.04382
27. Cummins N, Baird A, Schuller BW. Speech analysis for health: current state-of-the-art and the increasing impact of deep learning. *Methods*. (2018) 151:41–54. doi: 10.1016/j.ymeth.2018.07.007
28. Voleti R, Liss JM, Berisha V. A review of automated speech and language features for assessment of cognitive and thought disorders. *IEEE J Select Topics Signal Process*. (2019) 14:282–98. doi: 10.1109/JSTSP.2019.2952087
29. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int J Surg*. (2021) 88:105906. doi: 10.1016/j.ijsu.2021.105906
30. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*. (2020) 11:181–217. doi: 10.1002/jrsm.1378
31. Alghowinem S, Goecke R, Wagner M, Epps J, Breakspear M, Parker G. Detecting depression: a comparison between spontaneous and read speech. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Vancouver, BC: IEEE (2013). p. 7547–51.
32. Gratch J, Artstein R, Lucas G, Stratou G, Scherer S, Nazarian A, et al. The distress analysis interview corpus of human and computer interviews. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik (2014). p. 3123–8.
33. Jati A, Williams PG, Baucom B, Georgiou P. Towards predicting physiology from speech during stressful conversations: heart rate and respiratory sinus arrhythmia. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Calgary, AB: IEEE (2018). p. 4944–8.
34. Knibb JA, Woollams AM, Hodges JR, Patterson K. Making sense of progressive non-fluent aphasia: an analysis of conversational speech. *Brain*. (2009) 132:2734–46. doi: 10.1093/brain/awp207
35. Weiner J, Angrick M, Umesh S, Schultz T. Investigating the effect of audio duration on dementia detection using acoustic features. In: *Proceedings of Interspeech 2018*. Hyderabad (2018). p. 2324–8.
36. Barnish MS, Horton SM, Butterfint ZR, Clark AB, Atkinson RA, Deane KH. Speech and communication in Parkinson's disease: a cross-sectional exploratory study in the UK. *BMJ Open*. (2017) 7:e014642. doi: 10.1136/bmjopen-2016-014642
37. Liu Z, Li C, Gao X, Wang G, Yang J. Ensemble-based depression detection in speech. In: *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Kansas City, MO: IEEE (2017). p. 975–80. doi: 10.1109/BIBM.2017.8217789

38. Lefter I, Burghouts GJ, Rothkrantz LJ. An audio-visual dataset of human-human interactions in stressful situations. *J Multimodal User Interfaces*. (2014) 8:29–41. doi: 10.1007/s12193-014-0150-7
39. Fernandez R, Picard RW. Modeling drivers' speech under stress. *Speech Commun*. (2003) 40:145–59. doi: 10.1016/S0167-6393(02)00080-8
40. Palacios-Alonso D, Lázaro-Carrascosa C, López-Arribas A, Meléndez-Morales G, Gómez-Rodellar A, Loro-Álavez A, et al. Assessing an application of spontaneous stressed speech-emotions portal. In: *International Work-Conference on the Interplay Between Natural and Artificial Computation*. Springer (2019). p. 149–60.
41. Tsanas A, Little M, McSharry P, Ramig L. Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *Nat Prec*. (2009) 57:884–93. doi: 10.1038/npre.2009.3920.1
42. Ikeno A, Varadarajan V, Patil S, Hansen JH. UT-Scope: speech under lombard effect and cognitive stress. In: *2007 IEEE Aerospace Conference*. Big Sky, MT: IEEE (2007). p. 1–7.
43. Luz S. Longitudinal monitoring and detection of Alzheimer's type dementia from spontaneous speech data. In: *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*. Thessaloniki: IEEE (2017). p. 45–6.
44. Haider F, De La Fuente S, Luz S. An assessment of paralinguistic acoustic features for detection of Alzheimer's dementia in spontaneous speech. *IEEE J Select Top Signal Process*. (2019) 14:272–81. doi: 10.1109/JSTSP.2019.2955022
45. Zhang L, Duvvuri R, Chandra KK, Nguyen T, Ghomi RH. Automated voice biomarkers for depression symptoms using an online cross-sectional data collection initiative. *Depression Anxiety*. (2020) 37:657–69. doi: 10.1002/da.23020
46. Mendiratta A, Scibelli F, Esposito AM, Capuano V, Likforman-Sulem L, Maldonato MN, et al. Automatic detection of depressive states from speech. In: *Multidisciplinary Approaches to Neural Computing*. Cham: Springer (2018). p. 301–4.
47. Rodríguez-Parra M, Adrián J, Casado J. Voice therapy used to test a basic protocol for multidimensional assessment of dysphonia. *J Voice*. (2009) 23:304–18. doi: 10.1016/j.jvoice.2007.05.001
48. Khorram S, Jaiswal M, Gideon J, McInnis M, Provost EM. The priori emotion dataset: linking mood to emotion detected in-the-wild. *arXiv[Preprint].arXiv:180610658.* (2018). doi: 10.21437/Interspeech.2018-2355
49. Maxhuni A, Mu noz-Meléndez A, Osmani V, Perez H, Mayora O, Morales EF. Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients. *Pervasive Mobile Comput*. (2016) 31:50–66. doi: 10.1016/j.pmcj.2016.01.008
50. Khan T, Westin J, Dougherty M. Classification of speech intelligibility in Parkinson's disease. *Biocybernet Biomed Eng*. (2014) 34:35–45. doi: 10.1016/j.bbe.2013.10.003
51. Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgen F, Delil S, et al. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE J Biomed Health Inform*. (2013) 17:828–34. doi: 10.1109/JBHI.2013.2245674
52. Sapir S, Ramig LO, Spielman JL, Fox C. Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech. *J Speech Lang Hear Res*. (2010) 53:114–25. doi: 10.1044/1092-4388(2009/08-0184)
53. Wang J, Kothalkar PV, Cao B, Heitzman D. Towards automatic detection of amyotrophic lateral sclerosis from speech acoustic and articulatory samples. In: *Proceedings of Interspeech 2016*. San Francisco, CA (2016). p. 1195–9.
54. Bose A, van Lieshout P, Square PA. Word frequency and bigram frequency effects on linguistic processing and speech motor performance in individuals with aphasia and normal speakers. *J Neurolinguist*. (2007) 20:65–88. doi: 10.1016/j.jneuroling.2006.05.001
55. Dubey H, Goldberg JC, Mankodiya K, Mahler L. A multi-smartwatch system for assessing speech characteristics of people with dysarthria in group settings. In: *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*. Boston, MA: IEEE (2015). p. 528–33.
56. An K, Kim MJ, Teplansky K, Green JR, Campbell TF, Yunusova Y, et al. Automatic early detection of amyotrophic lateral sclerosis from intelligible speech using convolutional neural networks. In: *Proceedings of Interspeech 2018*. Hyderabad (2018). p. 1913–7.
57. Karan B, Sahu SS, Orozco-Arroyave JR, Mahto K. Hilbert spectrum analysis for automatic detection and evaluation of Parkinson's speech. *Biomed Signal Process Control*. (2020) 61:102050. doi: 10.1016/j.bspc.2020.102050
58. Patel R. Acoustic characteristics of the question-statement contrast in severe dysarthria due to cerebral palsy. *J Speech Lang Hear Res*. (2003) 46:1401–15. doi: 10.1044/1092-4388(2003/109)
59. Galaz Z, Mekyska J, Mzourek Z, Smekal Z, Rektorova I, Eliasova I, et al. Prosodic analysis of neutral, stress-modified and rhymed speech in patients with Parkinson's disease. *Comput Methods Progr Biomed*. (2016) 127:301–17. doi: 10.1016/j.cmpb.2015.12.011
60. Orozco-Arroyave JR, Arias-Londo no JD, Vargas-Bonilla JF, Gonzalez-Rátiva MC, Nöth E. New spanish speech corpus database for the analysis of people suffering from Parkinson's disease. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik (2014). p. 342–7.
61. Baird A, Amiriparian S, Berschneider M, Schmitt M, Schuller B. Predicting biological signals from speech: introducing a novel multimodal dataset and results. In: *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*. Kuala Lumpur: IEEE (2019). p. 1–5.
62. Ho AK, Iansek R, Bradshaw JL. Motor instability in parkinsonian speech intensity. *Cogn Behav Neurol*. (2001) 14:109–16.
63. Spielman J, Ramig LO, Mahler L, Halpern A, Gavin WJ. Effects of an extended version of the lee silverman voice treatment on voice and speech in Parkinson's disease. *Am J Speech Lang Pathol*. (2007) 16:95–107. doi: 10.1044/1058-0360(2007/014)
64. Kim Y, Choi Y. A cross-language study of acoustic predictors of speech intelligibility in individuals with Parkinson's disease. *J Speech, Lang Hear Res*. (2017) 60:2506–18. doi: 10.1044/2017_JSLHR-S-16-0121
65. Garcia-Gancedo L, Kelly ML, Lavrov A, Parr J, Hart R, Marsden R, et al. Objectively monitoring amyotrophic lateral sclerosis patient symptoms during clinical trials with sensors: observational study. *JMIR mHealth uHealth*. (2019) 7:e13433. doi: 10.2196/13433
66. Chmielińska J, Białek K, Potulska-Chromik A, Jakubowski J, Majda-Zdancewicz E, Nojszewska M, et al. Multimodal data acquisition set for objective assessment of Parkinson's disease. In: *Radioelectronic Systems Conference 2019, vol. 11442. International Society for Optics Photonics*. Jachranka (2020). p. 114420F.
67. Das B, Daoudi K, Klempir J, Rusz J. Towards disease-specific speech markers for differential diagnosis in Parkinsonism. In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brighton, UK: IEEE (2019). p. 5846–50.
68. Altay EV, Alatas B. Association analysis of Parkinson disease with vocal change characteristics using multi-objective metaheuristic optimization. *Medical Hypotheses*. (2020) 141:109722. doi: 10.1016/j.mehy.2020.109722
69. Tuncer T, Dogan S, Acharya UR. Automated detection of Parkinson's disease using minimum average maximum tree and singular value decomposition method with vowels. *Biocybernet Biomed Eng*. (2020) 40:211–220. doi: 10.1016/j.bbe.2019.05.006
70. Naranjo L, Perez CJ, Campos-Roca Y, Martin J. Addressing voice recording replications for Parkinson's disease detection. *Expert Syst Appl*. (2016) 46:286–92. doi: 10.1016/j.eswa.2015.10.034
71. Smekal Z, Mekyska J, Galaz Z, Mzourek Z, Rektorova I, Faundez-Zanuy M. Analysis of phonation in patients with Parkinson's disease using empirical mode decomposition. In: *2015 International Symposium on Signals, Circuits and Systems (ISSCS)*. Iasi: IEEE (2015). p. 1–4.
72. Prince J, Andreotti F, De Vos M. Multi-source ensemble learning for the remote prediction of Parkinson's disease in the presence of source-wise missing data. *IEEE Trans Biomed Eng*. (2018) 66:1402–11. doi: 10.1109/TBME.2018.2873252
73. Little M, McSharry P, Hunter E, Spielman J, Ramig L. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Nat Prec*. (2008) 56:1015. doi: 10.1038/npre.2008.2298.1
74. Slegers A, Filiou RP, Montembeault M, Brambati SM. Connected speech features from picture description in Alzheimer's disease: a systematic review. *J Alzheimers Dis*. (2018) 65:519–42. doi: 10.3233/JAD-170881

75. Mueller KD, Hermann B, Mecollari J, Turkstra LS. Connected speech and language in mild cognitive impairment and Alzheimer's disease: a review of picture description tasks. *J Clin Exp Neuropsychol*. (2018) 40:917–39. doi: 10.1080/13803395.2018.1446513
76. Clarke N, Barrick TR, Garrard P. A comparison of connected speech tasks for detecting early Alzheimer's disease and mild cognitive impairment using natural language processing and machine learning. *Front Comput Sci*. (2021) 3:634360. doi: 10.3389/fcomp.2021.634360
77. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. *Biol Psychiatry*. (2018) 3:223–30. doi: 10.1016/j.bpsc.2017.11.007
78. El Sharkawi A, Ramig L, Logemann J, Pauloski BR, Rademaker A, Smith C, et al. Swallowing and voice effects of Lee Silverman Voice Treatment (LSVT®): a pilot study. *J Neurol Neurosurg Psychiatry*. (2002) 72:31–6. doi: 10.1136/jnnp.72.1.31
79. Sale P, Castiglioni D, De Pandis M, Torti M, Dallrmi V, Radicati F, et al. The Lee Silverman Voice Treatment (LSVT®) speech therapy in progressive supranuclear palsy. *Eur J Phys Rehabil Med*. (2015) 51:569–74.
80. MacDonald B, Jiang PP, Cattiau J, Heywood R, Cave R, Seaver K, et al. Disordered speech data collection: lessons learned at 1 million utterances from project euphonia. In: *Proceedings of Interspeech 2021*. Brno (2021). p. 4833–7. doi: 10.21437/Interspeech.2021-697
81. Schwoebel JW, Schwartz J, Warrenburg L, Brown R, Awasthi A, New A, et al. A longitudinal normative dataset and protocol for speech and language biomarker research. *medRxiv [Preprint]*. (2021). doi: 10.1101/2021.08.16.21262125
82. Robin J, Harrison JE, Kaufman LD, Rudzicz F, Simpson W, Yancheva M. Evaluation of speech-based digital biomarkers: review and recommendations. *Digital Biomarkers*. (2020) 4:99–108. doi: 10.1159/000510820
83. Stasak B, Epps J. Differential performance of automatic speech-based depression classification across smartphones. In: *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. San Antonio, TX: IEEE (2017). p. 171–5.
84. Yordanova K. Challenges providing ground truth for pervasive healthcare systems. *IEEE Pervasive Comput*. (2019) 18:100–4. doi: 10.1109/MPRV.2019.2912261
85. Ramani Haulcy JG. CLAC: a speech corpus of healthy English speakers. In: *Proceedings of Interspeech 2021*. (2021). p. 2966–70. d
86. Hecker P, Pokorny FB, Bartl-Pokorny KD, Reichel U, Ren Z, Hantke S, et al. Speaking Corona? Human and machine recognition of COVID-19 from voice. In: *Proceedings of Interspeech 2021*. Brno (2021). p. 1029–33.
87. Altuwaiyan T, Hadian M, Rubel S, Liang X. Exploiting privacy-preserving voice query in healthcare-based voice assistant system. In: *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. Dublin: IEEE (2020). p. 1–6.
88. Dojchinovski D, Ilievski A, Gusev M. Interactive home healthcare system with integrated voice assistant. In: *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. Opatija: IEEE (2019). p. 284–8.
89. Wienrich C, Reitelbach C, Carolus A. The trustworthiness of voice assistants in the context of healthcare investigating the effect of perceived expertise on the trustworthiness of voice assistants, providers, data receivers, and automatic speech recognition. *Front Comput Sci*. (2021) 53:685250. doi: 10.3389/fcomp.2021.685250
90. Fleury A, Vacher M, Noury N. SVM-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results. *IEEE Trans Inf Technol Biomed*. (2009) 14:274–83. doi: 10.1109/TITB.2009.2037317
91. Brognara L, Palumbo P, Grimm B, Palmerini L. Assessing gait in Parkinson's disease using wearable motion sensors: a systematic review. *Diseases*. (2019) 7:18. doi: 10.3390/diseases7010018
92. Valstar M, Gratch J, Schuller B, Ringeval F, Lalanne D, Torres Torres M, et al. Avec 2016: Depression, mood, and emotion recognition workshop and challenge. In: *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. Amsterdam (2016). p. 3–10.
93. Ringeval F, Schuller B, Valstar M, Gratch J, Cowie R, Scherer S, et al. Avec 2017: real-life depression, and affect recognition workshop and challenge. In: *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*. Mountain View, CA (2017). p. 3–9.
94. Ringeval F, Schuller B, Valstar M, Cummins N, Cowie R, Tavabi L, et al. AVEC 2019 workshop and challenge: state-of-mind, detecting depression with AI, and cross-cultural affect recognition. In: *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*. Nice (2019). p. 3–12.
95. Ringeval F, Schuller B, Valstar M, Cowie R, Kaya H, Schmitt M, et al. AVEC 2018 workshop and challenge: Bipolar disorder and cross-cultural affect recognition. In: *Proceedings of the 2018 on Audio/Visual Emotion Challenge and Workshop*. Seoul (2018). p. 3–13.
96. Neumann M, Roesler O, Liscombe J, Kothare H, Suendermann-Oeft D, Pautler D, et al. Investigating the utility of multimodal conversational technology and audiovisual analytic measures for the assessment and monitoring of amyotrophic lateral sclerosis at scale. *arXiv[Preprint].arXiv:210407310*. (2021). doi: 10.21437/Interspeech.2021-1801

Conflict of Interest: PH, FE, and BS were employed by the company audEERING GmbH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Digital Health

TYPE Perspective
PUBLISHED 28 November 2022
DOI 10.3389/fdgth.2022.1063141

OPEN ACCESS

EDITED BY
Qingyu Chen,
National Center for Biotechnology Information, National Library of Medicine (NIH), United States

REVIEWED BY
Tian Kang,
Tempus Labs, United States

*CORRESPONDENCE
Jorge M. Rodríguez-Fernández
jormrodr@utmb.edu

SPECIALTY SECTION
This article was submitted to Health Informatics, a section of the journal Frontiers in Digital Health

RECEIVED 06 October 2022
ACCEPTED 10 November 2022
PUBLISHED 28 November 2022

CITATION
Rodríguez-Fernández JM, Loeb JA and Hier DB (2022) It's time to change our documentation philosophy: writing better neurology notes without the burnout.
Front. Digit. Health 4:1063141.
doi: 10.3389/fdgth.2022.1063141

It's time to change our documentation philosophy: writing better neurology notes without the burnout

Jorge M. Rodríguez-Fernández[1]*, Jeffrey A. Loeb[2] and Daniel B. Hier[2,3]

[1]Department of Neurology, University of Texas Medical Branch, Galveston, TX, United States, [2]Department of Neurology and Rehabilitation, University of Illinois at Chicago, Chicago, IL, United States, [3]Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO, United States

Succinct clinical documentation is vital to effective twenty-first-century healthcare. Recent changes in outpatient and inpatient evaluation and management (E/M) guidelines have allowed neurology practices to make changes that reduce the documentation burden and enhance clinical note usability. Despite favorable changes in E/M guidelines, some neurology practices have not moved quickly to change their documentation philosophy. We argue in favor of changes in the design, structure, and implementation of clinical notes that make them shorter yet still information-rich. A move from physician-centric to team documentation can reduce work for physicians. Changing the documentation philosophy from "bigger is better" to "short but sweet" can reduce the documentation burden, streamline the writing and reading of clinical notes, and enhance their utility for medical decision-making, patient education, medical education, and clinical research. We believe that these changes can favorably affect physician well-being without adversely affecting reimbursement.

KEYWORDS

electronic health records, documentation burden, clinician well-being, evaluation and management coding, medical decision-making

Introduction

A crisis in physician well-being and the mounting burden of clinical documentation drives the need for neurologists to change their documentation philosophy. Burnout is prevalent among all healthcare professionals (1) and neurologists in particular (2, 3). Documentation is a known contributor to burnout (1, 4, 5). An estimated 40% of physician time in the electronic health record (EHR) is devoted to documentation (6). In 2015, the American College of Physicians (7) emphasized the importance of writing "concise, history-rich notes" that prioritize information relevant to medical decision-making (MDM). Responding to physician complaints about the burden of EHR documentation, the AMA Current Procedural Terminology Editorial Panel recommended changes to evaluation and management (E/M) documentation guidelines that would reduce the clerical burden (8). On January 1, 2021, the Center

for Medicare and Medicaid Services (CMS) updated E/M coding guidelines for outpatient visits that would allow physicians to select a billing level based on the time or complexity of MDM and would reduce documentation requirements for the history and physical examination (including the unpopular bullet point requirements). These requirements were seen as both tedious and time-consuming. Furthermore, they were seen as reducing the time available for direct patient care (6). CMS plans to extend these changes to the inpatient setting (hospital inpatient, hospital observation, emergency department, and cognitive impairment assessment) by January 1, 2023 (9).

Time-based coding allows physicians to bill for time spent outside the patient visit on the day of evaluation, even if the patient is absent (chart review, documentation, orders, coordinating care, etc.) The complexity of medical decision-making (the main driver of billing level) is determined by the number and complexity of the problems addressed, the risk of complications (morbidity and mortality), and the complexity of the data reviewed. These changes simplify setting the correct level of service for billing purposes. For example, a patient with migraine headaches who is started on single-drug therapy has a low level of MDM; a patient with new-onset seizures who needs adjustments of anti-epileptic medications and imaging has a moderate level of MDM; and a patient with epilepsy and poorly-controlled seizures, a structural brain lesion, and who needs neuroimaging, extended electroencephalographic testing, and a surgical consultation has a high-level of MDM.

Initial studies suggest that these CMS changes will enhance reimbursement for Evaluation and Management (E/M) services (10) provided by non-surgical specialties but not by surgical specialties (11). Improvements have not yet been noted in time spent documenting in the EHR (10). Furthermore, to date, no study has determined that these changes in documentation guidelines have lightened provider documentation burden.

Why are our notes so bulky?

In the US, the volume of clinical documentation has increased over the past two decades. It is estimated that US clinicians do three times more documentation than clinicians in other medically advanced countries (12). The are several reasons for the bulkiness of our clinical notes in the US.

- We have been reimbursed more for bulkier notes. Since 1995, the level of service and reimbursement was linked to the number of coding elements (also known as "bullet points") documented (9).
- It's too easy to add bulk. Copy and paste functionality in electronic health records makes it easy to add bulk. Hyperlinks make it possible to add laboratory results, radiology findings, problem lists, medication lists, and other chart elements to a note with a single mouse click. One study of over 26,000 physician notes found that only 18% of the text was entered by the physician, 46% was copied from elsewhere in the EHR, and 35% was imported from other sections of the EHR (13). Copy and paste may document care that was never rendered or examinations that were never performed (14).
- We are trained to document negative findings. As Sinsky has observed, we need to move away from the dictum that "If it wasn't documented, it wasn't done" (15). Or, as Postal has wondered, do we have the time and energy to document all the negatives or should we stick to the salient positives (16)?
- We continue to document based on reimbursement and tradition rather than scientific evidence. In their essay subtitled "A farewell to the review of systems," Barry and Tseng argue for deleting the traditional review of systems and that documentation should be based on scientific evidence (17).

It is time to change our documentation philosophy

A change in documentation philosophy is needed to slim down bloated notes that are hard to read, hard to write, and often inaccurate (**Table 1**). Copying and pasting text from one note into another fosters bloat, redundancies, and inaccuracies. When long pre-completed templates are used to document the neurological exam, parts of the neurological examination may be documented as normal when these parts were not

TABLE 1 Suggested changes in documentation philosophy.

Old philosophy	New philosophy
Bigger is better	Less is more
	Short but sweet[a]
If it is not documented, it did not happen	Avoid excessive documentation of normal findings
Document all pertinent negative findings	Focus on abnormal findings
Import labs, radiology, allergies, medications, family history, and social history into every note	Maintain histories in one up to date central location
Longer notes with more bullet points are reimbursed at higher levels	Reimbursement is focused on medical decision-making
Documentation is the responsibility of the physician	Documentation is a team responsibility
Each physician is free to document as they please	Let's agree on a uniform approach to documentation
Patients will not read our notes	Patients can benefit from reading our notes
Notes are for patient care	Notes can be used for research and patient care

[a]Especially in neurology where the history of the event and quantitation of symptoms is often more important to make the diagnosis than MRI imaging and other testing. A detailed, chronological description of symptoms is critical to diagnostic accuracy in neurology.

examined. These documentation practices open the door to litigation. When test results (imaging, electroencephalography, electromyographic, etc.) are added to clinical notes, they should be addressed, discussed, and made relevant to the MDM.

Documentation needs to be clear, accurate, and concise. The emphasis should be on optimal patient care, not maximizing the billable level of care. Neurology departments need to discourage clinicians from the redundant documentation of information. We argue that certain types of medical data (e.g., social history, past medical history, allergies, surgical history, medication lists, laboratory results, and radiology results) are best housed in a central location in the EHR and should not be added to a note unless relevant to MDM. Clinicians need to be trained to document allergies, medications, past medical history, and social history in the appropriate place in the EHR and not redundantly in each note. This practice has multiple advantages: it allows patient care team members to share the work of documentation, it reduces duplicate work, and it reduces multiple and inconsistent versions of the same data.

We can build consistency and accuracy of documentation through standardized notes with quantitative longitudinal measures

Notes that have a consistent structure across the organization make notes easier to read and more predictable. Having a single consolidated note template for each department or subspeciality facilitates note maintenance. Detailed documentation of the history or the condition needed by a subspecialist can be collapsed in the EHR or linked elsewhere to not overwhelm the generalist user. In addition to text, developing innovative ways to visualize changes quantitatively in symptoms over time and their response to treatments can enhance the clinician's perspective on the disease course. An additional benefit of a single template is the option of providing organizational updates or reminders to all template users.

The general SOAP (subjective-objective-assessment-plan) format has been well accepted. With the growing emphasis on MDM, the APSO (assessment-and-subjective-objective) format has grown in popularity. We additionally recommend the following:

- Create an institutional culture that values concise information-rich notes.
- Encourage clinicians to use collapsible sections in their notes to prioritize which sections are visible.
- Standardize the adoption of the SOAP or APSO note format at the organizational level.
- Encourage providers to document pertinent negative and positive findings through direct entry into the EHR rather than by template or copy and paste.
- Discourage providers from using hyperlinks in the EHR to add laboratory and ancillary testing results to notes with unnecessary redundancy when not relevant to MDM.
- Encourage providers to focus on adding those findings to notes that are pertinent to medical decision-making.
- Implement "vanishing text" that allows clinicians to view findings in their notes to support note creation and have it deleted when the note is finalized.
- Discourage the use of pre-completed examination templates with all findings marked as "normal."
- If radiology or other reports are incorporated into a note, encourage the insertion of the "Impression" paragraph only.
- Look for help from the facility informatics department to create space-saving ways to represent bulky laboratory results as "fish bones" and other laboratory diagrams.
- Use hyperlinks, rather than text insertion, to connect notes to discrete data such as advanced directives or resuscitation status.
- Develop and implement policies that control copy and paste functionality, including highlighting of text that has been pasted into the note (14, 18).
- Use EHR metrics such as *note length*, *time in chart*, etc., to track changes in documentation practices by clinicians.

Let's engage providers and leadership in positive changes

The implementation of these recommendations depends upon proper organizational support. The engagement of key stakeholders, including departmental leadership, compliance officers, billing, coders, clinicians, and clinical informaticians, is critical. Principles and objectives for documentation change must be developed, agreed upon, and implemented. Documentation metrics are crucial to evaluating project success. Key documentation metrics include clinician time in the EHR, clinician time spent documenting outside of regular work hours, and time spent writing notes (19). We additionally suggest tracking mean note length over time. These metrics can demonstrate project success tangibly to leadership and clinicians.

The use of sprints or PDSA (plan-do-study-act) cycles can address implementation barriers before and after project roll-out. Departmental support is critical to assisting physicians in adjusting to new documentation methods. Iterative sprints and cycles are recommended to foster change. Although some documentation changes are driven by CMS, other regulatory agencies might have specific documentation requirements that require compliance, such as quality measures related to stroke (20). Interdisciplinary teams tasked with documentation change must address documentation compliance issues for each sub-specialty.

Let's get ready for OpenNotes

OpenNotes is coming. Federal legislation under the 21st Century Cures Act provides patients access to notes without delay and charge by April 5, 2021 (21). The OpenNotes initiative seeks to provide patients with access to their medical records (22) to improve their understanding of their condition and give patients more control of their treatment plan. Open communication with the patient during the visit combined with succinct understandable notes supports co-ownership of medical problems by patient and clinician (23–28). Long-term outcomes from the OpenNotes initiative on patient satisfaction and patient condition are still being evaluated.

As neurologists, we must recognize that patients will be reading our notes. For some sensitive conditions, this may be problematic. For example, in neurology, we often evaluate patients with functional disorders who have a limited understanding of the causes and nature of their condition. Neurologists must be open and should provide transparent communication with the patient at the time of evaluation and while creating their notes. Patients with functional disorders may find terms such as "non-physiological" or "no neurological correlate" confusing. These patients deserve a clear explanation of their symptoms in the office and in our notes (29, 30) Lastly, concerns on disclosure and result interpretation of sensitive testing for diseases such as Alzheimer's disease have been raised, and efforts to develop ethical and patient-centered policies for disclosure are needed (31).

Let's embrace team-based documentation

The goal of team-based documentation is to offload some of the documentation work from the physician to other team members. Team-based documentation may variably involve scribes, nurses, pharmacists, medical assistants, or artificial intelligence-based dictation devices (32). Some team-based care models distribute specific documentation tasks such as recording allergies, documenting past medical history, and reconciling medication to specific team members (33). When these tasks are done before the initial interaction between patient and clinician, clinician time and effort are conserved. Other team-based documentation models use scribes to free up clinician time at the point of patient contact (32).

We can flex our documentation to support education

Clinical documentation is central to the education of medical students, residents, and fellows (34). Although uniformity in documentation templates and methods is a stated goal, it is important to flex documentation expectations according to the level of training. Although it is reasonable to expect a medical student to document their neurological examination and history in great detail, the same is not true for an advanced fellow or experienced attending neurologist. While longer notes are de rigeur for medical students, we expect experienced clinicians to write concise notes with few notations about normal findings. Still, we should encourage medical students to focus on succinct formulations of the neurological examination and history. Academic institutions can take advantage of current E/M guideline changes to encourage trainees to document concisely, to prioritize MDM, and to avoid adding uninformative "bullet points" (35, 36). Documentation metrics can guide trainees and their mentors to adopt the best documentation strategies.

Clinical notes can support research

Although the primary purpose of physician notes in the EHR is to document care rendered, to support the billing for services provided, and to serve as a medical-legal record; electronic health records and free text physician notes have shown great potential for clinical research (37–39). For example, EHRs are being used to track disease severity, and progression in cohorts of patients with amyotrophic lateral sclerosis and multiple sclerosis (40–42). Natural language processing and other artificial intelligence algorithms are unlocking latent value in EHRs. Unlike skimpy information-poor notes or bloated information-poor notes, concise information-rich clinical notes can be of great value for clinical research.

Conclusion

For over two decades, our clinical notes have grown too long. They are a significant burden and contribute to physician burnout. This Perspective describes recommendations to simplify documentation that can be implemented because of changes in CMS guidelines for evaluation and management coding and billing. We argue that it is time to rethink our documentation to enhance communication, improve patient care, and reduce physician burnout. Although more work is needed to find optimal strategies to reduce the documentation burden, it is not too early to start creating notes that are easy to write and read yet are still information-rich. Bulky notes waste the time of the writer and the reader alike. Evidence is growing that when concerted efforts are made to simplify documentation, physician satisfaction with the EHR improves, and burnout is reduced (43–45).

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

Author contributions

Initial concept by JMRF and DBH. Concept elaboration by DBH, JAL, and JMRF. Initial draft by JMRF. Revisions and rewriting by JMRF, JAL, and DBH. All authors contributed to the article and approved the submitted version.

Funding

JAL was supported in part by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, through Grant Award Number UL1TR002003 (JAL). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al. Relationship between clerical burden, characteristics of the electronic environment with physician burnout, professional satisfaction. *Mayo Clin Proc.* (2016) 91:836–48. doi: 10.1016/j.mayocp.2016.05.007

2. Busis NA, Shanafelt TD, Keran CM, Levin KH, Schwarz HB, Molano JR, et al. Burnout, career satisfaction, and well-being among us neurologists in 2016. *Neurology.* (2017) 88:797–808. doi: 10.1212/WNL.0000000000003640

3. Levin KH, Shanafelt TD, Keran CM, Busis NA, Foster LA, Molano JRV, et al. Burnout, career satisfaction, and well-being among us neurology residents and fellows in 2016. *Neurology.* (2017) 89:492–501. doi: 10.1212/WNL.0000000000004135

4. Peccoralo LA, Kaplan CA, Pietrzak RH, Charney DS, Ripp JA. The impact of time spent on the electronic health record after work and of clerical work on burnout among clinical faculty. *J Am Med Inform Assoc.* (2021) 28:938–47. doi: 10.1093/jamia/ocaa349

5. Tajirian T, Stergiopoulos V, Strudwick G, Sequeira L, Sanches M, Kemp J, et al. The influence of electronic health record use on physician burnout: cross-sectional survey. *J Med Internet Res.* (2020) 22:e19274. doi: 10.2196/19274

6. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med.* (2016) 165:753–60. doi: 10.7326/M16-0961

7. Kuhn T, Basch P, Barr M, Yackel T, Medical Informatics Committee of the American College of Physicians. Clinical documentation in the 21st century: executive summary of a policy position paper from the American College of Physicians. *Ann Intern Med.* (2015) 162:301–3. doi: 10.7326/M14-2128

8. American Medical Association and others. CPT® evaluation and management (E/M) office or other outpatient (99202–99215) and prolonged services (99354: 99355, 99356, 99417) code and guideline changes (2021).

9. Centers for Medicare and Medicaid Services. Calendar year (CY) 2023 medicare physician fee schedule proposed rule (2022). Available at https://www.ama-assn.org/practice-management/sustainability/joy-medicine-health-system-recognition-program.

10. Apathy NC, Hare AJ, Fendrich S, Cross DA. Early changes in billing and notes after evaluation and management guideline change. *Ann Intern Med.* (2022) 175:499–504. doi: 10.7326/M21-4402

11. Francis DL, Cruddas BM. 2021 E/M code changes: forecasted impacts to gastroenterology practices. *Clin Gastroenterol Hepatol.* (2021) 19:2002–5. doi: 10.1016/j.cgh.2021.07.008

12. Downing NL, Bates DW, Longhurst CA. Physician burnout in the electronic health record era: are we ignoring the real cause? *Ann Intern Med.* (2018) 169:50–1. doi: 10.7326/M18-0139

13. Wang MD, Khanna R, Najafi N. Characterizing the source of text in electronic health record progress notes. *JAMA Intern Med.* (2017) 177:1212–3. doi: 10.1001/jamainternmed.2017.1548

14. Turchin A, Goldberg SI, Breydo E, Shubina M, Einbinder JS. Copy/paste documentation of lifestyle counseling and glycemic control in patients with diabetes: true to form? *Arch Intern Med.* (2011) 171:1393–400. doi: 10.1001/archinternmed.2011.219

15. Berg S. 3 ways to begin to reduce clinical documentation by 75% by 2025. Available at https://www.ama-assn.org/practice-management/sustainability/3-ways-begin-reduce-clinical-documentation-75-2025 (Accessed September 21, 2022).

16. Postal E. The impertinence of pertinent negatives. Available at https://www.diagnosticimaging.com/view/impertinence-pertinent-negatives (Accessed September 21, 2022).

17. Barry MJ, Tseng CW. Moving to more evidence-based primary care encounters: a farewell to the review of systems. *JAMA* (2022) 328:1495–6. doi: 10.1001/jama.2022.18346

18. Tsou AY, Lehmann CU, Michel J, Solomon R, Possanza L, Gandhi T. Safe practices for copy, paste in the EHR. *Appl Clin Inform.* (2017) 26:12–34. doi: 10.4338/ACI-2016-09-R-0150

19. Melnick ER, Ong SY, Fong A, Socrates V, Ratwani RM, Nath B, et al. Characterizing physician EHR use with vendor derived data: a feasibility study and cross-sectional analysis. *J Am Med Inform Assoc.* (2021) 28:1383–92. doi: 10.1093/jamia/ocab011

20. The Joint Commission. Stroke. Available at https://www.jointcommission.org/measurement/measures/stroke/ (Accessed September 21, 2022).

21. Office of the National Coordinator for Health Information Technology. The ONC Cures act final rule (2020). Available at https://www.healthit.gov/sites/default/files/page2/2020-03/TheONCCuresActFinalRule.pdf.

22. Arvisais-Anhalt S, Lau M, Lehmann CU, Holmgren AJ, Medford RJ, Ramirez CM, et al. The 21st century cures act, multiuser electronic health record access: potential pitfalls of information release. *J Med Internet Res.* (2022) 24:e34085. doi: 10.2196/34085

23. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. Opennotes after 7 years: patient experiences with ongoing access to their

clinicians' outpatient visit notes. *J Med Internet Res*. (2019) 21:e13876. doi: 10.2196/13876

24. DesRoches CM, Leveille S, Bell SK, Dong ZJ, Elmore JG, Fernandez L, et al. The views and experiences of clinicians sharing medical record notes with patients. *JAMA Netw Open*. (2020) 3:e201753. doi: 10.1001/jamanetworkopen.2020.1753

25. Blease C, Torous J, Hägglund M. Does patient access to clinical notes change documentation? *Front Public Health*. (2020) 8:577896. doi: 10.3389/fpubh.2020.577896

26. Klein JW, Jackson SL, Bell SK, Anselmo MK, Walker J, Delbanco T, et al. Your patient is now reading your note: opportunities, problems, and prospects. *Am J Med*. (2016) 129:1018–21. doi: 10.1016/j.amjmed.2016.05.015

27. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. *Ann Intern Med*. (2012) 157:461–70. doi: 10.7326/0003-4819-157-7-201210020-00002

28. DesRoches CM, Bell SK, Dong Z, Elmore J, Fernandez L, Fitzgerald P, et al. Patients managing medications and reading their visit notes: a survey of opennotes participants. *Ann Intern Med*. (2019) 171:69–71. doi: 10.7326/M18-3197

29. Keatley E, Molton I. A shift in approach: assessment and treatment of adults with functional neurological disorder. *J Health Serv Psychol*. (2022) 48:79–87. doi: 10.1007/s42843-022-00061-w

30. Stone J, Burton C, Carson A. Recognising and explaining functional neurological disorder. *BMJ*. (2020) 371:m3745. doi: 10.1136/bmj.m3745

31. Gale SA, Heidebrink J, Grill J, Graff-Radford J, Jicha GA, Menard W, et al. Preclinical alzheimer disease and the electronic health record: balancing confidentiality and care. *Neurology* (2022). doi: 10.1212/WNL.0000000000201347. https://n.neurology.org/content/neurology/early/2022/09/30/WNL.0000000000201347.full.pdf

32. Lin S. The present and future of team documentation: the role of patients, families, and artificial intelligence. *Mayo Clin Proc*. (2020) 95:852–5. doi: 10.1016/j.mayocp.2020.01.034

33. Hopkins K, Sinsky CA. Team-based care: saving time and improving efficiency. *Fam Pract Manag*. (2014) 21:23–9. Available at: https://pubmed.ncbi.nlm.nih.gov/25403048/

34. Wei M, Salgado E, Girard CE, Santoro JD, Lepore N. Your note, your way: how to write an inpatient progress note accurately and efficiently as an intern. *Postgrad Med J*. (2022). Available at: https://pmj.bmj.com/content/early/2022/07/07/postgradmedj-2022-141834.info

35. Oxentenko AS, West CP, Popkave C, Weinberger SE, Kolars JC. Time spent on clinical documentation: a survey of internal medicine residents and program directors. *Arch Intern Med*. (2010) 170:377–80. doi: 10.1001/archinternmed.2009.534

36. Varacallo MA, Wolf M, Herman MJ. Improving orthopedic resident knowledge of documentation, coding, and medicare fraud. *J Surg Educ*. (2017) 74:794–8. doi: 10.1016/j.jsurg.2017.02.003

37. Alzoubi H, Alzubi R, Ramzan N, West D, Al-Hadhrami T, Alazab M. A review of automatic phenotyping approaches using electronic health records. *Electronics*. (2019) 8:1235. doi: 10.3390/electronics8111235

38. Fu S, Chen D, He H, Liu S, Moon S, Peterson KJ, et al. Clinical concept extraction: a methodology review. *J Biomed Inform*. (2020) 109:103526. doi: 10.1016/j.jbi.2020.103526

39. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform*. (2017) 22:1589–604. doi: 10.1109/JBHI.2017.2767063

40. Yang Z, Pou-Prom C, Jones A, Banning M, Dai D, Mamdani M, et al. Assessment of natural language processing methods for ascertaining the expanded disability status scale score from the electronic health records of patients with multiple sclerosis: algorithm development and validation study. *JMIR Med Inform*. (2022) 10:e25157. doi: 10.2196/25157

41. Xia Z, Secor E, Chibnik LB, Bove RM, Cheng S, Chitnis T, et al. Modeling disease severity in multiple sclerosis using electronic health records. *PLoS ONE*. (2013) 8:e78927. doi: 10.1371/journal.pone.0078927

42. Karanevich AG, Weisbrod LJ, Jawdat O, Barohn RJ, Gajewski BJ, He J, et al. Using automated electronic medical record data extraction to model ALS survival and progression. *BMC Neurol*. (2018) 18:1–7. doi: 10.1186/s12883-018-1208-z

43. Sieja A, Markley K, Pell J, Gonzalez C, Redig B, Kneeland P, et al. Optimization sprints: improving clinician satisfaction and teamwork by rapidly reducing electronic health record burden. *Mayo Clin Proc*. (2019) 94:793–802. doi: 10.1016/j.mayocp.2018.08.036

44. Nguyen OT, Jenkins NJ, Khanna N, Shah S, Gartland AJ, Turner K, et al. A systematic review of contributing factors of and solutions to electronic health record–related impacts on physician well-being. *J Am Med Inform Assoc*. (2021) 28:974–84. doi: 10.1093/jamia/ocaa339

45. Eschenroeder Jr H, Manzione LC, Adler-Milstein J, Bice C, Cash R, Duda C, et al. Associations of physician burnout with organizational electronic health record support and after-hours charting. *J Am Med Inform Assoc*. (2021) 28:960–6. doi: 10.1093/jamia/ocab053

frontiers | Frontiers in Digital Health

TYPE Policy and Practice Reviews
PUBLISHED 30 November 2022
DOI 10.3389/fdgth.2022.1025086

OPEN ACCESS

EDITED BY
Premananda Indic,
University of Texas at Tyler, United States

REVIEWED BY
Peter Rehor,
Stenberg College, Canada
Joshua Chang,
The University of Texas at Austin, United States
Manisha Mantri,
Center for Development of Advanced Computing (C-DAC), India

*CORRESPONDENCE
Petr Brůha
pbruha@kiv.zcu.cz
Roman Mouček
moucek@kiv.zcu.cz

SPECIALTY SECTION
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

RECEIVED 22 August 2022
ACCEPTED 01 November 2022
PUBLISHED 30 November 2022

CITATION
Brůha P, Mouček R, Salamon J and Vacek V (2022) Workflow for health-related and brain data lifecycle.
Front. Digit. Health 4:1025086.
doi: 10.3389/fdgth.2022.1025086

Workflow for health-related and brain data lifecycle

Petr Brůha[1]*, Roman Mouček[1,2]*, Jaromír Salamon[1] and Vítězslav Vacek[1]

[1]Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Pilsen, Czech Republic, [2]New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Pilsen, Czech Republic

Poor lifestyle leads potentially to chronic diseases and low-grade physical and mental fitness. However, ahead of time, we can measure and analyze multiple aspects of physical and mental health, such as body parameters, health risk factors, degrees of motivation, and the overall willingness to change the current lifestyle. In conjunction with data representing human brain activity, we can obtain and identify human health problems resulting from a long-term lifestyle more precisely and, where appropriate, improve the quality and length of human life. Currently, brain and physical health-related data are not commonly collected and evaluated together. However, doing that is supposed to be an interesting and viable concept, especially when followed by a more detailed definition and description of their whole processing lifecycle. Moreover, when best practices are used to store, annotate, analyze, and evaluate such data collections, the necessary infrastructure development and more intense cooperation among scientific teams and laboratories are facilitated. This approach also improves the reproducibility of experimental work. As a result, large collections of physical and brain health-related data could provide a robust basis for better interpretation of a person's overall health. This work aims to overview and reflect some best practices used within global communities to ensure the reproducibility of experiments, collected datasets and related workflows. These best practices concern, e.g., data lifecycle models, FAIR principles, and definitions and implementations of terminologies and ontologies. Then, an example of how an automated workflow system could be created to support the collection, annotation, storage, analysis, and publication of findings is shown. The Body in Numbers pilot system, also utilizing software engineering best practices, was developed to implement the concept of such an automated workflow system. It is unique just due to the combination of the processing and evaluation of physical and brain (electrophysiological) data. Its implementation is explored in greater detail, and opportunities to use the gained findings and results throughout various application domains are discussed.

KEYWORDS

best practices, brain data, data lifecycle, health information system, health-related data, physical data, workflow, ontology

1. Introduction

Poor lifestyle leads potentially to chronic diseases and deteriorating physical and mental fitness. To overcome (at least partly) these troubles during aging, prior collection and evaluation of various health-related data accompanied by health interventions for those interested in them could help this unpleasant situation, which mainly affects developed societies. We can measure and analyze multiple aspects of physical and mental health, such as body parameters, health risk factors, degrees of motivation, and the overall willingness to change the current lifestyle in advance. In conjunction with data representing human brain activity, we can obtain and identify human health problems resulting from a long-term lifestyle more precisely and, where appropriate, improve the quality and length of human life.

However, the possibility of interpreting various health-related data and providing subsequent reasonable health interventions means first defining and collecting a large amount of various health-related data that can be processed automatically. It is impossible without using the results of standardization efforts and best practices applied across various domains of health-related data. These efforts and practices significantly impact the entire data collection, storage, processing, and interpretation lifecycle. As a result, these (infrastructure-related) issues need to be addressed, presented, and discussed in the scientific communities so that the experimental work is better reproducible and the data collected can be better analyzed across domains and scales. As technical solutions (technical means to collect, organize, store, annotate, and analyze data) are becoming less of a barrier, and it depends increasingly on knowledge and acceptance of partial solutions, existing standards, and best practices in different domains, this article offers a synthesis of some existing approaches to contribute to the interpretability of collected data and their actual use for communication and possible timely preventive adjustment of the lifestyle of individuals. It is done by providing an overview of some current "standards" and best practices and their integration into a proposed solution.

Health-related data accompanied by metadata are inherently heterogeneous; they are organized and stored in various structures, formats, and data repositories. Related metadata contains various written points ranging from precise data descriptions to only stated basic information based on experimenters' requirements and task circumstances. Also, metadata can be structured differently and stored in various formats, making the processing or recreating similar experiments somewhat tedious. As a result, retrieving the knowledge from these kinds of data is quite challenging (1). However, it is still no exception that metadata is written down on paper as notes without any used standards.

Recently, the popularity of Cyber-Physical Systems (CPSs) has been on the rise. The thought that wearables, small electronic devices (a fitness armband like FitBit is a good example of this) attached to the surface of the skin, collecting large quantities of medical data (with a sufficient degree of data quality and precision) and enhancing the lifestyle of a person, can be used on a day-to-day basis was adopted by many people. These CPSs can collect large amounts of health-related data. However, each of these CPS devices collects the data in various (generally self-made) data formats, for example, via connection to a Smartphone of the user (2).

Sharing collected data in a thoroughly described fashion is mainly left to the experimenter's best knowledge; it is up to the experimenter to assess how thoroughly or well-defined the objectives should be. The need to know how thoroughly the collected data should be described for a different party to reproduce the results is often to an open interpretation, which generally leads to different results. Some standards and conventions that apply to health-related data can also be applied to brain data. In our studies and this paper, we focus on the electrical activity of the human brain, i.e., on electroencephalography (EEG) and event-related potential (ERP) data. To bring some widely available standards into this domain, organizations like INCF have proposed how neuroscience data could be collected and stored, so they could be easily accessed and shared across the community (3).

This document emphasizes the best practices regarding the data lifecycle process, i.e., the collection, annotation, analysis, interpretation, and publication of data/results and offers them to a broad scientific audience. Our suggestions will cover the subjects ranging from the original experiment, data collection, storage, and description to processes on how to best store and publish the results. The benefit of a wider audience taking a look at one's raw data and findings might lead to a healthy debate about the achieved goals (highlighting errors or discovering new findings in the already collected data), as highlighted in (4). This was, for example, emphasized in win-win data sharing in neuroscience (5); there can be a lot of hidden benefits to being discovered when leading a procreative discussion of results. The data need to be stored to be understandable and easy to interpret to make the discussion as frictionless as possible. The general rules of practical data sharing that can be applied to either neuroscientific or physical health data were also mentioned in (6).

Inside the growing field, such as neuroscience, giving such "order" to the collected data is mostly used through the use of a dynamic "ever-evolving" ontology for the current subject (7). These ontologies precisely define the used terms inside the application domain, which again help in easier understanding and reuse of the once-collected data with new research goals. The dynamic ontology will help in this regard that the defined terminology may be used across the scope of multiple subjects and help thus to answer a variety of questions (8). Also, for a truly dynamic ontology, it is necessary to ensure how the changes will be propagated or added in the already existing

whole (9). There were already proposed multiple ways how it can be done, for example, through the usage of dynamic web ontology language (dOWL) (10).

In this document, we would like to show and help visualize our best practices focusing on all aspects of creating an experiment and help with the definition and categorization of results/findings through the use of widely used knowledge models (in this specific case, through a dynamic ontology) and publication of the results in an easy-to-understand way. Finally, various research groups worldwide can either discuss these findings or reuse these conclusions for their own specific research without the need to reinvent the wheel.

Since most of these steps seem too abstract, we would like to show a possible way on how such a data lifecycle might look, together with practical examples and underlying data published in widely accessible journals that followed these above-mentioned best practices. In this paper, we will cover the subjects ranging from experiment design, collection of generally heterogeneous data (e.g., heart rate, glucose, body proportions, physical strength with electroencephalography data, and many more), and the description of collected data (for example, by using ontologies) to the publication of both the findings and underlying raw data for a further verification/analysis done by the broad scientific audience.

2. Materials and methods

In this chapter, we will focus on showing the current state of the art and the technological background that was utilized during the conceptualization and development of the module architecture utilized by an information system for health-related data collection.

2.1. State of the art

The lifecycle (Section **2.1.1**) of any entity (such as software or health-related data) should follow key principles (Section **2.1.2**). We recognize the functional aspects (processing) and data (objects, subjects) stepping into the process. The descriptions of data and their organizations are various, and we prefer terminologies (Section **2.1.3**) and ontologies (Section **2.1.3**) when it comes to health-related data. The data properly identified and described are stored in the standardized and interchangeable data format (Section **2.1.4**).

2.1.1. Software engineering methodologies and data management lifecycles

The organization of software development and data processing as critical activities to achieve work effectiveness and efficiency has led to defining development methodologies and software/data lifecycles. These methodologies and lifecycles also create a primary platform to achieve another challenge—open, fair, and reproducible science.

Agile development methodologies have followed waterfall software development methodologies (11). At the same time as the agile methodologies, the era of big data began. It took significant importance in the last decade when cheap data storage and computational power increased exponentially. Agile software development has evolved into a complementary set of practices called DevOps (12–14), where software development (Dev, Software Engineering), IT (technology) operations (Ops), and quality assurance (QA) are present (**Figure 1**, left part). Processwise, DevOps represents a typical chain for delivering software solutions; it includes software development, building, testing, deployment, and running (**Figure 2**, upper part).

However, the DevOps does not correspond to the specific needs of big data; thus, DataOps (15) has been introduced. DataOps includes other data-driven disciplines like data engineering, data integration, data security, and data quality (**Figure 1**, right part). It represents a complete data lifecycle from data preparation and gathering over the transformation to reporting. It brings a bridge between data analytics teams and IT operations. DataOps focuses highly on data pipeline orchestration, data quality, and continuous integration/delivery. It provides the chance to get a consistent and reliable source for data ingestion and reporting and advanced analytics represented by machine learning (ML) models and artificial intelligence (AI) solutions. Next to the processwise qualities, DataOps provides capabilities about data lifecycle, data annotation (relations, the meaning given by ontology, versioning), and data lineage (auditability, explainability).

Processwise, DataOps extends the processing chain to focus more on the data-related products instead of being software-centric. It adds sandbox management for implementing data prototype products, replaces the build process, runs with orchestration, and adds monitoring at the end (**Figure 2**, lower part).

For machine learning, DataOps moved even further and evolved into the MLOps (18) lifecycle, which covers specific needs of data science. It represents the practice of collaboration and communication among data scientists and operations professionals to help manage the production ML (or deep learning) lifecycle. The movement of DataOps to MLOps and later to AIOps (19) for artificial intelligence was natural since there was technical debt.

Tom et al. (20) explain the technical debt as follows: "Technical debt is a metaphor that refers to the consequences of poor software development. Cunningham (1992), who introduced the concept of technical debt, described how 'shipping first time code is like going into debt. A little debt speeds development so long as it is paid back promptly with a rewrite.' Since then, the suitability of debt as a way of explaining the various drivers of increasing costs throughout the life of a software system has been affirmed by the software development community (21–25). On the other hand, debt is

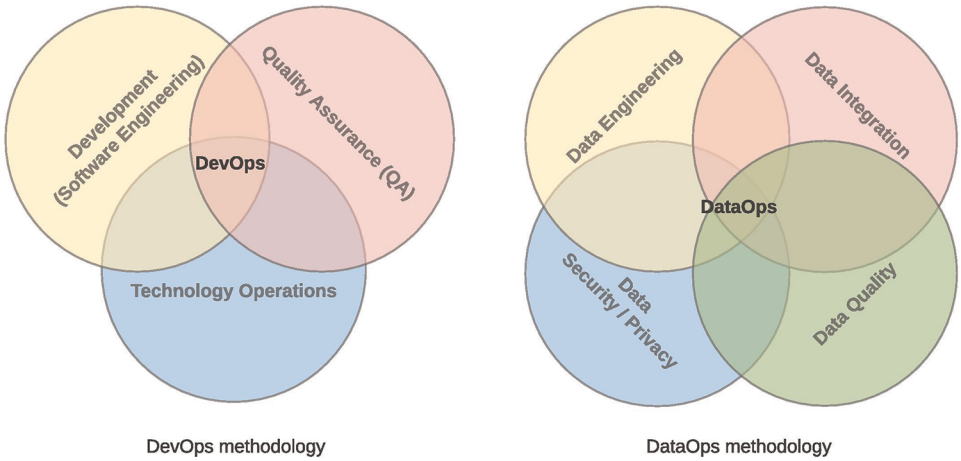


FIGURE 1
DevOps (16) and DataOps in the enterprise (16).

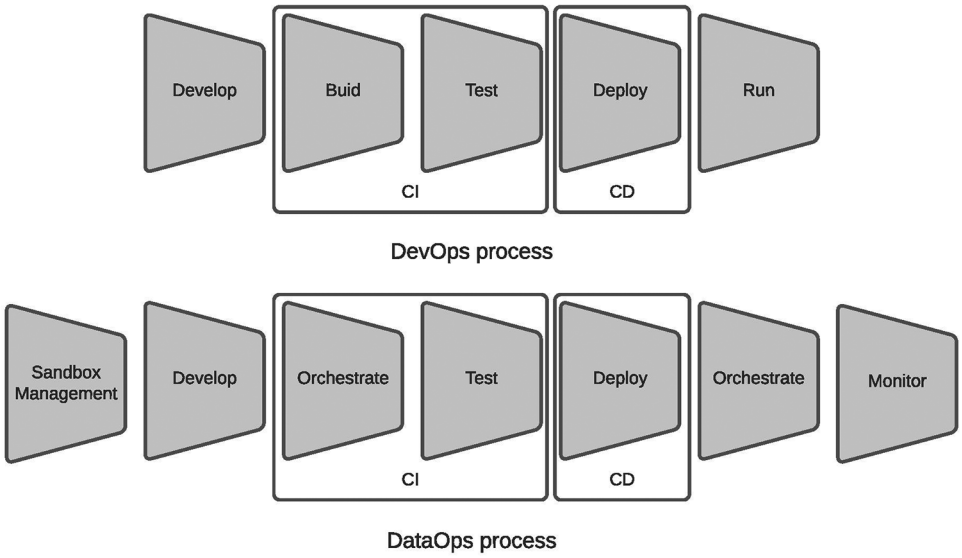


FIGURE 2
DevOps and DataOps processes (17).

not necessarily 'bad'—a small level of debt can help developers speed up the development process in the short term (21). In any case, this consequence may be felt in the longer term if the project is highly 'geared' (which implies onerous debt repayments), leading to slower development and killing of productivity."

Science has been evolving to be more open, fair, and reproducible (Section **2.2.4**) in the last years. Data are published on various platforms and processed in on-premise, private, or public cloud storage and services. The importance of sharing data across scientific fields has been raised.

The technical debt can thus also be considered for research. The research systems should provide functionality like data preprocessing and sharing, analytical tools, reporting tools, and a complex methodology and ecosystem that consider all those steps part of a unified lifecycle. Then, ResearchOps[1] "provides the roles, tools and processes needed to support

[1]Available at: https://researchops.community.

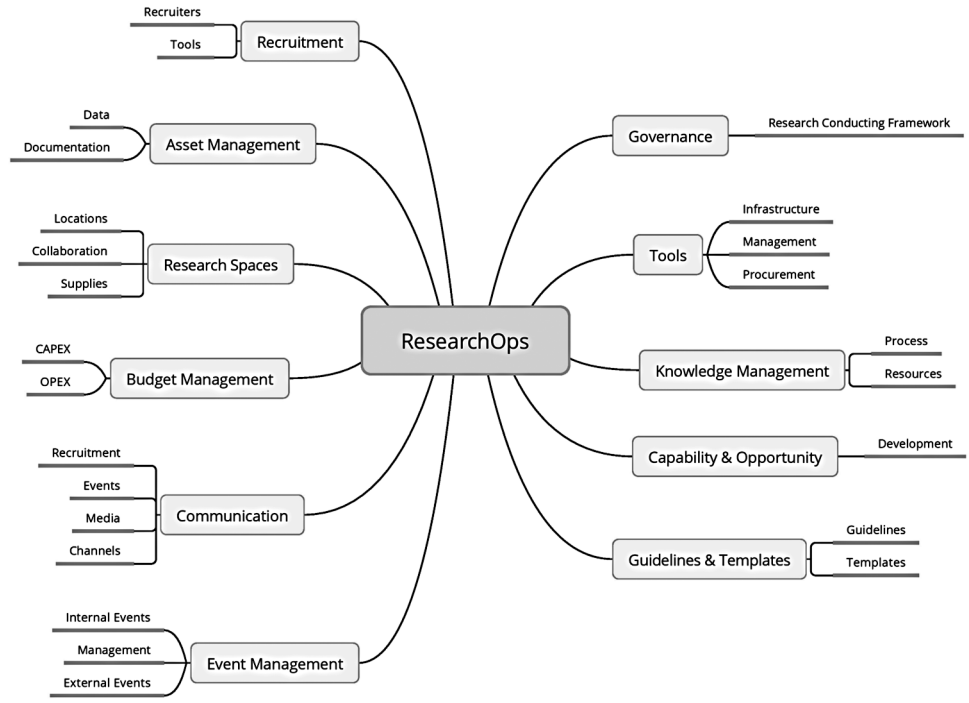


FIGURE 3
ResearchOps Community-separated resources for each of the topics. The link to community- made tools that support each aspect of these open topics can be found in the footnote below.[2]

researchers in delivering and scaling the impact of the craft across an organization."

The following topics (**Figure 3**) are important whenever research is managed. *Governance* defines safe, ethical, and legal research; *Guidelines & Templates* frame it formally. *Tools* are necessary for doing research, its management, and operations. *Knowledge management* deals with data and documentation (which are part of *Asset management*) and provides resources for *Capability & Opportunity* to develop career or capabilities necessary for a particular research project. *Research spaces* have to be maintained, and research staff and subjects must be recruited (*Recruitment*). This all needs to be published, promoted, presented, and advertised through the events within *Event management* and communicated (*Communication*) via various channels. All needs to be managed and maintained within *Budget management*.

Considering scientific research's disciplines, processes, and requirements, we need to adapt existing ResearchOps or define our operational process for the health-related data lifecycle.

2.1.2. FAIR principles

There is a plenitude of guidelines and principles available that can be used to store, maintain, and disclose open scientific data. However, the FAIR (an acronym for Findable, Accessible, Interoperable, and Reusable) principles (26) have become popular and widely accepted within scientific communities. They make data for computational systems easier to find, access, interoperate, and reuse, without any or just with minimal human intervention. The four intertwined categories describe how data, metadata, and resources should be described, stored, and made available to a broad audience.

The Findable principle declares that it should be easy to find data and metadata by both humans and machines. It can be achieved, for example, by assigning a globally unique and persistent identifier to data, describing data with rich metadata, and registering or indexing them in a searchable source.

The Accessible principle deals with data access since not all data have to be strictly open. If possible, metadata should be accessible even when the data are no longer available.

The Interoperable principle focuses on integrating data with other kinds of information. It is generally achieved using domain-wide agreed data formats, languages, and

[2]Available at: https://researchops.community/resources (Accessed 2022-06-08).

vocabularies. Also, qualified references to other metadata and data are included.

The Reusable principle ensures that data are easy to reuse, i.e., they can be well-replicated or combined into different settings. In this regard, data should be richly described with many accurate and relevant attributes and released with a clear and accessible license.

The FAIR principles do not state how they should be achieved; they represent recommendations that keep the data open and independent of the application domain. Multiple initiatives promote these principles across scientific fields, like the FAIR Data & Services (IFDS[3].) or the European Open Science Cloud (EOSC[4].).

2.1.3. Terminologies and ontologies

Terminologies and ontologies are popular for modeling domain knowledge in many scientific disciplines. Roche (27) explained that an ontology is not a terminology and a terminology is not an ontology and that terminology relies on two different semiotic systems (the linguistic one, which is directly linked to the "Language for Special Purposes" and the conceptual system that describes the domain knowledge), whereas ontology does not take into account the linguistic dimension of terminology. Zemmouchi-Ghomari and Ghomari (28) state that building ontologies is considered much time-consuming and costly than building terminologies with regard to ontology complexity and formality, two major differences between these types of resources. They also claim that terminologies can be considered as preliminary attempts to model particular domains by their respective experts. Then, terminologies are intended for human users, while ontologies are mainly developed for knowledge sharing between both humans and artificial agents.

Gruber (29) formulated the definition of an ontology: "*An ontology is an explicit specification of a conceptualization*." Borst (30) modified the definition as: "*a formal specification of shared conceptualization*." In other words, ontologies are formalized vocabularies of terms. Le Franc et al. (31) defined ontologies in an alternative way: "*Ontologies are formal models of knowledge in a particular domain and composed of classes that represent concepts defining the field as well as the logical relations that link these concepts together*." Designing an ontology is a long process in which it is necessary to understand the area and compile a list of used terms (and creating terminology).

The conceptualization of the real world can be also translated as taking real existing things and creating standard terms and their classes to categorize them into a hierarchical structure. However, many questions need to be answered before any description can be made.

The first question is how detailed or abstract the ontology should be since varying degrees of details are desirable. A detailed ontology can be overdefined when it describes and tracks every available detail. It can cause uncertainties when new changes are added to the ontology. On the other hand, overabstracting (generalization) leads to uncertainty in the definitions of the terms and their classes when many instances fall into more classes. The next question deals with the relations between the terms and their classes. The hierarchy and number of relationships must be carefully defined; otherwise, overdefinition or overabstracting can also occur.

Dynamic ontologies (10) add a layer for adjusting or "evolving" the ontology according to the project's needs over time. The changes need to be accommodated once the project grows, and the used terms and relations will need to be expanded or adjusted to address these changing needs in the already existing and published ontology. Examples of such "actions" can include adding/deleting existing relations between terms, adding a new property, changing the ontologies hierarchy, and reusing certain aspects or portions of other published ontologies.

Especially domain ontologies are popular since more general ontologies are very difficult to define (suffer from overabstracting). There are hundreds of biomedical ontologies and millions of classes (uploaded to Bioportal). The list of published ontologies steadily increases.

Popular languages for the implementation of ontologies include, e.g., the Web Ontology Language (OWL) of the Semantic web or dOWL, an extension to OWL, which consists of a set of elements that can be used to model these evolutionary changes in an ontology (32).

There are a lot of web-based systems to support ontology reuse (e.g., Bioportal,[5] OntoFox,[6] Ontobee,[7] Neuroscience Information Framework,[8] and Ontology Lookup Service[9]).

Although the popularity of terminologies and ontologies is still high, the requirement for an analytical definition of the part of the world is their limiting factor. It is a time-consuming task requiring not only the definition and implementation of

[3]Available at: https://www.go-fair.org/resources/internet-fair-data-services/ (Accessed 2022-06-08).

[4]Available at: https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud (Accessed 2022-06-08).

[5]Available at: https://bioportal.bioontology.org/ (Accessed 2022-06-08).

[6]Available at: http://ontofox.hegroup.org/ (Accessed 2022-06-08).

[7]Available at: http://www.ontobee.org/ (Accessed 2022-06-08).

[8]Available at: https://neuinfo.org/ (Accessed 2022-06-08).

[9]Available at: https://www.ebi.ac.uk/ols/index (Accessed on 2022-06-08).

the terminology or even ontology itself but also its acceptance by the wider community when the ontology should become a standard formal description of a domain. This step is crucial; hundreds of existing biomedical ontologies and systems that use them illustrate this issue well. There is some hope for expanding deep learning methods, which have lower requirements for data organization and could significantly alleviate problems with overdefined ontologies.

There are many broad terminologies defined in published ontologies like the National Cancer Institute Thesaurus (NCIT)[10] (33) describing set of terms and their relations. The main NCIT focus was on providing a controlled vocabulary used by specialists in the various subdomains of oncology. Across neuroscience, there exist projects that include terms related to event-related potentials, also containing MEG (magnetoencephalographic) or EEG (electroencephalographic) terminology.

The NEMO project (Neural ElectroMagnetic Ontologies) (34) provides an ontology that contains descriptions of classes of event-related brain potentials together with their properties, including spatial, temporal, and functional (cognitive/behavioral) attributes.[11]

Minimal Information for Neural ElectroMagnetic Ontologies (MINEMO) is the minimum set of experimental metadata required for datasets that are used in the NEMO project (35). MINEMO specifies the key information that should be provided when an ERP experiment is uploaded to the NEMO database. MINEMO terms are explicated in the NEMO ontology, a formal semantic system created for the ERP domain. There were also developed web applications (the NEMO portal) and a database aligned with the MINEMO checklist and ontology. The checklist, ontology, and database are intended to support the first complete, cross-laboratory meta-analysis for the ERP domain.

While creating new terminology (where the reuse of already existing terms is much endorsed), reusing only its essential parts may be easier than including the entire terminology. A recommended set of guidelines MIREOT (Minimum Information to Reference an External Ontology Term) (36) was created. It describes the necessary minimum of information that needs to be overtaken.

For ontologies, we have used recommendations by large and long-running standardization bodies like The World Wide Web Consortium (W3C) while including various most commonly used or recommended practices across the ontology lifecycle. It is the case with the iterative evolution, expansion, and enhancement of dynamic ontologies.

2.1.4. Data formats

Storing and processing health-related data are difficult because hardware devices and software drivers usually provide data in proprietary formats. Specific neurophysiological data-storing formats can severely hamper the collaboration between the researchers, as there is a need to have the same (and usually licensed) processing tools that support these data formats (37).

The main goal of neurophysiology data standardization initiatives (38) is to create a unified data model and storage format and tools to convert existing data stored in the proprietary data formats. These standardization efforts and their results (data models/formats) can be found in the following.

In this case, we have selected a list of the data formats endorsed [Brain Imaging Data Structure (BIDS), Neuroscience Information Exchange (NIX), Neurodata Without Borders: Neurophysiology version 2.0 (NWB:N 2.0)] or submitted for endorsement (Open Metadata Markup Language, odML) to the INCF Standards and Best Practices Committee. The endorsement process consists of an expert review against an established set of criteria, a community review, and a final committee review that considers comments received during the expert and community reviews (39). As for the remaining recommended format, JavaScript Object Notation for Linked Data (JSON/LD) is one of the few data and metadata formats used by large technological giants like Google.

2.1.4.1. Neuroscience Information Exchange format

The NIX data model (40) allows storing fully annotated scientific datasets, i.e., the data together with rich metadata and their relations in a consistent, comprehensive format. Although developed initially for electrophysiology data, neither the data model nor the metadata model are domain-specific. Both models can be linked to predefined or custom terminologies. It enables the user to give elements of the models a domain-specific, semantic context. In contrast to most other approaches, NIX achieves flexibility with a minimum set of data model elements. The NIX project includes native I/O libraries for C++ and Python, language bindings for Java and MATLAB, and a viewer for NIX data files, although the HDF5 (41) viewer can also be used.[12]

2.1.4.2. Open Metadata Markup Language

odML is a format for storing metadata in an organized human- and machine-readable way (42, 43). It does not constrain the metadata content while providing a common schema to

[10]The NCIT ontology is available at https://bioportal.bioontology.org/ontologies/NCIT?p=summary (Accessed 2022-06-08).

[11]The NEMO ontology is available at http://bioportal.bioontology.org/ontologies/NEMO (Accessed 2022-06-08).

[12]More information is available at https://github.com/G-Node/nix/wiki/Model-Definition (Accessed 2022-06-08).

integrate metadata from various sources. odML facilitates and encourages standardization by providing terminologies[13] for metadata.[14] An example of the odML use when collecting and exchanging metadata in an automated, computer-based fashion is described in (44). Currently, the odML is included in the NIX data model.

2.1.4.3. JavaScript Object Notation for Linked Data

JSON-LD (45) is a lightweight syntax to serialize Linked Data in JSON. Its design allows existing JSON to be interpreted as Linked Data with minimal changes. JSON-LD is primarily intended to be a way to use Linked Data in Web-based programming environments, build interoperable Web services, and store Linked Data in JSON-based storage engines. Since JSON-LD is 100% compatible with JSON, many JSON parsers and libraries can be reused. In addition to all the features that JSON provides, JSON-LD introduces, e.g., a universal identifier mechanism for JSON objects via the use of Internationalized Resource Identifiers (IRIs), disambiguation of keys shared among different JSON documents, a mechanism in which a value in a JSON object may refer to a resource on a different Web site, or the ability to annotate strings with their language.

2.1.4.4. Brain Imaging Data Structure

The BIDS is a standard endorsed by INCF prescribing a formal way to name and organize MRI data and metadata in a file system that simplifies communication and collaboration between users. There also exists an extension onto the BIDS format called EEG-BIDS, which is specifically designed to store the electroencephalography data. If you would be interested in learning more about the EEG-BIDS format, you can find it in (46).

In both variants, it enables easier data validation and software development by using consistent paths and naming for data files. BIDS is strict regarding file organization, naming, and metadata, but to support broad adoption, it permits substantial flexibility in the details of how other dataset metadata are described within the standard (47).

2.1.4.5. Neurodata Without Borders: Neurophysiology version 2.0

NWB is a data standard enabling sharing, archiving, using, and building analysis tools for neurophysiology data. NWB is designed to store various neurophysiology data, including data from intracellular and extracellular electrophysiology experiments, data from optical physiology experiments, and tracking and stimulus data (48). NWB:N 2.0 defines an ecosystem for standardizing neurophysiology data.

2.2. Motivation

In this paper, we focus on the health-related data lifecycle. It mainly includes data description for further processing, the richness of data/metadata from the subject perspective, correlation, and causality research. We aim to achieve that via four pillars—ontologies (**2.2.1**), 360-degree overview (**2.2.2**) of the research subject from data perspective, standardized lifecycle (**2.2.3**) for health-related data, and research reliability (**2.2.4**) through reproducibility and repeatability.

2.2.1. Using ontologies

First, let us start with what are the ontologies good for (49). Their first and foremost advantage is to capture the used terminology inside any application domain (or the research subject) and map the definitions, attributes, and relations of these terms to one another. Since many terms can have multiple meanings, their precise definition helps even a newcomer to the application domain better understand the used terms and their relations; this extended vocabulary maps relations between defined terms.

An additional benefit to ontologies is that they enable easy understanding of multiple application domains and can be reused easily. Their reuse helps reduce the redefinition of the terms. When an ontology becomes widely available, it increases its value. The ontology can be expanded and corrected further down the line to a more detailed and sufficiently defined result.

2.2.2. 360-degree overview

In most cases, we are talking about EEG/ERP data (50–53). These types of data bring crucial information about the measured subject, but we cannot forget to record also the data related to the subject, outside environment, and the experiment itself. These additional data provides a 360-degree overview of the measured subject and give additional potential to better understand the foundation during the analytical process.

In some literature studies (54, 55) are such data neglected, and the main focus is on neuroscientific data. It works well for narrow field research for single-purpose data collection during the experiment and making the conclusion published through single paper. However, with the greater goal, we need to collect as much data as we can, so it can be later used for multiple research use cases.

We cannot consider our work to be frontier-bringing such idea since some literature studies (56, 57) present collection of

[13]More information about the odML terminologies can be found at https://github.com/G-Node/odml-terminologies (Accessed 2022-06-08).

[14]More information can be found at https://g-node.github.io/python-odml/ (Accessed 2022-06-08).

data, Metadata, and scenarios of experiments, but we would like to define standards that can be adapted as is or with extensions or adjustments.

2.2.3. Standardized lifecycle

Inspired by ResearchOps and DataOps (2.1.1), we derived the subset of disciplines useful for health-related data (Figure 4).

We kept in mind disciplines and processes necessary to cover complete neuroinformatics data lifecycle from asset management, subjects recruitment, guidelines and templates, knowledge management, and data governance.

2.2.4. Research reliability

The main idea about the ontology-driven (2.2.1) system is to provide a platform for reproducibility and repeatability (58). These two major principles of scientific methods for research supporting are very important to ensure research reliability.

2.2.4.1. Replication crisis

In 2005, an essay was published in PLoS Medicine by Professor John Ioannidis at the Stanford School of Medicine (59), who argued that a large number, if not the majority, of published medical research papers contain results that cannot be replicated. This is practically the foundation for later-defined term replication crisis, respective replicability, or reproducibility crisis.

The crisis itself has longer roots, but it started to be significantly used in the early 2010s (60) as part of growing awareness of the problem (61, 62).

2.2.4.2. Reproducibility and repeatability

The meaning of reproducibility is to achieve the results of the experiment again with a high degree of agreement when the experiment is replicated with the same methodology by the different researchers. When the reproducibility is achieved once or several times, the experiment can be considered a valid contribution to scientific research.

The repeatability is defined as one (test-retest Reliability) of the four general classes of reliability estimates (61) from the theory of reliability (62) we know:

- Inter-rater or inter-observer reliability used to assess the degree to which different raters/observers give consistent estimates of the same phenomenon.
- Test-retest reliability used to assess the consistency of a measure from one time to another.
- Parallel-forms reliability used to assess the consistency of the results of two tests constructed in the same way from the same content domain.
- Internal consistency reliability used to assess the consistency of results across items within a test.

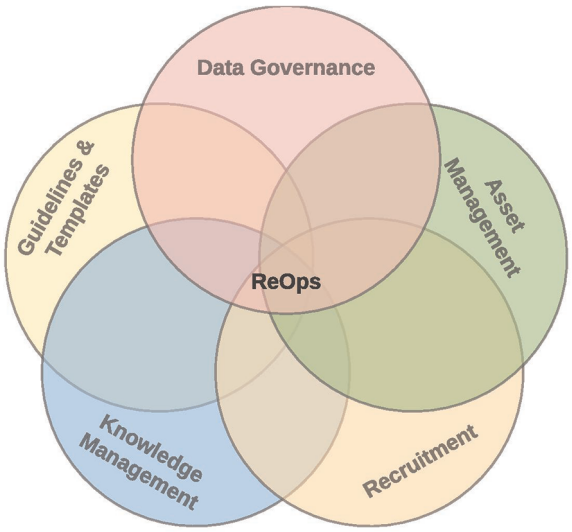


FIGURE 4
ResearchOps for health-related data.

2.2.4.3. Terms' ambiguity

In the scientific research exists the ambiguity of reproducibility and repeatability (62, 63). The usage of the recurrent terms reproduce and replicate often means different things but sometimes interchangeable.

As (62) claims, the terminology can be classified as, First, make no distinction between the words reproduce and replicate or, second, use them distinctly. This two-term direct substitution leads to the weight issue that might be solved by various attempts to invent the terminology across disciplines and establishment of patterns that help us resolve the contradictions.

2.3. Summary

In the following section, we mention some of the best practices and pieces of advice that were recommended by the wider scientific audience and used within the Body In Numbers project:

- **FAIR principles**—It is the utilization of the FAIR principles across all processes, making the collected data and metadata easily accessible and shareable.
- **Size and scope of the new ontology**—It is necessary to define who will and how to use the ontology. Examples of questions that might be asked are as follows: How abstract or detailed should the ontology be? or What subjects will it cover?
- **Learning from already existing ontologies**—In case you have not much experience with creating ontologies, it is best to go through existing and published ontologies

inside the same field of study when planning to create a new ontology. In this case, it is necessary to focus on how each used term is being defined and what annotation attributes are used to devise rules for the ontology creation process.

- **Appropriate annotation properties**—It is a good practice to maximize the reuse from already published ontologies. Relevant ontologies for this task can, for example, be the Information Artifact Ontology (IAO). If there are no adequate replacements, new ones can be created.
- **Naming terms**—Using plain English in the term names is strongly advised. CamelCase or Under_Score notations should be avoided. If the term has any notable synonyms or shortcuts (e.g., acronyms), they are stated. If any dedicated annotation properties are used, the rdfs:label can be used instead.
- **A unique identifier**—In the case of overtaken terms, the original unique identifier from the source ontology should be retained; otherwise, a unique ID for each new term has to be created. Organizations can generate a persistent URL that enforces the uniqueness requirement for the primary identifier, e.g., http://purl.obolibrary.org/obo/OBI˙0000185.
- **Textual definition of each term**—Textual definition needs to best describe the meaning of the term under which it is used inside the ontology.
- **Reuse (import) of external terms**—When any existing term is overtaken, the attributes and ID should be identical to the source ontology. Rules for the import of the term can be further specified inside the source ontology under the annotation property rdfs:comment.
- **Ontology open to collaboration**—Any collaboration with the community can enhance the overall quality of the ontology.
- **Ontology license**—The Creative Commons license in its latest version is advised for most of the open source ontologies. For monetized ontologies, when the ontology can be used (under which circumstances), in which projects or how to obtain access to the ontology needs to be specified.
- **Serialization of ontology**—Common formats, e.g., OWL, RDF/XML, or OBO, are defined for the publication of the finalized ontology.
- **Incremental expansion of ontology**—Certain terms might not have been properly defined as in the original plan. Small incremental additions of new terms can make the ontology overall more well-defined.
- **Data formats**—Suitable, possibly free, and widely accessible data formats are used within the tools that can be used to explore the collected data (e.g., for MRI, EEG, ERP, and more).

The methodology used in the Body in Numbers ("BiN" for short) project is mainly based on the recommendations mentioned above.

A thesaurus of all related terms used inside the project was created in the initial phase of the BiN project. The terms were related to data collection or the subsequent data processing phases. Each of these terms had the name and basic definition stated in plain English.

As a next step, these terms were separated into categories (classes of terms) and related notation properties, further used to describe all remaining terms in the ontology. An example of the annotation property is the author of the definition, textual definition, synonyms and shortcuts, and name of the term.

After this phase, it was necessary to maximize the reuse out of any already existing and published ontological sources, so that we would limit the amount of "reinventing the wheel" so to say. At this step, ontological portals were of great help (like BioPortal, OntoBee, and OBO Foundry).

In the next step, it was necessary to distinguish overtaken terms from other ontologies and newly defined terms as each needed to contain different annotation properties. In the case of overtaken definitions, all attributes were overtaken into the annotation properties equivalents of the devised ontology. For the newly defined terms, it was necessary to define a bare minimum of information that the term needed to contain (like the synonyms, known shortcuts, and textual definition). Once the process for the newly added terms was refined, a dynamic ontology was created. The ontology was then used in the next steps of the data flow.

However, the workflow for the health-related data lifecycle goes behind the created ontology. Next to the ontology, we also provide a 360-degree data overview (2.2.2). This includes our proposal for standardized lifecycle (2.2.3) inspired by DevOps, DataOps, and derived from ResearchOps. Then, we included research reliability (2.2.4) consisting of two principles—reproducibility and repeatability.

3. Results

In the following section, we will take a closer look onto both the abstract and also on the implementation aspect of an undertaken project from the University of West Bohemia called "Body in Numbers." We will also describe the entire process on how the data were acquired, stored, processed, successively analyzed, and published. The unique part about this process is a specialized support module architecture developed in tandem to help with each step of the data cycle.

3.1. Main overview (cube)

The implementation of neuroinformatics experiments is a conceptually complex system that solves every aspect of the data flow in races to a given category of interest within

each of its functionalities. In this case, it is the age category of the participants in the experiments. The analytical model of the system is shown in Figure 5. This system determines any way, for example, how the collected data from children and adults are handled. Other variants of research addendum agreements and questionnaires could be used to measure children and adults. These templates are used to change measurement procedures depending on whether a child or an adult participates in the experiment. Later in the same year, a new change needed to be implemented related to the analysis, namely, our goal was to find a statistical significance between the already collected data and metadata.

When it comes to evaluation of the efficacy in the presented lifecycle (in the form of the "Cube"), the chosen dimensions separate the larger process into small tasks that can be at least partially (based on the situation) automated outright, or it is possible to create support processes that overall make the step easier (or faster) across the data flow. For example, data quality can be ensured during the data collection process by filling out all mandatory fields and pairing included metadata. Data duplication can be prevented (at least partially) in this step. As for the mentioned support processes, the analysis part of the data flow can automatically prepare overview statistics about the collected data that help classify or further analyze the data and metadata.

The chance of errors in the filled-out data and metadata is dramatically decreased when steps eliminate the "human factor" from highly repetitive or easy-to-automate tasks. Aside from that, the time necessary to spend in each data flow step will also decrease, with much of the validation steps being eliminated from the equation.

Some of the proposed metrics for evaluating this lifecycle were defined as follows:

- percentage of complete data entries inside the data collection phase for the dataset,
- percentage of described data entries inside the annotation phase for the dataset,
- number of newly required metadata definitions (number of new terms) from the dataset, for inclusion into the ontology,
- time spent in each of the data flow phases, and
- age range histogram for all measured participants inside the dataset.

3.1.1. Data flow

This dimension helps streamline and potentially automate repeatable experiments, minimizing room for errors and increasing overall efficiency. This part improves the description of the experiment. Experimenters can make quicker and smarter decisions. Researchers are empowered to collaborate in a more productive and agile way.

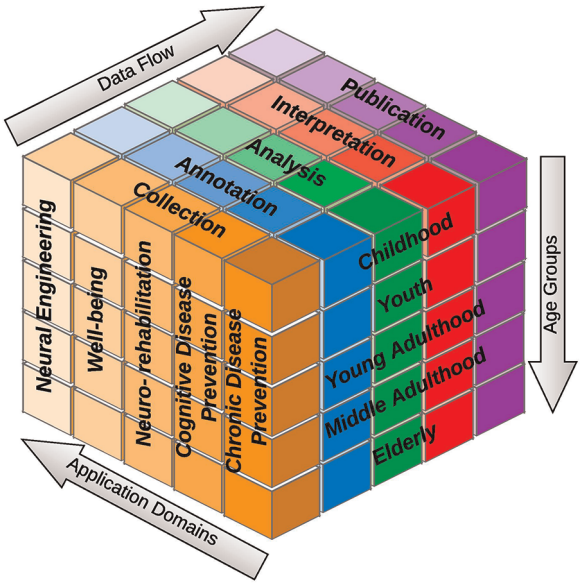


FIGURE 5
"Cube" overview: each cube part can be imagined as a box with specific needs, which have to be accounted for and supported by the underlying system.

The data flow phases are shown in Figure 5. The collection phase deals primarily with acquiring data from a BiN project participant; the data are subsequently preprocessed (more details are in Section 3.3).

The preprocessing consists of cleaning the data and adjusting it for the next step by converting it into the format expected in the analysis phase.

The analysis phase is currently aimed at determining statistical significance between each segment of the measured data, the values achieved by participants, and respective questionnaire answers. The summary tables and graphs are created and used for finding further subsequent research activities.

During the interpretation phase, new hypotheses are outlined and revised by scientists to refuse or confirm them.

In the last phase of the data cycle, anonymized data are published utilizing raw data. The results obtained from the initial analysis and the pilot experiment are included. The data published in this way help the broader scientific community answer further questions and hypotheses.

3.1.2. Application domains

There already exist conceptually close domains. The Chronic Disease Prevention (4.2.2) domain is focused on nutrition counseling. The food balance system is calculated to reduce the user's food consumption. This system can monitor the users using smart bracelets (such as FitBit) when information is sent to a mobile tracking application.

The Cognitive Disease Prevention (4.2.2) domain works with cognitive games that improve memory, attention, speed, and problem-solving abilities.

Both domains are indirectly related to brain activity data (e.g., EEG, ERP), which however can be used to further help with either of the previously mentioned application domains.

3.1.3. Dimensions (age groups)

The project focuses on the age range from 11 to 60 years. Beyond this age limit, the measured results are distorted by specific errors related to either young or advanced age. In the case of preschool children, there may be problems with the numbers (color vision, number recognition) and hand reaction times (the children might be unable to reach the upper buttons on the table due to the minimum required height). In older age, the problems may be responding quickly to stimuli using legs (leg reaction times) or hands.

The monitored age categories are shown in Figure 5 (age groups), namely, childhood, youth, young adulthood, middle adulthood, and the elderly. Of course, the consent and completion of the questionnaire for a young child will be different and proceed differently than for adults (in the case of young children, their legal representatives or parents must approve the participation in the project and The General Data Protection Regulation data processing).

3.2. Data flow dimension

3.2.1. Data flow semantics

The first step in data preprocessing is to identify the individual parts of interest and categorize them so that the data are transferable and shared between different working groups. Creating the ontologies schematizing relations between each part of the cube is necessary to make the data more shareable. Therefore, the task was to develop a system that would be able to preprocess the data and make it easier to share with members of the scientific community. The project aims to create a uniquely annotated collection of heterogeneous health-related data available for further analysis. The Body in Numbers system helps collect additional metadata from questionnaires combined with the measured health-related data (e.g., weight, height, and blood pressure) and EEG data from brain–computer interface (BCI) experiments.

The data collected are anonymized and published within data articles. They are converted into one of the commonly used RDF formats, and a backing ontology is created; it may define additional properties.

The Body in Numbers terminology has included and used the best practices presented in Section 2.3. The main set of terms was defined (the definition under which it is used in the BiN terminology) and compared against the existing definitions from relevant sources. If the term was already described inside other ontologies, but the description did not match the meaning utilized in the BiN project, a new definition of the term was created. Otherwise, the existing definition was overtaken, and a citation was attached appropriately.

The basic set of key terms are given in Figure 6.

3.2.2. Data flow detail

Considering the cube as a modular representation, we sliced its particular layers with the following topics (Figure 7).

Body in Numbers (BiN) uses its specific terminology:

1. Tools—Devices and tools used during experiments (e.g., pressure gauge for measuring pressure and pulse, and spirometer for measuring lung capacity)
2. Experiments—Participants examination (e.g. measuring blood sugar requires a finger prick by taking blood on the measuring strip and then evaluating the results with a glucometer; the experiment's output is blood sugar = 6.4 mmol/l).
3. Locations—A locally determined measurement that may contain one or two more experiments.
4. Measurements—Definition of all sites, experiments, and assigned aids.
5. Scheduling—Definition of what (e.g., glucose level, weight, height), how (e.g., glucosemeter, scale, meter), and what (specific type of the measuring instrument, i.e., specific glucosemeter type) is used for measurements.

The created ontology[15] contained 141 classes of terms, of which 56 classes of terms were overtaken from already existing ontologies (with their original definition), and for the remaining 85 terms, new definitions were created (the definitions from existing published ontologies were insufficient or the terms were not yet defined). The ontology also contains 30 annotation properties where every single one was overtaken from existing sources. An example of subclasses visualizations is available in Figure 8.

3.2.2.1. Data collection

The following subflow is typical for the data collection process:

1. The experiment and its environment (Section 3.2.2.2) are defined.
2. The experiment is introduced (64) to each participant; they sign the consent agreement.
3. The participant is registered to the system [(64), more information is available in Section 2.2].

[15]The BiN ontology can be found here: https://bioportal.bioontology.org/ontologies/BIN/?p=summary (Accessed 2022-06-08).

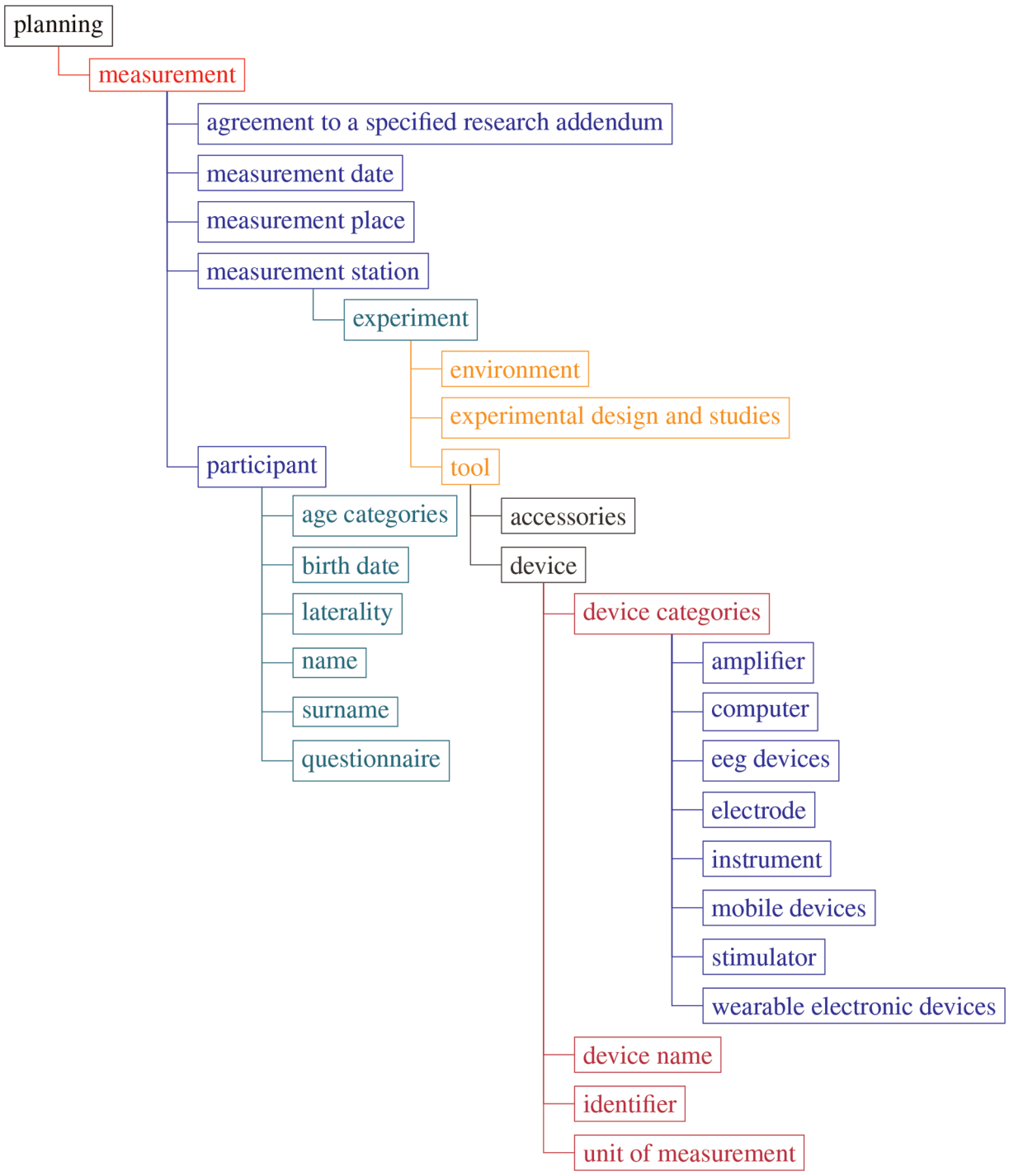


FIGURE 6
Body in Numbers terminology—examples of the root terms. Note that tree visualization only contains a few top degrees of the tree. For more details, see the link to the terminology in Section 3.2.2.

4. The participant fills in the questionnaire [(64), more information is available in Section 2.3] related to a particular measurement or multiple measurements.
5. The measurement is performed.
6. The collected data are stored (Section 2.1.4) in a standardized format together with their metadata collected during the registration and questionnaire phases (Sections 3.3.2 and 3.3.4).

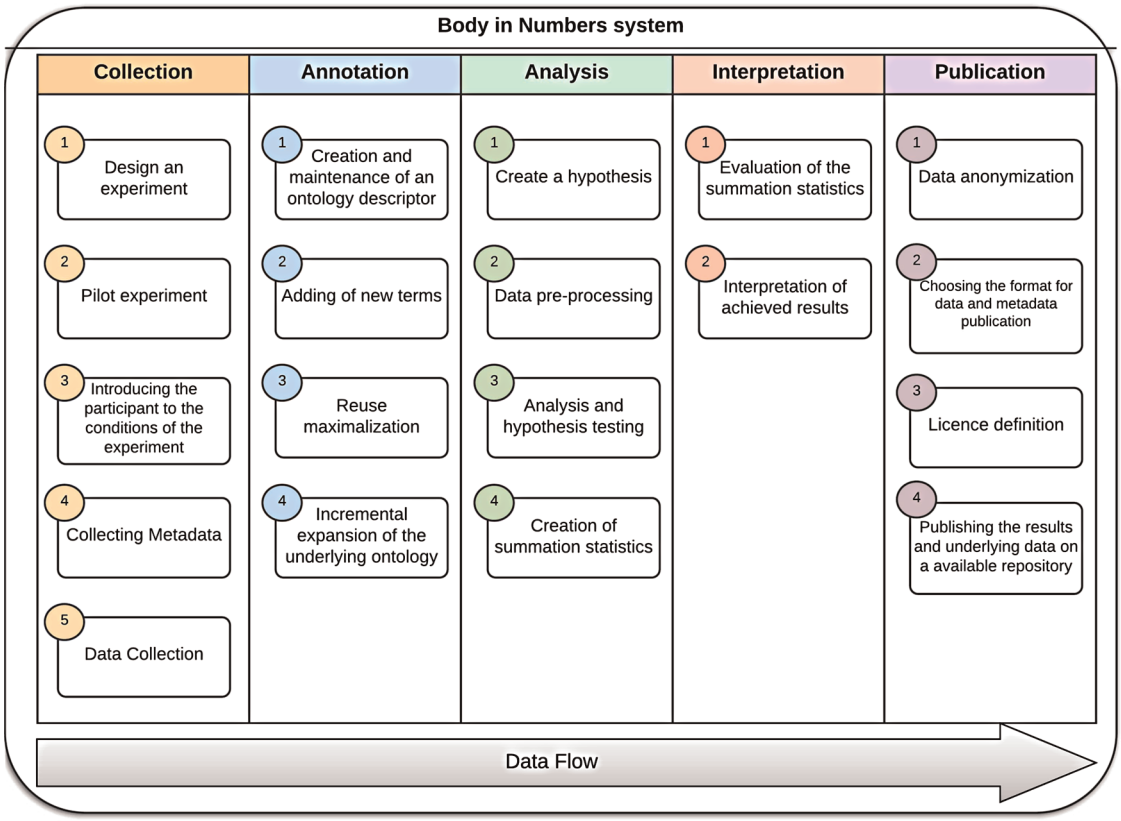


FIGURE 7
Data flow–layer details.

7. Data analysis is performed (Section 3.3.3).
8. The raw data are prepared for publishing (Section 3.3.5).

At the beginning of the experiment, the participant is acquainted with the project's goals and the necessary conditions under which the data are used, and if they agree, the questionnaire part of the investigation is filled in. After completing these steps, the participant completes the selected (or all available) measurement sites. The collected data are then exported into a .csv or .xlsx file, further used during the preprocessing and later in the statistical module. After the unification and purification of preprocessed data, an analysis follows. In this study, graphs of interest [age, body mass index (BMI), and others] and chosen statistical parameters are used to evaluate statistical dependencies and whether they are essential. From the statistics compiled, signs of some interesting dependencies can be found.

3.2.2.2. Environment

The measurements can occur in various environments with limited control of outside disturbances. It affects possible (and even substantial) modifications of the experiment. The typical environments are as follows:

- **Hospital environment**—The hospital operation's limitations and the participant's health status are usually significant; the experimental procedure needs to be adjusted to these limitations.
- **Laboratory environment**—The laboratory environment is generally highly controlled; only minor modifications of the experimental procedure are usually required.
- **Participant home environment**—The most suitable environment for the participant; usually minor to averaged modifications of experimental procedure are required.
- **Public environment**—A high range and variety of unwanted disruptions are present; these create unwanted side effects—substantial and *ad hoc* modifications of the experimental procedure are common.

The environment is defined within the Body in Numbers system as a part of the measurement site.

3.2.2.3. Annotation

While recommendations and opinions on what a proper ontology should contain are widespread, there is no unified opinion on what each defined ontology term should contain (in terms of granularity and detail).

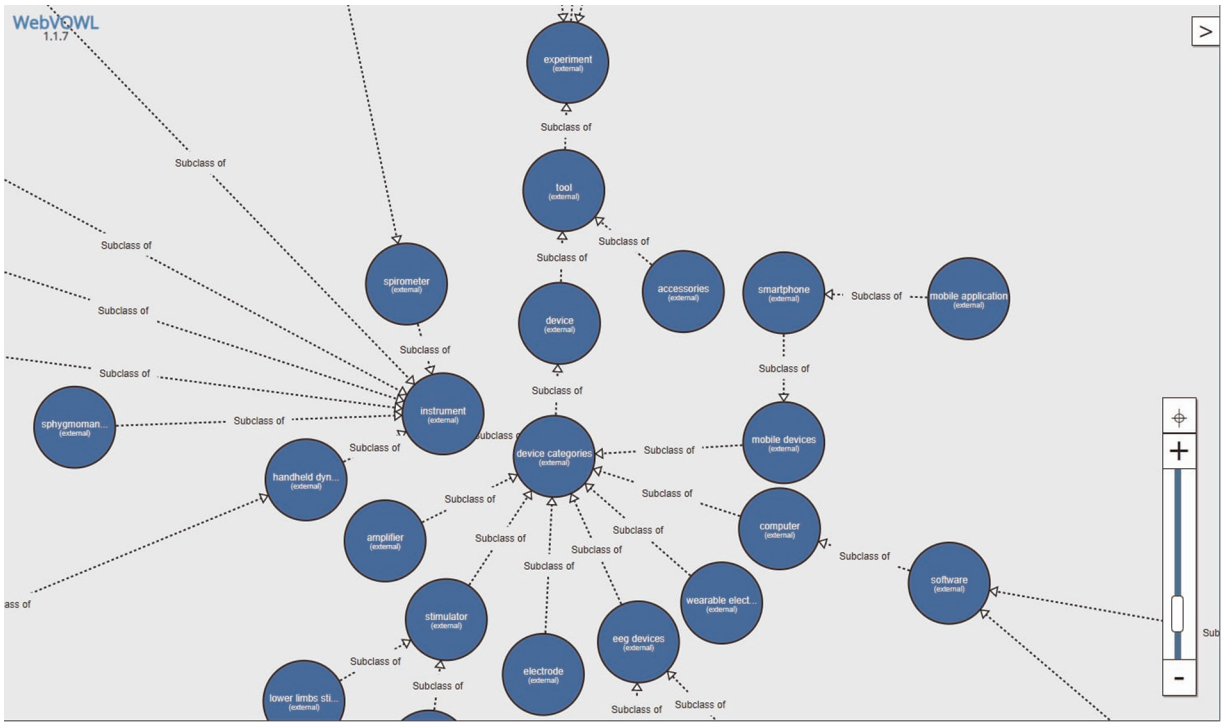


FIGURE 8
Body in Numbers ontology—the parent node "Device categories" contains links to its child and sibling nodes. Only the subclasses are visualized (without any additional defined properties). The ontology was visualized with WebVOWL: Web-based Visualization of Ontologies, version 1.1.7, available at http://vowl.visualdataweb.org/webvowl.html.

The authors of ontologies generally do not use identical defining terms, and ontologies often contain similar definitions of synonyms even inside the same field of expertise. Functionally identical terms, like a term defining the author of a dataset, can be marked differently in each ontology (e.g., author, original_author, creator, and more).

Consequently, global organizations (such as OBO Foundry, Open Semantic Framework, or W3C) bring recommendations and standardization efforts. These standardization efforts include, e.g., rules that help define importing procedures for terms already defined in published ontologies, the necessity to define nomenclature, and providing text definitions for each contained term inside the ontology.

3.2.2.4. *Analysis*

In analysis, we created a combination of hypothesis testing and basic data overview variables. The basic data overview consisted of summation values like the number of artifacts within the EEG dataset, averages of various kinds, e.g., box plots of BMI separated by specific categories, reaction times of either upper or lower limbs, and age intervals. Another part of this information also consisted of more EEG data-related variables, mainly the P300 visibility, detection, and latency.

3.2.2.5. *Interpretation*

A statistically significant relationship was examined between the metadata (questionnaire) and the measured data. For example, if a participant answered that he/she enjoyed physical activity, he/she was supposed to answer faster when performing the reaction time experiment.

When performing statistics, we might find repeated patterns in the measured data. If such patterns are identified, it is worth studying what is causing them to appear. A typical example of a searched pattern is the P300 component, which is prominent during visual stimulation in most circumstances.

3.2.2.6. *Publication*

During data publication, an open and widely accepted data format for data storage and the availability of tools to work with the collected data are considered. For example, table-like data are suitable to be published in CSV or Excel Spreadsheets (rather than in RDF). EEG data are preferred

to be published in an open format, like NIX, rather than in a proprietary format.

3.3. Body in Numbers

We will now show a practical example how the project Body in Numbers system was developed to satisfy the previously defined abstract necessities (i.e., the "Cube"). Initially, we needed to design a system for fast health-related data collection. This can be partially remedied by an individualized exercise and wellness program that integrates basic knowledge domains: lifestyle, sports and fitness, nutrition, and personal/environmental health. However, collecting, managing, and analyzing data and metadata related to these domains is demanding and time-consuming. Moreover, the appropriate **annotation** of raw data is crucial for their subsequent processing. A proposed software infrastructure included the P300 module, a specialized module for data collection of ERP data (3.3.1), a subsection for **data collection**, a semantic module (3.3.2) for **data annotation**, a statistical module (3.3.3) for **data analysis**, an evaluation module (3.3.4) for **data interpretation**, and a publishing module (3.3.5) for **data publication**. A part of this infrastructure, namely, the P300 module, was developed and tested outside laboratory conditions. This software prototype allows experimenters to collect various heterogeneous health-related data in a highly organized and efficient way. Data are then evaluated, and users can view relevant information related to their health and fitness (65). The structure of these specialized modules is available in Figure 9.

3.3.1. P300 module

The P300 module provides basic information and statistics (like noise percentages from whole dataset and average values, i.e., response time) related to the EEG signal. When the EEG signal is cleaned and preprocessed, the P300 components latency and amplitude are extracted, with plots of averaged ERP components and blinking artifacts. The details can be seen in Figure 10.

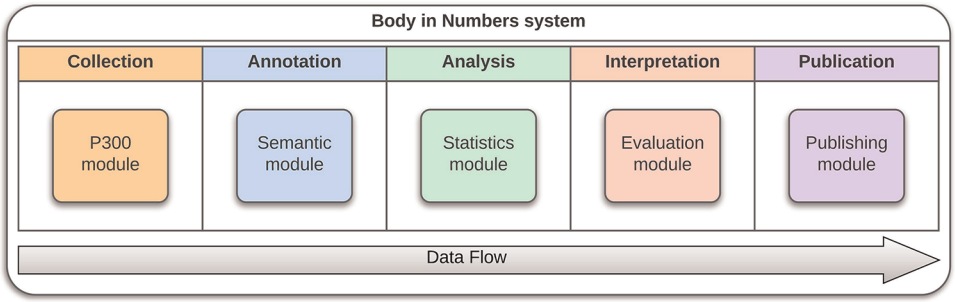


FIGURE 9
BiN lifecycle—For each of the data flow steps, module was developed that assisted with their respective tasks independently of the remaining architecture.

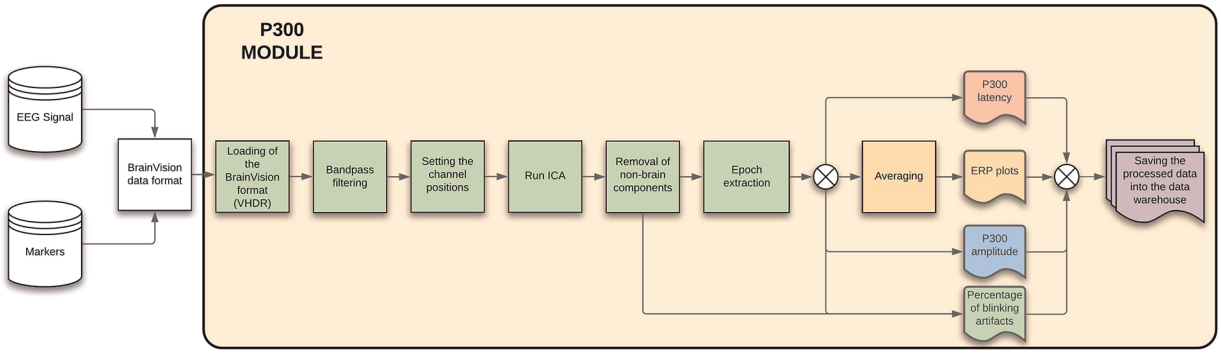


FIGURE 10
P300 module—The figure describes each of the steps that will be done automatically by the module, showing also the process inputs and outputs.

3.3.2. Semantic module

The semantic module ensures that the BiN ontology is updated when needed. This is achieved by varying approaches based on whether the newly encountered term is already known.

The data are enriched based on the used ontology if the term is known. For example, a list of measurement sites with the used devices is created; when heart rate is measured, both the units and device type are automatically added.

If the term is not known, a curator has to decide whether a definition for this new term already exists or if a community poll approach needs to be initiated (for the most useful and up-to-date definition). Both of these approaches eventually define a new term for the BiN ontology. This module description is available in Figure 11.

3.3.3. Statistical module

The statistical module looks for statistical significance between the metadata (questionnaire) and the measured data. The data are first preprocessed and split into sections (based on the used questionnaire), which are then compared against each category separately using a stepwise regression on a 5% significance level. The statistical module is further described in Figure 12.

For example, if there is a question about sport in the questionnaire, the stepwise regression will show it as statistically significant to participant's reaction times (both hands and legs), as the reaction times are shorter when the participant is doing some form of a sport.

3.3.4. Evaluation module

The resulting graphs generally contain averaged or summed statistic values, for example, top five fastest

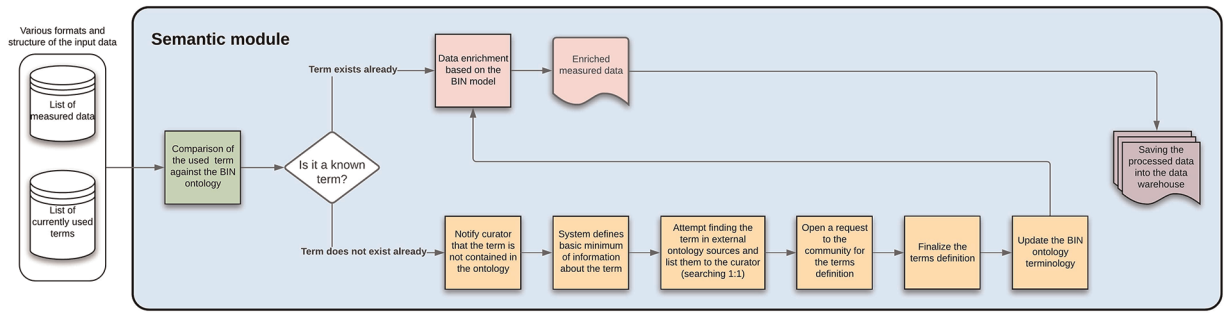


FIGURE 11
Semantic module—the figure describes each of the steps that will be done automatically by the module, showing also the process inputs and outputs.

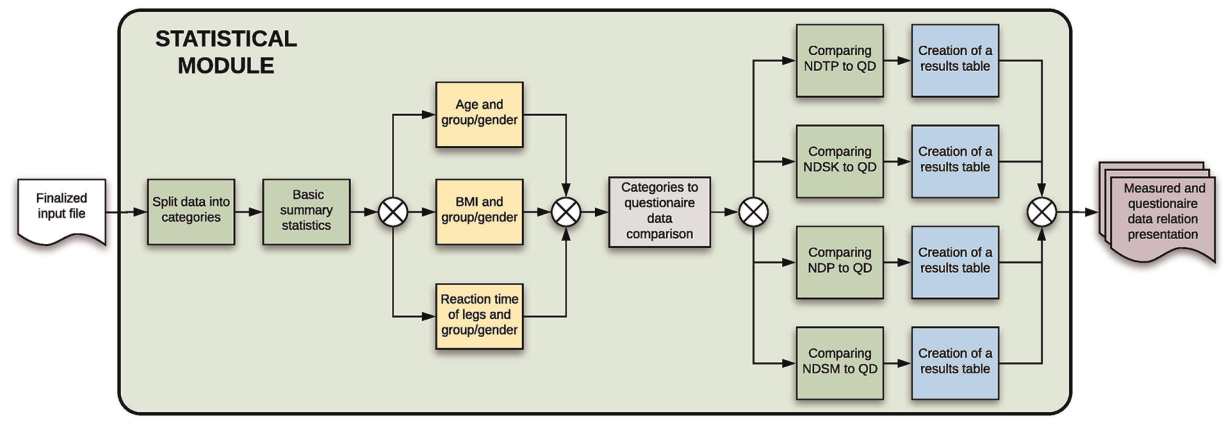


FIGURE 12
Statistical module—the figure describes each of the steps that will be done automatically by the module, showing also the process inputs and outputs.

participants' reaction times, average reaction time, number of participating smokers vs. nonsmokers, etc. If the researcher finds anything interesting they would like to publish or further investigate, they can let the system create an ID link to the dataset and, if necessary, convert the measured data/metadata into a preferred format. Details of this module are available in Figure 13.

3.3.5. Publishing module

The publishing module is mainly responsible for the anonymization of data that are going to be published within a journal and creates a LaTex template that is filled directly with the measured data, attached to the template, or linked with the generated ID from the evaluation module. A license is stated (open or closed variant) within this template. The output is then for the researcher to be further filled out. This module is shown in Figure 14.

3.4. Dimensions

The 360-degree overview (Section 2.2.2) is beneficial for further data processing and analysis. It gives context to health-related data sources and allows researchers to work with them from different perspectives, e.g., filter them to get a significant subset or cluster them into different cohorts.

What kind of sets of data and metadata would be useful to collect as part of the undertaken experiments have been

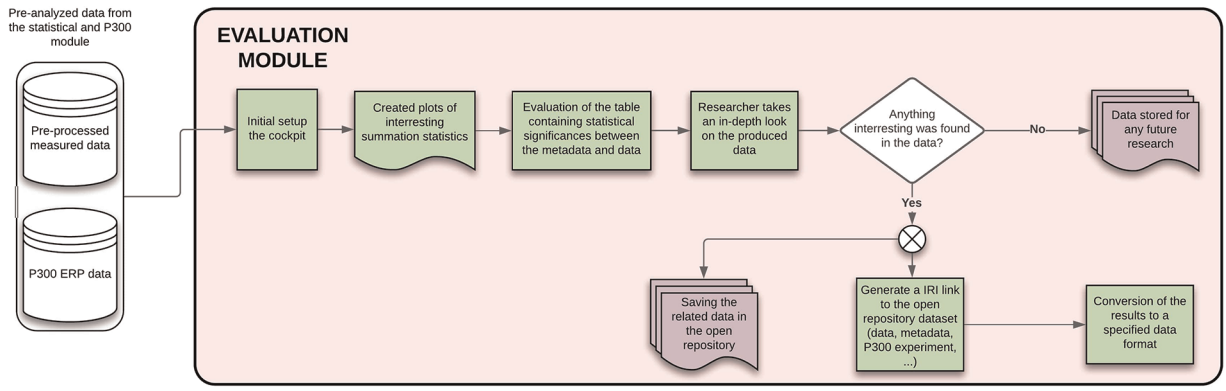


FIGURE 13
Evaluation module – The figure describes each of the steps that will be done automatically by the module, showing also the process inputs and outputs.

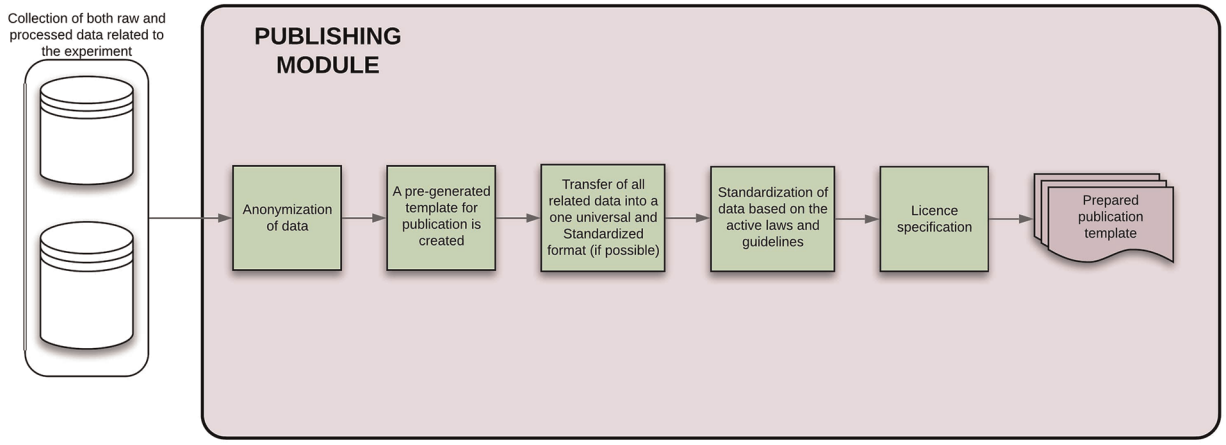


FIGURE 14
Publishing module—the figure describes each of the steps that will be done automatically by the module, showing also the process inputs and outputs.

discussed extensively. In our case, we have looked for useful sets of metadata recommendations to collect in similarly undertaken experiments.

This EEG experiment (66) was aimed mainly at motor imagery BCI; additional metadata were collected to be used for further analysis and dataset evaluation. Recommendations about what metadata could be helpful as a part of an EEG dataset were described by (67) back in 2000.

The above-mentioned literature considers the following dimensions to be the most significant:

- age groups,
- gender,
- handedness,
- BMI,
- eye defects, and
- geolocation.

For example, brain neuroplasticity changes with age. Older people are more likely to have problems with some of their motor skills (and the reaction times of either hands/legs grow as a result). Most of the collected data are more precisely evaluated based on age groups instead of age.

Gender was chosen instead of the "easier-to-manage" sex. Handedness (which hand is used by the participant in most tasks) is more valuable than laterality (superior development of one side of the body or brain).

BMI is valuable as it can be used partly to distinguish between the more bodily challenged participants, as these generally have a higher chance of being affected by chronic diseases. Eye defects are essential since most of the experiments relied on visual stimuli.

Participants feel more comfortable under different weather and temperature conditions (hot, cold, humid, or dry environment). Even the measurement's geolocation affects the data quality and the participant's physical and mental condition. The uncomfortable outside environment leads to more artifacts in the measured data. The participant's home can positively affect his/her mood and concentration.

4. Discussion

This part discusses how the Body in Numbers system can be used as a template to create an instance of the real system. Further modules of the system are also described.

4.1. Validation use case

The Body in Numbers system was initially designed to rapidly collect heterogeneous health-related data for chronic disease prevention (obesity, diabetes). The original idea was first published in 2017. However, as the project has grown, the system expanded from collecting pure data to accommodating cognitive and chronic disease prevention and started to relate to neurorehabilitation and BCIs.

As the number of subjects has grown, the need to have well-annotated data has arisen. It would allow for creating different datasets based on the investigative needs (e.g., creating a study covering the youth—a need to narrow down the dataset to only people between 15 and 24 years). The earliest Body in Numbers system has been expanded to accommodate other data dimensions and has allowed researchers to collect and analyze data in different environments.

A working example of such a system, accommodating a wide range of functions, is shown.

4.1.1. Proof of concept

As the next step, we defined, collected, and annotated human reaction times and relevant health-related data and metadata for further human physical and cognitive performance analysis. A collection of human reaction times and supporting health-related data was obtained from two groups comprising together 349 people of all ages—the visitors of the Days of Science and Technology 2016 held on the Pilsen central square and members of the Mensa Czech Republic visiting the neuroinformatics lab at the University of West Bohemia. Each provided dataset contains a complete or partial set of data from the following measurements: hand and leg reaction times, color vision, spirometry, electrocardiography, blood pressure, blood glucose, body proportions, and flexibility. It also provides a sufficient set of metadata (age, gender, and summary of the participant's current lifestyle and health) to allow researchers to analyze further.

A well-annotated collection of human reaction times and health-related data suitable for further lifestyle analysis and human cognitive and physical performance was provided. This data collection was complemented with a preliminarily statistical evaluation. A procedure for efficiently acquiring human reaction times and supporting health-related data in nonlaboratory and laboratory conditions was also presented (64).

Thanks to the success and daily use of the system, new requirements related to better security, Scalability, and maintainability of its architecture have emerged. The next work presented advances and changes in the architecture of the Body In Numbers health strategy framework, mainly focusing on a new definition of user roles, optimization of the system deployment, and orchestration of the system components. A Kubernetes cluster prototype was used as proof of concept to demonstrate the improved architectural solution (68).

4.2. Application domains

Relations between the person's daily life and predispositions toward specific health-related issues have been proven (e.g., (69) that covers health-related topics related to insomnia, (70) for the noise sensitivity and its effect on health, and (71) for the movement amount and its impact on health). Most well known are the issues connected with jobs with minimal movement or jobs that heavily impact a person's body.

The question of prevention is, most of the time, a decisive factor in evading any health-threatening problems that might occur, based on the person's daily lifestyle when aging. The topic of prevention is vast; five notable topics were selected:

- **Chronic disease prevention** (4.2.1)—it is related to recommended movement activities throughout the day (generally concerns jobs with large amount of sitting).
- **Cognitive disease prevention** (4.2.2)—it is related to maintaining healthy mind and cognitive abilities
- **Neurorehabilitation** (4.2.3)—when an accident happens that hinders brain functions (and subsequently, e.g., speech and mobility), neurorehabilitation helps with the person's recovery.
- **Overall wellbeing** (4.2.4)—chatbots help reduce the overall effect of depression symptoms.
- **Neural engineering** (4.2.5)—it is a discipline aimed at creating Brain–computer interfaces to control machines by using a person's mind.

4.2.1. Chronic diseases prevention

Smoking, excessive drinking, overeating, and physical inactivity are well-established risk factors that decrease human physical performance and increase the incidence of chronic diseases. Moreover, epidemiological work has identified modifiable lifestyle factors, such as poor diet and physical and cognitive inactivity, associated with the risk of reduced cognitive performance (72). Chronic diseases present an enormous burden to society by increasing medical costs and human suffering. The Body in Numbers system aims to influence such modifiable lifestyle risk factors in voluntarily enrolled individuals, thus decreasing the incidence of chronic diseases.

The Body in Numbers system enables the collection of large amounts of heterogeneous data and related metadata in relatively short periods, enabling repeated measurements of participants, data processing, and evaluation. The system output is a list of information for participants (recommended exercises) and management (employees in various physical "fitness" categories).

This service is provided to firms and institutions in a series of steps:

- Participants are registered into the system.
- The participant fills in the entry questionnaires containing questions aimed at food consumption habits (dietary and drinking habits), current lifestyle, significant health issues (diseases), intensity of medical checkups, smoking habit, and the intensity of smoking, whether or not the person is currently under stress and subjectively quantifies it, quantity and quality of sleep and more (filling out the questionnaires takes about 3–5 min).
- The system organization allows to manage 16 participants at once (they are eight measurement sites).
- The output from this system includes raw data (e.g., participants' blood pressure, cholesterol levels, and BMI), and the participants' categories based on their willingness to participate in exercises.
- The participants receive advice on which exercises are most beneficial to them individually and a set of eating habit adjustments (based on their favorite food instead of replacing them with something different).
- Currently, all these properties and calculations were evaluated based on the Czech Republic Healthcare System using the standardized SI units and measurements.

4.2.2. Cognitive disease prevention

The Body in Numbers system also helps with cognitive impairments by offering "brain games," which motivate users through various memory, attention, visuospatial perception, language and speech, or problem-solving exercises. These games are primarily intended for mobile devices, as the exercises can be done almost anywhere with little effort and setup time.

Brain games can be adjusted to the user's experience and needs. Some of these games can also be used with a special neurogear (the NeuroSky Mobile headset).

In the "Dangerous path" game (one of the memory-oriented exercises), the user has to memorize the location of "good" and "bad" objects at the beginning of the game. After this initial startup phase, the path is obfuscated by darkness, and the user needs to find a way from start to finish while evading all the "bad" objects.

The "Save the princess" game uses additional gear. The user is equipped with the NeuroSky Mobile sensor, which measures the levels of concentration and meditation. The goal of this game is to save a captured princess somewhere on the mobile screen while destroying "bad" objects through the usage of a cannon that shoots in a straight line within a time limit. The catch is that the shot from the cannon is only as strong as the users' concentration level. With low concentration, shooting the "bad" objects might be necessary multiple times.

The primary goal of these games is to motivate users to challenge their cognitive skills throughout short repeated sessions, working as a preventive measure.

4.2.3. Neurorehabilitation

In neurorehabilitation, it is necessary to adjust the exercise to the needs of each user to motivate them for repeatable sessions. This can be generally achieved by appealing to the user's hobbies and dynamically adjusting the rehabilitation task's difficulty level.

The "Smart Train" model was created in the Body in Numbers project. This is a handcrafted railway model that manipulates the train using the BCI. The user drives the train using a NeuroSky Mindwave Mobile headset (similar to the previous "Save the princess" game). The supported operations are as follows:

- The speed of the train model is affected by the concentration level achieved in the last 3 s (numeric average); this selects one of the four defined speeds.
- Change of direction (achieved by blinking twice in succession within 2 s).
- The train is stopped by achieving a meditation level of 100%.
- Lights on the train are turned on/off by blinking once.
- If 80% of concentration is achieved, the train starts to hoot.
- If the locomotive is kept stopped for 10 straight seconds, the conductor starts whistling.
- When the headset signal is classified as good, the locomotive starts up and starts to move based on the concentration levels.
- If the headset signal is low (e.g., due to poor electrode contact with skin), then the locomotive stops and turns off the engine.

The "Smart Train" neuroexercise challenges include stopping the train at a specific location (e.g., on the boarding platform) or achieving the maximum speed for a certain number of seconds. The rewarding motivational aspects of this exercise ("locomotive hooting," "conductors whistle," and "locomotive lights") reward and inform the user about the current state of the concentration levels and blinking. Also, the effort required to achieve high concentration levels can be modified manually inside the application, making these tasks easier or demanding based on the user's current needs.

Another example of a rehabilitation exercise is a ball being moved using a controller and a NeuroSky Mindwave Mobile headset. The direction of movement is set using the controller, and the speed is controlled by the users' concentration levels measured by the headset. In this case, the goal is to navigate the ball through a modifiable labyrinth of passages as fast as possible.

4.2.4. Overall wellbeing

In recent years, wellbeing has been an interesting topic for scientific research. Wellbeing fits various applications, from healthy eating to mindful living. Such services can be easily provided by various applications that notify the user of repetitive activities or maintain healthy habits. When the repetitive notifications become annoying, there is a chance to increase adherence and attrition by using gamification or psychology in natural language conversation.

The natural conversation can be scaled and automatized through dialog systems called chatbots. For example, a food tracking chatbot named Nombot (73) was developed. The dialog systems Woebot (74), Wysa (75), and Youper (76) serve as a treatment for people with symptoms of depression and anxiety. Lark (77) was designed to promote weight loss and other health behaviors related to diabetes prevention.

The approaches differ; Woebot utilizes gamification with various motivation types like points collection or higher-level unlocking. The remaining two are built on top of cognitive behavioral therapy. It combines behavioral techniques with cognitive psychology, the scientific study of mental processes, such as perception, memory, reasoning, decision making, and problem solving, to replace maladaptive behavior and faulty cognition with thoughts and self-statements that promote adaptive behavior (78).

Furthermore, although these dialog systems are far from the perfection of full human intervention, they are simple to use and available 24x7 with significant results, for instance, in the reduction of depression symptoms with randomized controlled trials (74).

In the Body in Numbers project, a wellbeing module is under development. It is interconnected with various projects mentioned above like "Smart Train" and investigated within research activities (79).

4.2.5. Neural engineering

The created BCI ERP Experiment, the "Guess the Number," uses a visual stimulation where the participant picks one number between 1 and 9 and focuses on it throughout the experiment without telling the experimenters (e.g., this number effectively becomes the target stimulus). In this case, the experimenters must correctly guess the number the participant thought.

Throughout the experiment, the participant is exposed to single pictures of each number between 1 and 9 (shown in random order) on the screen while the EEG signal and stimuli markers are recorded. Concurrently, experimenters observe average ERP waveforms for each number, search for the P300 component, and try to guess the number thought. During this time, the same "guessing" is done with a software component, which automatically identifies the number thought. The guess is verified at the end of the experiment when the participant is asked to reveal the number thought (80, 81).

A variation of this experiment was also done with a picture of nine tasks (e.g., opening a window, eating, calling the nurse), where instead of the numbers, one of the tasks was being used as the target stimulus. This variation was initially created to demonstrate the use of BCI to the public while also serving

users with limited mobility (e.g., people who cannot move their bodies freely and cannot talk) to communicate or accomplish some basic tasks.

4.3. Experience with the implementation of best practices

The devised Body in Numbers ontology was successfully overtaken and used within the components of the Body in Numbers project while maximizing the reusability of existing ontologies. The same was applied to the overtaken terms and the annotation properties, where all the properties were reused from existing ontologies. The resulting ontology was published on the Bioportal for community reuse and in various data formats: OWL, CSV, and RDF/XML.[16]

So far, the current use of the ontology is limited to the team of Body in Numbers, which accounts for only tens of people at a time. The community feedback is severely limited, even though the incremental processes for expanding the existing ontology are in place.

4.4. Future work

The following works are planned:

- Extension of terminology and ontology used in other application domains (cognitive disease prevention, neurorehabilitation, overall wellbeing, and neural engineering).
- Completion of individual application domains (like neurorehabilitation, overall wellbeing and neural engineering), data collection, and involvement of machine learning in the evaluation of results.
- Expansion of the used terms based on community feedback.
- Increase in the number of indexers on which the ontology is available.

It is clear to us that this paper touches on the current absence of best practices in the health-related data lifecycle and presents, in particular, technical and methodological solutions that contribute to the sharing of annotated data from different providers. However, we have not practically touched on the nontechnical issues that accompany the possible use of these best practices.

In general, using best practices that contribute to reproducible data collection, annotation, analysis, interpretation, and publication is a laborious process that requires extensive knowledge and community involvement from all stakeholders. The resulting rewards (i.e., any additional benefits from providing well-described and reproducible data and results) are often perceived as remote and uncertain. Such a process is then necessarily more compromised in its implementation. Thus, the acquisition of clean or well-annotated data (there is no substitute for clean data), the use of shared terminologies, ontologies, and data formats (sharing them contributes to mutual understanding; AI methods better handle extensive data collections), or software and data engineering practices (they contribute to reducing technical debt) are goals that can only be gradually achieved if the environment supports and rewards such efforts. Providing best practice methods, resources, and tools for the health-related data lifecycle presented in this paper is one support in this effort. Another important support may be a gradual culture change regarding reproducibility and openness of health-related data and results and the valuing and demanding of this culture by scientific journals (which is already happening today), coupled with the provision of additional means of technical and methodological support [e.g., repositories for long-term preservation of data such as (82) or repositories of community-contributed protocols used in data acquisition process such as (83)].

Another, but probably limited, solution is to rely on artificial intelligence methods that can extract some relevant information even from large amounts of raw, noisy data.

Author contributions

PB, RM, JS, and VV wrote and edited the text. All authors contributed to the article and approved the submitted version.

Funding

This publication was supported by the UWB grant SGS-2022-016, Advanced methods of data processing and analysis. This work was published with the financial support of the European Union, as part of the project entitled Development of capacities and environment for boosting the international, intersectoral and interdisciplinary cooperation at UWB, Project Reg. No. CZ.02.2.69/0.0/0.0/18_054/0014627.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

[16]The BiN ontology can be found here: https://bioportal.bioontology.org/ontologies/BIN/?p=summary (Accessed 2022-06-08).

Publisher's note

References

1. Salem R, Elsharkawy B, Kader HA. An ontology-based framework for linking heterogeneous medical data. *International Conference on Advanced Intelligent Systems, Informatics*. Cham (Switzerland): Springer (2016). p. 836–45.

2. Mavrogiorgou A, Kiourtis A, Kyriazis D. A plug "n" play approach for dynamic data acquisition from heterogeneous IoT medical devices of unknown nature. *Evol Syst*. (2020) 11:269–89. doi: 10.1007/s12530-019-09286-5

3. Abrams MB, Bjaalie JG, Das S, Egan GF, Ghosh SS, Goscinski WJ, et al. A standards organization for open, fair neuroscience: the international neuroinformatics coordinating facility. *Neuroinformatics* (2021) 20:1–12. doi: 10.1007/s12021-020-09509-0

4. Adel E, El-Sappagh S, Barakat S, Elmogy M. Ontology-based electronic health record semantic interoperability: a survey. *U-healthcare monitoring systems*. Amsterdam: Elsevier (2019). p. 315–52.

5. Ascoli GA, Maraver P, Nanda S, Polavaram S, Armañanzas R. Win–win data sharing in neuroscience. *Nat Methods*. (2017) 14:112–6. doi: 10.1038/nmeth.4152

6. Meyer MN. Practical tips for ethical data sharing. *Adv Meth Pract Psychol Sci*. (2018) 1:131–44. doi: 10.1177/2515245917747656

7. Pittet P, Cruz C, Nicolle C. *Guidelines for a dynamic ontology—integrating tools of evolution, versioning in ontology*. Paris: SciTePress (2011). p. 173–9. doi: 10.5220/0003653201730179

8. Eshghishargh A. *Ontologies in neuroscience, their application in processing questions* [Ph.D. thesis]. Melbourne: The University of Melbourne (2019).

9. Fudholi DH, Rahayu W, Pardede E, Hendrik. A data-driven approach toward building dynamic ontology. *Information, Communication Technology—EurAsia Conference*. Berlin, Heidelberg: Springer (2013). p. 223–32.

10. Bonacin R, Hornung H, Dos Reis JC, Pereira R, Baranauskas MCC. Pragmatic aspects of collaborative problem solving: towards a framework for conceptualizing dynamic knowledge. *International Conference on Enterprise Information Systems*. Berlin, Heidelberg: Springer (2012). p. 410–26.

11. Beck K, Beedle M, Van Bennekum A, Cockburn A, Cunningham W, Fowler M, et al. Manifesto for agile software development (2001). Available at: https://agilemanifesto.org/ (Accessed 2022-06-08).

12. Dyck A, Penners R, Lichter H. Towards definitions for release engineering and devops. *2015 IEEE/ACM 3rd International Workshop on Release Engineering*. Florence: IEEE (2015). p. 3.

13. Jabbari R, Petersen K, Tanveer B. What is devops? A systematic mapping study on definitions and practices. *Proceedings of the Scientific Workshop Proceedings of XP2016*. New York, NY, United States: Association for Computing Machinery (2016). p. 1–11.

14. Erich F, Amrit C, Daneva M. A qualitative study of devops usage in practice. *J Softw Evol Process*. (2017) 29:e1885. doi: 10.1002/smr.1885

15. Liebmann L. Reasons why dataops is essential for big data success. *IBM Big Data & Analytics Hub. Retrieved October*. (2014) 28:2020.

16. Palmer A. *From devops to dataops*. Tamr Inc (2015).

17. DataKitchen. *Dataops is not just devops for data*. Medium (2018).

18. Sculley D, Holt G, Golovin D, Davydov E, Phillips T, Ebner D, et al. Hidden technical debt in machine learning systems. *Advances in neural information processing systems*. Curran Associates, Inc. (2015). p. 2503–11.

19. Prasad P, Cappelli W, Fletcher C. Market guide for AIOps platforms (2017). Available at: https://www.gartner.com/en/documents/3772124 (Accessed 2022-06-08).

20. Tom E, Aurum A, Vidgen R. An exploration of technical debt. *J Syst Softw*. (2013) 86:1498–516. doi: 10.1016/j.jss.2012.12.052

21. Guo Y, Seaman C, Gomes R, Cavalcanti A, Tonin G, Da Silva FQ, et al. Tracking technical debt—an exploratory case study. *2011 27th IEEE international conference on software maintenance (ICSM)*. Williamsburg: IEEE (2011). p. 528–31.

22. Klinger T, Tarr P, Wagstrom P, Williams C. An enterprise perspective on technical debt. *Proceedings of the 2nd Workshop on Managing Technical Debt*. New York, NY, United States: Association for Computing Machinery (2011). p. 35–8. (Accessed 2022-05-12).

23. Seaman C, Guo Y, Zazworka N, Shull F, Izurieta C, Cai Y, et al. Using technical debt data in decision making: potential decision approaches. *2012 Third International Workshop on Managing Technical Debt (MTD)*. Zurich: IEEE (2012). p. 45–8.

24. Lim E, Taksande N, Seaman C. A balancing act: what software practitioners have to say about technical debt. *IEEE Softw*. (2012) 29:22–7. doi: 10.1109/MS.2012.130

25. Snipes W, Robinson B, Guo Y, Seaman C. Defining the decision factors for managing defects: a technical debt perspective. *2012 Third International Workshop on Managing Technical Debt (MTD)*. Zurich: IEEE (2012). p. 54–60.

26. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The fair guiding principles for scientific data management, stewardship. *Sci Data*. (2016) 3:1–9. doi: 10.1038/sdata.2016.18

27. Roche C. Ontoterminology: how to unify terminology, ontology into a single paradigm. *LREC 2012, Eighth International Conference on Language Resources and Evaluation*. European Language Resources Association (2012). p. 2626.

28. Zemmouchi-Ghomari L, Ghomari AR. Ontology versus terminology, from the perspective of ontologists. *Int J Web Sci*. (2012) 1:315–31. doi: 10.1504/IJWS.2012.052531

29. Gruber TR. A translation approach to portable ontology specifications. *Knowl Acquis*. (1993) 5:199–220. doi: 10.1006/knac.1993.1008

30. Borst WN. Construction of engineering ontologies for knowledge sharing and reuse (1997). University of Twente. Available at: https://research.utwente.nl/en/publications/construction-of-engineering-ontologies-for-knowledge-sharing-and- (Accessed 2022-08-22).

31. Le Franc Y, Davison AP, Gleeson P, Imam FT, Kriener B, Larson SD, et al. Computational neuroscience ontology: a new tool to provide semantic meaning to your models. *BMC Neurosci*. (2012) 13:1–2. doi: 10.1186/1471-2202-13-S1-P149

32. Avery J, Yearwood J. DOWL: a dynamic ontology language. *ICWI*. (2003) 2003:985–8. Available at: https://www.researchgate.net/publication/220970000DOWL_A_Dynamic_Ontology_Language

33. Kumar A, Smith B. Oncology ontology in the NCI thesaurus. *Conference on Artificial Intelligence in Medicine in Europe*. Berlin, Heidelberg: Springer (2005). p. 213–20.

34. Frishkoff G, LePendu P, Frank R, Liu H, Dou D. Development of neural electromagnetic ontologies (NEMO): ontology-based tools for representation and integration of event-related brain potentials. *Nat Preced*. (2009) 1. doi: 10.1038/npre.2009.3458.1

35. Frishkoff G, Sydes J, Mueller K, Frank R, Curran T, Connolly J, et al. Minimal information for neural electromagnetic ontologies (MINEMO): a standards-compliant method for analysis and integration of event-related potentials (ERP) data. *Stand Genomic Sci*. (2011) 5:211–23. doi: 10.4056/sigs.2025347

36. Courtot M, Gibson F, Lister AL, Malone J, Schober D, Brinkman RR, et al. Mireot: The minimum information to reference an external ontology term. *Appl Ontol*. (2011) 6:23–33. doi: 10.3233/AO-2011-0087

37. Stead M, Halford JJ. A proposal for a standard format for neurophysiology data recording and exchange. *J Clin Neurophysiol*. (2016) 33:403. doi: 10.1097/WNP.0000000000000257

38. Teeters JL, Godfrey K, Young R, Dang C, Friedsam C, Wark B, et al. Neurodata without borders: creating a common data format for neurophysiology. *Neuron*. (2015) 88:629–34. doi: 10.1016/j.neuron.2015.10.025

39. INCF. Standards and best practices portfolio (2022). Available at: https://www.incf.org/resources/sbps (Accessed 2022-09-05).

40. Martone M, Gerkin R, Moucek R, Das S, Goscinski W, Hellgren-Kotaleski J, et al. Nix-neuroscience information exchange format. *F1000Research* (2020) 9:1–8. doi: 10.7490/f1000research.1117858.1

41. Folk M, Heber G, Koziol Q, Pourmal E, Robinson D. An overview of the HDF5 technology suite, its applications. *Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases* (2011). p. 36–47.

42. Stoewer A, Kellner CJ, Benda J, Wachtler T, Grewe J. File format, library for neuroscience data and metadata. *Front Neuroinform*. (2014) 8:10–3389. doi: 10.3389/conf.fninf.2014.18.00027

43. Sprenger J, Zehl L, Pick J, Sonntag M, Grewe J, Wachtler T, et al. odMLtables: a user-friendly approach for managing metadata of neurophysiological experiments. *Front Neuroinform*. (2019) 13:62. doi: 10.3389/fninf.2019.00062

44. Grewe J, Wachtler T, Benda J. A bottom-up approach to data annotation in neurophysiology. *Front Neuroinform*. (2011) 5:16. doi: 10.3389/fninf.2011.00016

45. Sporny M, Longley D, Kellogg G, Lanthaler M, Champin P-A, Lindström N. *JSON-LD 1.1—a JSON-based serialization for linked data* [Ph.D. thesis]. W3C (2019).

46. Pernet CR, Appelhoff S, Gorgolewski KJ, Flandin G, Phillips C, Delorme A, et al. EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Sci Data*. (2019) 6:1–5. doi: 10.1038/s41597-019-0104-8

47. Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data*. (2016) 3:1–9. doi: 10.1038/sdata.2016.44

48. Ruebel O, Tritt A, Dichter B, Braun T, Cain N, Clack N, et al. NWB:N 2.0: an accessible data standard for neurophysiology. *ResearchGate* (2019). doi: 10.1101/523035

49. Horrocks I. What are ontologies good for? *Evolution of semantic systems*. Springer (2013). p. 175–188.

50. Johnstone SJ, Barry RJ, Clarke AR. Ten years on: a follow-up review of ERP research in attention-deficit/hyperactivity disorder. *Clin Neurophysiol*. (2013) 124:644–57. doi: 10.1016/j.clinph.2012.09.006

51. Enriquez-Geppert S, Huster RJ, Herrmann CS. EEG-neurofeedback as a tool to modulate cognition, behavior: a review tutorial. *Front Hum Neurosci*. (2017) 11:51. doi: 10.3389/fnhum.2017.00051

52. Tian Y, Xu W, Zhang H, Tam KY, Zhang H, Yang L, et al. The scalp time-varying networks of N170: reference, latency, information flow. *Front Neurosci*. (2018) 12:250. doi: 10.3389/fnins.2018.00250

53. Qin Y, Jiang S, Zhang Q, Dong L, Jia X, He H, et al. BOLD-fMRI activity informed by network variation of scalp EEG in juvenile myoclonic epilepsy. *NeuroImage: Clin*. (2019) 22:101759. doi: 10.1016/j.nicl.2019.101759

54. Taylor JR, Williams N, Cusack R, Auer T, Shafto MA, Dixon M, et al. The Cambridge centre for ageing and neuroscience (Cam-CAN) data repository: structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *Neuroimage*. (2017) 144:262–9. doi: 10.1016/j.neuroimage.2015.09.018

55. Zuo X-N, Xing X-X. Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci Biobehav Rev*. (2014) 45:100–18. doi: 10.1016/j.neubiorev.2014.05.009

56. Gongora M, Nicoliche E, Magalhães J, Vicente R, Teixeira S, Bastos VH, et al. Event-related potential (p300): the effects of levetiracetam in cognitive performance. *Neurol Sci* (2020) 42:1–8. doi: 10.1007/s10072-020-04786-8

57. DiStefano C, Dickinson A, Baker E, Jeste SS. EEG data collection in children with ASD: the role of state in data quality and spectral power. *Res Autism Spectr Disord*. (2019) 57:132–44. doi: 10.1016/j.rasd.2018.10.001

58. Fisher RA. et al *The design of experiments*. London and Edinburgh: Oliver and Boyd (1960).

59. Ioannidis JP. Why most published research findings are false. *PLoS Med*. (2005) 2:e124. doi: 10.1371/journal.pmed.0020124

60. Trochim WM. Theory of reliability. *Web Center for Social Research Methods*. Research Methods of Knowledge (2006).

61. Trochim W. Types of reliability. Research methods knowledge base. *Web Center for Social Research Methods* (2006). Available at: http://www.socialresearchmethods.net/kb/reltypes.php (Accessed 2022-06-08).

62. Barba LA. Terminologies for reproducible research [Preprint] *arXiv*. 2018). Available at: http://arxiv.org/1802.03311 (Accessed 2022-06-08).

63. Liberman M. Replicability vs. reproducibility—or is it the other way around. *The Language Log*. Language Log (2015).

64. Brůha P, Mouček R, Vacek V, Šnejdar P, Černá K, Řehoř P. Collection of human reaction times and supporting health related data for analysis of cognitive and physical performance. *Data Br*. (2018) 17:469–511. doi: 10.1016/j.dib.2018.01.025

65. Bruha P, Moucek R, Snejdar P, Bohmann D, Kraft V, Rehor P. Exercise and wellness health strategy framework - software prototype for rapid collection and storage of heterogeneous health related data. *HEALTHINF*. Setúbal: SciTePress (2017). p. 477–83.

66. Cho H, Ahn M, Ahn S, Kwon M, Jun SC. EEG datasets for motor imagery brain–computer interface. *GigaScience*. (2017) 6:gix034. doi: 10.1093/gigascience/gix034

67. Picton T, Bentin S, Berg P, Donchin E, Hillyard S, Johnson R Jr., et al. Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*. (2000) 37:127–52. doi: 10.1111/1469-8986.3720127

68. Brůha P, Mouček R, Volf P, Šimečkové L, Šťáva O. On architecture of bodyinnumbers exercise and wellness health strategy framework. *Proceedings of the 4th International Conference on Medical and Health Informatics (ICMHI 2020)*. New York, NY: Association for Computing Machinery (2020). p. 145–49. Available at: http://doi.org/10.1145/3418094.3418102

69. LeBlanc M, Beaulieu-Bonneau S, Mérette C, Savard J, Ivers H, Morin CM. Psychological and health-related quality of life factors associated with insomnia in a population-based sample. *J Psychosom Res*. (2007) 63:157–66. doi: 10.1016/j.jpsychores.2007.03.004

70. Shepherd D, Welch D, Dirks KN, Mathews R. Exploring the relationship between noise sensitivity, annoyance and health-related quality of life in a sample of adults exposed to environmental noise. *Int J Environ Res Public Health*. (2010) 7:3579–94. doi: 10.3390/ijerph7103580

71. Sampasa-Kanyinga H, Standage M, Tremblay MS, Katzmarzyk P, Hu G, Kuriyan R, et al. Associations between meeting combinations of 24-h movement guidelines and health-related quality of life in children from 12 countries. *Public Health*. (2017) 153:16–24. doi: 10.1016/j.puhe.2017.07.010

72. Bruha P, Moucek R, Vacek V, Snejdar P, Vareka L, Kraft V, et al. Advances in building bodyinnumbers exercise and wellness health strategy framework. *HEALTHINF*. Setúbal: SciTePress (2018). p. 548–54.

73. Graf B, Krüger M, Müller F, Ruhland A, Zech A. Nombot: simplify food tracking. *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*. New York, NY, United States: Association for Computing Machinery (2015). p. 360–3.

74. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (WOEBOT): a randomized controlled trial. *JMIR Ment Health*. (2017) 4:e19. doi: 10.2196/mental.7785

75. Inkster B, Sarda S, Subramanian V, Inkster B, Sarda S, Subramanian V, et al. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth*. (2018) 6:e12106. doi: 10.2196/12106

76. Mehta A, Niles AN, Vargas JH, Marafon T, Couto DD, Gross JJ, et al. Acceptability and effectiveness of artificial intelligence therapy for anxiety and depression (youper): Longitudinal observational study. *J Med Internet Res*. (2021) 23:e26771. doi: 10.2196/26771

77. Stein N, Brooks K. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. *JMIR diabetes*. (2017) 2:e8590. doi: 10.2196/diabetes.8590

78. Beck AT. *Cognitive therapy and the emotional disorders*. City of Westminster, London: Penguin (1979).

79. Salamon J. Influencing an artificial conversational entity by information fusion (2020). Available at: http://www.kiv.zcu.cz/site/documents/verejne/vyzkum/publikace/technicke-zpravy/2020/Rigo˙Salamon˙2020˙04.pdf (Accessed 2022-06-08).

80. Vareka L. *Methods for signal classification and their application to the design of brain-computer interfaces* [Ph.D. thesis]. Pilsen: Faculty of Applied Sciences, University of West Bohemia (2018).

81. Vařeka L, Ladouce S. Prediction of navigational decisions in the real-world: a visual p300 event-related potentials brain-computer interface. *Int J Hum Comput Interact*. (2021) 37:1375–89. doi: 10.1080/10447318.2021.1888510

82. Nature Portfolio. Protocol exchange (2022). Available at: https://protocolexchange.researchsquare.com (Accessed 2022-06-08).

83. Harvard College. Harvard dataverse (2022). Available at: https://dataverse.harvard.edu (Accessed 2022-06-08).

frontiers | Frontiers in Digital Health

TYPE Brief Research Report
PUBLISHED 08 December 2022
DOI 10.3389/fdgth.2022.1065581

OPEN ACCESS

EDITED BY
Harry Hochheiser,
University of Pittsburgh, United States

REVIEWED BY
Terri Elizabeth Workman,
George Washington University, United States,
Md Adnanul Islam,
Monash University, Australia

*CORRESPONDENCE
Sima Azizi
azizi.sima.67@gmail.com

SPECIALTY SECTION
This article was submitted to Health Informatics, a section of the journal Frontiers in Digital Health

RECEIVED 09 October 2022
ACCEPTED 21 November 2022
PUBLISHED 08 December 2022

CITATION
Azizi S, Hier DB and Wunsch II DC (2022) Enhanced neurologic concept recognition using a named entity recognition model based on transformers.
Front. Digit. Health 4:1065581.
doi: 10.3389/fdgth.2022.1065581

Enhanced neurologic concept recognition using a named entity recognition model based on transformers

Sima Azizi[1]*, Daniel B. Hier[1,2] and Donald C. Wunsch II[1,3]

[1]Applied Computational Intelligence Laboratory, Department of Electrical & Computer Engineering, Missouri University of Science & Technology, Rolla, MO, United States, [2]Department of Neurology and Rehabilitation, University of Illinois at Chicago, Chicago, IL, United States, [3]National Science Foundation, ECCS Division, Arlington, VA, United States

Although deep learning has been applied to the recognition of diseases and drugs in electronic health records and the biomedical literature, relatively little study has been devoted to the utility of deep learning for the recognition of signs and symptoms. The recognition of signs and symptoms is critical to the success of deep phenotyping and precision medicine. We have developed a named entity recognition model that uses deep learning to identify text spans containing neurological signs and symptoms and then maps these text spans to the clinical concepts of a neuro-ontology. We compared a model based on convolutional neural networks to one based on bidirectional encoder representation from transformers. Models were evaluated for accuracy of text span identification on three text corpora: physician notes from an electronic health record, case histories from neurologic textbooks, and clinical synopses from an online database of genetic diseases. Both models performed best on the professionally-written clinical synopses and worst on the physician-written clinical notes. Both models performed better when signs and symptoms were represented as shorter text spans. Consistent with prior studies that examined the recognition of diseases and drugs, the model based on bidirectional encoder representations from transformers outperformed the model based on convolutional neural networks for recognizing signs and symptoms. Recall for signs and symptoms ranged from 59.5% to 82.0% and precision ranged from 61.7% to 80.4%. With further advances in NLP, fully automated recognition of signs and symptoms in electronic health records and the medical literature should be feasible.

KEYWORDS

named entity recognition, clinical concepts, concept extraction, phenotype, transformers, natural language processing, annotation

I. Introduction

Several factors have accelerated interest in the automated recognition of clinical concepts in unstructured text held in electronic health records and electronic publications (1). First, most paper medical records have been converted to electronic health records (EHRs) (2) with as much as 80% of the data held as unstructured text

(3). Second, most medical journals are available electronically (4). Third, the deep phenotyping and precision medicine initiatives have made the detailed description of patient signs and symptoms a key piece of data (5,6). Fourth, automated clinical concept recognition is an important area of natural language processing (NLP) research. Automated concept recognition is closely related to the NLP problems of text mining and named entity recognition. Other important NLP research areas include machine translation, text classification, text clustering, speech recognition, question answering, text summarization, sentiment analysis, picture captioning, and natural language understanding (7–14).

Krauthammer and Nenadic (1) have divided concept recognition (variously called term identification, concept extraction, and information extraction) into three steps: term recognition (identification of the text span corresponding to the clinical concept), term classification (identification of the class membership of the term, i.e., drug, disease, sign, symptom, etc.), and term mapping (linking of the term to an entry in a standard vocabulary with an identification code which is also known as "concept normalization" (15)). Clinical concept recognition is closely related to the NLP problem of named entity recognition (NER) in which text spans referring to named entities (people, places, organizations, etc.) are tagged and mapped to dictionaries, gazetteers, or other registries (16).

Text spans that encode clinical concepts (diseases, drugs, signs, symptoms, etc.) can be mapped (normalized) to hierarchical ontologies that include SNOMED CT with 352,000 concepts, the Human Phenotype Ontology (HPO) with 20,000 concepts, the Online Mendelian Inheritance in Man ontology (OMIM) with 97,000 concepts, or the UMLS Metathesaurus with 4.6 million concepts (17–20). The NLM UMLS Metathesaurus maintains interchangeable machine-readable codes for SNOMED CT, UMLS, HPO, and the OMIM.

Initial NER systems for clinical concept recognition were either dictionary-based, or rule-based (1,21,22). Some second-generation NER systems were based on machine learning algorithms such as conditional random fields, support vector machines, and hidden Markov models (23,24). Other second-generation NER systems developed as an outgrowth of advances in semantic and syntactic analysis (25,26). MetaMap utilizes linguistic analysis and statistical algorithms to identify clinical concepts in unstructured text and maps them to machine-readable codes in the UMLS (27,28). The UMLS has grown from 900,000 concepts, and 2 million names in 2004 (29) to 4.6 million concepts and 17 million names in 2022 (20). MetaMap tokenizes text input, finds sentence boundaries, and uses lexical and syntactic analysis to identify candidate phrases for mapping to concepts in the UMLS. Candidate phrases are compared to target strings in the UMLS, lists of potential clinical concepts are generated, and scored by statistical algorithms. MetaMap can recognize abbreviations, acronyms, and negation, can generate word variants, and can perform word sense disambiguation (27). In a preliminary study, we found that MetaMap can identify signs and symptoms in neurological case histories with an accuracy of 55–84% (30). Most MetaMap errors were false negatives due to a failure to recognize neurological concepts that had been expressed as descriptions (e.g., *reflexes were absent*) as opposed to those expressed as discrete lexical items (e.g., *hyporeflexia*). In their 2017 literature review of automated information extraction, Wang et al. (31) reviewed 263 information extraction studies and found most centered on identifying diseases or drugs. The most common systems used were MetaMap, MedLEE, and cTAKES (32–36) followed by traditional machine learning algorithms (conditional random fields, support vector machines, random forests, decision trees, and naive Bayes).

Third-generation systems for NER are built on deep learning (37–40). Lample et al. suggested a model for named entity recognition based on an RNN (recurrent neural network) with bidirectional LSTM (long short term memory) and conditional random fields (CRFs). Vani et al. (41) proposed a "grounded" RNN to predict medical diagnoses based on text from patient discharge summaries. Liu et al. (42) found that on a task to label protected health information in medical records that RNNs based on bidirectional LSTM outperformed those that used CRFs. An LSTM NER model with conditional random fields (CRFs) has been used to identify five classes of chemicals, species, genes/proteins, cell lines, and diseases (43). Hybrid methods that combine rule-based and machine learning-based methods have been proposed to identify protected health information (PHI) in clinical discharge summaries (44). Liu et al. (42) developed a hybrid system to identify clinical information by ensemble learning that combined the instances predicted from a bidirectional LSTM, a CRF model, and a rule-based system (45,46). Gehrmann et al. (47) used a convolutional neural network (CNN) for ten phenotyping tasks and compared it with other common NLP models. Arbabi et al. (48) have created a neural concept recognizer (NCR) that uses CNNs and word embedding to recognize clinical concepts in unstructured text. The NCR uses an encoder to convert input phrases to word vectors and word embedding to convert entries in the target ontology into word vectors. The similarity between the input phrases and concepts in the target ontology is calculated by the dot product. For concept recognition in PubMed abstracts or clinical notes, the NCR outperformed the NCBO Annotator and BioLark (49). RNNs and variants can handle long-term dependency in text, but only for a limited span length. The deep learning architecture transformers can process longer text spans and has shown improved performance on NLP tasks (50). Bidirectional encoder representations from transformers (BERT) have outperformed other neural network architectures on named

entity recognition (51,50). For clinical concept recognition, BERT models that are pre-trained on the medical literature (BioBERT) or clinical notes (ClinicalBERT) outperform BERT models pre-trained on general corpora by at least 1% (52–55).

A. Proposed approach

Although considerable work has been done on automated concept identification of drugs and diseases, less work has been done on the automated identification of signs and symptoms (52). Identifying signs and symptoms is critical to precision medicine and deep phenotyping (56). To make the problem tractable, we limited the signs and symptoms to the specialty of neurology and restricted the target ontology to a neuro-ontology with 1,600 concepts (57). Automating the recognition of signs and symptoms is more challenging than automating the recognition of diseases or drugs for three reasons. First, many neurological signs and symptoms have multiple synonyms; something that is not typical with diseases or drugs. For example, an expressionless face may be described as a "masked face," or "hypomimia." Second, physicians variably choose to record signs and symptoms as *descriptions* or as *names*. For example, a patient with diplopia can be described as "seeing double" or a patient with nausea can be described as "sick to their stomach." In contrast, physicians uniformly identify drugs and diseases by name and not by description. Third, the meaning of a term may depend on context. For example, to a neurologist *ptosis* is a droopy eyelid, but to a gynecologist, *ptosis* is a prolapsed uterus.

We propose to identify and normalize the neurological signs and symptoms found in the unstructured text in two steps: first, we have trained a neural network-based named entity recognition model to identify text spans that contain clinical concepts (signs and symptoms). Second, we have normalized identified text spans by mapping them to clinical concepts in a neuro-ontology using a look-up table and similarity metric (**Figure 1**).

Since neurologic signs and symptoms can be extracted from both the medical literature and electronic health records, we have tested the concept identification pipeline on three corpora: case histories from neurological textbooks, neurological clinical synopses from the Online Mendelian Inheritance of Man (OMIM), and physician neurological notes from an electronic health record. With this work, we propose to address four questions:

1. Does writing style differ by corpus?
2. Does the accuracy of concept recognition differ by corpus?
3. Is the accuracy of clinical concept recognition reduced with longer text spans?
4. Does concept recognition based on BERT outperforms concept recognition based on CNNs?

Although the superiority of BERT over other neural networks for concept identification is well-established, the contribution of this work is to demonstrate that the accuracy of concept identification depends upon text span length and corpus writing style.

II. Methods

A. Corpora

We identified signs and symptoms (clinical concept identification) in three corpora: neurological case histories from five neurological textbooks (referred to as *Textbook Corpus* (58–62), clinical synopses of neurological disease from the Online Mendelian Inheritance of Man (referred to as *OMIM Corpus*) (18), and neurology physician notes from the electronic health record of the University of Illinois at Chicago (referred to as *EHR Corpus*). The use of de-identified physician notes was approved by the Institutional Review Board of the University of Illinois at Chicago. Corpora were converted to plain text files and pre-processed using python. Email addresses, URLs, HTML, special characters, and unnecessary punctuation were removed using regular expressions in python. Contractions were replaced with the expanded form. Misspelled words, separated words, and

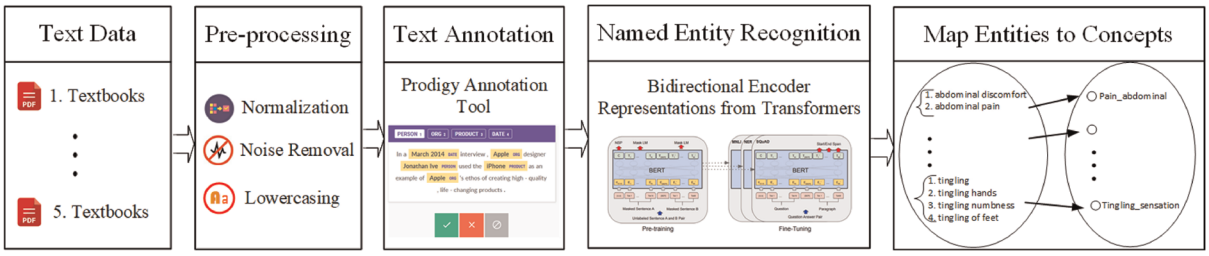


FIGURE 1
Overview of the pipeline that recognizes text spans that are clinical concepts in three corpora: Textbook neurology case histories, EHR physician notes, and Clinical synopses in the OMIM. Text spans are normalized by mapping to clinical concepts in a neuro-ontology.

hyphenated words were corrected manually using the spelling correction tool in Microsoft Word. Abbreviations were not edited. The pre-processed files were manually inspected for errors and converted to JSONL files.

B. Text annotation

Signs and symptoms in JSONL files were annotated by a neurologist using the Prodigy annotation tool (63,64). An inter-rater reliability study with two other raters based on fifteen neurology notes showed an unadjusted agreement rate for text span annotation of 89% and a kappa statistic of 0.85 (65).

Each sign or symptom was tagged as a unigram, bigram, trigram, tetragram, extended, compound, or tabular concept. Unigrams were signs and symptoms of length one-word such as *alexia*, *hyperreflexia*, or *bradykinesia*. Bigrams were signs and symptoms of length two-words such as *double vision*, *facial weakness*, and *poor balance*. Trigrams were signs of symptoms of length three-words such as *absent ankle reflex*, *impaired hand dexterity*, or *weak ankle dorsiflexors*. Tetragrams were four-word signs and symptoms such as *relative afferent pupil defect* and *Hoffman sign was present*. Text spans were tagged as *extended* when signs and symptoms were more than four words, such as *hand grip was very weak* and *barely able to lift his legs off the bed*. Text spans were tagged as *compound* when more than one sign or symptom was combined in a single text span such as *decreased vibratory sensation, joint position, and pinprick below the knees*. *Tabular* concepts with separate columns for the right and left sides of the body were found only in the EHR notes. Examples of concepts in table form included biceps weakness represented as *[biceps strength 3 3]* (meaning that biceps strength was 3/5 on both right and left sides) or knee hyperreflexia represented as *[knee reflexes 4+ 4+]* (meaning that the knee reflex was 4+ on both right and left sides). Text span annotations were stored in an SQLite database and exported in JSONL format for further processing in the *spaCy* (Explosion, Berlin, Germany) python programming environment.

TABLE I. Performance of CNN and BERT neural networks on concept extraction task.

Corpus	NN	*F*	Precision	Recall
EHR	CNN	57.5	65.6	51.2
	BERT	61.7	64.0	59.5
Textbook	CNN	69.0	70.1	67.9
	BERT	73.0	73.6	72.3
OMIM	CNN	76.2	78.8	73.7
	BERT	80.4	79.0	82.0

C. NN model training and evaluation

Two neural network models were trained to recognize text spans that encoded clinical concepts in text corpora. Both models were based on NER pipelines. NER pipelines identify a named entity in a text span and assign the named entity to a predefined category. Each NN model assigned text spans to one of the seven defined categories of clinical concepts (unigram, bigram, trigram, tetragram, extended, compound, and tabular). For each corpus, 80% of the instances were used for training and 20% for evaluation. The baseline NN was the default spaCy named entity recognition model based on a four-layer convolutional neural network (CNN) that looks at four words on either side of each token using the NER pipeline and *tok2vec* with an initial learning rate 1×10^{-3}. The standard word vectors included with spaCy were used for word embedding.

The second named entity recognition model was based on BERT (51). The BERT base model was implemented in spaCY (66) and consisted of 12 layers of transformer encoder, 12 attention heads, 786 hidden size, and 100 M parameters. The BERT model was pre-trained with publicly available weights and fine-tuned using our training set. We used the Adam optimizer with a learning rate of 5×10^{-5}, $\beta_1 = 0.9$, $\beta_2 = 0.99$, a learning rate warm-up over the first 500 steps, and a linear decay learning rate. The dynamic batch size was set according to the longest sequence in the batch. The training was conducted over 20,000 steps. The mini-batch size dynamically changed according to the longest sequence in the batch. The largest padded size for batch sequences was 2,000, and the buffer was 256. A GELU activation function was used. For each corpus and each model, the *F* score, precision, and recall were computed (**Table I**).

D. Mapping text spans to concepts in the neuro-ontology (normalization)

Candidate text spans identified by the CNN and BERT models were mapped to neurological concepts in the target neuro-ontology. The neuro-ontology (57) is a hierarchical ontology with 1,600 concepts constructed with the Protégé ontology editor (67). All concepts map to terms and CUIs (unique concept identifiers) from the UMLS (20). The highest levels of neuro-ontology correspond to the main elements of the neurological examination: mental status, cranial nerves, motor, sensory, reflexes, and symptoms. The neuro-ontology is available for download in CSV or OWL format at the National Center for Biomedical Ontologies BioPortal (https://bioportal.bioontology.org/ontologies/NEO).

We manually created a look-up table by mapping 3,500 potential target phrases to concepts in the neuro-ontology. Similarities between the candidate text spans (from either the

CNN or BERT models) and target phrases in the lookup table were calculated using the *doc.similarity* method from spaCy (66). Both the candidate text span and the target phrase were converted to *doc* objects using the spaCy NLP pipeline (https://spacy.io/api/doc/#similarity), which converts each token in the phrase into a word vector. The similarity is the cosine distance between the word vectors from the two phrases and ranges between 0.0 (least similar) and 1.0 (most similar). We mapped the candidate text span to its most similar target text span in the look-up table and retrieved the corresponding concept name and UMLS CUI from the neuro-ontology (57).

III. Results

A. Writing style and accuracy varied by corpus

The OMIM corpus used more unigrams and digrams to encode signs and symptoms and had shorter spans of text annotations than the EHR corpus or the Textbook corpus (**Figure 2**). The length of annotations (histogram insets, **Figure 2**) was longer for the EHR corpus. Extended annotations were more frequent in the EHR corpus and Textbook corpus. Only the EHR corpus had tabular annotation (clinical concepts expressed in table format). Performance on the concept identification task differed by corpus; *F*, precision, and recall were highest for the OMIM corpus and lowest for the EHR corpus (**Table 1**).

B. Performance of NER model decreased with the increasing text span length

For all three corpora, the recognition of clinical concepts as measured by *F* scores was better for shorter text spans (**Figures 3**A,B). This applied to both the CNN and the BERT models for concept identification (**Table 1**). *F* was highest for unigrams (one-word concepts like *ataxia*, *diplopia*, *aphasia*) for all three corpora. In general, performance on bigrams was better than trigrams, and performance on trigrams was better than tetragrams. Performance tended to be worse for text spans greater than four words (extended), or text spans with compound constructions such as *weakness of the biceps, triceps, and deltoids*.

C. Performance varied by neural network model

For all three corpora, BERT outperformed the CNN neural network for the recall of clinical concepts. Precision in clinical concept identification was about the same for all three corpora when BERT was compared to the CNN model (**Table 1**).

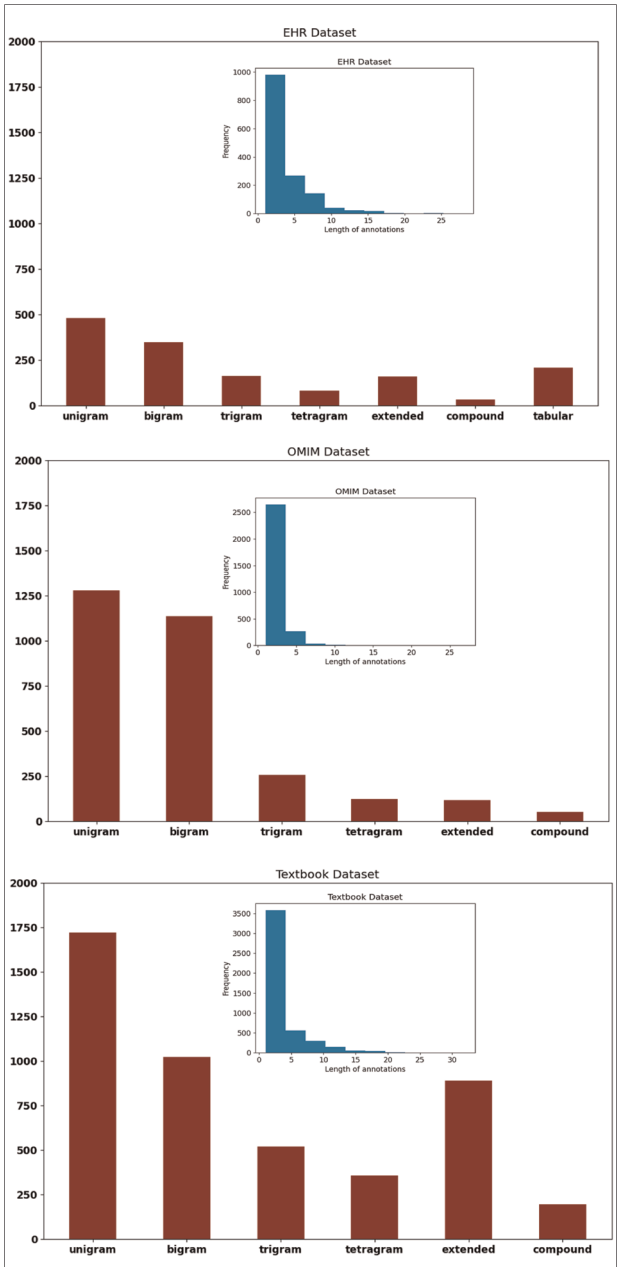


FIGURE 2
Text spans that identified clinical concepts were longer in the EHR corpus and shortest in the OMIM corpus (see blue inset histograms). Proportionately, the OMIM corpus used the most unigrams and bigrams as compared to the EHR corpus and the Textbook corpus (see red bar charts).

IV. Discussion

Named entity recognition models based on deep learning can recognize neurologic signs and symptoms in the biomedical literature and electronic health records (**Table 1**). Previous work has shown that BERT outperforms CNNs on recognizing drugs and diseases in annotated test corpora (52,55). We

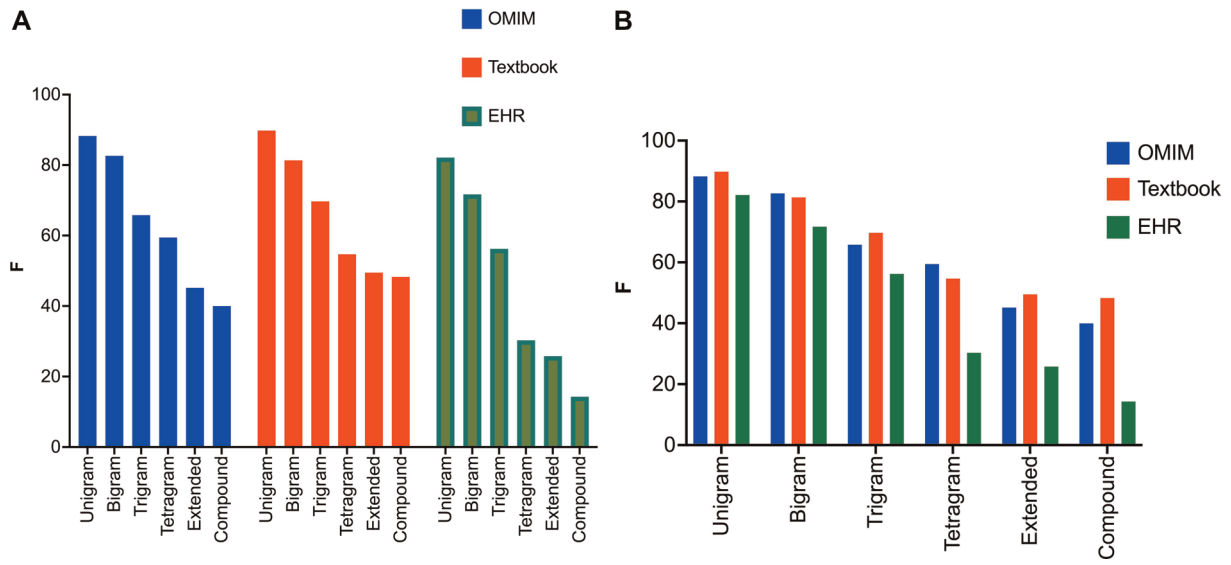


FIGURE 3
(**A**) *F* values for the BERT NER model sorted by text span label type. The BERT NER model performs similarly on the three corpora for unigrams and bigrams, but *F* values lag for the EHR corpora for the tetragrams, extended text spans, and compound text spans. Note that the BERT NER model performs significantly worse on the EHR corpus for tetragrams, extended text spans, and compound text spans when compared to the Textbook or OMIM corpus. (**B**) *F* values for the BERT NER model sorted by corpus. All three corpora show the same pattern with declining *F* values with the increasing length of the text span.

extend these observations to demonstrate the superiority of BERT over CNNs for recognizing neurological signs and symptoms in electronic health records and biomedical literature.

A significant finding was that the accuracy of recognition of signs and symptoms fell with increasing text span length (**Figures 3A,B**). Increased variability in longer text spans likely poses greater difficulty for NER pipelines, regardless of whether they are based on linguistic/symbolic methods like MetaMap or deep learning like BERT or CNNs. Longer text spans are more likely to be descriptions of named entities (e.g., "the patient fell to the left when standing with eyes closed" rather than more concise named entities themselves (e.g., "Romberg sign positive"). Normalization of longer text spans (mapping to suitable concepts in the ontology) may pose additional challenges. The successful mapping (normalization) of "wavering with eyes closed" to "Romberg sign positive" may require vectorization (word embedding) of terms in an ontology, as well as the synonyms and definitions of these terms (48,55).

Another significant observation was that recall of neurologic signs and symptoms was lower in the EHR corpus than in the OMIM corpus or Textbook corpus. The Textbook and the OMIM corpus were written by professional writers and had undergone careful editing and correction. The EHR corpus was written by physicians who were not professional writers. The EHR corpus was marred by irregular spelling, irregular abbreviations, typographical errors, grammatical errors, and other irregularities absent from the OMIM corpus and the Textbook corpus. Others have noted the high frequency of irregular abbreviations, spelling, grammatical, and other writing errors in the clinical notes created by physicians (68–72) The general approach of the writers of the OMIM corpus was brevity. OMIM writers tended to use lists of clinical concepts such as "the patient had optic disk pallor, miosis, anisocoria, and a relative afferent pupil defect." The general approach of the writers of the Textbook corpus was didactic and explanatory so that a relative afferent pupil defect might be described as "the swinging flashlight test was abnormal and the pupil dilated when the light was placed over the abnormal pupil and the pupil constricted when the light was moved to the normal pupil." The EHR corpus was characterized by brevity but irregular spellings, abbreviations, and syntax so that the same patient might be described as "RAPD present on R."

The lower accuracy for recognition of signs and symptoms in the EHR corpus (physician notes) deserves further comment. One way to improve automated recognition of signs and symptoms in physician notes is to encourage them to use structured rather than unstructured documentation (73). However, given physician burnout associated with clinical documentation (74), and physician distaste for structured documentation (75), it seems unlikely that physicians will adopt structured documentation for recording signs and symptoms. Furthermore, given that by training, physicians are often asked to describe findings rather than name findings, it seems unlikely that physicians can be converted to using short names instead of lengthy descriptions of signs and symptoms. Rather, improvements in NLP are needed to identify better

clinical concepts held as lengthier texts spans or represented as descriptions of named entities rather than as the named entity itself.

NLP models that extract clinical concepts from free text must recognize negation successfully. The sentence "the patient has ataxia" has a clinical concept whereas the sentence "ataxia is absent" denies ataxia (76–78). Negation makes it difficult to determine if a sign or symptom is present and suggests that strategies based on regular expressions (REGEX) will fail. The patient who complains of tremor, who is tremulous, or is observed to have a tremor must be distinguished from the patient who denies tremor, is not tremulous, or has no tremor. MetaMap uses the NEGEX algorithm to recognize negation (27). We relied on examples to train the neural networks to recognize negated concepts for our BERT and CNN models. Further work is needed on handling negated concepts accurately and efficiently (77). Another challenge is word disambiguation (79). The sentence the "patient has had a fall" may contain a valid neurological concept, whereas the sentence "the patient was seen in the Fall" does not. Word disambiguation is another area of continuing research in NLP (79).

This study has several limitations. The study was limited to the domain of neurology (neurological signs and symptoms). Furthermore, the text span annotations were done by a single annotator. We have planned an inter-rater agreement study (65). We limited the target ontology to 1,600 neurological concepts. Whether our methods can be generalized to more complex domains and larger ontologies is uncertain. Although we achieved a recall of 80% to 90% with shorter text span lengths, the recall was lower for longer text span lengths. To make automated high throughput neuro-phenotyping practical, we estimate that a recall of at least 90% is needed depending on the application (i.e., research versus patient care). Identifying clinical concepts in complex grammatical structures remains challenging for even the best NLP algorithms. For example, identifying the concepts *biceps weakness*, *triceps weakness*, and *hand weakness* in the sentence *the patient had 3+/5 strength in the biceps, 2+/5 strength in the triceps, and 1/5 hand grip strength* remains problematic. Efficient NLP algorithms that simplify grammar and syntax are an area of evolving research (80,81). Another limitation of the study is the small corpus used for training. Our NER models would likely have improved with more training annotations.

In conclusion, given the burden of physician documentation (74), patient signs and symptoms will likely continue in electronic health records as unstructured text. The automated identification of these signs and symptoms is critical to the success of deep phenotyping, and precision medicine initiatives (5,6). Advances in NLP based on word embedding and deep learning make the automated identification of signs and symptoms in unstructured text increasingly feasible.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board of the University of Illinois at Chicago. The patients/participants provided their written informed consent to participate in this study.

Author's contributions

Concept and design by SA and DBH. Model parameters and computations by SA. Data interpretation, drafting, revising, and final approval by SA, DBH, and DCW II.All authors contributed to the article and approved the submitted version.

Funding

The research was partially sponsored by the Mary K. Finley Missouri Endowment, the Missouri S&T Intelligent Systems Center, the National Science Foundation, and the Leonard Wood Institute in cooperation with the U.S. Army Research Laboratory. It was accomplished under Cooperative Agreement Number W911NF-14-2-0034. The views, opinions, findings, recommendations, or conclusions contained in this document are those of the authors. They should not be interpreted as representing the views or official policies expressed or implied by the Leonard Wood Institute, the Army Research Laboratory, the National Science Foundation, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

Acknowledgments

The authors thank Dr. Michael D. Carrithers for providing the physician EHR notes.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Krauthammer M, Nenadic G. Term identification in the biomedical literature. *J Biomed Inform*. (2004) 37:512–26. doi: 10.1016/j.jbi.2004.08.004

2. Office of the National Coordinator for Health Information Technology. Adoption of electronic health records by hospital service type 2019–2021, Health IT Quick Stat #60 (2022). Available from: https://www.healthit.gov/data/quickstats/adoption-electronic-health-records-hospital-service-type-2019-2021.

3. Banda JM, Seneviratne M, Hernandez-Boussard T, Shah NH. Advances in electronic phenotyping: from rule-based definitions to machine learning models. *Annu Rev Biomed Data Sci*. (2018) 1:53. doi: 10.1146/annurev-biodatasci-080917-013315

4. Tenopir C, Grayson M, Zhang Y, Ebuen M, King DW, Boyce PB. Patterns of journal use by scientists through three evolutionary phases. *D-Lib* (2003) 9:1–15. doi: 10.1045/may2003-king

5. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. (2015) 372:793–5. doi: 10.1056/NEJMp1500523

6. Robinson PN. Deep phenotyping for precision medicine. *Hum Mutat*. (2012) 33:777–80. doi: 10.1002/humu.22080

7. Fu S, Chen D, He H, Liu S, Moon S, Peterson KJ, et al. Clinical concept extraction: a methodology review. *J Biomed Inform*. (2020) 109:103526. doi: 10.1016/j.jbi.2020.103526

8. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med*. (2019) 25:24–9. doi: 10.1038/s41591-018-0316-z

9. Chowdhary K. Natural language processing. *Fundam Artif Intell*. (2020):603–49.

10. Hirschberg J, Manning CD. Advances in natural language processing. *Science*. (2015) 349:261–6. doi: 10.1126/science.aaa8685

11. Islam MA, Anik MSH, Islam ABMAA. Towards achieving a delicate blending between rule-based translator, neural machine translator. *Neural Comput Appl*. (2021) 33:12141–67. doi: 10.1007/s00521-021-05895-x

12. Islam MA, Mukta MSH, Olivier P, Rahman MM. Comprehensive guidelines for emotion annotation. *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*, 2022 Sep. New York, NY, USA: Association for Computing Machinery (2022). p. 1–8.

13. Mohammad S. A practical guide to sentiment annotation: challenges, solutions. *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment, Social MEDIA Analysis*, 2016 Jun. San Diego, California: Association for Computational Linguistics (2016). p. 174–9.

14. Hasan HM, Islam MA, Hasan MT, Hasan MA, Rumman SI, Shakib MN. A spell-checker integrated machine learning based solution for speech to text conversion. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)* (2020). p. 1124–30.

15. Gonzalez-Hernandez G, Sarker A, O'Connor K, Savova G. Capturing the patient's perspective: a review of advances in natural language processing of health-related text. *Yearb Med Inform*. (2017) 26:214–27. doi: 10.15265/IY-2017-029

16. Bird S, Klein E, Loper E. *Natural language processing with python*. Sebastopol, CA: O'Reilly Media (2009). Available from: https://www.nltk.org/book/.

17. SNOMED CT. *NCBO BioPortal* (2022). Available from: https://bioportal.bioontology.org/ontologies/SNOMEDCT/ (Accessed October 5, 2022).

18. Online Mendelian Inheritance in Man. *NCBO BioPortal* (2022). Available from: https://bioportal.bioontology.org/ontologies/OMIM (Accessed October 5, 2022).

19. Human Phenotype Ontology. *NCBO BioPortal* (2022). Available from: https://bioportal.bioontology.org/ontologies/HP (Accessed October 5, 2022).

20. UMLS Metathesaurus Browser. *National Library of Medicine* (2022). Available from: https://uts.nlm.nih.gov/uts/umls/home (Accessed October 5, 2021).

21. Eltyeb S, Salim N. Chemical named entities recognition: a review on approaches, applications. *J Cheminform*. (2014) 6:1–12. doi: 10.1186/1758-2946-6-17

22. Quimbaya AP, Múnera AS, Rivera RAG, Rodríguez JCD, Velandia OMM, Peña AAG, et al. Named entity recognition over electronic health records through a combined dictionary-based approach. *Procedia Comput Sci*. (2016) 100:55–61. doi: 10.1016/j.procs.2016.09.123

23. Hirschman L, Morgan AA, Yeh AS. Rutabaga by any other name: extracting biological names. *J Biomed Inform*. (2002) 35:247–59. doi: 10.1016/S1532-0464(03)00014-5

24. Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions,, relations in clinical text. *J Am Med Inform Assoc*. (2011) 18:552–6. doi: 10.1136/amiajnl-2011-000203

25. Funk C, Baumgartner W, Garcia B, Roeder C, Bada M, Cohen KB, et al. Large-scale biomedical concept recognition: an evaluation of current automatic annotators and their parameters. *BMC Bioinf*. (2014) 15:1–29. doi: 10.1186/1471-2105-15-59

26. Shah NH, Bhatia N, Jonquet C, Rubin D, Chiang AP, Musen MA. Comparison of concept recognizers for building the open biomedical annotator. *BMC Bioinf*. (2009) 10:1–9. doi: 10.1186/1471-2105-10-S2-S1

27. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc*. (2010) 17:229–36. doi: 10.1136/jamia.2009.002733

28. Lindberg DA, Humphreys BL, McCray AT. The unified medical language system. *Yearb Med Inform*. (1993) 2:41–51. doi: 10.1055/s-0038-1637976

29. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res*. (2004) 32:D267–70. doi: 10.1093/nar/gkh061

30. Hier DB, Yelugam R, Azizi S, Carrithers MD, Wunsch II DC. High throughput neurological phenotyping with MetaMap. *Eur Sci J*. (2022) 18:37–49. doi: 10.19044/esj.2022.v18n4p37

31. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. *J Biomed Inform*. (2018) 77:34–49. doi: 10.1016/j.jbi.2017.11.011

32. Sevenster M, Van Ommering R, Qian Y. Automatically correlating clinical findings and body locations in radiology reports using MedLEE. *J Digit Imaging*. (2012) 25:240–9. doi: 10.1007/s10278-011-9411-0

33. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc*. (2010) 17:507–13. doi: 10.1136/jamia.2009.001560

34. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc*. (2004) 11:392–402. doi: 10.1197/jamia.M1552

35. Friedman C, Shagina L, Socratous SA, Zeng X. A web-based version of MedLEE: a medical language extraction and encoding system. *Proceedings of the AMIA Annual Fall Symposium*. American Medical Informatics Association (1996). p. 938.

36. Friedman C. A broad-coverage natural language processing system. *Proceedings of the AMIA Symposium*. American Medical Informatics Association. (2000). p. 270–4.

37. Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging [Preprint] (2015). Available at: http://arxiv.org/1508.01991.

38. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural architectures for named entity recognition [Preprint] (2016). Available at: http://arxiv.org/1603.01360.

39. Chiu JP, Nichols E. Named entity recognition with bidirectional LSTM-CNNs. *Trans Assoc Comput Linguist*. (2016) 4:357–70. doi: 10.1162/tacl_a_00104

40. Peters ME, Ammar W, Bhagavatula C, Power R. Semi-supervised sequence tagging with bidirectional language models [Preprint] (2017). Available at: http://arxiv.org/1705.00108.

41. Vani A, Jernite Y, Sontag D. Grounded recurrent neural networks [Preprint] (2017). Available at: http://arxiv.org/1705.08557.

42. Liu Z, Tang B, Wang X, Chen Q. De-identification of clinical notes via recurrent neural network and conditional random field. *J Biomed Inform*. (2017) 75:S34–S42. doi: 10.1016/j.jbi.2017.05.023

43. Habibi M, Weber L, Neves M, Wiegandt DL, Leser U. Deep learning with word embeddings improves biomedical named entity recognition. *Bioinformatics*. (2017) 33:i37–i48. doi: 10.1093/bioinformatics/btx228

44. Dehghan A, Kovacevic A, Karystianis G, Keane JA, Nenadic G. Combining knowledge-and data-driven methods for de-identification of clinical narratives. *J Biomed Inform*. (2015) 58:S53–9. doi: 10.1016/j.jbi.2015.06.029

45. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. (1997) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735

46. Lafferty J, McCallum A, Pereira FC. Conditional random fields: probabilistic models for segmenting and labeling sequence data. *Proceedings of the 18th International Conference on Machine Learning 2001*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. (2001).

47. Gehrmann S, Dernoncourt F, Li Y, Carlson ET, Wu JT, Welt J, et al. Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. *PLoS ONE*. (2018) 13:e0192360. doi: 10.1371/journal.pone.0192360

48. Arbabi A, Adams DR, Fidler S, Brudno M, et al. Identifying clinical terms in medical text using ontology-guided machine learning. *JMIR Med Inform*. (2019) 7:e12596. doi: 10.2196/12596

49. Groza T, Köhler S, Doelken S, Collier N, Oellrich A, Smedley D, et al. Automatic concept recognition using the human phenotype ontology reference and test suite corpora. *Database*. (2015) 2015:1–13. doi: 10.1093/database/bav005

50. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst*. (2017) 30:5998–6008.

51. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding [Preprint] (2018). Available at: http://arxiv.org/1810.04805.

52. Zhu R, Tu X, Huang JX. Utilizing BERT for biomedical, clinical text mining. *Data Analytics in Biomedical Engineering, Healthcare*. Elsevier (2021). p. 73–103. Available from: https://doi.org/10.1016/B978-0-12-819314-3.00005-7

53. Yu X, Hu W, Lu S, Sun X, Yuan Z. Biobert based named entity recognition in electronic medical record. *2019 10th international conference on information technology in medicine and education (ITME)*. New York NY: IEEE (2019). p. 49–52.

54. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. (2020) 36:1234–40.

55. Ji Z, Wei Q, Xu H. Bert-based ranking for biomedical entity normalization. *AMIA Summits Transl Sci Proc*. (2020) 2020:269.

56. Weng C, Shah NH, Hripcsak G. Deep phenotyping: embracing complexity and temporality-towards scalability, portability, and interoperability. *J Biomed Inform*. (2020) 105:103433. doi: 10.1016/j.jbi.2020.103433

57. Hier DB, Brint SU. A neuro-ontology for the neurological examination. *BMC Med Inform Decis Mak*. (2020) 20:1–9. doi: 10.1186/s12911-020-1066-7

58. Gondolo T. *Neurology study guide: oral board examination review*. Cham Switzerland: Springer Nature (2005).

59. Ubogu EE. *Neurology oral boards review*. New York NY: Humana Press (2005).

60. Alpert JN. *The neurologic diagnosis: a practical bedside approach*. Cham Switzerland: Springer (2018).

61. Kung D, Nguyen T. *Absolute case-based neurology review*. Oxford UK: Springer (2019).

62. Macleod M, Pal S, Simpson M. *Neurology clinical cases uncovered*. San Francisco CA: Wiley-Blackwell (2011).

63. Neves M, Ševa J. An extensive review of tools for manual annotation of documents. *Brief Bioinformatics*. (2021) 22:146–63. doi: 10.1093/bib/bbz130

64. Montani I, Honnibal M. Prodigy: a new annotation tool for radically efficient machine teaching. *Artif Intell*. (2018). Available from: https://explosion.ai/blog/prodigy-annotation-tool-active-learning.

65. Oommen C, Howlett-Prieto Q, Carrithers MD, Hier DB. Inter-Rater Agreement for the Annotation of Neurologic Concepts in Electronic Health Records. *medRxiv* (2022). Available from: http://doi.org/10.1101/2022.11.16.22282384.

66. Vasiliev Y. *Natural language processing with Python and Spacy*. San Francisco CA: No Starch Press (2020).

67. Noy NF, McGuinness DL. Ontology development 101: a guide to creating your first ontology. *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05* (2001).

68. Assale M, Dui LG, Cina A, Seveso A, Cabitza F. The revival of the notes field: leveraging the unstructured content in electronic health records. *Front Med*. (2019) 6:66. doi: 10.3389/fmed.2019.00066

69. Shilo G, Shilo L. Writing style of young physicians in the computer and internet era. *Int J Med Educ*. (2014) 5:82. doi: 10.5116/ijme.534a.a3e2

70. Pagano MP, Mair D. Writing medical records. *J Tech Writ Commun*. (1986) 16:331–41. doi: 10.2190/WY9T-634E-V2JT-JDVQ

71. Zisowitz ML. Teaching medical students and physicians to write. *Acad Med*. (1964) 39:481–4.

72. Hamiel U, Hecht I, Nemet A, Pe'er L, Man V, Hilely A, et al. Frequency, comprehension and attitudes of physicians towards abbreviations in the medical record. *Postgrad Med J*. (2018) 94:254–8. doi: 10.1136/postgradmedj-2017-135515

73. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *J Am Med Inform Assoc*. (2011) 18:181–6. doi: 10.1136/jamia.2010.007237

74. Thomas Craig KJ, Willis VC, Gruen D, Rhee K, Jackson GP. The burden of the digital environment: a systematic review on organization-directed workplace interventions to mitigate physician burnout. *J Am Med Inform Assoc*. (2021) 28:985–97. doi: 10.1093/jamia/ocaa301

75. Han H, Lopp L. Writing and reading in the electronic health record: an entirely new world. *Med Educ Online*. (2013) 18:18634. doi: 10.3402/meo.v18i0.18634

76. Shivade C, de Marneffe MC, Fosler-Lussier E, Lai AM. Extending negex with kernel methods for negation detection in clinical text. *Proceedings of the Second Workshop on Extra-Propositional Aspects of Meaning in Computational Semantics (ExProM 2015)* (2015). p. 41–46.

77. Wu S, Miller T, Masanz J, Coarr M, Halgrim S, Carrell D, et al. Negation's not solved: generalizability versus optimizability in clinical natural language processing. *PLoS ONE*. (2014) 9:e112774. doi: 10.1371/journal.pone.0112774

78. Elkin PL, Brown SH, Bauer BA, Husser CS, Carruth W, Bergstrom LR, et al. A controlled trial of automated classification of negation from clinical notes. *BMC Med Inform Decis Mak*. (2005) 5:1–7. doi: 10.1186/1472-6947-5-13

79. Navigli R. Word sense disambiguation: a survey. *ACM Comput Surv*. (2009) 41:1–69. doi: 10.1145/1459352.1459355

80. Shardlow M. A survey of automated text simplification. *Int J Adv Comput Sci Appl*. (2014) 4:58–70.

81. Al-Thanyyan SS, Azmi AM. Automated text simplification: a survey. *ACM Comput Surv*. (2021) 54:1–36. doi: 10.1145/3442695

frontiers | Frontiers in Digital Health

TYPE Original Research
PUBLISHED 20 December 2022
DOI 10.3389/fdgth.2022.991814

OPEN ACCESS

EDITED BY
Renato Anghinah,
University of São Paulo, Brazil

REVIEWED BY
Mohamed-Amine Choukou,
University of Manitoba, Canada
Rosalia Dacosta-Aguayo,
University of Barcelona, Spain

*CORRESPONDENCE
Melissa Duff
melissa.c.duff@vanderbilt.edu

SPECIALTY SECTION
This article was submitted to Human Factors and Digital Health, a section of the journal Frontiers in Digital Health

RECEIVED 04 October 2022
ACCEPTED 31 October 2022
PUBLISHED 20 December 2022

CITATION
Zhao F, Lim H, Morrow EL, Turkstra LS, Duff MC and Mutlu B (2022) Designing evidence-based support aids for social media access for individuals with moderate-severe traumatic brain injury: A preliminary acceptability study. Front. Digit. Health 4:991814.
doi: 10.3389/fdgth.2022.991814

Designing evidence-based support aids for social media access for individuals with moderate-severe traumatic brain injury: A preliminary acceptability study

Fangyun Zhao[1,2], Hajin Lim[1], Emily L. Morrow[3], Lyn S. Turkstra[4], Melissa C. Duff[3]* and Bilge Mutlu[1]

[1]Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI, United States, [2]Department of Psychology, University of Wisconsin-Madison, Madison, WI, United States, [3]Department of Hearing & Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, United States, [4]School of Rehabilitation Science, McMaster University, Hamilton, ON, United States

Background: Adults with traumatic brain injury (TBI) report significant barriers to using current social media platforms, including cognitive overload and challenges in interpreting social cues. Rehabilitation providers may be tasked with helping to address these barriers.
Objectives: To develop technological supports to increase social media accessibility for people with TBI-related cognitive impairments and to obtain preliminary data on the perceived acceptability, ease of use, and utility of proposed technology aids.
Methods: We identified four major barriers to social media use among individuals with TBI: sensory overload, memory impairments, misreading of social cues, and a lack of confidence to actively engage on social media platforms. We describe the process of developing prototypes of support aids aimed at reducing these specific social media barriers. We created mock-ups of these prototypes and asked 46 community-dwelling adults with TBI (24 females) to rate the proposed aids in terms of their acceptability, ease of use, and utility.
Results: Across all aids, nearly one-third of respondents agreed they would use the proposed aids frequently, and the majority of respondents rated the proposed aids as easy to use. Respondents indicated that they would be more likely to use the memory and post-writing aids than the attention and social cue interpretation aids.
Conclusions: Findings provide initial support for social-media-specific technology aids to support social media access and social participation for adults with TBI. Results from this study have design implications for future development of evidence-based social media support aids. Future work should develop and deploy such aids and investigate user experience.

KEYWORDS

technology aids, social media, traumatic brain injury, social participation, accessibility

Introduction

The prevalence of social media and computer-mediated communication (CMC) platforms have altered how people establish social connections, engage in social events, obtain information, and maintain effective collaboration in daily life (1–3). A growing body of research shows that engagement in social media and CMC, particularly *via* Facebook, increases social connectedness and decreases loneliness, plays a critical role in friendship maintenance, and promotes health and well-being (4, 5). For individuals with health-related concerns, social media platforms have provided an important mechanism to find health information, participate in support groups, and share their experiences (6, 7). Individuals with traumatic brain injury (TBI) may particularly benefit from social media, given that they often report social isolation (8) and friendship loss (9), along with physical and cognitive limitations that make in-person social interactions difficult (10, 11). Previous research suggested that social media can promote mental well-being among individuals with TBI and allow them to keep or increase opportunities for social participation (12, 13). Individuals with TBI want to use social media platforms such as Facebook and Twitter as much as their uninjured peers (14). However, these individuals may experience cognitive impairments and have reported significant barriers to using current social media platforms, including cognitive overload and challenges in interpreting social cues (12, 14–17), so the potential benefits of social media are often not accessible to them. Rehabilitation professionals see social media use as a way to reduce social isolation following brain injury, and such professionals may play a future role in addressing barriers to increase social media participation (18).

Social media platforms have provided limited support for increasing accessibility to individuals with TBI and other populations with cognitive impairments (19–21). Accessibility features of social media platforms mostly focus on supporting individuals with sensory disabilities such as hearing or vision impairments (22–24). These features include allowing voice-over gestures for navigating social media sites and providing automatically generated image captions (22, 25). There are no parallel supports for individuals with cognitive impairments such as those routinely observed in individuals with TBI.

The current study is part of a broader effort to develop technological supports to increase social media accessibility for people with TBI-related cognitive impairments. A long-term goal of this line of work is to also understand individual differences that may influence who is willing to use, and who would benefit from, technological support to increase social media accessibility. Here, we report on the process of designing four social media support aids that address challenges in using social media platforms associated with social and cognitive impairments in adults with TBI reported in the literature and those we have observed in the clinical experiences of the author team. The future success of any technological support to improve accessibility and social media use, however, depends on their acceptability and perceived utility and benefit to individuals with TBI (26). Thus, as a first step in this process, we obtained preliminary feedback from individuals with TBI on the acceptability and potential use of these aids to guide future development.

In the following sections, we review previous literature on social and cognitive impairments in individuals with moderate-severe TBI that would affect use of social media platforms and identify four main barriers. We describe potential technological support aids to address these barriers and the process of designing prototypes of these aids. Finally, we report on a survey study where we presented mock-ups of these aids to gain acceptability data and perceptions of the utility of the aids.

Background

Social and cognitive impairments in individuals with TBI

Individuals with TBI have a range of deficits that make it difficult to navigate the social world. Impairments in social communication skills are a hallmark of TBI, including impairments in recognizing and interpreting social cues (14, 15, 27–29); missing implied meanings such as sarcasm and jokes; and losing track of topics in a conversation (30–33). These social communication deficits are thought to be a major contributor to the negative social outcomes reported by many adults with TBI (34–36). Indeed, as a group, adults with TBI report having fewer friends and social contacts overall (19), and less social participation with, and more social isolation from, their uninjured peers (20). These negative outcomes in turn affect mental health and wellbeing, not only for the person with TBI but also for their caregivers (37, 38). Impairments in basic cognitive functions are also common following TBI in domains such memory (39), attention (40, 41), decision-making (42–44), and executive functioning (45, 46). These social communication and cognitive deficits have typically been examined and reported in face-to-face, in-person interactions, but recent work suggests that they might extend to computer-mediated communication on social media platforms (18, 47).

Four evidence-based social media barriers among individuals with TBI

The literature on barriers and challenges to social media use among individuals with TBI, together with the clinical experiences of some of our team members, identifies four

major barriers to social media use among individuals with TBI: sensory overload, memory impairments, misreading of social cues, and a lack of confidence to actively engage on social media platforms (14, 47, 48).

Sensory overload

Social media platforms can place high demands on sensory processing and attention. Individuals with TBI report difficulty navigating social media sites, keeping up with rapid feeds, and managing sensory overload (12, 17, 27). Some individuals report going through a try-and-fail process to get familiar with the social media platforms due to lack of instructions (14, 15, 27), being overwhelmed, and going offline. In one study, individuals with TBI reported that they found the information on Twitter meaningless and random due to information overload (14). Difficulty managing attention and disrupted information processing are well documented challenges in face-to-face interactions for individuals with TBI (39, 49, 50). These challenges are consistent with the reports of being overwhelmed and overloaded and ultimately abandoning online sessions. A potential solution to this challenge might include restricting the amount of content displayed at any given time by, for example, discretizing the information that is shown in the form of an "infinite scroll" that is widely used by social media platforms.

Memory impairments

Social media platforms can place high demands on working and declarative long-term memory. Social media users must quickly identify the owner of the message or post and recall previous events and histories to interpret a given message, as well as quickly integrate and update memory as new information becomes available. Working and declarative memory impairments commonly follow TBI, and these deficits are likely to pose a challenge for using social media platforms. Indeed, declarative memory impairments affected how individuals with TBI process information on social media (47) and decreased their social media use (17). Providing memory assistance that consolidates previous messages to facilitate comprehension of a current message may help individuals with TBI manage the memory demands of using social media.

Misreading of social cues

Computer mediated communication requires users to read social cues from a variety of single and integrated sources including faces, videos, text, and emoji. Deficits in reading social cues in individuals with TBI are well documented. Individuals with TBI have difficulty reading cues in social interaction and managing turn taking (14, 15) and, relative to uninjured peers, are less accurate in facial affect recognition (51, 52) and less sensitive to text-based social cues (53). Such deficits in social communication are consistent with reports of individuals with TBI misreading social cues in social media and experiencing negative consequences (12, 17). Providing users with information about the general sentiment of a post might help individuals with TBI in reading social cues on social media platforms.

Lack of confidence to actively engage on social media platforms

Individuals with TBI reported a lack of confidence in engaging in online social activities on social media platforms. They also report using Facebook more passively than actively, i.e., being less likely to post status updates or send direct messages to others on social media compared to uninjured peers (17, 20). In particular, individuals with TBI reported worrying about misreading conversations or making mistakes (17, 28). Support tools that allow individuals with TBI to monitor their messages and get feedback before posting could increase confidence when engaging on social media. If so, increased confidence may result in more active participation, which could in turn provide more opportunities to experience the benefits of social media use reported by neurotypical individuals.

Guided by the literature described above, we designed four aids to address: (1) sensory overload, (2) memory impairments, (3) misreading social cues, and (4) a lack of confidence to actively engage on social media platforms. Our overarching strategy was to design aids that reduced the cognitive or sensory load (e.g., memory) or that provided assistance in meeting the social or cognitive demands (e.g., reading social cues) reported by individuals with TBI as barriers to CMC and that the use of the aids would be as simple and intuitive as possible.

After the initial conceptual design, we engaged in ideation and iterative design to develop specific interface solutions that can be implemented as interface augmentations. We ensured that our designs were technologically feasible using available user interface software (e.g., react.js) and commercial text and visual analysis toolbox (e.g., Watson Natural language understanding, IBM Visual Insights) for future implementation. We then created a mock-up of each design for acceptability testing of the concepts of these aids. The mock-ups were created by capturing screenshots of the Facebook interface and modifying its visual elements to represent the design of our aids.

Design of social media aids for Facebook

We created the mock-ups for Facebook's platform, as adults with TBI cited it as their most commonly used platform (21, 54). In this section, we present the design rationale and development for the social media aids.

Attention aids

To address sensory and information overload reported by individuals with TBI, we aimed to reduce the visual and

technical complexity in the current Facebook interface (15, 54) (**Figure 1**). A traditional Facebook page contains many elements including color side bars, newsfeed posts, friend lists, and advertisements. Arfaa and Wang proposed that grouping and highlighting necessary information could facilitate easier navigation of a website layout for older adults (24). We expected that their suggestion would also be helpful for adults with TBI.

In conceptualizing this aid, we first aimed to reduce visual complexity by putting a transparent gray overlay over the Facebook newsfeed, so users can pay attention to and read one post at a time. Also, to guide better navigation of Facebook, we grouped and labeled each area of a page by its primary purpose (i.e., post status updates, send messages) (see **Figure 1**). We also created a "Next Post" button to enable users to bring the next post to focus.

Memory aids

To address impairments in memory, we designed an aid that automatically searches and consolidates related posts from a user's profile and presents them to the user. When users see a post that builds on context or information from previous posts, the memory aid retrieves related posts and presents them in a section, titled "explain more," so users have context for the current post. For example, as illustrated in **Figure 2**, previous posts about a stage performance were combined into a thread in the "explain more" button. A user could then see how the current post related to previous posts, which provides information implied in the current post and thus supports understanding of that post based on context that may be missing for individuals with a memory deficit (i.e., that it refers to a celebration for the performance).

Social cue interpretation aids

Based on evidence of impaired social communication and misreading of social cues in adults with TBI, we designed an aid to facilitate social cue interpretation. We suspect that reading social cues may be particularly difficult when the demands include integrating the information from text and images. As illustrated in **Figure 3**, the social cue interpretation aid automatically extracts the main sentiment and topic from the text and/or image from the target post and presents a short summary of the post.

Message production aids

Individuals with TBI report worrying about misreading conversations and then making mistakes (17, 28). As a consequence, they report being less active on social media. To address this barrier, we designed a message production aid. In addition to providing feedback on spelling and grammar, we expected that this tool would serve as a "Theory-of-Mind check;" that is, it provides feedback on how a recipient would likely interpret that message. As shown in **Figure 4**, the aid provides feedback on grammar and sentiment of the message before a user posts it, which would give users opportunities to fix grammatical errors and monitor the tone and emotion conveyed by their posts before they are sent. When no error

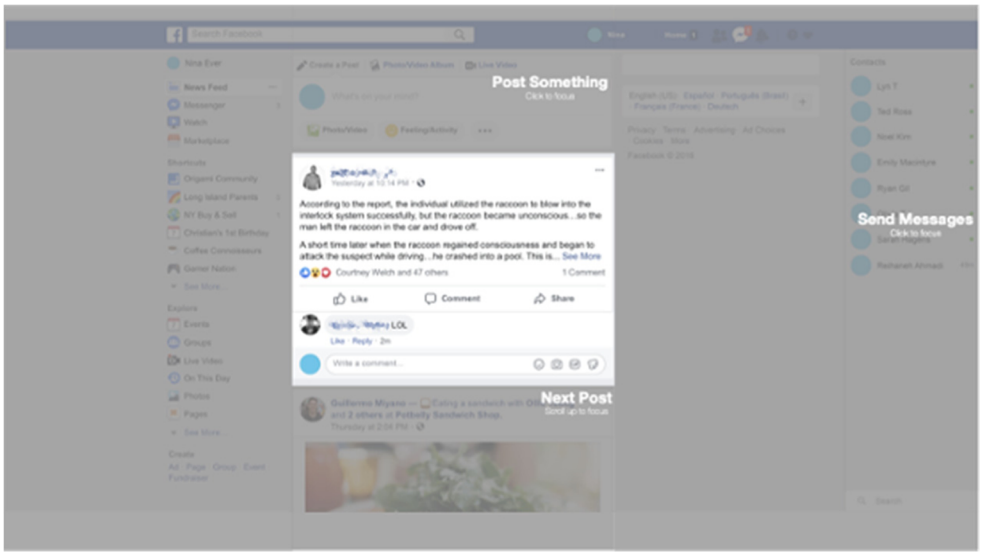


FIGURE 1
An example of modified Facebook page after using the attention aid. Identifying information such as names and faces have been blurred to protect confidentiality.

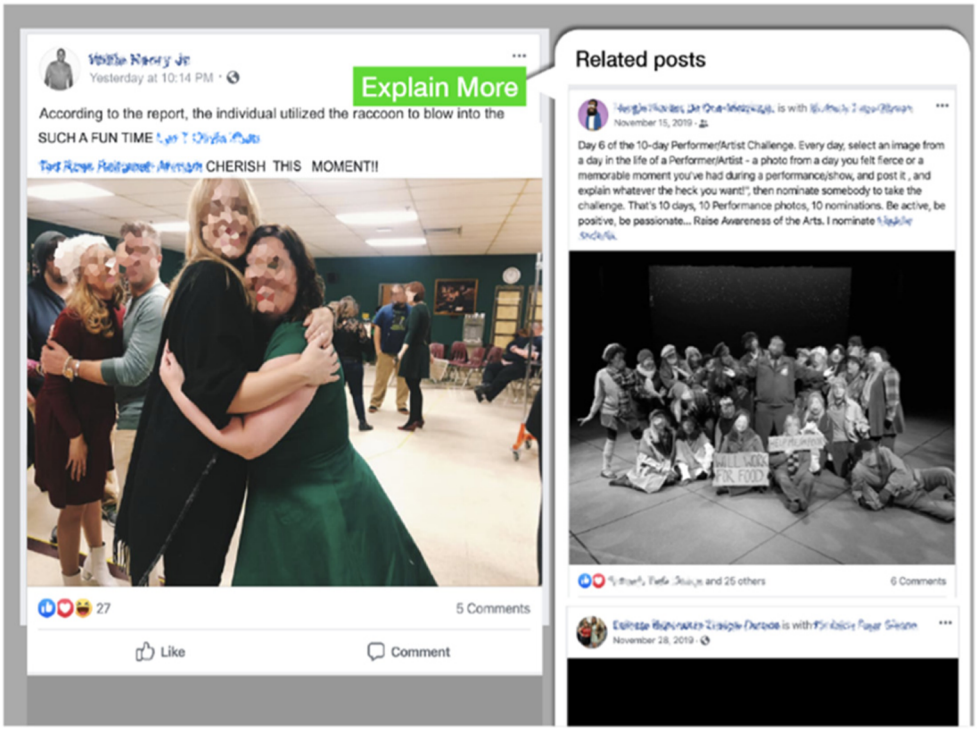


FIGURE 2
An example of the modified Facebook page using the memory aid. Identifying information such as names and faces have been blurred to protect confidentiality.

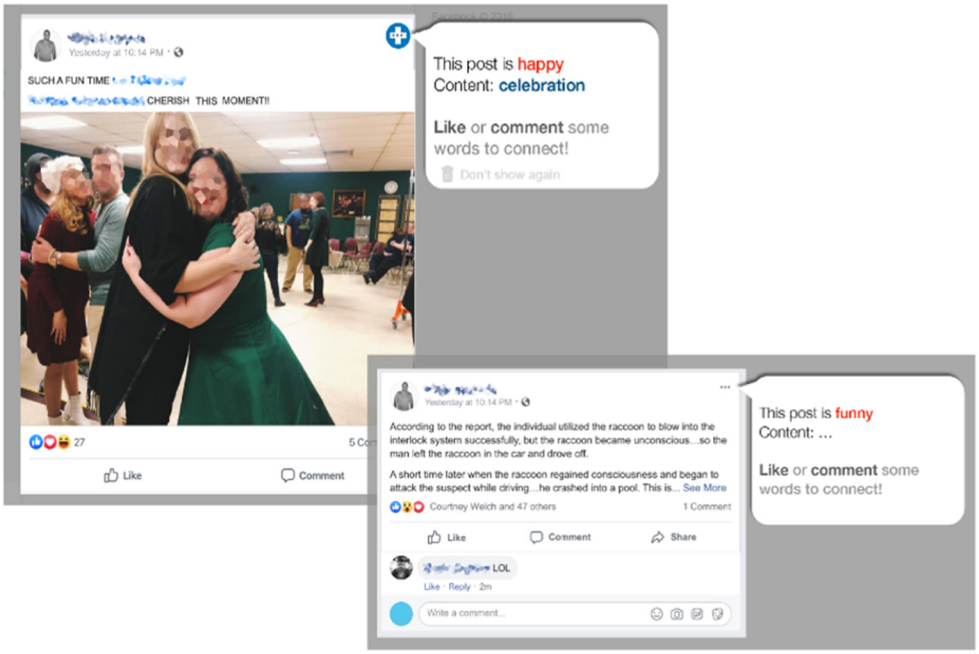


FIGURE 3
An example of the modified Facebook page using the social cue aid. Identifying information such as names and faces have been blurred to protect confidentiality.

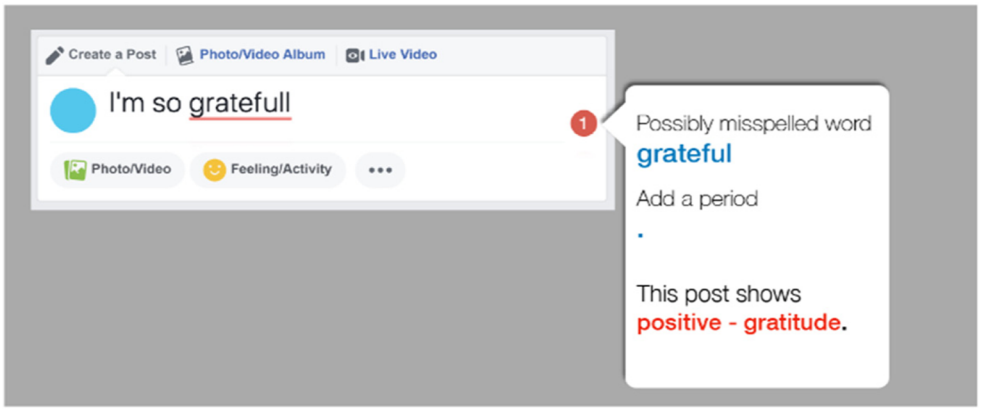


FIGURE 4
An example of the modified Facebook page using the message production aid.

is detected, the system provides positive encouragement to generate their content.

In summary, we identified four key barriers to social media use for individuals with TBI and designed a set of aids aimed at reducing these barriers. Each of these aids included specific design features and functionality to address a key social-cognitive barrier to social media use by individuals with TBI, as summarized in **Table 1**. Before beginning software development and implementation of the social media support aids, we created mock-ups of our designs to determine if adults with TBI would find these tools acceptable and to obtain additional design suggestions for future social media support tools for individuals with TBI. To obtain this information, we conducted a survey to gauge the acceptability of these aids and perceived use and benefit to individuals with TBI.

Methods for acceptability study

Participants

Participants were recruited through the Vanderbilt Brain Injury Patient Registry (55) and were a subset of individuals with TBI surveyed by Morrow and colleagues (21; see below). The study by Morrow et al. included 53 adults (28 females) with moderate-severe TBI, but we excluded seven participants from that sample who reported that they never had a Facebook account. The final sample was 46 participants with moderate-severe TBI (24 females, $M = 38.0$ years old, SD = 9.6). Participants with TBI had 14.9 years of education (SD = 2.3), on average.

All participants with TBI were in the chronic phase of injury (>6 months post-injury) and sustained their injuries in adulthood (i.e., after age 18). Thus, participants' neuropsychological profiles were in the chronic and stable phase (56). Average time since injury was 71.8 months (SD = 64.0). Participants with TBI did not have a history of neurological or cognitive disabilities before the qualifying brain injury. TBI severity was determined using the Mayo Classification System (57). Participants were classified as having sustained a moderate-severe TBI if at least one of the following criteria was met: (1) Glasgow Coma Scale (GCS) < 13 within 24 h of acute care admission (i.e., moderate or severe injury according to the GCS); (2) positive neuroimaging findings (acute CT findings, or lesions visible on a chronic MRI); (3) loss of consciousness (LOC) > 30 min;

TABLE 1 Summary of four aids and their design features.

Type of Aid	Target Barrier	Design Goal	Design Features
Attention Aids	Sensory overload	Reduce visual complexity	A semi-transparent overlay to cover the page; a summary of the functions of each area; highlighting of targeted area of interest
Memory Aids	Memory impairments	Consolidate and present previous posts to help users comprehend the current post	An added button that retrieves related posts
Social Cue Interpretation Aids	Misreading of social cues	Provide a summary of the sentiment of targeted post	An added button that shows the topic and the sentiment of the current message
Message Production Aids	Lack of confidence in actively engaging on social media platforms	Reduce grammatical errors; provide a preview of the sentiment of the message	A button that indicates errors in the current message being produced; when clicked, a message appears that includes errors, a suggested fix, and the sentiment of the current message

or (4) post-traumatic amnesia (PTA) > 24 h. Injury-related information was collected from available medical records and a semi-structured interview with participants.

GCS was available for 38 participants (ranging from 3 to 15); loss of consciousness (LOC) information was available for 42 participants; PTA information was available for 44 participants; acute imaging information was available for 44 participants (43 with positive findings). Causes of injury were motor vehicle accidents (25), falls (6), motorcycle or snowmobile accidents (4), being hit by a car as a pedestrian (4), assault (3), non-motorized vehicle accidents (1), being hit by a moving object (1), or other (3).

Survey & procedures

The data for the acceptability study were collected as part of a larger project investigating social media use among individuals with TBI. Participants received a link to complete the survey online *via* the Research Electronic Data Capture System (REDCap; 29). The full survey consisted of up to 280 questions, depending on participants' responses. Participants with TBI first answered questions related to their general social media use (reported in ref. 21), how their Facebook usage changed after the injury, their current experience with Facebook, and their perceived social support and social connectedness on Facebook. The results reported here were from the second part of the survey, in which we presented the mock-up images of the prototype designs for the four aids, with explanations of their features, and asked participants about their perceptions of each prototype design. That is, participants were presented with screen shot images to give the sense of the visual appearance and functionality of the aids, but participants could not click on or interact with the aids during this phase of testing.

Measure

For each aid, participants were asked to complete a 10-item questionnaire. The first five items were from the System Usability Scale (SUS; 58), which has widely been used as a reliable method to measure usability of software products. Aids were referred to as "modifications." We also modified the wording for question five to make it more relevant to the current study. SUS items were: #1: *I would use this modification frequently*; #2: *I found this modification unnecessarily complex*; *#3: This modification looks easy to use*; #4: *I would need technical support to use this modification;* #5: *Most people with TBI would learn to use this modification very quickly.*

The five subsequent items asked participants to rate the perceived benefits of each aid, particularly how that aid could help them more actively engage in Facebook social interaction. The items were: #6: *I would post and/or share more things with this modification;* #7: *I would click on the content shared by my friends more with this modification*; *#8: I would comment more with this modification*; #9: *I would spend more time on Facebook with this modification*; #10: *I would send more messages to my friends with this modification.*

Participants were asked to indicate how much they agreed or disagreed with each of the 10 statements using a three-item scale (Disagree, Neither agree nor disagree, Agree). In addition, participants were asked whether they noticed any changes in the way they used Facebook after brain injury by answering either "yes" or "no." Participants also answered two open-ended questions regarding their changes in Facebook use since their injury and their recommendations for modifications to the existing Facebook platform.

Data analysis

The goal of this study was to explore how individuals with TBI perceive the aids we designed to address their reported challenges in using social media. Consistent with this exploratory goal, we primarily used descriptive statistics to analyze participants' responses. We expect that findings would serve as the foundation for future hypothesis-driven research on technology-based social media interventions for individuals with TBI (59). Consistent with this goal, we also performed *ad hoc* exploratory analyses to investigate if individual characteristics such as age, sex, or education influenced the ratings of the aids.

Results

Responding to individual questions was voluntary, thus, not all participants answered all questions. The number of individuals who responded to a given question is listed in parentheses.

Overall attitudes towards the aids

Before examining the participants' responses for each aid type separately, we first summed responses for all aids together (**Table 2**). Overall, 29.7% of respondents agreed that they would use the proposed aids frequently; 33.5% were neutral; and 36.8% disagreed. Most respondents agreed that the aids looked easy to use (59.2%) and that they would not require any technical support (69.2%). Only 24.0% of respondents indicated that the aids appeared unnecessarily complex, and 10.9% indicated that they would struggle to learn how to use them.

In regard to Facebook functions, 11.6% of respondents agreed that the proposed aids would help them become more

TABLE 2 Summary of participants' responses for all types of aids.

#	Item	Agree % (count)	Disagree % (count)	Neutral % (count)	Total % (count)
1	I would use this modification frequently.	29.7% (54)	36.8% (67)	33.5% (61)	100% (182)
2	I find this modification unnecessarily complex. (Reversed)	38.8% (71)	24.0% (44)	37.2% (68)	100% (183)
3	This modification looks easy to use.	59.2% (109)	10.9% (20)	29.9% (55)	100% (184)
4	I would need technical support to use this modification. (Reversed)	69.2% (126)	5.5% (10)	25.3% (46)	100% (182)
5	Most people with TBI would learn to use this modification very quickly.	39.6% (72)	11.5% (21)	48.9% (89)	100% (182)
6	I would post and/or share more things with this modification.	14.8% (27)	38.5% (70)	46.7% (85)	100% (182)
7	I would comment more with this modification.	14.4% (26)	40.3% (73)	45.3% (82)	100% (181)
8	I would spend more time on Facebook with this modification.	11.6% (21)	46.4% (84)	42.0% (76)	100% (181)
9	I would send more messages to my friends with this modification.	14.4% (26)	42.2% (76)	43.3% (78)	100% (180)
10	I would click on the content shared by my friends more on this modification.	21.5% (39)	33.7% (61)	44.8% (81)	100% (181)

Note: % is the percentage of respondents who endorsed a statement. Count is the number of respondents who endorsed a statement. Total count is the total number of respondents who answered a given item. The variability in total count reflects that not all respondents answered all questions. Maximum total count is 184 (46 respondents and four aid types).

active on Facebook; 14.8% agreed they would post more, 14.4% agreed they would comment more; 14.4% agreed they would send more messages to friends; and 21.5% agreed they would click on others' content more often if using the proposed aids. The remaining responses were relatively equally divided between "neutral" and "disagree."

Attitudes towards the support aids by type

For the five survey questions about how the aids might affect participants' Facebook use (i.e., spending time on Facebook, posting, commenting, messaging, clicking on content), responses that were not "agree" were largely divided between "neutral" and "disagree." Thus, in the interest of clarity and brevity, the results for each tool presented in the text include only the percent that agreed with the statement. Percentages in the other two categories are listed in the tables for each aid.

Attitudes towards the attention aid

In regard to ease of use, 19.6% agreed that they would use the attention aid often; 26.1% agreed that they found it unnecessarily complex; 54.3% agreed that it would be easy to use; 6.5% agreed they would need technical assistance; and 45.5% agreed that people with TBI would learn to use the tool quickly. In regard to Facebook functions, 6.5% agreed that they would spend more time on Facebook if they used the tool; 8.7% agreed that they would post or share more; 8.7% agreed that they would comment more; 6.8% agreed that they would send more messages to friends, and 17.8% agreed that they would click on others' content more. Results are summarized in **Table 3**.

Attitudes towards the memory aid

In regard to ease of use, 37.0% of respondents agreed that they would use the memory aid frequently; 28.9% agreed that the memory aid looked complex; 8.7% agreed that it would be difficult to use; 10.9% agreed that they would need technical assistance to use it; and 32.6% agreed that most people with TBI would learn to use it quickly. In regard to Facebook functions, 13.3% agreed that the memory aid could help them spend more time on Facebook; 13.0% agreed that they would post or share more; 8.9% agreed that they would comment more; 13.0% agreed that they would send more messages; and 21.7% agreed that they would click on content shared by others more. Results are summarized in **Table 4**.

Attitudes towards the social cue interpretation aid

In regard to ease of use, 22% of participants agreed that they would use the social interpretation aid frequently; 26.1% agreed that it was unnecessarily complex; 17.4% agreed that it would take a long time to learn; 56.5% agreed that it was easy to use; and 75.0% agreed that they would not require technical support to use it. In regard to Facebook functions, 8.9% agreed that they would spend more time on Facebook if they had the social cue aid; 8.9% agreed that they would comment more; 17.8% agreed that they would post or share more; 13.3% agreed that they would send more messages to friends; and 24.4% agreed that they would click on more content by others if they had the aid. Results are summarized in **Table 5**.

TABLE 3 Summary of participants' responses for attention aid.

#	Item	Agree % (count)	Disagree % (count)	Neutral % (count)	Total % (count)
1	I would use this modification frequently.	19.6% (9)	37.0% (17)	43.5% (20)	100.0% (46)
2	I find this modification unnecessarily complex. (Reversed)	30.4% (14)	26.1% (12)	43.5% (20)	100.0% (46)
3	This modification looks easy to use.	54.3% (25)	10.9% (5)	34.8% (16)	100.0% (46)
4	I would need technical support to use this modification. (Reversed)	60.9% (28)	6.5% (3)	32.6% (15)	100.0% (46)
5	Most people with TBI would learn to use this modification very quickly.	45.5% (20)	4.5% (2)	50.0% (22)	100.0% (44)
6	I would post and/or share more things with this modification.	8.7% (4)	41.3% (19)	50.0% (23)	100.0% (46)
7	I would comment more with this modification.	8.7% (4)	43.5% (20)	47.8% (22)	100.0% (46)
8	I would spend more time on Facebook with this modification.	6.5% (3)	45.7% (21)	47.8% (22)	100.0% (46)
9	I would send more messages to my friends with this modification.	6.8% (3)	40.9% (18)	52.3% (23)	100.0% (44)
10	I would click on the content shared by my friends more on this modification.	17.8% (8)	31.1% (14)	51.1% (23)	100.0% (45)

Note: % is the percentage of respondents who endorsed a statement. Count is the number of respondents who endorsed a statement. Total count is the total number of respondents who answered a given item. The variability in total count reflects that not all respondents answered all questions. Maximum total count is 46 (46 respondents).

TABLE 4 Summary of participants' responses for memory aid.

#	Item	Agree % (count)	Disagree % (count)	Neutral % (count)	Total % (count)
1	I would use this modification frequently.	37.0% (17)	32.6% (15)	30.4% (14)	100.0% (46)
2	I find this modification unnecessarily complex. (Reversed)	37.8% (17)	28.9% (13)	33.3% (15)	100.0% (45)
3	This modification looks easy to use.	58.7% (27)	8.7% (4)	32.6% (15)	100.0% (46)
4	I would need technical support to use this modification. (Reversed)	63.0% (29)	10.9% (5)	26.1% (12)	100.0% (46)
5	Most people with TBI would learn to use this modification very quickly.	32.6% (15)	15.2% (7)	52.2% (24)	100.0% (46)
6	I would post and/or share more things with this modification.	13.0% (6)	41.3% (19)	45.7% (21)	100.0% (46)
7	I would comment more with this modification.	8.9% (4)	40.0% (18)	51.1% (23)	100.0% (45)
8	I would spend more time on Facebook with this modification.	13.3% (6)	42.2% (19)	44.4% (20)	100.0% (45)
9	I would send more messages to my friends with this modification.	13.0% (6)	43.5% (20)	43.5% (20)	100.0% (46)
10	I would click on the content shared by my friends more on this modification.	21.7% (10)	39.1% (18)	39.1% (18)	100.0% (46)

Note: % is the percentage of respondents who endorsed a statement. Count is the number of respondents who endorsed a statement. Total count is the total number of respondents who answered a given item. The variability in total count reflects that not all respondents answered all questions. Maximum total count is 46 (46 respondents).

Attitudes towards post-writing aid

In regard to ease of use, 40% agreed that they would use the post-writing aid frequently; 15.2% agreed that the attention aid looked complex; 67.4% agreed that it would be easy to use; and 78.3% agreed that they would not require technical support to use it; and 43.5% agreed that most people with TBI would be able to easily learn to use it. In regard to Facebook functions, 17.8% agreed that they would spend more time on Facebook if they used this tool; 68.9% agreed that they would post more; 22.2% agreed that they would comment more; 24.4% agreed that they would send more messages to friends; and 22.2% agreed that they would click on others' comments more. Results are summarized in **Table 6**.

Ad hoc exploration of individual characteristics and types of aids

We were next interested in whether there were individual characteristics (age, sex, education) that influenced the ratings of the aids. To conduct this *ad hoc* exploratory analysis, we converted the response options to numeric values (i.e., Disagree = −1, Neutral = 0, Agree = 1) and conducted an exploratory factor analysis on data from the ten items. Factor analysis with Varimax rotation indicated the presence of two factors: one corresponding to the potential utility of the aids and another corresponding to ease of use (see **Table 7**). These factors accounted for 43.9% and 21.9% of the variance, respectively. By averaging the items loading on each factor, we

TABLE 5 Summary of participants' responses for social cue interpretation aid.

#	Item	Agree % (count)	Disagree % (count)	Neutral % (count)	Total % (count)
1	I would use this modification frequently.	22.2% (10)	48.9% (22)	28.9% (13)	100.0% (45)
2	I find this modification unnecessarily complex. (Reversed)	32.6% (15)	26.1% (12)	41.3% (19)	100.0% (46)
3	This modification looks easy to use.	56.5% (26)	13.0% (6)	30.4% (14)	100.0% (46)
4	I would need technical support to use this modification. (Reversed)	75.0% (33)	0.0% (0)	25.0% (11)	100.0% (44)
5	Most people with TBI would learn to use this modification very quickly.	37.0% (17)	17.4% (8)	45.7% (21)	100.0% (46)
6	I would post and/or share more things with this modification.	17.8% (8)	40.0% (18)	42.2% (19)	100.0% (45)
7	I would comment more with this modification.	17.8% (8)	44.4% (20)	37.8% (17)	100.0% (45)
8	I would spend more time on Facebook with this modification.	8.9% (4)	51.1% (23)	40.0% (18)	100.0% (45)
9	I would send more messages to my friends with this modification.	13.3% (6)	48.9% (22)	37.8% (17)	100.0% (45)
10	I would click on the content shared by my friends more on this modification.	24.4% (11)	37.8% (17)	37.8% (17)	100.0% (45)

Note: % is the percentage of respondents who endorsed a statement. Count is the number of respondents who endorsed a statement. Total count is the total number of respondents who answered a given item. The variability in total count reflects that not all respondents answered all questions. Maximum total count is 46 (46 respondents).

TABLE 6 Summary of participants' responses for post-writing aid.

#	Item	Agree % (count)	Disagree % (count)	Neutral % (count)	Total % (count)
1	I would use this modification frequently.	40.0% (18)	28.9% (13)	31.1% (14)	100.0% (45)
2	I find this modification unnecessarily complex. (Reversed)	54.3% (25)	15.2% (7)	30.4% (14)	100.0% (46)
3	This modification looks easy to use.	67.4% (31)	10.9% (5)	21.7% (10)	100.0% (46)
4	I would need technical support to use this modification (Reversed)	78.3% (36)	4.3% (2)	17.4% (8)	100.0% (46)
5	Most people with TBI would learn to use this modification very quickly.	43.5% (20)	8.7% (4)	47.8% (22)	100.0% (46)
6	I would post and/or share more things with this modification.	20.0% (9)	31.1% (14)	48.9% (22)	100.0% (45)
7	I would comment more with this modification.	22.2% (10)	33.3% (15)	44.4% (20)	100.0% (45)
8	I would spend more time on Facebook with this modification.	17.8% (8)	46.7% (21)	35.6% (16)	100.0% (45)
9	I would send more messages to my friends with this modification.	24.4% (11)	35.6% (16)	40.0% (18)	100.0% (45)
10	I would click on the content shared by my friends more on this modification.	22.2% (10)	26.7% (12)	51.1% (23)	100.0% (45)

Note: % is the percentage of respondents who endorsed a statement. Count is the number of respondents who endorsed a statement. Total count is the total number of respondents who answered a given item. The variability in total count reflects that not all respondents answered all questions. Maximum total count is 46 (46 respondents).

created measures of "potential utility of the aids" (Cronbach's α = .91) and "ease of use" (Cronbach's α = .70).

Correlation analyses were conducted to explore relationships among the measures for potential utility and ease-of-use ratings (from the factor analysis) for the four types of aids and individual characteristics such as age, sex, education, time since onset (TSO), and frequency of Facebook use. Results are summarized in **Table 8**.

There was no significant correlation between type of aid and either potential utility or ease of use (r = .09,.11 respectively, $p > .05$), so we did not conduct *post hoc* correlational analyses for each aid type separately. Age was significantly correlated with both potential utility and ease of use ($r = .23$,.27 respectively, $p < .01$). Education was significantly, but negatively, correlated with potential utility ($r = -.15$, $p < .05$), but not ease of use ($r = .14$, $p > .05$). Sex, TSO, and frequency of Facebook use were not significantly correlated with either ease of use or potential utility ($r = -.03$, $-.01$ respectively, $p > .05$).

Discussion

The goal of this study was to elicit feedback on prototype aids designed to reduce social media access barriers for adults with TBI. The aids were based on evidence of barriers to social media use by adults with TBI (e.g., 12, 15, 17, 47, 54) and known cognitive challenges in this population (e.g., 39–42, 45), including challenges we had discovered in studies

TABLE 7 Factor analysis of the ten item measures.

#	Item	Factor 1 (Potential Utility)	Factor 2 (Ease of Use)
1	I would use this modification frequently.	.68	
2	I find this modification unnecessarily complex. (Reversed)		.63
3	This modification looks easy to use.		.81
4	I would need technical support to use this modification. (Reversed)		.75
5	Most people with TBI would learn to use this modification very quickly.		.60
6	I would post and/or share more things with this modification.	.85	
7	I would comment more with this modification.	.86	
8	I would spend more time on Facebook with this modification.	.84	
9	I would send more messages to my friends with this modification.	.87	
10	I would click on the content shared by my friends more on this modification.	.82	

TABLE 8 Pearson correlation table among variables.

	(Factor 1) Potential Utility	(Factor 2) Ease of Use
Sex	−.03	−.01
Age	.23**	.27**
Education	−.15*	.14
TSO	.14	.09
Frequency of FB use	.00	−.07
Type of Aid	.09	.11
Potential Utility	1	.40**
Ease of Use	.40**	1

*Correlation is significant at the 0.05 level.
**Correlation is significant at the 0.01 level.

leading up to this project (e.g., 51, 60–63). The four key barriers to social media use included cognitive overload, memory impairments, deficits in social cognition and communication, and a lack of confidence to actively engage on social media platforms. We designed prototypes of aids to address these barriers, presented mock-ups of these aids to participants with TBI, and asked participants to rate the aids' potential utility and ease of use. While the proposed aids are unlikely to address all barriers to successful social media use, to our knowledge this was the first evidence-based study to introduce the concept of social media aids for people with TBI. The findings here provide a foundation for future development of technological supports to enable individuals with TBI to fully access and participate on social media platforms.

Across all aids, nearly one-third of respondents agreed that they would use the proposed aids frequently. The majority of respondents also agreed that all of the aids would be easy to use without technical support and that most people with TBI could learn to use them quickly. These are positive findings given the cognitive demands of adopting new technology and known cognitive challenges of individuals with TBI.

Among the four aids, respondents indicated that they would be more likely to use the memory and post-writing aids than the attention and social cue interpretation aids. The post-writing aid was rated by users as the most helpful of the four aids and easiest to use. Brunner and colleagues noted that many individuals with TBI already rely on writing supports such as Grammarly to produce messages on social media (17). That familiarity might have contributed to acceptance of the post-writing aid, as it includes traditional spelling and grammar support.

While the memory aid was rated as potentially useful, about 30% of participants found the user interface unnecessarily complex. As shown in **Figure 3**, the memory aid consolidated previous posts and presented them to the user all at once, which inadvertently added visual complexity to the interface and increased the amount of information presented at once. Further investigation is needed to evaluate the tradeoff between memory-recovery benefit vs. visual and informational complexity cost associated with such aids.

One potential reason for the low agreement on utility of the attention and social cue interpretation aids is that the mock-ups did not fully convey the aids' functionalities and did not offer the experience of seeing the aid in operation while using Facebook. In the context of the attention aid, as seeing the aid function while being presented a large amount of self-relevant information might be necessary to effectively experience the aid's functioning. In the context of the social cue interpretation aid, users might have to experience the difficulty of understanding or interpreting content to appreciate the potential value of such an aid. A second reason, and one that might underlie many of the results, is that individuals with TBI often underestimate their own cognitive challenges "in the moment," (64, 65) and thus might not have appreciated that they had challenges that the aids could help overcome. While the memory and post-writing aids were similar to what individuals without TBI might use (e.g., commercial products like Grammarly or smart phone apps)) and thus would have face validity without the individual needing to be aware of their own challenges, the social cue perception aid in particular would have been novel to participants and thus might have seemed

unnecessary. In the future, it would be helpful to collect subjective and objective measures of participants' cognitive abilities, including social cognition, to determine if insight into one's deficit is a factor in perceived utility of technology aids.

Consistent with the well documented heterogeneity among people with TBI (e.g., 66), the individual differences in attitudes toward the technology aids that we identified in our survey is reflective of the wide range of challenges, needs, and preferences of individuals with TBI. We argue that rehabilitation professionals will play a key role in personalizing the social media use of each individual based on the unique deficits, use patterns, and preferences. Rehabilitation professionals already report that they see social media use as a way to reduce social isolation following brain injury and may play an important supporting role in addressing social media barriers and participating safely on social media platforms (e.g., avoiding online scams) (18). We envision that rehabilitation professionals may also play a critical future role in helping individuals with TBI determine if they might benefit from the type of social media aids reported here and personalizing the social media use of each individual based on their cognitive profile and social media use goals. Indeed, rehabilitation professionals, including speech-language pathologists, are particularly well positioned to help individuals with TBI understand how cognitive-communication deficits that are present in face-to-face interactions can extend to computer-mediated communication and can provide training on the features and functionality of future social media aids.

Understanding the utility of these aids requires information about which individuals may most benefit from or be most willing to try social media aids. In an attempt to obtain some preliminary data on individual differences, we conducted an *ad hoc* exploratory analysis on the relation between individual demographic characteristics and potential utility and ease of use of the aids. Participant age had a significant positive correlation with perceived ease of use and utility of social media support tools. An individual's ability to adopt new technology decreases with age later in life (67), which might predict a negative correlation of age with ratings, but older adults in the U.S. are as active on Facebook as younger adults (68). The correlation with age merits replication in the future. Finally, despite the unique cognitive-communicative challenges individuals with TBI face in social media use, younger individuals might more readily accept social media platforms as designed, and older users might see themselves more as benefiting from aids that facilitate their use.

Although there is some evidence of a female advantage in social perception skills in adults with TBI (69), we did not find any effect of sex on perceptions of the social media supports in our study. There is evidence that women are more likely to seek help for healthcare-related concerns (70), but to our knowledge there is no evidence that this tendency extends to cognitive supports such as the aids proposed here.

Future directions

The current study presents several opportunities for future investigation with the proposed aids. First, based on the initial evaluation on different types of aids, we can prioritize the development of post-writing and memory aids over other types of aids. To extend the potential interests and adoption of these aids to individuals with TBI who are less conscious about their social media use after injury, future studies should consider intervention or tests that can raise awareness on one's social media use patterns and TBI symptoms. Second, in the current study, we found that age might have contributed to the acceptability of aids. However, due to the small sample size, we did not find how the effect of age differed with each aid. For example, we suspect that older adults with TBI might show a stronger interest in memory aids than younger adults with TBI. Future research should seek to better understand individual differences in attitudes towards social media aids with a larger study population. The potential utility of such aids can also be assessed in genuine clinical settings where rehabilitation specialists match the set of aids used by each individual to their cognitive profile, personalize these aids to their needs and preferences, and provide the appropriate training in their use. In this way, individuals with TBI would opt in or opt out of specific aids in the same way social media users can select among other display and security features for personalization.

Limitations

The study described here was an exploratory study that aimed to assess initial acceptance of the proposed social media aids. We conducted an online survey with static images of the design mock-ups. As a result, respondents might not have been able to fully understand the design concept and engage with the potential functionality of the aids. The critical next step, currently underway, is for participants to test and use the aids over time, to see the costs and benefits in real time. The study also was a relatively small sample of 46 individuals with TBI, and thus our results might not be representative of the general TBI population. The general findings, however, were similar to those reported in previous studies, and the sample was similar to those in others studies in regard to age, sex, social media experience, race, and socioeconomic status of participants. Finally, our findings are necessarily shaped by the specific decisions we have made in designing and creating mock-ups of the aids. Iterative

improvement, expert feedback, and usability testing of our designs can ensure future aids that are more effective and widely accepted.

Conclusions

Adults with TBI report significant barriers to using current social media platforms. We are working to develop technological supports to increase social media accessibility for people with TBI-related cognitive impairments. Here, we found initial support for social-media-specific technology aids to support social media access and social participation for adults with TBI. Future work should develop and deploy such aids and investigate user experience. Future work should also investigate the role of rehabilitation providers in personalizing the social media use of each individual based on the unique deficits, use patterns, and preferences.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Vanderbilt University Medical Center. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MCD, BM, and LST designed the study. ELM and MCD collection the data. FZ and HL analyzed the data. All authors contributed to the article and approved the submitted version.

Funding

This work is supported by NIH R01 HD071089 to MCD, BM, and LST.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Ellison N, Vitak J. Social network site affordances and their relationship to social capital processes. *Handbook Psychol Commun Technol.* (2015) 32:205–28. doi: 10.1002/9781118426456.ch9

2. de Zúñiga H G, Jung N, Valenzuela S. Social Media use for news and Individuals' social capital, civic engagement and political participation. *J Comput Mediat Commun.* (2012) 17(3):319–36. doi: 10.1111/j.1083-6101.2012.01574.x

3. Herring SC. Computer-mediated communication on the internet. *Annu Rev Inf Sci Technol.* (2002) 36(1):109–68. doi: https://doi.org/10.1002/aris.1440360104

4. Grieve R, Indian M, Witteveen K, Anne Tolan G, Marrington J. Face-to-face or Facebook: can social connectedness be derived online? *Comput Human Behav.* (2013) 29(3):604–9. doi: 10.1016/j.chb.2012.11.017

5. Veltri NF, Krasnova H, Günther O, Koroleva K. It's all about networking! Empirical investigation of social capital formation on social network sites. (Published online 2011). doi: 10.7892/BORIS.47120

6. Househ M, Borycki E, Kushniruk A. Empowering patients through social media: the benefits and challenges. *Health Informatics J.* (2014) 20(1):50–8. doi: 10.1177/1460458213476969

7. Eckler P, Worsowicz G, Rayburn JW. Social Media and health care: an overview. *PM&R.* (2010) 2(11):1046–50. doi: 10.1016/j.pmrj.2010.09.005

8. Mukherjee D, Reis JP, Heller W. Women living with traumatic brain injury: social isolation, emotional functioning and implications for psychotherapy. *Women Ther.* (2003) 26(1–2):3–26. doi: 10.1300/J015v26n01_01

9. Salas CE, Casassus M, Rowlands L, Pimm S, Flanagan DAJ. 'Relating through sameness': a qualitative study of friendship and social isolation in chronic traumatic brain injury. *Neuropsychol Rehabil.* (2018) 28(7):1161–78. doi: 10.1080/09602011.2016.1247730

10. Hoofien D, Gilboa A, Vaki E. Traumatic brain injury (TBI) 10?20 years later: a comprehensive outcome study of psychiatric symptomatology, cognitive abilities and psychosocial functioning. *Brain Inj.* (2001) 15(3):189–209. doi: 10.1080/026990501300005659

11. Turkstra LS, Williams W, Tonks J, Frampton I. Measuring social cognition in adolescents: implications for students with T…: eBSCOhost. *NeuroRehabilitation.* (2008) 23(6):501–9. doi: 10.3233/nre-2008-23606

12. Brunner M, Hemsley B, Palmer S, Dann S, Togher L. Review of the literature on the use of social media by people with traumatic brain injury (TBI). *Disabil Rehabil.* (2015) 37(17):1511–21. doi: 10.3109/09638288.2015.1045992

13. Eghdam A, Hamidi U, Bartfai A, Koch S. Facebook As communication support for persons with potential mild acquired cognitive impairment: a content and social network analysis study. *PLoS One.* (2018) 13(1):e0191878. doi: 10.1371/journal.pone.0191878

14. Brunner M, Palmer S, Togher L, Dann S, Hemsley B. "If I knew what I was doing on twitter then I would use it more": twitter experiences and networks of people with traumatic brain injury (TBI). *Brain Impair*. (2020) 21(1):1–18. doi: 10.1017/BrImp.2019.12

15. Tsaousides T, Matsuzawa Y, Lebowitz M. Familiarity and prevalence of Facebook use for social networking among individuals with traumatic brain injury. *Brain Inj*. (2011) 25(12):1155–62. doi: 10.3109/02699052.2011.613086

16. Feuston JL, Marshall-Fricker CG, Piper AM. *The social lives of individuals with traumatic brain injury. Proceedings of the 2017 CHI conference on human factors in computing systems*. ACM (2017). p. 182–94. doi: 10.1145/3025453.3025784

17. Brunner M, Palmer S, Togher L, Hemsley B. 'I kind of figured it out': the views and experiences of people with traumatic brain injury (TBI) in using social media—self-determination for participation and inclusion online. *Int J Lang Commun Disord*. (2019) 54(2):221–33. doi: 10.1111/1460-6984.12405

18. Brunner M, Togher M, Palmer S, Dann S, Hemsley B. Rehabilitation professionals' views on social media use in traumatic brain injury rehabilitation: gatekeepers to participation. *Disabil Rehabil*. (2021) 43(14):1955–64. doi: 10.1080/09638288.2019.1685604

19. Flynn MA, Mutlu B, Duff MC, Turkstra LS. Friendship quality, friendship quantity, and social participation in adults with traumatic brain injury. *Semin Speech Lang*. (2018) 39(5):416–26. doi: 10.1055/s-0038-1670672

20. Flynn MA, Rigon A, Kornfield R, Mutlu B, Duff MC, Turkstra LS. Characterizing computer-mediated communication, friendship, and social participation in adults with traumatic brain injury. *Brain Inj*. (2019) 33 (8):1097–104. doi: 10.1080/02699052.2019.1616112

21. Morrow E, Zhao F, Turkstra L, Toma C, Mutlu B, Duff MC. Computer-mediated communication in adults with and without moderate-severe traumatic brain injury: a survey of social media use. *J Med Int Res Rehabil Assist Technol*. (2021) 8(3):e26586. doi: 10.2196/26586

22. Facebook Accessibility. Available at: https://www.facebook.com/accessibility/posts/we-are-happy-to-announce-new-voiceover-gestures-for-accessing-the-delete-mute-an/806063152770745/ (Accessed June 29, 2021).

23. Arfaa J. An Improved Website Design for Elders Utilizing Social Networking Sites. (Published online 2015). p. 12.

24. Arfaa J, Wang Y. An accessibility evaluation of social Media websites for elder adults. In: G Meiselwitz, editors. *Social computing and social Media. Lecture Notes in Computer Science*. Cham: Springer International Publishing (2014). Vol. 8531. p. 13–24. doi: 10.1007/978-3-319-07632-4_2

25. Wu S, Wieland J, Faivar O, Schiller J. *Automatic alt-text: computer-generated image descriptions for blind users on a social network service. Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. ACM (2017). p. 1180–92. doi: 10.1145/2998181.2998364

26. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Admin Pol Ment Health*. (2011) 38(2):65–76. doi: 10.1007/s10488-010-0319-7

27. Vaccaro M, Hart T, Whyte J, Buchhofer R. Internet use and interest among individuals with traumatic brain injury: a consumer survey. *Disabil Rehabil Assist Technol*. (2007) 2(2):85–95. doi: 10.1080/17483100601167586

28. Ahmed O, Sullivan SJ, Schneiders A, Moon S, McCrory P. Exploring the opinions and perspectives of general practitioners towards the use of social networking sites for concussion management. *J Prim Health Care*. (2013) 5:36–42. doi: 10.1071/HC13036

29. 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. *J Biomed Inform*. (2009) 42(2):377–81. doi: 10.1016/j.jbi.2008.08.010

30. Martin I, Mcdonald S. Evaluating the causes of impaired irony comprehension following traumatic brain injury. *Aphasiology*. (2005) 19(8): 712–30. doi: 10.1080/02687030500172203

31. Channon S, Pellijeff A, Rule A. Social cognition after head injury: sarcasm and theory of mind. *Brain Lang*. (2005) 93(2):123–34. doi: 10.1016/j.bandl.2004.09.002

32. Martín-Rodríguez JF, León-Carrión J. Theory of mind deficits in patients with acquired brain injury: a quantitative review. *Neuropsychologia*. (2010) 48 (5):1181–91. doi: 10.1016/j.neuropsychologia.2010.02.009

33. Turkstra LS, Politis AM, Forsyth R. Cognitive–communication disorders in children with traumatic brain injury. *Dev Med Child Neurol*. (2015) 57(3):217–22. doi: 10.1111/dmcn.12600

34. McDonald S, Flanagan S. Social perception deficits after traumatic brain injury: interaction between emotion recognition, mentalizing ability, and social communication. *Neuropsychology*. (2004) 18(3):572–9. doi: 10.1037/0894-4105.18.3.572

35. Genova HM, Haight A, Natsheh JY, DeLuca J, Lengenfelder J. The relationship between social communication and social functioning in pediatric TBI: a pilot study. *Front Neurol*. (2019) 10(850). doi: 10.3389/fneur.2019.00850 Available at: https://www.frontiersin.org/articles/10.3389/fneur.2019.00850/full (Accessed June 28, 2021).

36. Dahlberg C, Hawley L, Morey C, Newman J, Cusick C, Harrison-Felix C. Social communication skills in persons with post-acute traumatic brain injury: three perspectives. *Brain Inj*. (2006) 20:425–35. doi: 10.1080/02699050600664574

37. Engberg AW, Teasdale TW. Psychosocial outcome following traumatic brain injury in adults: a long-term population-based follow-up. *Brain Inj*. (2004) 18 (6):533–45. Available at: https://www-tandfonlinecom.ezproxy.library.wisc.edu/doi/abs/10.1080/02699050310001645829 (Accessed June 7, 2021). doi: 10.1080/02699050310001645829

38. Douglas JM, Spellacy FJ. Correlates of depression in adults with severe traumatic brain injury and their carers. *Brain Inj*. (2000) 14(1):71–88. doi: 10.1080/026990500120943

39. Vakil E. The effect of moderate to severe traumatic brain injury (TBI) on different aspects of memory: a selective review. *J Clin Exp Neuropsychol*. (2005) 27(8):977–1021. doi: 10.1080/13803390490919245

40. Mathias JL, Wheaton P. Changes in attention and information-processing speed following severe traumatic brain injury: a meta-analytic review. *Neuropsychology*. (2007) 21(2):212–23. doi: 10.1037/0894-4105.21.2.212

41. Yeates KO, Armstrong K, Janusz J, Taylor HG, Wade S, Stancin T, et al. Long-Term attention problems in children with traumatic brain injury. *J Am Acad Child Adolesc Psychiatry*. (2005) 44(6):574–84. doi: 10.1097/01.chi.0000159947.50523.64

42. Knox L, Douglas JM, Bigby C. "I won't be around forever": understanding the decision-making experiences of adults with severe TBI and their parents. *Neuropsychol Rehabil*. (2016) 26(2):236–60. doi: 10.1080/09602011.2015.1019519

43. Cotrena C, Branco LD, Zimmermann N, Cardoso CO, Grassi-Oliveira R, Fonseca RP. Impaired decision-making after traumatic brain injury: the Iowa gambling task. *Brain Inj*. (2014) 28(8):1070–5. doi: 10.3109/02699052.2014.896943

44. Knox L, Douglas JM, Bigby C. 'The biggest thing is trying to live for two people': spousal experiences of supporting decision-making participation for partners with TBI. *Brain Inj*. (2015) 29(6):745–57. doi: 10.3109/02699052.2015.1004753

45. Krpan KM, Levine B, Stuss DT, Dawson DR. Executive function and coping at one-year post traumatic brain injury. *J Clin Exp Neuropsychol*. (2007) 29 (1):36–46. doi: 10.1080/13803390500376816

46. Mozeiko J, Le K, Coelho C, Krueger F, Grafman J. The relationship of story grammar and executive function following TBI. *Aphasiology*. (2011) 25:825–36. doi: 10.1080/02687038.2010.543983

47. Brunner M, Hemsley B, Togher L, Stephen D, Palmer S. Social Media and people with traumatic brain injury: a metasynthesis of research informing a framework for rehabilitation clinical practice, policy, and training. *Am J Speech Lang Patho*. (2021) 30(1):19–33. doi: 10.1044/2020_AJSLP-20-00211

48. Bosyj C, Baath S, Mutlu B, Duff MC, Turkstra L. Cognitive demands of computer-mediated communication: a scoping review. *Neuropsychol Rehabil*. (under review).

49. Palacios EM, Sala-Llonch R, Junque C, Fernandez-Espejo D, Roig T, Tormos J, et al. Long-term declarative memory deficits in diffuse TBI: correlations with cortical thickness, white matter integrity and hippocampal volume. *Cortex*. (2013) 49(3):646–57. doi: 10.1016/j.cortex.2012.02.011

50. Mathias JL, Mansfield KM. Prospective and declarative memory problems following moderate and severe traumatic brain injury. *Brain Inj*. (2005) 19 (4):271–82. doi: 10.1080/02699050400005028

51. Turkstra LS, Kraning SG, Riedeman SK, Mutlu B, Duff M, VanDenHeuvel S. Labeling facial affect in context in adults with and without TBI. *Brain Impair*. (2017) 18(1):49–61. doi: 10.1017/BrImp.2016.29

52. Rigon A, Voss MW, Turkstra LS, Mutlu B, Duff MC. Different aspects of facial affect recognition impairment following traumatic brain injury: the role of perceptual and interpretative abilities. *J Clin Exp Neuropsychol*. (2018) 40 (8):805–19. doi: 10.1080/13803395.2018.1437120

53. Turkstra LS, Duff MC, Politis AM, Mutlu B. Detection of text-based social cues in adults with traumatic brain injury. *Neuropsychol Rehabil*. (2019) 29 (5):789–803. doi: 10.1080/09602011.2017.1333012

54. Baker-Sparr C, Hart T, Bergquist T, Bogner J, Dreer L, Juengst S, et al. Internet and social media use after traumatic brain injury: a traumatic brain injury model systems study. *J Head Trauma Rehabil*. (2018) 33(1):E9–17. doi: 10.1097/HTR.0000000000000305

55. Duff MC, Morrow E, Edwards M, McCurdy R, Clough S, Patel N, et al. The value of patient registries to advance basic and translational research in the area of *traumatic brain injury*. *Front Behav Neurosci*. (2022) 16:846919. doi: 10.3389/fnbeh.2022.846919

56. Salmond CH, Menon DK, Chatfield DA, Pickard JD, Sahakian BJ. Changes over time in cognitive and structural profiles of head injury survivors. *Neuropsychologia*. (2006) 44:1995–8. doi: 10.1016/j.neuropsychologia.2006.03.013

57. Malec JF, Brown AW, Leibson CL, Flaada JT, Mandrekar JN, Diehl NN, et al. The mayo classification system for traumatic brain injury severity. *J Neurotrauma*. (2007) 24:1417–24. doi: 10.1089/neu.2006.0245

58. Brooke J. Sus: a "quick and dirty'usability. *Usability Eval Ind*. (1996) 189. p. 189–94. doi: 10.1201/9781498710411

59. Scheel A, Tiokhin L, Isager P, Lakens D. Why hypothesis testers should spend less time testing hypotheses. *Perspect Psychol Sci*. (2021) 16(4):744–55. doi: 10.1177/1745691620966795

60. Byom L, Duff M, Mutlu B, Turkstra L. Facial emotion recognition of older adults with traumatic brain injury. *Brain Inj*. (2019) 33(3):322–32. doi: 10.1080/02699052.2018.1553066

61. Rigon A, Voss MW, Turkstra LS, Mutlu B, Duff MC. Functional neural correlates of facial affect recognition impairment following TBI. *Brain Imaging Behav*. (2019) 13(2):526–40. doi: 10.1007/s11682-018-9889-x

62. Turkstra LS, Norman RS, Mutlu B, Duff MC. Impaired theory of mind in adults with traumatic brain injury: a replication and extension of findings. *Neuropsychologia*. (2018) 111:117–22. doi: 10.1016/j.neuropsychologia.2018.01.016

63. Rigon A, Turkstra LS, Mutlu B, Duff MC. Facial-affect recognition deficit as a predictor of different aspects of social communication impairment in traumatic brain injury. *Neuropsychology*. (2018) 32(4):476–83. doi: 10.1037/neu0000368

64. Dromer E, Kheloufi L, Azouvi P. Impaired self-awareness after traumatic brain injury: a systematic review. Part 2. Consequences and predictors of poor self-awareness. *Ann Phys Rehabil Med*. (2021) 64(5):101542. doi: 10.1016/j.rehab.2021.101542

65. Dromer E, Kheloufi L, Azouvi P. Impaired self-awareness after traumatic brain injury: a systematic review. Part 1: assessment, clinical aspects and recovery. *Ann Phys Rehabil Med*. (2021) 64(5):101468. doi: 10.1016/j.rehab.2020.101468

66. Covington N, Duff MC. Heterogeneity is a hallmark of traumatic brain injury, not a limitation: a new perspective on study design in rehabilitation research. *Am J Speech Lang Pathol*. (2021) 30(2S):974–85. doi: 10.1044/2020_AJSLP-20-00081

67. Yang K-C, Shih P-H. Cognitive age in technology acceptance: at what age are people ready to adopt and continuously use fashionable products? *Telemat Inform*. (2020) 51:101400. doi: 10.1016/j.tele.2020.101400

68. Pew Research Center. Social media fact sheet. (Published April 7, 2021). https://www.pewresearch.org/internet/fact-sheet/social-media/ (Accessed June 23, 2022).

69. Turkstra LS, Mutlu B, Ryan CW, Despins Stafslien EH, Richmond EK, Hosokawa E, et al. Sex and gender differences in emotion recognition and theory of mind after TBI: a narrative review and directions for future research. *Front Neurol*. (2020) 11:59. doi: 10.3389/fneur.2020.00059

70. Thompson AE, Anisimowicz Y, Miedema B, Hogg W, Wodchis WP, Aubrey-Bassler K. The influence of gender and other patient characteristics on health care-seeking behaviour: a QUALICOPC study. *BMC Fam Pract*. (2016) 17(1):38. doi: 10.1186/s12875-016-0440-0

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1058817

OPEN ACCESS

EDITED BY
Hehuang Xie,
Virginia Tech, United States

REVIEWED BY
Alessandro Didonna,
East Carolina University, United States
Ewoud Ewing,
Karolinska Institutet (KI), Sweden
Ivan Kiselev,
Medical Research Center for Cardiology, Ministry of Health of the Russian Federation, Russia

*CORRESPONDENCE
Michael D. Carrithers,
mcar1@uic.edu

RECEIVED 03 October 2022
ACCEPTED 28 November 2022
PUBLISHED 04 January 2023

CITATION
Bingen JM, Clark LV, Band MR, Munzir I and Carrithers MD (2023), Differential DNA methylation associated with multiple sclerosis and disease modifying treatments in an underrepresented minority population.
Front. Genet. 13:1058817.
doi: 10.3389/fgene.2022.1058817

Differential DNA methylation associated with multiple sclerosis and disease modifying treatments in an underrepresented minority population

Jeremy M. Bingen[1,2], Lindsay V. Clark[3], Mark R. Band[3], Ilyas Munzir[1] and Michael D. Carrithers[1,2,4]*

[1]Neurology, University of Illinois College of Medicine, Chicago, IL, United States, [2]Physiology and Biophysics, University of Illinois College of Medicine, Chicago, IL, United States, [3]High Performance Biological Computing, and Roy J Carver Biotechnology Center, University of Illinois, Champaign, IL, United States, [4]Neurology, Jesse Brown Veterans Administration Hospital, Chicago, IL, United States

Black and Hispanic American patients frequently develop earlier onset of multiple sclerosis (MS) and a more severe disease course that can be resistant to disease modifying treatments. The objectives were to identify differential methylation of genomic DNA (gDNA) associated with disease susceptibility and treatment responses in a cohort of MS patients from underrepresented minority populations. Patients with MS and controls with non-inflammatory neurologic conditions were consented and enrolled under an IRB-approved protocol. Approximately 64% of donors identified as Black or African American and 30% as White, Hispanic-Latino. Infinium MethylationEPIC bead arrays were utilized to measure epigenome-wide gDNA methylation of whole blood. Data were analyzed in the presence and absence of adjustments for unknown covariates in the dataset, some of which corresponded to disease modifying treatments. Global patterns of differential methylation associated with MS were strongest for those probes that showed relative demethylation of loci with lower M values. Pathway analysis revealed unexpected associations with shigellosis and amoebiasis. Enrichment analysis revealed an over-representation of probes in enhancer regions and an under-representation in promoters. In the presence of adjustments for covariates that included disease modifying treatments, analysis revealed 10 differentially methylated regions (DMR's) with an FDR <1E-77. Five of these genes (ARID5B, BAZ2B, RABGAP1, SFRP2, WBP1L) are associated with cancer risk and cellular differentiation and have not been previously identified in MS studies. Hierarchical cluster and multi-dimensional scaling analysis of differential DNA methylation at 147 loci within those DMR's was sufficient to differentiate MS donors from controls. In the absence of corrections for disease modifying treatments, differential methylation in patients treated with dimethyl fumarate was associated with immune regulatory pathways that regulate cytokine and chemokine signaling, axon guidance, and adherens junctions. These results demonstrate possible associations of

gastrointestinal pathogens and regulation of cellular differentiation with MS susceptibility in our patient cohort. This work further suggests that analyses can be performed in the presence and absence of corrections for immune therapies. Because of their high representation in our patient cohort, these results may be of specific relevance in the regulation of disease susceptibility and treatment responses in Black and Hispanic Americans.

KEYWORDS

epigenetics, biomarker, black, african American, hispanic, latino, dimethyl fumarate

Introduction

Multiple sclerosis (MS) is a major cause of non-traumatic neurologic disability in young adults. The prevalence of MS is increasing worldwide and is more common in underrepresented minority groups than previously thought (Weinstock-Guttman et al., 2003; Cree et al., 2004; Chinea et al., 2012; Caldito et al., 2018; Wallin et al., 2019). Although non-Hispanic Whites still have the highest prevalence rate for MS in the US, the demographics of newly diagnosed MS are also changing. One study of patients in the US demonstrated that Black American women had the highest incidence of MS and that Black men had a similar incidence as compared to White, non-Hispanic men (Langer-Gould et al., 2013). Analysis of the Gulf War military-veteran cohort also demonstrated a higher incidence of MS in Black Americans than other demographic groups (Wallin et al., 2012).

In addition, multiple studies have demonstrated increased disease severity and risk of long-term disability in Black American patients (Cree et al., 2004; Caldito et al., 2018; Wallin et al., 2018). Although studies in the modern era suggest that disease modifying treatments and improved diagnosis are associated with decreased long-term severity of MS (Sorensen et al., 2020), these observations may not be relevant to minority populations. These disparities in clinical outcomes and treatment responses may reflect social and environmental determinants of health as has been shown for other chronic diseases.

These determinants of health may impact the epigenome. One example is the regulation of DNA methylation, which is a dynamic process throughout the lifetime of an individual (Li and Zhang, 2014). The rationale for the study of epigenetic mechanisms in MS is that environmental factors such as stress, diet, and environmental exposures are all known modulators of DNA methylation. Some of these epigenetic mechanisms are associated with chronic inflammatory states (Celarain and Tomas-Roig, 2020). Most prior studies of global DNA methylation in MS have focused on individuals of Northern European ancestry. As in genome wide association studies (GWAS), the strongest association between MS and differential DNA methylation occurs at the HLA-DRB locus (Kular et al., 2018).

The approach in this study was to evaluate differential DNA methylation in a cohort of patients that are predominantly from underrepresented minority groups. This cohort is from our clinical practice at the University of Illinois, Chicago where approximately 55% of patients identify as Black or African American and 25% as Hispanic or Latino. The primary goal of this work was to identify epigenetic markers and related cellular signaling mechanisms that are associated with disease susceptibility in our patient population. In addition, challenges for the characterization of epigenetic biomarkers in a real-world setting is that most patients are on disease modifying treatments which may also regulate DNA methylation. An additional goal was to demonstrate the feasibility of identifying epigenetic biomarkers of disease and treatment in parallel analyses.

Results

Clinical phenotype variance in the MS cohort

MS patients (n = 29) and controls (n = 18) were recruited from our clinical practice at the University of Illinois, Chicago. A summary of demographic data for each group is shown in Table 1, and more detailed demographic and clinical data for each MS patient are shown in Supplmentary Table S1. Phenotypic variance of this patient cohort is shown in Supplmentary Figure S1 based on Functional Systems Scores. More extensive clinical phenotyping using network analysis has been performed on a larger number of patients from the same cohort (Howlett-Prieto et al., 2022).

Differential DNA methylation between MS and controls at specific probe sites

The next goal was to analyze patterns of differential DNA methylation between control and MS donors. As described in Methods, probes were filtered (n = 788,804) and adjusted for gender, age, and unknown covariates, some of which corresponded to disease modifying treatment. Adjusted

TABLE 1 Donor demographics for methylomic studies.

Group	N =	Age±SD	% Female (%)	%Black or african american (%)	%White, hispanic american (%)	%White, non-hispanic (%)
MS	29	43 ± 11	69	73	24	3
Control	18	43 ± 14	67	50	39	11

M-values were used to generate Mean Difference (MD) plots (Figure 1) (Su et al., 2017).

Analysis revealed distinct patterns of global differential methylation (Figure 1A, blue represents loci that are demethylated in MS *versus* controls and red increased methylation). For those DMP's (differentially methylated probes for a specific CpG region) with the greatest fold change differences [Log (FC) > 0.7 or < -0.7, fold change of greater 5x, FDR<0.01], these differences primarily reflected decreased methylation of those DMP's with lower average M-values (M value <0; blue, left lower quadrant). There were 174 DMP's that met these criteria in the left lower quadrant, 10 in the upper left, 1 in the lower right, and 4 in the upper right (Supplementary Table S2). These results suggested that those probes with the greatest differences between MS and controls were associated with demethylation of loci that have relatively low levels of methylation across all donors.

In contrast, global differential methylation patterns showed a more normalized distribution of relative increases or decreases of probe methylation in a comparison of all Black donors *versus* all White (Hispanic and non-Hispanic) individuals (Figure 1B). Global patterns of methylation were also analyzed in demographic subgroups (Figures 1C,D). These data suggested that the global pattern observed for MS *versus* Control (Figure 1A) occurs in both MS comparator groups but is most marked in the Hispanic-Latino group (Figure 1C).

Comparison of differentially methylated probes between racial and ethnic groups

The top 10,000 DMP's (Supplementary Tables S3–S5) for each comparator group (MS *versus* Control for all patients, MS *versus* Control Black American only, and MS *versus* Control Hispanic-Latino only) were analyzed to assess common and distinct probe sets (Supplementary Figure S2). There were 20,518 probes that were present in at least one of the comparator groups. We further identified 4395 probes unique for the Black American group and 6,025 probes for the Hispanic-Latino group (Supplementary Tables S6,S7). These data were not adjusted for unknown covariates.

KEGG pathway analysis was performed on the Top 10,000 DMP's for each comparator group (Supplementary Tables S8–S10). In the comparison between all patients, there was a possible association with sphingolipid and T cell signaling pathways (FDR = 0.01). In the Hispanic-Latino group, there was a possible association with apelin signaling (FDR = 0.07) which was not observed in the other comparator groups. Pathway analysis of the probes unique for the subgroups (Supplementary Figure S1) did not yield any statistically significant associations (FDR>0.18 for the Hispanic-Latino group and FDR>0.68 for the Black American group).

More robust results were obtained for pathway analysis following the removal of CpG regions associated with methylation quantitative trait loci (mQTL). Although known SNP regions were filtered prior to analysis, many probes remain in the data set that are associated with genetic variation at CpG loci (Min et al., 2021). For this reason, we performed analysis on subsets of probes that are associated mQTL and those that are not. The MeQTL Epic database (https://epicmeqtl.kcl.ac.uk) was utilized to identify mQTL associated with probes for the Illumina Infinium MethylationEPIC array (Villicana and Bell, 2021). In the comparison between all patients, 6,577 loci were identified that were not associated with mQTL (Supplementary Table S11). KEGG pathway analysis revealed 20 pathways potentially associated with differential methylation in MS (Table 2; $p < 0.005$, FDR<0.10). These pathways included those related to immune function such as hematopoietic cell lineage and chemokine signaling and unexpected associations with bacterial invasion of epithelia, amoebiasis, and shigellosis. Notably, no statistically significant associations were found for viral infections such as Epstein Barr ($p = 0.64$, FDR = 0.90). Significant associations were not observed for loci associated with mQTL or in the demographic subgroup analyses (FDR>0.10).

Comparison of differentially methylated probes with genome wide association studies

The International MS Genetics Consortium (IMSGC) reported a detailed analysis of currently available GWAS data identified a list of 551 non-MHC genes considered to be of high priority that are associated with peripheral immune function and microglia (International, 2019). We compared this gene list with our list of top DMP's (n = 20,518). 43 SNP regions from the prioritized gene list were present in genes that also contained DMP's at other loci (Supplementary Table S13).

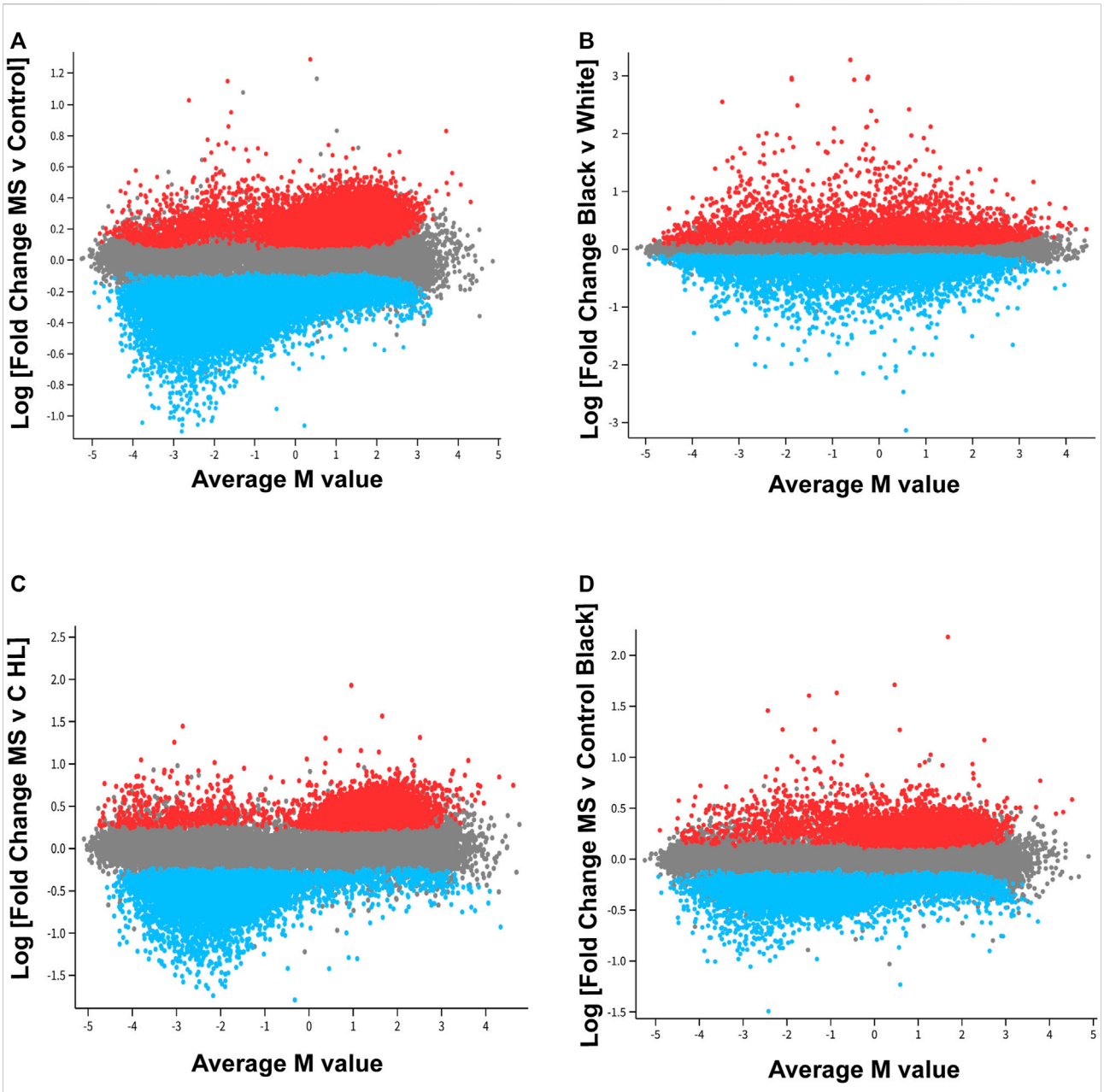


FIGURE 1
Distinct pattern of differential methylation of genomic DNA associated with MS. In the mean difference (MD) plots, average M-values across all donors is plotted on the x-axis. A negative number on the x-axis indicates decreased methylation for that locus [differentially methylated probe, DMP; CpG region), whereas as a positive number designates increased methylation as compared to other loci. Log fold change is plotted on the y-axis. **(A)**] For the MS *versus* control comparison, 52,295 DMP's were included in this analysis (FDR<0.01). An additional 20,000 probes were randomly selected for inclusion in the plot (gray). Red points designate probes that showed a relative increase in methylation MS as compared to controls, and blue points represent probes that demonstrated a relative decrease in methylation in the MS group. Two trends were observed. For those probes with the greatest fold change differences [Log (FC) > 0.7 or < −0.7, fold change of greater 5×, FDR<0.01], these differences primarily reflected decreased methylation of those DMP's with lower average M-values (M value <0; blue, left lower quadrant). If a less stringent cut off was used for fold change, the results suggested a tendency for significant DMP's that had an average methylation score over 50% (Average M value >0) to be hypermethylated in MS patients, and significant probes that had an average methylation below 50% to be hypomethylated in MS patients. **(B)** These trends were not observed in a comparison of all Black donors *versus* all White donors. The racial differences showed a more normalized distribution of relative increases or decrease in probe methylation as compared to the MS *versus* control analysis. **(C,D)**. MD plots were also performed in racial and ethnic subgroups. The trend observed in the MS *versus* control comparison was most pronounced in the Hispanic-Latino subgroup.

TABLE 2 KEGG pathway analysis of differential methylation in MS at loci not associated with mQTL.

KEGG pathway	Description	N (loci)	DE	P.DE	FDR
path:hsa04640	Hematopoietic cell lineage	91	29	2.91E-05	0.010
path:hsa05100	Bacterial invasion of epithelial cells	76	32	0.0001	0.020
path:hsa04611	Platelet activation	123	44	0.0003	0.021
path:hsa04062	Chemokine signaling pathway	189	55	0.0003	0.021
path:hsa04071	Sphingolipid signaling pathway	118	42	0.0003	0.021
path:hsa05418	Fluid shear stress and atherosclerosis	137	41	0.0003	0.021
path:hsa04973	Carbohydrate digestion and absorption	45	18	0.0007	0.035
path:hsa05146	Amoebiasis	98	33	0.0009	0.035
path:hsa04725	Cholinergic synapse	112	42	0.0009	0.035
path:hsa05131	Shigellosis	238	64	0.0014	0.044
path:hsa04912	GnRH signaling pathway	91	33	0.0014	0.044
path:hsa04750	Inflammatory mediator regulation of TRP channels	97	36	0.0015	0.044
path:hsa04660	T cell receptor signaling pathway	99	34	0.0021	0.056
path:hsa04014	Ras signaling pathway	228	68	0.0024	0.059
path:hsa04072	Phospholipase D signaling pathway	144	50	0.0026	0.060
path:hsa04668	TNF signaling pathway	109	32	0.0029	0.065
path:hsa05200	Pathways in cancer	515	132	0.0034	0.067
path:hsa05144	Malaria	49	15	0.0034	0.067
path:hsa04722	Neurotrophin signaling pathway	114	38	0.0041	0.076
path:hsa05221	Acute myeloid leukemia	64	24	0.0043	0.076

DE: discrete elements (genes); FDR: false detection rate.

TABLE 3 Cell type composition analysis.

Condition	CD8	CD4	NK	B cell	Monocyte	Neutrophil
Control	0.12 ± 0.04	0.13 + 0.03	0.06 + 0.02	0.07 ± 0.04	0.09 ± 0.02	0.56 ± 0.08
Multiple Sclerosis	0.09 ± 0.04	0.10 ± 0.06	0.05 ± 0.02	0.06 ± 0.04	0.10 + 0.02	0.64 ± 0.11

Values are Mean ± standard deviation.

Analysis of gene regulatory regions reveals over-representation of enhancer regions in MS associated DMP's

Enrichment analysis was performed to determine if there was an over-representation of enhancer or promoter regions among those DMP's that were associated with MS. For these analyses, the top 10,000 statistically significant DMP's (Supplementary Tables S3–S5; n = 20,518) in the 3 comparator groups (MS *versus* Control for All, Black and Hispanic American subgroups) were compared to the proportion of gene regulatory elements in the full data set (n = 788,804). The following databases and regions were analyzed: FANTOM5 (functional annotations of the mammalian genome, version 5) enhancers (Andersson et al., 2014), ENCODE (encyclopedia of DNA elements) annotations for promoter and enhancer regions (Gerstein et al., 2012), TSS200 (transcriptional start site within 200 bp), and TSS1500 (transcriptional start site within 1,500 bp.

These analyses demonstrated over-representation of enhancer regions and reduced frequency of promoter regions in the MS datasets (Table 3). The most striking findings were for over-representation of FANTOM5 enhancer regions (odds ratio 3.90, $p < 1e\text{-}15$, Fisher's Exact Test for Count Data) and under-representation of ENCODE promoter regions (odds ratio 0.26, p <

TABLE 4 Enrichment analysis for DNA regulatory regions.

Group	DNA region	DMP proportion		Odds ratio*	*p*-value*
		Top DMP's in MS n = 10,000	All DMP's n = 788,804		
MS *versus* Control (All)	Enhancer FANTOM5	0.1295	0.0332	3.90	<1e-15
	Enhancer	0.2101	0.1535	1.47	<1e-15
	ENCODE/X450K				
	TSS200	0.0216	0.0433	0.49	<1e-15
	TSS1500	0.0393	0.0631	0.61	<1e-15
	Promoter associated	0.0325	0.1298	0.26	<1e-15
	ENCODE				
MS *versus* Control (Black or African American)	Enhancer FANTOM5	0.0915	0.0332	2.94	<1e-15
	Enhancer	0.2000	0.1535	1.38	4.33e-13
	ENCODE/X450K				
	TSS200	0.0249	0.0433	0.56	<1e-15
	TSS1500	0.0461	0.0631	0.72	<1e-15
	Promoter associated	0.0362	0.1298	0.25	<1e-15
	ENCODE				
MS *versus* Control (Hispanic or Latino)	Enhancer FANTOM5	0.0766	0.0332	2.42	<1e-15
	Enhancer	0.2053	0.1535	1.42	<1e-15
	ENCODE/X450K				
	TSS200	0.0388	0.0433	0.89	0.026
	TSS1500	0.0510	0.0631	0.80	3.78e-07
	Promoter associated	0.0442	0.1298	0.31	<1e-15
	ENCODE				

* Fisher's Exact Test for Count Data.
Abbreviations: TSS, transcriptional start site; DMP, differentially methylated probe; FANTOM5, functional annotations of the mammalian genome, version 5; ENCODE, encyclopedia of DNA, elements.

1e-15) in the MS *versus* Control (all donors) comparator group. Similar results were observed in the demographic subgroups.

Cell composition analysis

Because whole blood methylomics were utilized in this study, some differences may reflect differences in subsets of peripheral blood cells. For that reason, we also performed cell type composition analysis. Methylation data from flow sorted whole blood was used to estimate cell type composition for each sample (Salas et al., 2018). Although there was a trend toward a modest increase in neutrophils and a decrease in CD8 T lymphocytes in the MS group, these differences were not statistically significant (Table 4, Figure 2). Neutrophils were the predominant subtype in both MS and controls.

Differential methylation in HLA-DRB1 region is not associated with MS in our patient population

In addition to these comparisons, we performed analysis of 14 CpG loci within the HLA-DRB1 region which had previously

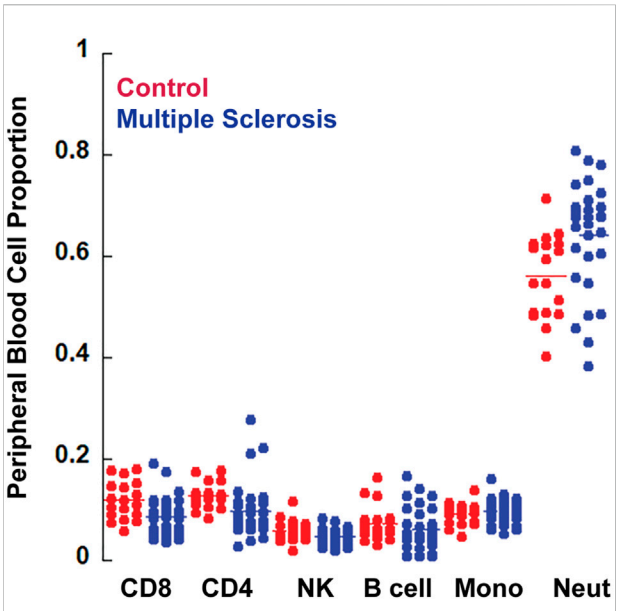


FIGURE 2
Cell composition analysis. The CellCounts2 function in the FlowSorted.Blood.EPIC package in R software was used to estimate differential peripheral blood cell counts. No statistically significant differences were observed between MS and controls.

been shown to be differentially methylated in MS patients recruited in Scandinavia and Germany (Kular et al., 2018). Unfiltered and unadjusted M values were utilized for this analysis because ten of the CpGs had to be retrieved from the dataset before filtering. Six of them were within known SNPs, three were known to be cross-reactive, and two (including one of the cross-reactive probes) were not detected in all samples. Out of the 19 CpGs in the region identified by Kular et al., 2018, two were not analyzed here due to being absent from the Infinium MethylationEPIC bead array chip, and three are not shown here because they did not show significant methylation differences.

Differential methylation was observed at the remaining 14 loci but did not necessarily indicate disease state (Figure 3). Four out of 18 control samples were hypomethylated in this region, whereas 14 out of 29 MS samples were hypomethylated. These findings primarily reflect differences in HLA-DRB1 genotype in our patient population. Hypomethylation in this region (14 MS and 4 controls) likely signifies that they are HLA-DRB1*15:01 positive.

Identification of gene-level biomarkers by DMR analysis

The next goal was to identify gene level differences between MS and controls. The DMRcate package was used to identify differentially methylated regions (DMR's) (Peters et al., 2015). DMR's contain multiple CpG loci that may be differentially methylated within a particular gene. This analysis increases the statistical power. As with the MD plots (Figure 1), the adjusted M values were used for this analysis and included corrections for disease modifying treatments.

We first analyzed differences in a comparison of all MS patients *versus* all Controls, irrespective of race or ethnicity. This analysis revealed 10,450 regions of interest (Table S14; FDR<1.93e-6, HMFDR ≤0.005). Using hierarchal clustering analysis, a subset of 147 DMP's (Supplementary Table S15) within the top 10 DMR's (FDR<1E-77, HMFDR≤1.15<e-6) was sufficient to differentiate MS from controls (Table 5; Figure 4A). Gene regions included: ARID5B, BAZ2B, CDK2AP1, CLU, CTSZ, RAB34, RABGAP1, SFRP2, TNFSF12-TNFSF-13, and WBP1L. These genes were not found to be differentially methylated in a comparison of all Black American donors *versus* all White donors (not shown). ARID5B, BAZ2B, RABGAP1, SFRP2, and WBP1L have not been previously associated with MS risk, and all are associated with neoplastic diseases and cellular proliferation.

In the hierarchal cluster analysis, approximately 86% of loci (123/143) demonstrated relative demethylation in the MS group as compared to controls. Relative demethylation at these DMR's was observed for 9/10 of the gene regions (maximal and mean differences in M-values, Table 5). Taken together with the distribution of probes in the MD plot (Figure 1), these results suggested a tendency toward relative demethylation of DMR's in the MS group compared to controls (Table 6).

Multi-dimensional scaling (MDS) was also used to assess similarities in the DMR datasets of differential methylation based on disease state (MS *versus* controls, Figure 4B). As with the hierarchal cluster analysis, the 143 probes in the top 10 DMR were sufficient to differentiate MS from controls (Figure 4B). The goodness of fit (GOF) for this MDS analysis was 0.72.

As described for the analysis of the top DMP's, we also performed mQTL analysis of the 143 probes within the top 10 DMR. There were 66 regions associated with mQTL (Supplementary Table S16) and 77 that were not (Supplementary Table S17). Hierarchal cluster analysis and MDS plots are shown in Supplementary Figures S3,S4. The GOF was 0.71 for those associated with mQTL and 0.74 for those not. These analyses also showed differentiation of MS from controls. Plots of the eigenvalues for the MDS plots are shown in Supplementary Figure S5.

Identification of gene-level biomarkers by DMR analysis in racial and ethnic subgroups

DMR analysis was also performed in racial and ethnic subgroups. In the comparison of MS *versus* Control in the Black American subgroup, 3127 regions of interest were

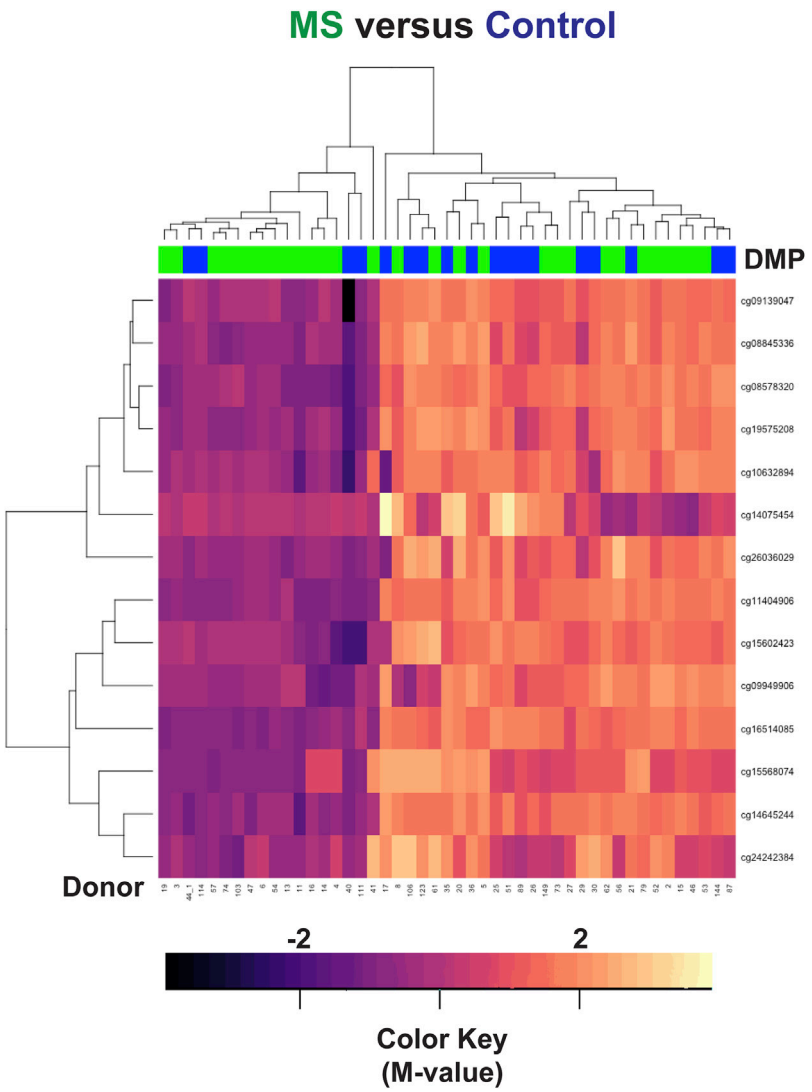


FIGURE 3
Hierarchal clustering analysis of the HLA-DRB1 region. Differences in methylation were observed at 14 CpG regions within HLA-DRB1 but did not necessarily correlate with disease state. In the heatmap, MS donors are designated by green, and controls by blue at the top of the heatmap. Blue/purple designates relative demethylation and orange/red increased methylation at a specific CpG site.

identified (Supplementary Table S18, FDR<1.94e-5). The top 10 DMR's included 6 regions that were present in the analysis of all donors (ARID5B, CDK2AP1, CLU, CTSZ, RABGAP1, and TNFSF12-TNFSF-13) and 4 other regions that reached statistical significance in the analysis of all donors but that were not in the top 10 regions (DUSP6, FOXI2, GPX6, and SPI2). The hierarchal cluster analysis for 72 DMP's (Supplementary Table S19) within the top 10 DMR's is shown in Figure 5A and the MDS plot in Figure 5B (GOF = 0.81).

In the Hispanic-Latino subgroup, 2285 regions of interest were identified (FDR<1.01e-10; Supplementary Table S20). The top 10 DMR's were: HOXD8, HPS4, KCNIP4, mir124-2, PTCHD4, PHYHIPL, RAB32, TREML2, UNC5, and WBSCR17. In the hierarchal clustering analysis of this subgroup, 55 probes within the top 10 DMR's was used (Supplementary Table S21). One control outlier was observed in the MS cluster (Figure 5C). However, MDS analysis showed differentiation between MS and controls (GOF = 0.75, Figure 5D).

Confirmation of DMR results by pyrosequencing

Three DMR regions (BAZ2B, CLU, and RABGAP1) were selected for confirmation by pyrosequencing. The common

TABLE 5 Top 10 differentially methylated regions (DMR) associated with multiple sclerosis.

DMR	Chromosome	Start	End	#CpGs	FDR (min smoothed)	HMFDR	Max difference M Value (MSvCon)	Mean difference M Value (MSvCon)	Overlapping genes
1	chr8	27467783	27470225	14	1.20E-136	2.43E-07	−0.05091	−0.03652	CLU
2	chr17	27044169	27045894	21	6.15E-111	1.79E-05	−0.06709	−0.03536	RAB34
3	chr9	1.26E+08	1.26E+08	14	1.89E-107	9.11E-07	−0.07037	−0.0436	RABGAP1
4	chr10	63807168	63809170	17	1.50E-94	3.11E-06	0.056923	0.040503	ARID5B
5	chr17	7460485	7462249	15	8.03E-94	3.34E-06	−0.05216	−0.02834	TNFSF12-TNFSF13
6	chr12	1.24E+08	1.24E+08	14	2.76E-90	1.35E-06	−0.06444	−0.03567	CDK2AP1
7	chr20	57581529	57583709	27	1.91E-89	1.58E-05	−0.0482	−0.0146	CTSZ
8	chr10	1.05E+08	1.05E+08	14	8.26E-82	8.14E-07	−0.04491	−0.02297	WBP1L
9	chr4	1.55E+08	1.55E+08	36	1.98E-81	1.97E-05	−0.04727	−0.01471	SFRP2
10	chr2	1.6E+08	1.6E+08	16	6.13E-78	1.15E-06	−0.07885	−0.03049	BAZ2B

DMR: differentially methylated region, CpG: 5′-cytosine-phosphate-guanine-3′, FDR: false detection rate, HMFDR: harmonic mean of individual CpG FDR's.

feature of these regions is that they all contain multiple loci in close proximity that demonstrated relative demethylation in the TSS1500 region (1,500 bp upstream of the transcriptional start site) in the MS group. The regions and assay details are shown in Table 7. A representative example of the pyrosequencing analysis is shown for the BAZ2B gene region in Figure 6. The sequences of interest for the 3 genes (BAZ2B, CLU, and RABGAP1) contain 3 CpG sites, and data were pooled for analysis of each of the differentially methylated regions. There was a statistically significant reduction in relative percent methylation at the CpG sites within each of the analyzed regions for BAZ2B ($p < 0.0001$), CLU ($p < 0.0001$), and RABGAP1 ($p = 0.0004$) (Table 7).

Identification of gene-level differentially methylated regions associated with dimethyl fumarate treatment

Analysis was also performed in the MS subgroup (n = 29) to compare differential methylation of those patients treated with dimethyl fumarate (n = 12) *versus* all other individuals with MS (n = 17, 8 on glatiramer acetate, 6 on ocrelizumab, 1 on beta-interferon, and 1 untreated). Probes were filtered as described above, and adjustments were made for gender, age, race, and latent variables. However, unlike the prior analyses, disease modifying treatments were listed as a factor to be preserved. This analysis showed 1485 DMP's (Supplementary Table S22) with an FDR<0.01 (p < 2E-5) and 12,915 with an FDR<0.1 ($p < 0.01$). KEGG pathway analysis of those probes identified possible associations with cytokine receptor interactions, adherens junction regulation, chemokine signaling, and axonal guidance (Table 8).

Hierarchal cluster and MDS analysis were performed on 77 DMP's (Supplementary Table S23) within the top 10 DMR's (FDR<2.5E-40). These gene regions included: CLASP2, CLU, DOK3, GPR146, PARVB, PARVG, RAB34, SLC11A2, TAGLN3, and WBP1L. Four of these genes (CLASP2, PARVB, PARVG, and TAGLN3) regulate the cytoskeleton, and three of them were also identified in the top 10 DMR's for the MS *versus* controls comparison (CLU, RAB34, and WBP1L). As shown in Table 8 (maximal and mean differences between groups), all these regions showed relative demethylation in the DMF group as compared to those not on DMF. As shown in the heatmap of the hierarchal clustering analysis (Figure 7A) and in the MDS plot (Figure 7B), analysis of these regions was sufficient to distinguish those individuals on dimethyl fumarate *versus* all other MS patients (GOF = 0.81 for the MDS analysis).

Discussion

This epigenome-wide association study demonstrated unique patterns of global and gene level differential DNA methylation in our MS patient population. To our knowledge, this study is the first to focus on differential DNA methylation in an underrepresented population of MS patients in the United States. Notable findings included distinct global patterns of differential demethylation in MS, a preferential association with enhancer regions rather than promoters, and identification of novel gene level biomarkers associated with MS and disease

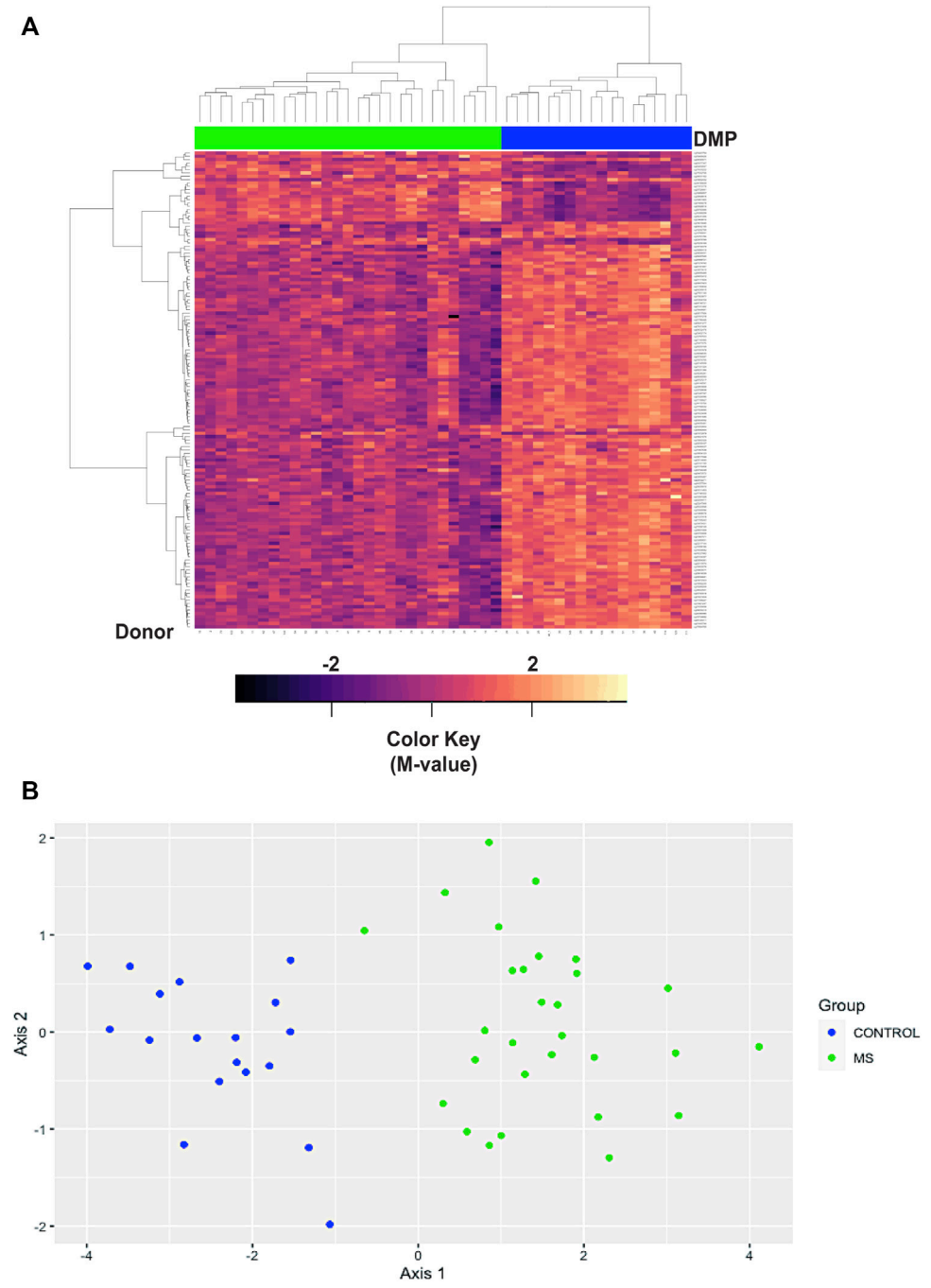


FIGURE 4
Hierarchal clustering and multi-dimensional scaling (MDS) analysis of differentially methylated regions associated with MS. **(A)** Hierarchal clustering analysis was performed on 147 differentially methylated loci within the top 10 differentially methylated regions (Supplementary Table S15). In the heatmap, MS donors are designated by green, and controls by blue at the top of the heatmap. Blue/purple designates relative demethylation and orange/red increased methylation at a specific CpG site. Gene regions included: ARID5B, BAZ2B, CDK2AP1, CLU, CTSZ, RAB34, RABGAP1, SFRP2, TNFSF12-TNFSF-13, and WBP1L. **(B)** MDS analysis was performed on the same data set and showed a goodness of fit (GOF) of 0.72.

modifying treatments. There was a tendency for many of the differentially methylated regions to demonstrate relative demethylation in MS. In addition, pathway analysis suggested possible associations of epigenetic biomarkers of cellular differentiation, Shigellosis, and amoebiasis in our patient cohort.

TABLE 6 Pyrosequencing analysis.

Chromosome	Assay ID#	Position	Strand	Name	Gene	Region	CpG#
chr2	PM00685013	160473461	+	cg17503977	BAZ2B	TSS1500	3
chr9	PM00685139	125795935	+	cg14115756	RABGAP1	TSS1500	3
chr8	PM00683935	27469338	+	cg13488078	CLU	TSS1500	3
Condition	**BAZ2B**	**CLU**	**RABGAP1**				
	%Methylation	**%Methylation**	**%Methylation**				
MS (n = 6)	24.9 ± 2.3**	17.1 ± 1.5**	15.3 + 1.2*				
Control (n = 4)	39.6 ± 2.0**	25.6 ± 1.4**	21.2 + 1.4*				

*$p = 0.004$, **$p < 0.001$.

The most notable observation in the analysis of global differential DNA methylation in MS was a tendency toward demethylation of probe regions that demonstrated relatively low levels of methylation across all donors. Although exceptions exist, there is a tendency for CpG islands to be hypomethylated in normal cells and hypermethylated in neoplastic cells (Bird and Wolffe, 1999; Baylin et al., 2001; Michal and Wojtas, 2019). Some cancers, such as high-grade pediatric gliomas, are associated with DNA hypomethylation (Bender et al., 2013). Many prior studies suggested that hypomethylated regions in a variety of cancers occur at introns and intergenic regions (Wilson et al., 2007). More recent studies in chronic lymphocytic leukemia and other hematologic malignancies revealed an association of hypomethylation of promoter regions (Upchurch et al., 2016). This observation may be most relevant to this study because we examined whole blood methylomics. Taken together with our pathway analysis results, these data suggest a possible relationship between the pathogenesis of MS in some patient populations and hematological malignancies.

Consistent with those results, we also observed increased representation of enhancer regions and decreased frequency of associations with promoters among differentially methylated probe regions. These differential DNA methylation patterns of enhancers have been associated with neoplastic transformation, metastasis of solid tumors, and myelodysplastic diseases (Bell et al., 2016; Ordonez et al., 2019). In addition, gene level analysis suggested a pattern of hypomethylation of a subset of genes in putative promoter regions (TSS1500). For example, one of the genes analyzed by pyrosequencing analysis, BAZ2B, regulates chromatin structure and hematopoietic cell development (Arumugam et al., 2020). These results will need to be confirmed in larger data sets.

Pathway analysis revealed unexpected associations with gastrointestinal infections due to bacteria and parasites but not with viral infection. The specific pathways identified were for Shigellosis, amoebiasis, and bacterial invasion of epithelia. Shigellosis may be particularly relevant to our patient population because frequent outbreaks have been identified in Chicago (Jones et al., 2006). In addition, one prior genetic study suggested an association of the Shigellosis pathway with MS and Crohn's disease (Restrepo et al., 2016). The possible association with amoebiasis may be relevant to our Hispanic-Latino population who have emigrated from Mexico and Central America and those who have relocated from Puerto Rico. Although these findings need to be examined in greater detail, they suggest that prior bacterial and parasitic gastrointestinal infections may be more relevant to MS susceptibility in our patient cohort than prior viral infections such as Epstein Barr Virus (Bjornevik et al., 2022).

Prior studies of differential DNA methylation in MS have focused primarily on patients of Northern European ancestry (Kulakova et al., 2016; Kular et al., 2018; Souren et al., 2019; Kiselev et al., 2021). A prior study demonstrated relative hypomethylation of the HLA-DRB1 region in MS patients (Kular et al., 2018). We examined this region in our patient cohort and observed a subset of individuals that had hypomethylation of this region, but it did not correlate with disease state. Although these results may suggest that epigenetic biomarkers of MS may differ between racial and ethnic groups, these data likely reflect differences between HLA-DRB1 haplotypes in our patient population. Several other recent studies of differential methylation in MS have focused on specific immune cell populations. These include analyses of $CD4^+$ T lymphocytes (Ewing et al., 2019; Roostaei et al., 2021), $CD8^+$ T lymphocytes (Li et al., 2017; Deng et al., 2019; Ewing et al., 2019), monocytes (Ewing et al., 2019; Diniz et al., 2021), and $CD19^+$ B lymphocytes (Maltby et al., 2018a).

For gene level analysis of differential methylation associated with MS in our patient cohort, we focused on the top 10 DMR's: ARID5B, BAZ2B, SFRP2, WBP1L, CDK2AP1, CLU, CTSZ, RAB34, RABGAP1, and TNFSF12/TNFSF13. One of these gene regions, CTSZ, was previously reported to be hypomethylated in MS in post-mortem brain tissue (Huynh

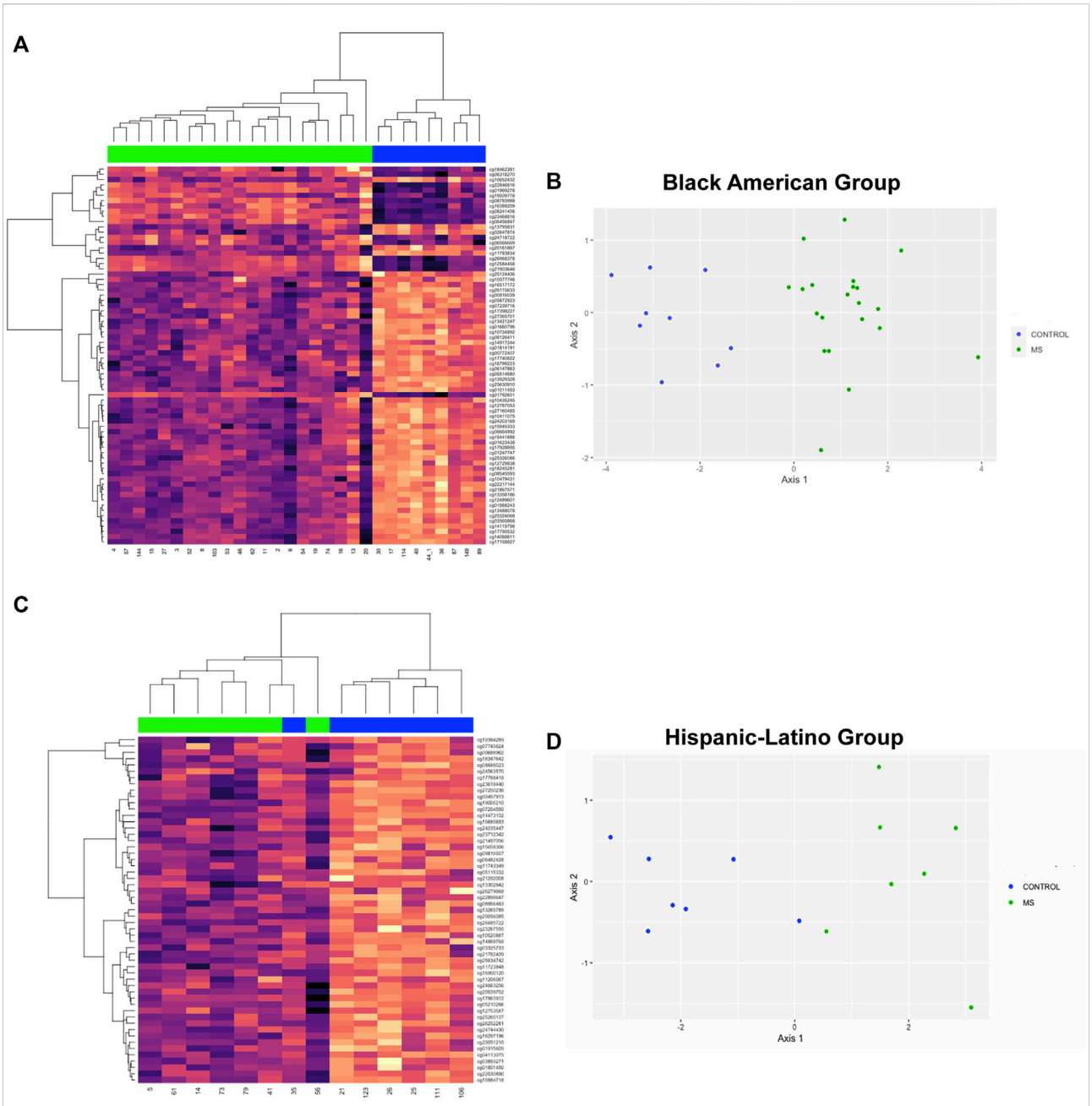


FIGURE 5
Hierarchal clustering and multi-dimensional scaling (MDS) analysis of differentially methylated regions associated with MS in racial and ethnic subgroups. **(A)** The top 10 DMR's for the comparison of MS *versus* Controls for Black Americans included 6 regions that were present in the analysis of all donors (ARID5B, CDK2AP1, CLU, CTSZ, RABGAP1, and TNFSF12-TNFSF-13) and 4 other regions that reached statistical significance in the analysis of all donors but that were not in the top 10 regions (DUSP6, FOXI2, GPX6, and SPI2). 72 DMP's (Supplementary Table S19) within the top 10 DMR's were used for hierachal cluster analysis. **(B)** MDS plot showed a GOF = 0.81. **(C)** In the Hispanic-Latino subgroup, the top 10 DMR's were: HOXD8, HPS4, KCNIP4, mir124-2, PTCHD4, PHYHIPL, RAB32, TREML2, UNC5, and WBSCR17. 55 probes within the top 10 DMR's were used for hierarchal cluster analysis (Supplementary Table S21). One control outlier was observed in the MS cluster. **(D)**) MDS revealed a GOF = 0.75.

et al., 2014) and to also be associated with risk of systemic sclerosis (Zhu et al., 2018). This latter observation may be of relevance to the current study because Black Americans have a higher incidence of systemic sclerosis and are at increased risk of a more severe disease course (Steen et al., 2012; Steen et al., 1963-1982). One potential mechanism for CTSZ to mediate pro-inflammatory effects is through increased interleukin 1β secretion by antigen presenting cells (Allan et al., 2017).

TABLE 7 KEGG pathway analysis of differential methylation with dimethyl fumarate treatment.

KEGG pathway	Description	N (loci)	DE	P.DE	FDR
path:hsa04060	Cytokine receptor interaction	280	85	0.00043	0.0756
path:hsa04360	Axon guidance	175	90	0.00052	0.0756
path:hsa04520	Adherens junction	70	41	0.00067	0.0756
path:hsa04062	Chemokine signaling pathway	189	79	0.00088	0.0756

DE, discrete elements (genes); FDR, false detection rate.

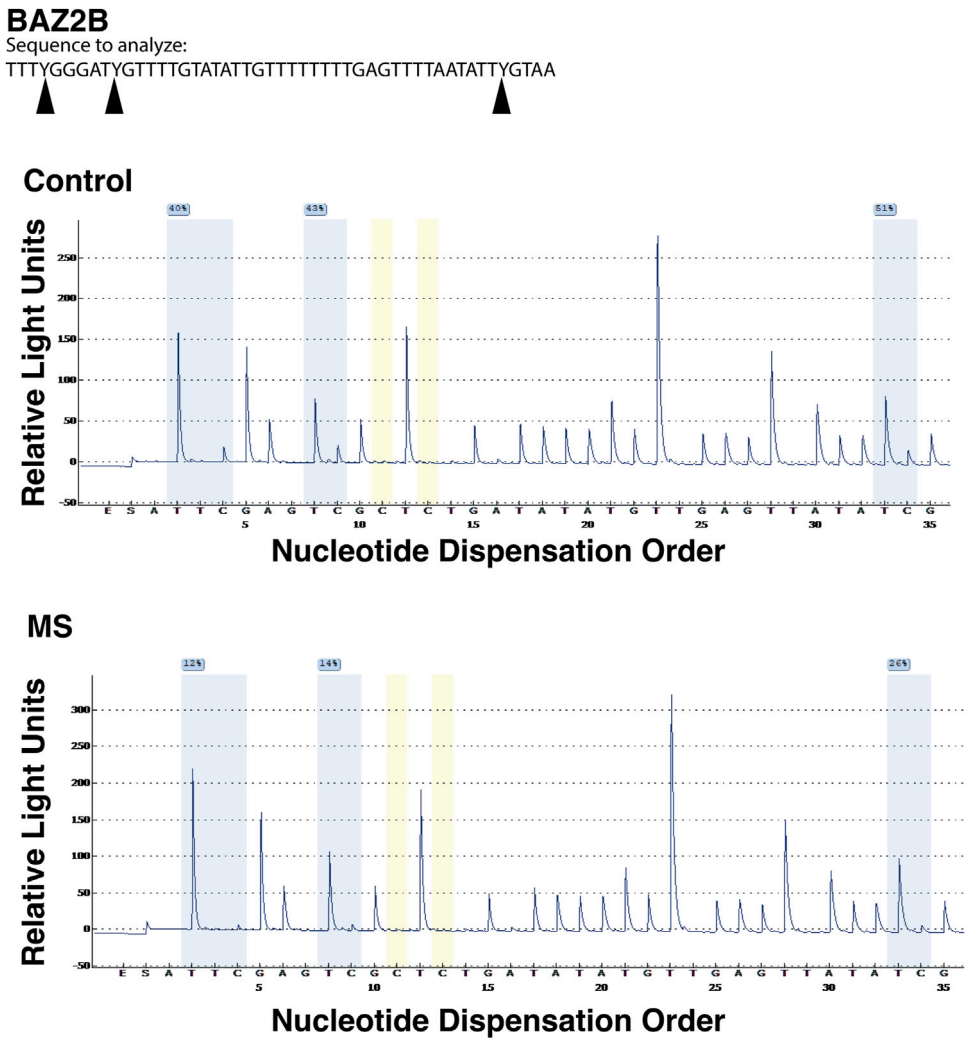


FIGURE 6
Pyrosequencing analysis of differentially methylated regions. Bead array data were confirmed by pyrosequencing following bisulfite conversion of the differentially methylated region of interest. In this representative analysis, the region of interest is in the promoter region (TSS1500) of BAZ2B gene (sequence is at the top of the figure). CpG regions are highlighted in blue, the regions utilized for bisulfite conversion controls are shown in yellow. For each of the 3 CpG regions analyzed, there was a relative reduction in the %methylation for the MS donor (bottom pyrogram) compared to the control (top).

Although the other regions have not been shown to be differentially methylated in prior studies, some of these genes have been associated with MS in genomic, transcriptomic, and proteomic studies. For example, CDK2AP1, a cell cycle regulator, was previously identified as an MS risk allele that correlated with reduced RNA expression in lymphoblast cells and peripheral

TABLE 8 Top 10 differentially methylated regions (DMR) associated with dimethyl fumarate treatment.

DMR	Chromosome	Start	End	#CpGs	FDR (min smoothed)	HMFDR	Max difference M Value (DMFvOther)	Mean difference M Value (DMFvOther)	Overlapping genes
1	chr22	44463707	44465038	10	2.57E-64	0.001067	−0.07046	−0.0306	PARVB
2	chr17	27044169	27045894	21	2.19E-55	0.01804	−0.07319	−0.03276	RAB34
3	chr10	1.05E+08	1.05E+08	14	5.76E-53	0.003279	−0.04699	−0.01822	WBP1L
4	chr3	1.12E+08	1.12E+08	15	1.42E-52	0.007249	−0.04847	−0.02573	TAGLN3
5	chr22	44568203	44568812	9	9.40E-51	0.004387	−0.05	−0.03259	PARVG
6	chr5	1.77E+08	1.77E+08	13	2.84E-50	0.003224	−0.04945	−0.02524	DOK3
7	chr12	51403056	51403966	6	5.95E-48	0.000422	−0.05586	−0.02471	SLC11A2
8	chr7	1094263	1096387	12	2.54E-46	0.004215	−0.04168	−0.02055	GPR146
9	chr8	27467783	27469673	13	1.19E-44	0.0076	−0.04168	−0.02666	CLU
10	chr3	33700962	33701707	9	2.44E-40	0.003399	−0.03639	−0.02407	CLASP2

*12 patients on dimethyl fumarate, 8 on glatiramer, 6 on ocrelizumab, 1 on interferon, and 1 on natalizumab.
DMR, differentially methylated region; CpG, 5′-cytosine-phosphate-guanine-3′; FDR, false detection rate; DMF, dimethyl fumarate; HMFDR, harmonic mean of individual CpG FDR's.

blood mononuclear cells (PBMC's) (IMSGC, 2010). It was also identified as a susceptibility gene for MS in the genomic map IMSGC study (Supplementary Table S13) (International, 2019). Other studies demonstrated increased RNA expression of CLU, a calcium binding protein, in peripheral blood from MS patients (Razia et al., 2022) and increased protein levels in cerebrospinal fluid (van Luijn et al., 2016). RAB34 (Liu et al., 2022) and the TNFSF12-TNFSF13 (Krumbholz et al., 2008) have also been associated with increased MS risk.

The more novel findings in this study were in genes (ARID5B, BAZ2B, RABGAP1, SFRP2, and WBP1L) that are associated with cancer risk and cellular differentiation. For example, ARID5B, has been associated with leukemia (Treviño et al., 2009), prostate cancer (Davalieva et al., 2015), gastric cancer (Lim et al., 2014), and endometrial carcinoma (Kandoth et al., 2013). As discussed above, BAZ2B regulates chromosome structure, hematopoietic cell development, and reprogramming of pluripotent stem cells (Arumugam et al., 2020). RABGAP1 regulates mitosis, cell migration, and mTOR signaling (Oh et al., 2022). SFRP2 is a tumor suppressor protein that can induce cell apoptosis, and differential methylation of its promoter region has been associated with leukemias and renal cancer (Jost et al., 2008; Konac et al., 2013; Li and Luo, 2018). WBP1L regulates proapoptotic pathways in myeloid cell leukemia (Morishima et al., 2011). Differential methylation of these 5 gene regions further suggests an association between regulation of neoplasia and cellular proliferation in our patient population. These epigenetic determinants may be mediated by environmental exposures that increase the risk of some cancers and autoimmune diseases (Dor and Cedar, 2018).

This study also demonstrated the feasibility of performing parallel analyses to detect associations with MS and disease modifying treatments. This analysis is important because it may inform treatment decisions based on biomarkers of medication responders and non-responders. In addition, in real world practice, monitoring of biomarkers over the patient's disease course requires approaches that allow ongoing disease modifying treatment to continue. In this study, we focused on differential DNA methylation associated with dimethyl fumarate treatment because that group represented the largest treatment cohort in this pilot study. Analysis of probes within the top DMR's demonstrated relative hypomethylation of these loci in the dimethyl fumarate group. In a prior study of the effect of dimethyl fumarate in CD4 T lymphocytes, four differentially methylated loci were observed, SNORD1A, SHTN1, MZB1 and TNF (Maltby et al., 2018b). We observed differential methylation of SHTN1 in our comparison of the top 10,000 DMP's for the MS *versus* Control comparison (Supplementary Tables S5–S7), but not in the comparison of dimethyl fumarate *versus* other treatments. One common feature is that differential methylation of TNF was observed in that study and in our own. Another study assessed differential methylation in monocytes and CD4 T lymphocytes prior to initiation of dimethyl fumarate and following treatment (Carlstrom et al., 2019). A potentially important observation between our study and theirs is an association of differential methylation with cytokine

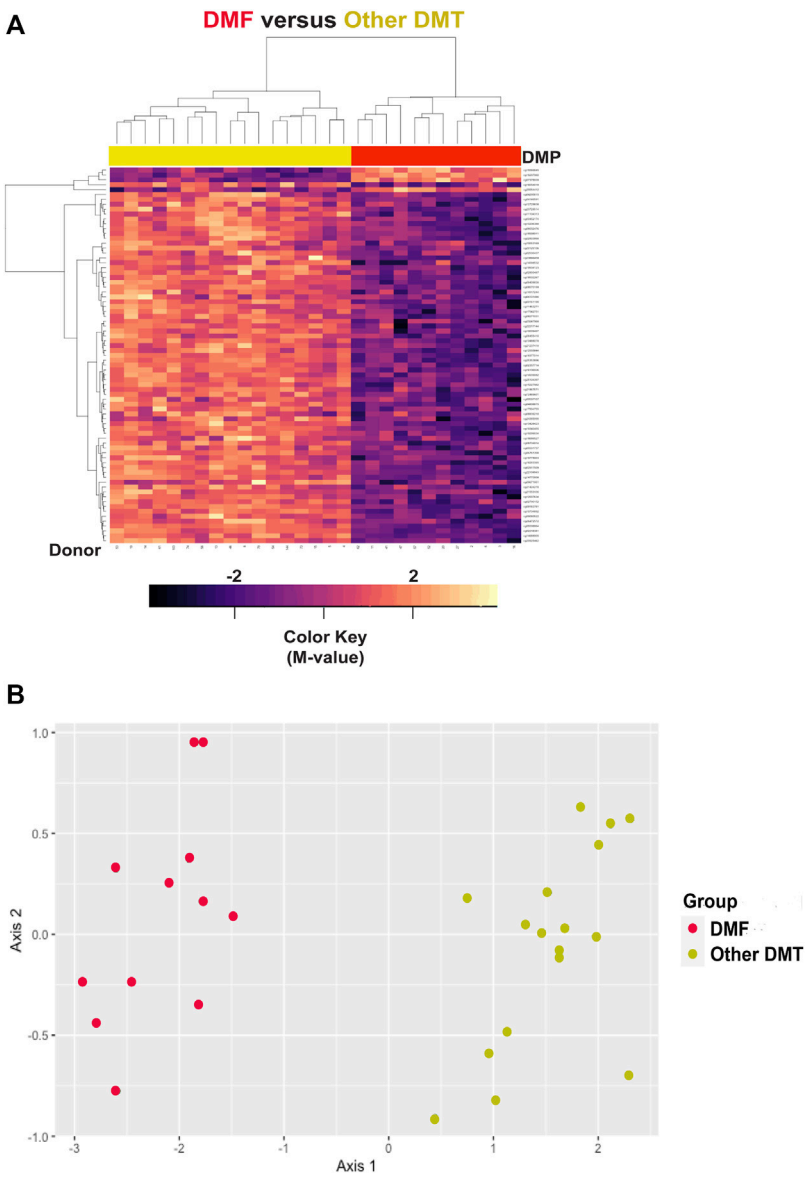


FIGURE 7
Hierarchal clustering and multi-dimensional scaling (MDS) analysis of differentially methylated regions associated with dimethyl fumarate treatment. **(A)** Heatmap is shown for hierarchal clustering analysis performed on differentially methylated loci within 77 differentially methylated loci within the top 10 differentially methylated regions (DMR) (Supplementary Table S23). The gene regions are: CLASP2, CLU, DOK3, GPR146, PARVB, PARVG, RAB34, SLC11A2, TAGLN3, and WBP1L. In the heatmap, dimethyl fumarate treatment is designated by red (n = 12), and controls by yellow at the top of the heatmap (n = 17). Blue/purple designates relative demethylation and orange/red increased methylation at a specific CpG site. **(B)** MDS analysis showed a goodness of fit of 0.81.

pathways, including IL6 and IL17 regulated signaling. In a study that focused only on global patterns of differential methylation, INFβ treatment significantly reduced global methylation in monocytes but not in lymphocyte of MS patients (Diniz et al., 2021). Additional studies are required in the future to assess potential biomarkers associated with other treatments.

This study has several limitations. One is that it is a pilot study on a limited number of patients from our clinical practice. Further studies are required with larger numbers of patients. In addition, some of the findings may reflect regional differences due to unique environmental exposures, and additional studies are necessary in other patient cohorts. However, even if some of the differential DNA methylation are due to regional differences, it is important to identify

those biomarkers associated with geographic location to better understand disease heterogeneity on a national level. Another limitation is that this study focused on whole blood methylomics, and many relevant methylation differences in the CNS may be missed. However, pathway and gene level analysis, including CLU and RAB34, revealed associations with axonal regulatory pathways and gene regions identified in prior studies of CNS tissue (Liu et al., 2022) and CSF (van Luijn et al., 2016). Our results are consistent with a prior study that demonstrated the feasibility of detecting CNS relevant differential methylation in peripheral blood samples. That study suggested that analysis of peripheral blood samples can detect approximately 20%–30% of differential methylation observed in live brain tissue (Braun et al., 2019). In addition, an important feature of using whole blood methylomics is that it can be assessed using a minimally invasive approach, which is important for longitudinal assessments in real world clinical practice. Integrated analysis with single cell approaches such as RNAseq also can be used to assess the relevance of differential DNA methylation in specific immune cell subtypes. An additional limitation of the current study is that HLA typing and genome-wide genotyping were not performed. Integrated analysis of these data with methylomics will need to be performed in future studies of our patient cohort. In addition, it will be important to analyze the associations of differentially methylated regions with environmental and social determinants of health, particularly for those regions not associated with SNP, eQTL, or mQTL regions.

Despite the limitations of this study, the results allow us to develop a working model to postulate possible pathobiological differences of MS susceptibility in select populations. Overall, the results suggest that DNA hypomethylation of many gene regions previously associated with neoplastic regulation are associated with MS susceptibility in Black and Hispanic American patients in our cohort. Additional studies are required to assess the relevance of these findings to the proliferation, invasiveness, and pathogenicity of specific immune cell populations.

Methods

Subjects

This was a cross-sectional, case-control study. 29 subjects with multiple sclerosis (MS) and 18 controls with non-inflammatory neurological disease were enrolled (Supplementary Tables S1). All subjects were followed at the University of Illinois-Neurosciences Center and were enrolled in the University of Illinois at Chicago (UIC) Neuroimmunology Biobank between August, 2018 and October, 2019. The UIC Neuroimmunology Biobank is approved by the Institutional Review Board (IRB) of the University of Illinois College of Medicine. All subjects provided informed written consent at enrollment.

Inclusion and exclusion criteria

MS donors met the following criteria: 1) age between 18 and 80 years at the time of enrollment, 2) a diagnosis of relapsing-remitting MS (RRMS) based on the McDonald criteria 2017 (Thompson et al., 2018), 3) no history of relapse(s) 30 days prior to the sample collection, 4) no history of receiving steroids within 30 days prior to the sample date, 5) no MRI activity within 30 days prior to the sample collection date (if MRI available), and 6) availability of the clinical data at the time of sample collection. The control group met these criteria: 1) between 18 and 80 years at the time of sampling, 2) presentation with a neurological complaint other than a neuro-inflammatory or neurodegenerative disorder, 3) no history of a recent ischemic stroke within the 6 months prior to the sample date, 4) no history of a systemic autoimmune disease, and 5) ambulatory without assistance at the time of sampling. Exclusion criteria for both MS and control subjects were: 1) failing to meet the inclusion criteria or 2) being on an immunomodulating or immunosuppressant agent other than the disease modifying treatments for MS within 6 months prior to the sample date.

Whole blood methylomics of genomic DNA

Whole blood genomic DNA (gDNA) was isolated from whole blood using EZ1 Advanced XL automated instrument (Qiagen Cat. No. 9001875) using EZ1&2 DNA blood 350 ul kit (Qiagen Cat. No.951054). Infinium MethylationEPIC bead arrays (Illumina) were utilized to characterize whole blood genomic DNA (gDNA) methylation of MS patients and controls Samples were randomized on the chip. All samples had very high CpG detection rates, and, therefore, none needed to be removed from the analysis.

Normalization and filtering

Analysis was performed in R software (version 4.0.3). The session information and packages utilized are shown in Supplementary Table S24. Data were normalized using the preprocessQuantile function from the minfi R package (Aryee et al., 2014). Probes were filtered and removed from analysis for low detection value in one or more samples (4,726 probes), sex differences (19,072 probes on the X and Y chromosomes), CpG sites associated with known SNP's (single nucleotide polymorphisms; 28,567 probes), and probe cross-reactivity (24,690 probes) (Chen et al., 2013). 788,804 probes remained for further analysis.

Linear model of differential methylation

Using the lmFit and eBayes functions from the limma package (Ritchie et al., 2015), preliminary models were fit as:

$$Y = Group + Race + Gender + Age$$

For all 47 samples, and:

$$Y = Group + Gender + Age$$

for the 30 Black American samples, and separately in the 14 Hispanic-Latino samples, where Y is the M-value indicating the degree of methylation at a given CpG, and Group indicates the MS or control group.

Removal of Unwanted Variation (Jacob et al., 2016) was used to identify any latent variables that should be included in analysis. The method of Buja and Eyuboglu (Buja and Eyuboglu, 1992) was implemented in the num.sv function of the sva package (Leek and Storey, 2007) indicated that 6 latent variables should be included in the model with all 47 samples, 4 latent variables should be included in the model with the 30 Black American samples, and 3 latent variables should be included in the model with the 14 Hispanic-Latino samples. Negative control probes were selected as those that had a *p*-value >0.5 for every effect in each of the respective above models, which yielded 27413 probes for the model with all 47 samples, 80572 probes for the model with the 30 Black American samples, and 78369 probes for the model with the 14 Hispanic-Latino samples. The iterateRUV function from the RUVnormalize package (Jacob et al., 2016) was then used to estimate latent variables under default parameters. There was no obvious relationship between the latent variables and demographic cofactors.

Models were then re-run as:

$$Y = Group + Race + Gender + Age + W$$

For all 47 samples, and:

$$Y = Group + Gender + Age + W$$

for the 30 Black American samples, and separately for the 14 Hispanic-Latino samples, where W represents the matrix of latent variables estimated for the respective model.

Adjustment of M-values for covariates, including disease modifying treatments

The RUVnormalize R package (Jacob et al., 2016) was used to estimate unknown covariates in the dataset, some of which corresponded to disease modifying treatments. The removeBatchEffects function from the limma package (Ritchie et al., 2015) was used to adjust M-values. Group, race, and intercept were listed as factors to be preserved. Slide was indicated as a batch effect. Gender, age, and latent variables were indicated as covariates to be adjusted for. Data were analyzed in the presence and absence of adjustments for disease modifying treatments as a covariate. Adjusted M-values were used to draw Mean Difference (MD) plots using the glMDPlot function in the Glimma package (Su et al., 2017). These values were also used for analysis of DMP's within DMR's.

Analysis of differentially methylated regions

The DMRcate package was used to identify differentially methylated regions (DMR's) based on the *p*-values used to detect differentially methylated probes (DMP's) (Peters et al., 2015). For each of the three models used for detecting DMP's, an FDR threshold was determined at which probes with a *p*-value of 0.001 or lower would be captured. The dmrcate function was run with lambda = 1,000 and C = 2.

KEGG pathway and cell composition analysis

The gometh function of missMethyl was used to identify enriched KEGG terms among differentially methylated genes (Phipson et al., 2016). Cell composition analysis was performed using the CellCounts2 function in the FlowSorted. Blood.EPIC package in R software (Salas et al., 2018).

Pyrosequencing

Bisulfite conversion of gDNA was performed (EpiTect Bisulfide Kit (Qiagen Cat. No. 59104), and regions of interest were amplified by PCR (Qiagen Pyromark Custom Assays, Qiagen PyroMark kit Cat. No. 978703) using a QuantStudio 3 real time PCR system. Pyrosequencing and analysis were performed at the Stanford University School of Medicine's Beckman Center for Molecular and Genetic Medicine. The regions analyzed and assay numbers are shown in Table 4. Data were analyzed in SPSS (version 28, IBM).

Data availability statement

The data presented in this study was deposited in the GEO repository, accession number GSE219293.

Ethics statement

The studies involving human participants were reviewed and approved by University of Illinois Chicago IRB. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Design and concept by MC, LC, and MB. Data acquisition by MB and JB. Computations by LC and MC. Data analysis, data interpretation, and writing by MC, LC, IM, and JB. Revisions and approval by MC, LC, MB, IM, and JB.

Funding

This work was funded by the Evelyn Chinea Garcia Multiple Sclerosis Research Fund at the University of Illinois, Chicago and by the Department of Veterans Affairs (BLR&D Merit Award BX000467). The content is solely the responsibility of the authors and does not represent the official views of the U.S. Department of Veterans Affairs.

Acknowledgments

We thank the University of Illinois Health Biorepository for technical assistance with the processing and storage of donor samples; Rajeev Ranjan of the University of Illinois, Chicago and the Stanford University School of Medicine's Beckman Center for Molecular and Genetic Medicine for assistance with pyrosequencing; and the University of Illinois, Urbana-Champaign, Carver Biotechnology Center for processing Illumina EPIC arrays and the High Performance Biological Computing Center for assistance with data analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1058817/full#supplementary-material

References

Allan, E. R. O., Campden, R. I., Ewanchuk, B. W., Tailor, P., Balce, D. R., McKenna, N. T., et al. (2017). A role for cathepsin Z in neuroinflammation provides mechanistic support for an epigenetic risk factor in multiple sclerosis. *J. Neuroinflammation* 14, 103. doi:10.1186/s12974-017-0874-x

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461. doi:10.1038/nature12787

Arumugam, K., Shin, W., Schiavone, V., Vlahos, L., Tu, X., Carnevali, D., et al. (2020). The master regulator protein BAZ2B can reprogram human hematopoietic lineage-committed progenitors into a multipotent state. *Cell Rep.* 33, 108474. doi:10.1016/j.celrep.2020.108474

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., et al. (2014). Minfi: A flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30, 1363–1369. doi:10.1093/bioinformatics/btu049

Baylin, S. B., Esteller, M., Rountree, M. R., Bachman, K. E., Schuebel, K., and Herman, J. G. (2001). Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum. Mol. Genet.* 10, 687–692. doi:10.1093/hmg/10.7.687

Bell, R. E., Golan, T., Sheinboim, D., Malcov, H., Amar, D., Salamon, A., et al. (2016). Enhancer methylation dynamics contribute to cancer plasticity and patient mortality. *Genome Res.* 26, 601–611. doi:10.1101/gr.197194.115

Bender, S., Tang, Y., Lindroth, A. M., Hovestadt, V., Jones, D. T., Kool, M., et al. (2013). Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas. *Cancer Cell* 24, 660–672. doi:10.1016/j.ccr.2013.10.006

Bird, A. P., and Wolffe, A. P. (1999). Methylation-induced repression--belts, braces, and chromatin. *Cell* 99, 451–454. doi:10.1016/s0092-8674(00)81532-9

Bjornevik, K., Cortese, M., Healy, B. C., Kuhle, J., Mina, M. J., Leng, Y., et al. (2022). Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* 375 (2022), 296–301. doi:10.1126/science.abj8222

Braun, P. R., Han, S., Hing, B., Nagahama, Y., Gaul, L. N., Heinzman, J. T., et al. (2019). Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Transl. Psychiatry* 9, 47. doi:10.1038/s41398-019-0376-y

Buja, A., and Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivar. Behav. Res.* 27, 509–540. doi:10.1207/s15327906mbr2704_2

Caldito, N. G., Saidha, S., Sotirchos, E. S., Dewey, B. E., Cowley, N. J., Glaister, J., et al. (2018). Brain and retinal atrophy in african-Americans versus caucasian-Americans with multiple sclerosis: A longitudinal study. *Brain* 141, 3115–3129. doi:10.1093/brain/awy245

Carlstrom, K. E., Ewing, E., Granqvist, M., Gyllenberg, A., Aeinehband, S., Enoksson, S. L., et al. (2019). Therapeutic efficacy of dimethyl fumarate in relapsing-remitting multiple sclerosis associates with ROS pathway in monocytes. *Nat. Commun.* 10, 3081. doi:10.1038/s41467-019-11139-3

Celarain, N., and Tomas-Roig, J. (2020). Aberrant DNA methylation profile exacerbates inflammation and neurodegeneration in multiple sclerosis patients. *J. Neuroinflammation* 17, 21. doi:10.1186/s12974-019-1667-1

Chen, Y. A., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., et al. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 8, 203–209. doi:10.4161/epi.23470

Chinea, A., Perez, N., Perez-Canabal, A., Rojas, F., Torres, J., and Poser, C. (2012). The Puerto Rico study for the prevalence of multiple sclerosis. *Bol. Asoc. Med. P. R.* 104, 4–9.

Cree, B. A., Khan, O., Bourdette, D., Goodin, D. S., Cohen, J. A., Marrie, R. A., et al. (2004). Clinical characteristics of African Americans vs Caucasian Americans with multiple sclerosis. *Neurology* 63, 2039–2045. doi:10.1212/01.wnl.0000145762.60562.5d

Davalieva, K., Kostovska, I. M., Kiprijanovska, S., Markoska, K., Kubelka-Sabit, K., Filipovski, V., et al. (2015). Proteomics analysis of malignant and benign prostate tissue by 2D DIGE/MS reveals new insights into proteins involved in prostate cancer. *Prostate* 75, 1586–1600. doi:10.1002/pros.23034

Deng, Q., Luo, Y., Chang, C., Wu, H., Ding, Y., and Xiao, R. (2019). The emerging epigenetic role of CD8+T cells in autoimmune diseases: A systematic Review. *Front. Immunol.* 10, 856. doi:10.3389/fimmu.2019.00856

Diniz, S. N., da Silva, C. F., de Almeida, I. T., da Silva Costa, F. E., and de Oliveira, E. M. L. (2021). INFβ treatment affects global DNA methylation in monocytes of patients with multiple sclerosis. *J. Neuroimmunol.* 355, 577563. doi:10.1016/j.jneuroim.2021.577563

Dor, Y., and Cedar, H. (2018). Principles of DNA methylation and their implications for biology and medicine. *Lancet* 392, 777–786. doi:10.1016/S0140-6736(18)31268-6

Ewing, E., Kular, L., Fernandes, S. J., Karathanasis, N., Lagani, V., Ruhrmann, S., et al. (2019). Combining evidence from four immune cell types identifies DNA methylation patterns that implicate functionally distinct pathways during Multiple Sclerosis progression. *EBioMedicine* 43, 411–423. doi:10.1016/j.ebiom.2019.04.042

Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K. K., Cheng, C., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature* 489, 91–100. doi:10.1038/nature11245

Howlett-Prieto, Q., Carrithers, M. D., Wunsch, D. C., and Hier, D. B. (2022). Subtypes of relapsing-remitting multiple sclerosis identified bypass network analysis. medRxiv. doi:10.1101/2022.11.16.22282420

Huynh, J. L., Garg, P., Thin, T. H., Yoo, S., Dutta, R., Trapp, B. D., et al. (2014). Epigenome-wide differences in pathology-free regions of multiple sclerosis-affected brains. *Nat. Neurosci.* 17, 121–130. doi:10.1038/nn.3588

Imsgc, I. M. S. G. C. (2010). IL12A, MPHOSPH9/CDK2AP1 and RGS1 are novel multiple sclerosis susceptibility loci. *Genes Immun.* 11, 397–405. doi:10.1038/gene.2010.28

International, C. (2019). Multiple Sclerosis Genetics, Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* 365.

Jacob, L., Gagnon-Bartsch, J. A., and Speed, T. P. (2016). Correcting gene expression data when neither the unwanted variation nor the factor of interest are observed. *Biostatistics* 17, 16–28. doi:10.1093/biostatistics/kxv026

Jones, R. C., Liberatore, M., Fernandez, J. R., and Gerber, S. I. (2006). Use of a prospective space-time scan statistic to prioritize shigellosis case investigations in an urban jurisdiction. *Public Health Rep.* 121, 133–139. doi:10.1177/003335490612100206

Jost, E., Schmid, J., Wilop, S., Schubert, C., Suzuki, H., Herman, J. G., et al. (2008). Epigenetic inactivation of secreted Frizzled-related proteins in acute myeloid leukaemia. *Br. J. Haematol.* 142, 745–753. doi:10.1111/j.1365-2141.2008.07242.x

Kandoth, C., Schultz, N., Cherniack, A. D., Akbani, R., Liu, Y., Shen, H., et al. (2013). Integrated genomic characterization of endometrial carcinoma. *Nature* 497, 67–73. doi:10.1038/nature12113

Kiselev, I. S., Kulakova, O. G., Boyko, A. N., and Favorova, O. O. (2021). DNA methylation as an epigenetic mechanism in the development of multiple sclerosis. *Acta Naturae* 13, 45–57. doi:10.32607/actanaturae.11043

Konac, E., Varol, N., Yilmaz, A., Menevse, S., and Sozen, S. (2013). DNA methyltransferase inhibitor-mediated apoptosis in the Wnt/β-catenin signal pathway in a renal cell carcinoma cell line. *Exp. Biol. Med.* 238, 1009–1016. doi:10.1177/1535370213498984

Krumbholz, M., Faber, H., Steinmeyer, F., Hoffmann, L. A., Kümpfel, T., Pellkofer, H., et al. (2008). Interferon-beta increases BAFF levels in multiple sclerosis: implications for B cell autoimmunity. *Brain* 131, 1455–1463. doi:10.1093/brain/awn077

Kulakova, O. G., Kabilov, M. R., Danilova, L. V., Popova, E. V., Baturina, O. A., Tsareva, E. Y., et al. (2016). Whole-genome DNA methylation analysis of peripheral blood mononuclear cells in multiple sclerosis patients with different disease courses. *Acta Naturae* 8, 103–110. doi:10.32607/20758251-2016-8-3-103-110

Kular, L., Liu, Y., Ruhrmann, S., Zheleznyakova, G., Marabita, F., Gomez-Cabrero, D., et al. (2018). DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. *Nat. Commun.* 9, 2397. doi:10.1038/s41467-018-04732-5

Langer-Gould, A., Brara, S. M., Beaber, B. E., and Zhang, J. L. (2013). Incidence of multiple sclerosis in multiple racial and ethnic groups. *Neurology* 80, 1734–1739. doi:10.1212/WNL.0b013e3182918cc2

Leek, J. T., and Storey, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* 3, 1724–1735. doi:10.1371/journal.pgen.0030161

Li, E., and Zhang, Y. (2014). DNA methylation in mammals. *Cold Spring Harb. Perspect. Biol.* 6, a019133. doi:10.1101/cshperspect.a019133

Li, X., Xiao, B., and Chen, X. S. (2017). DNA methylation: A new player in multiple sclerosis. *Mol. Neurobiol.* 54, 4049–4059. doi:10.1007/s12035-016-9966-3

Li, Z., and Luo, J. (2018). Research on epigenetic mechanism of SFRP2 in advanced chronic myeloid leukemia. *Biochem. Biophys. Res. Commun.* 501, 64–72. doi:10.1016/j.bbrc.2018.04.149

Lim, B., Park, J. L., Kim, H. J., Park, Y. K., Kim, J. H., Sohn, H. A., et al. (2014). Integrative genomics analysis reveals the multilevel dysregulation and oncogenic characteristics of TEAD4 in gastric cancer. *Carcinogenesis* 35, 1020–1027. doi:10.1093/carcin/bgt409

Liu, Y., Zhou, Y., Yue, H., Dou, H., Rang, X., Wang, X., et al. (2022). Identification of potential key genes and immune infiltration in Multiple sclerosis. *Mult. Scler. Relat. Disord.* 60, 103748. doi:10.1016/j.msard.2022.103748

Maltby, V. E., Lea, R. A., Graves, M. C., Sanders, K. A., Benton, M. C., Tajouri, L., et al. (2018). Genome-wide DNA methylation changes in CD19(+) B cells from relapsing-remitting multiple sclerosis patients. *Sci. Rep.* 8, 17418. doi:10.1038/s41598-018-35603-0

Maltby, V. E., Lea, R. A., Ribbons, K. A., Sanders, K. A., Kennedy, D., Min, M., et al. (2018). DNA methylation changes in CD4(+) T cells isolated from multiple sclerosis patients on dimethyl fumarate. *Mult. Scler. J. Exp. Transl. Clin.* 4, 2055217318787826. doi:10.1177/2055217318787826

Michal, J. D., and Wojtas, B. (2019). Global DNA methylation patterns in human gliomas and their interplay with other epigenetic modifications. *Int. J. Mol. Sci.* 20, 3478. doi:10.3390/ijms20143478

Min, J. L., Hemani, G., Hannon, E., Dekkers, K. F., Castillo-Fernandez, J., Luijk, R., et al. (2021). Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat. Genet.* 53, 1311–1321. doi:10.1038/s41588-021-00923-x

Morishima, N., Nakanishi, K., and Nakano, A. (2011). Activating transcription factor-6 (ATF6) mediates apoptosis with reduction of myeloid cell leukemia sequence 1 (Mcl-1) protein via induction of WW domain binding protein 1. *J. Biol. Chem.* 286, 35227–35235. doi:10.1074/jbc.M111.233502

Oh, R. Y., Deshwar, A. R., Marwaha, A., Sabha, N., Tropak, M., Hou, H., et al. (2022). Biallelic loss-of-function variants in RABGAP1 cause a novel neurodevelopmental syndrome. *Genet. Med.*

Ordonez, R., Martinez-Calle, N., Agirre, X., and Prosper, F. (2019). DNA methylation of enhancer elements in myeloid neoplasms: Think outside the promoters? *Cancers (Basel)* 11, 1424. doi:10.3390/cancers11101424

Peters, T. J., Buckley, M. J., Statham, A. L., Pidsley, R., Samaras, K., Clark, S. J., et al. (2015). De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* 8, 6. doi:10.1186/1756-8935-8-6

Phipson, B., Maksimovic, J., and Oshlack, A. (2016). missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* 32, 286–288. doi:10.1093/bioinformatics/btv560

Razia, R., Majeed, F., Amin, R., Mukhtar, S., Mehmood, K., and Baig, D. N. (2022). The analysis of dynamic gene expression patterns in peripheral blood of multiple sclerosis patients indicates possible diagnostic and prognostic biomarkers. *Mol. Immunol.* 147, 147–156. doi:10.1016/j.molimm.2022.05.002

Restrepo, N. A., Butkiewicz, M., McGrath, J. A., and Crawford, D. C. (2016). Shared genetic etiology of autoimmune diseases in patients from a biorepository linked to de-identified electronic health records. *Front. Genet.* 7, 185. doi:10.3389/fgene.2016.00185

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47. doi:10.1093/nar/gkv007

Roostaei, T., Klein, H. U., Ma, Y., Felsky, D., Kivisäkk, P., Connor, S. M., et al. (2021). Proximal and distal effects of genetic susceptibility to multiple sclerosis on the T cell epigenome. *Nat. Commun.* 12, 7078. doi:10.1038/s41467-021-27427-w

Salas, L. A., Koestler, D. C., Butler, R. A., Hansen, H. M., Wiencke, J. K., Kelsey, K. T., et al. (2018). An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol.* 19, 64. doi:10.1186/s13059-018-1448-7

Sorensen, P. S., Sellebjerg, F., Hartung, H. P., Montalban, X., Comi, G., and Tintore, M. (2020). The apparently milder course of multiple sclerosis: Changes in the diagnostic criteria, therapy and natural history. *Brain* 143, 2637–2652. doi:10.1093/brain/awaa145

Souren, N. Y., Gerdes, L. A., Lutsik, P., Gasparoni, G., Beltrán, E., Salhab, A., et al. (2019). DNA methylation signatures of monozygotic twins clinically discordant for multiple sclerosis. *Nat. Commun.* 10, 2094. doi:10.1038/s41467-019-09984-3

Steen, V. D., Oddis, C. V., Conte, C. G., Janoski, J., Casterline, G. Z., and Medsger, T. A., Jr. (1963-19821997). Incidence of systemic sclerosis in Allegheny County, Pennsylvania. A twenty-year study of hospital-diagnosed cases. *Arthritis Rheum.* 40, 441–445. doi:10.1002/art.1780400309

Steen, V., Domsic, R. T., Lucas, M., Fertig, N., and Medsger, T. A., Jr. (2012). A clinical and serologic comparison of African American and Caucasian patients with systemic sclerosis. *Arthritis Rheum.* 64, 2986–2994. doi:10.1002/art.34482

Su, S., Law, C. W., Ah-Cann, C., Asselin-Labat, M. L., Blewitt, M. E., and Ritchie, M. E. (2017). Glimma: Interactive graphics for gene expression analysis. *Bioinformatics* 33, 2050–2052. doi:10.1093/bioinformatics/btx094

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., et al. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet. Neurol.* 17, 162–173. doi:10.1016/S1474-4422(17)30470-2

Treviño, L. R., Yang, W., French, D., Hunger, S. P., Carroll, W. L., Devidas, M., et al. (2009). Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat. Genet.* 41, 1001–1005. doi:10.1038/ng.432

Upchurch, G. M., Haney, S. L., and Opavsky, R. (2016). Aberrant promoter hypomethylation in CLL: Does it matter for disease development? *Front. Oncol.* 6, 182. doi:10.3389/fonc.2016.00182

van Luijn, M. M., van Meurs, M., Stoop, M. P., Verbraak, E., Wierenga-Wolf, A. F., Melief, M. J., et al. (2016). Elevated expression of the cerebrospinal fluid disease markers chromogranin A and clusterin in astrocytes of multiple sclerosis white matter lesions. *J. Neuropathol. Exp. Neurol.* 75, 86–98. doi:10.1093/jnen/nlv004

Villicana, S., and Bell, J. T. (2021). Genetic impacts on DNA methylation: Research findings and future perspectives. *Genome Biol.* 22, 127. doi:10.1186/s13059-021-02347-6

Wallin, M. T., Culpepper, W. J., Campbell, J. D., Nelson, L. M., Langer-Gould, A., Marrie, R. A., et al. (2019). The prevalence of MS in the United States: A population-based estimate using health claims data. *Neurology* 92, e1029–e1040. doi:10.1212/WNL.0000000000007035

Wallin, M. T., Culpepper, W. J., Coffman, P., Pulaski, S., Maloni, H., Mahan, C. M., et al. (2012). The Gulf war era multiple sclerosis cohort: Age and incidence rates by race, sex and service. *Brain* 135, 1778–1785. doi:10.1093/brain/aws099

Wallin, M. T., Culpepper, W. J., Maloni, H., and Kurtzke, J. F. (2018). The Gulf war era multiple sclerosis cohort: 3. Early clinical features. *Acta Neurol. Scand.* 137, 76–84. doi:10.1111/ane.12810

Weinstock-Guttman, B., Jacobs, L. D., Brownscheidle, C. M., Baier, M., Rea, D. F., Apatoff, B. R., et al. (2003). Multiple sclerosis characteristics in african American patients in the New York state multiple sclerosis Consortium. *Mult. Scler.* 9, 293–298. doi:10.1191/1352458503ms909oa

Wilson, A. S., Power, B. E., and Molloy, P. L. (2007). DNA hypomethylation and human diseases. *Biochim. Biophys. Acta* 1775, 138–162. doi:10.1016/j.bbcan.2006.08.007

Zhu, H., Zhu, C., Mi, W., Chen, T., Zhao, H., Zuo, X., et al. (2018). Integration of genome-wide DNA methylation and transcription uncovered aberrant methylation-regulated genes and pathways in the peripheral blood mononuclear cells of systemic sclerosis. *Int. J. Rheumatol.* 2018, 7342472. doi:10.1155/2018/7342472

frontiers | Frontiers in Digital Health

TYPE Brief Research Report
PUBLISHED 11 January 2023
DOI 10.3389/fdgth.2022.1063264

OPEN ACCESS

EDITED BY
Lina F. Soualmia,
Université de Rouen, France

REVIEWED BY
Leif Simmatis,
University Health Network (UHN), Canada,
A. N. M. Bazlur Rashid,
Edith Cowan University, Australia

*CORRESPONDENCE
Daniel B. Hier
✉ hierd@mst.edu

SPECIALTY SECTION
This article was submitted to Health Informatics, a section of the journal Frontiers in Digital Health

RECEIVED 06 October 2022
ACCEPTED 22 December 2022
PUBLISHED 11 January 2023

CITATION
Howlett-Prieto Q, Oommen C, Carrithers MD, Wunsch II DC and Hier DB (2023) Subtypes of relapsing-remitting multiple sclerosis identified by network analysis.
Front. Digit. Health 4:1063264.
doi: 10.3389/fdgth.2022.1063264

Subtypes of relapsing-remitting multiple sclerosis identified by network analysis

Quentin Howlett-Prieto[1], Chelsea Oommen[1], Michael D. Carrithers[1], Donald C. Wunsch II[2] and Daniel B. Hier[1,2]*

[1]Department of Neurology and Rehabilitation, University of Illinois at Chicago, Chicago, IL, United States, [2]Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO, United States

We used network analysis to identify subtypes of relapsing-remitting multiple sclerosis subjects based on their cumulative signs and symptoms. The electronic medical records of 113 subjects with relapsing-remitting multiple sclerosis were reviewed, signs and symptoms were mapped to classes in a neuro-ontology, and classes were collapsed into sixteen superclasses by subsumption. After normalization and vectorization of the data, bipartite (subject-feature) and unipartite (subject-subject) network graphs were created using NetworkX and visualized in Gephi. Degree and weighted degree were calculated for each node. Graphs were partitioned into communities using the modularity score. Feature maps visualized differences in features by community. Network analysis of the unipartite graph yielded a higher modularity score (0.49) than the bipartite graph (0.25). The bipartite network was partitioned into five communities which were named fatigue, behavioral, hypertonia/weakness, abnormal gait/sphincter, and sensory, based on feature characteristics. The unipartite network was partitioned into five communities which were named fatigue, pain, cognitive, sensory, and gait/weakness/hypertonia based on features. Although we did not identify pure subtypes (e.g., pure motor, pure sensory, etc.) in this cohort of multiple sclerosis subjects, we demonstrated that network analysis could partition these subjects into different subtype communities. Larger datasets and additional partitioning algorithms are needed to confirm these findings and elucidate their significance. This study contributes to the literature investigating subtypes of multiple sclerosis by combining feature reduction by subsumption with network analysis.

KEYWORDS

multiple sclerosis, phenotype, network analysis, communities, modularity, subtype, feature reduction, subsumption

Introduction

Multiple sclerosis (MS) is one of several immune-mediated demyelinating diseases of the central nervous system that includes transverse myelitis, optic neuritis, neuromyelitis optical, acute disseminated encephalomyelitis, and acute hemorrhagic leukoencephalopathy (1). MS has traditionally been divided into four clinical course phenotypes that include relapsing-remitting multiple sclerosis (RRMS), primary progressive multiple sclerosis (PPMS), secondary progressive multiple sclerosis (SPMS), and relapsing progressive multiple sclerosis

(RPMS) (2). In 2013, the criteria for MS phenotypes were revised to remove RPMS (3–6). More recently, MS has been additionally classified by activity (active or inactive) or phase (relapsing or progressive) (4, 5, 7, 8). Another way of subtyping neurologic diseases is by deep phenotyping where signs and symptoms are recorded in detail and are mapped to a restricted terminology (9–13). Patients can then be subtyped according to patterns of signs and symptoms.

MS may have a variable onset with the diverse symptoms of optic neuritis, facial pain, hemifacial spasm, Lhermitte's sign, transverse myelitis, limb weakness, limb numbness, urinary retention, dysmetria, intention tremor, incoordination, dysarthria, hearing loss, color blindness, gait disturbance, and diplopia (14). MS variably involves the optic nerve (painful loss of vision), the spinal cord (sphincter dysfunction, monoparesis, hemiparesis, hypoesthesia, paresthesia), the brainstem and cerebellum (diplopia, oscillopsia, vertigo, ataxia, tremor, facial weakness), or the cerebral hemispheres (hemiparesis, hemihypoesthesia) (8). Subtypes of multiple sclerosis based on clinical presentation (signs and symptoms) are recognized (15–18) including tremor (19), ataxia (20), visual disturbances (21, 22), sensory symptoms (numbness and paresthesias) (23–26), pyramidal tract findings (weakness, hyperreflexia, spasticity, and hypertonia) (27–29), or spinal cord findings (paraparesis, sphincter dysfunction, and sensory levels) (30, 31). Other MS subjects show cognitive impairment (32, 33), dysarthria (34), dysautonomia (35), depression (36), imbalance (37), paroxysmal symtoms (38), or fatigue (39, 40).

The Kurtzke Functional System Score (FSS) (41) is useful in rating sensory, visual, sphincter, mental, pyramidal, cerebellar, and brainstem dysfunction in MS. However, there is a limited ability to categorize MS subjects based on their predominant clinical presentation. A network analysis of subjects with MS based on their signs and symptoms could assist in identifying clinically significant subtypes of MS.

This paper is organized as follows. We first review prior work on finding subtypes of multiple sclerosis based on signs and symptoms. We then describe our proposed approach to finding subtypes of multiple sclerosis based on deep phenotyping, subsumption of phenotype classes into superclasses, and network analysis. In the Methods section, we describe how deep phenotyping was performed, how the features were collapsed into superclasses, and how the networks were created and partitioned. In the Results section, we report the partitioning of the networks into five communities of MS subjects. In the Discussion section, we discuss the identified communities as possible clinical subtypes of MS. Finally, we discuss the limitations of network analysis as a method of finding MS subtypes.

Prior work

Although network analysis has not been used to identify clinical subtypes of MS, other work is relevant to this undertaking (**Table 1**). Depression and anxiety have been reported in MS in about 27% of patients, but no specific phenotype has been described (42, 43, 62). Zhang et al. (63) examined 13 common symptoms of MS in 1985 MS subjects and found that depression, pain, and walking difficulties were the strongest predictors of impaired quality of life. Cognitive impairment is frequent in MS, possibly affecting 40–80% of subjects (33, 45–50). Pure cognitive subtypes (cognitive impairment without other major neurological signs) have been described in a small minority of MS patients (47, 50). In their review of functional connectivity based on functional MRI, Tahedl et al. (64) suggested that cognitive impairment in MS was associated with disruptions of the default-mode network of the brain, whereas sensory-motor deficits were associated with disruptions of the sensory-motor networks of the brain. Although fatigue is frequent in MS, specific subtypes have not been described (51).

Although uncommon, spinal MS (leg weakness, sphincter dysfunction, sensory levels, spasticity, and hyperreflexia), as well as opticospinal MS (combining spinal MS with optic nerve involvement), are recognized forms of MS (65–67). Opticospinal MS must be differentiated from neuromyelitis optica, a similar but etiologically different disease from MS. Cree et al. (60) have suggested that spinal MS and opticospinal MS may be more common in Blacks than Whites. Nociti et al. (30) reported spinal MS in 2.3% of their cohort of subjects.

Cerebellar and brainstem phenotypes of MS have been reported (44) with prominent ataxia and cranial nerve deficits. Naismith et al. (17) compared 79 Black subjects with MS to 80 White subjects (17) and found more tremor, ataxia, and need for assistive walking devices in the Black MS subjects. They speculated that the optico-spinal, cognitive, and ataxic-spastic phenotypes are more common in Black than White subjects. In a small study, Ayache et al. (19) found tremor in 56% of their cohort of MS subjects but did not identify a specific phenotype.

Sensory symptoms are common in MS, including pain, hypesthesias hyperesthesias, band-like sensations, and paresthesias (26, 56); however, no specific sensory phenotype has been described. Optic neuritis is common in MS but generally recovers fully or partially. Gerbis et al. (61) describe 5 subjects from a cohort of 550 MS who had severe unilateral optic neuritis without recovery, and suggest that these cases may represent a subtype of MS subjects.

Functional Systems Scores (FSS) are a good candidate for identifying clinical subtypes of MS. It is widely used in MS clinical trials and is divided into seven domains (pyramidal, cerebellar, brainstem, sensory, bowel and bladder, visual, and cerebral) (68). An asymmetric distribution of scores in these domains could identify subtypes of MS. Yang et al. (69) used a combination of a convolutional neural network and a rule-based natural language algorithm to accurately predict

TABLE 1 Summarization of prior work relevant to subtyping multiple sclerosis by phenotypic feature.

Author	Domain	*N*	Cite	Findings	Limitations
Donnchadha	Behavioral	33	(42)	27% had anxiety	Small N, no correlations with other features
Koch	Behavior	1,376	(43)	27% depressed, stable over time	No correlation with EDSS
Ganesvaran	Brainstem or cerebellum	20	(44)	80% of 20 MS patients with psoriasis 80% had brainstem or cerebellar lesions	Small N
Naismith	Cerebellar and gait	166	(17)	Compared to Whites, Blacks had more gait and cerebellar deficits	Small N
Rocca	Cognitive	227	(45)	Loss of connectivity predicts memory and attention deficits	Correlation with cognitive impairment
Hancock	Cognitive	1,281	(46)	Cognitive domain impairment: 48% intact, 22% 1-domain, 24% 2-domain, 15% multi-domain cognitive deficits	Only examined cognitive impairment
De Meo	Cognitive	1,212	(33)	5 cognitive subtypes identified Only 19.5% completely normal	Only examined cognitive impairment
Staff	Cognitive	23	(47)	Mayo Clinic reported 23 MS patients with isolated cognitive impairment	Population not reported
Leavitt	Cognitive	128	(48)	43.7% cognitively impaired Memory, Processing speed, or both	Convenience sample
Cabeça	Cognitive	35	(49)	Discriminant analysis found reaction time best discriminator	Study did not identify cognitive subtypes
Zurawski	Cognitive	2,302	(50)	2.6% had pure cognitive phenotype	Only examined cognition
Beckerman	Fatigue	264	(51)	88 with low and 174 with high fatigue physical from mental fatigue correlated	Did not correlate fatigue with other features
Bove	FSS[a]	1,028	(52)	Median and range FSS provided	No classification by subtype
Kalincik	FSS[a]	14,969	(53)	Increased disability with relapse on on pyramidal, cerebellar, sphincter FSS	Not classified by phenotype
Stewart	FSS[a]	19,504	(54)	Pyramidal, cerebellar, sphincter add to disability with relapse	Not classified by phenotype
Scott	FSS[a]	1,173	(55)	On followup, most worsening on pyramidal, sensory, cerebellar, and sphincter FSS scales	Not classified by phenotype
Revil	Pain	112	(56)	40 pain free with normal sensation 44 central pain with hyposensitivity	Only examined pain
Tsantes	Relapse phenotype	199	(57)	47% of relapses recurred at initial optic, spinal, brainstem-cerebellum sites	Small N
Mowry	Relapse phenotype	195	(58)	Relapse more likely to recur at optic nerve spinal cord, or brainstem-cerebellum	Small N
Deen	Relapse phenotype	190	(59)	Relapse recurred at optic nerve brainstem-cerebellum, spinal cord	Small N
Nociti	Spinal cord	563	(30)	13/563 had spinal MS (2.3%)	Retrospective study
Sanders	Sensory	127	(26)	84% had paresthesias	Small N
Cree	Spinal cord	1490	(60)	Compared to White, Black MS subjects more corticospinal and transverse myelitis	
Ayache	Tremor	32	(19)	56% with tremor	Not population-based
Gerbis	Vision	550	(61)	5 of 550 patients had severe optic neuritis that never recovered	Only examined optic neuritis

Author is the first author, N is the number of subjects in the study, cite is the reference number. Studies are sorted by domain.
[a]Functional system score.

Kurtzke Functional System Scores (FSS) from the EHR notes of 4906 multiple sclerosis subjects. SUMMIT (Serially Unified Multicenter Multiple Sclerosis Investigation) is an international effort to create a repository of deeply phenotyped MS subjects utilizing standardized neurological examinations and the Kurtzke FSS (12). However, no subtypes based on FSS have been reported. Similarly, Dahlke et al. (70) examined the clinical course in 34,987 MS patients who had

entered into clinical trials (31,863 with relapsing-remitting MS, 1873 with secondarily progressive MS, and 986 with primary progressive MS) but did not characterize MS subjects further as to clinical phenotype. Other ongoing longitudinal studies have been undertaken to characterize MS clinical phenotypes (16, 71) but they have not yet yielded new subtypes.

The increment in neurological deficits after MS relapses has been investigated (53–55, 57–59). Increasing disability in some subjects has been linked to the accumulation of pyramidal, sensory, cerebellar, and sphincter abnormalities (53–55). Furthermore, in some subjects relapses tend to occur at the same anatomical site as previous attack, and this is especially so for the optic nerve, spinal cord, brainstem, and cerebellum sites (57–59), suggesting that neurological signs and symptoms could accumulate at those affected areas. If relapses recur at sites of the previous attacks, this could foster subtypes of MS based on repeated relapses at the same anatomic site.

A network (also called a graph) is an assembly of nodes that are interconnected by edges (52). When all connected nodes come from the same class, the graph is unipartite. When each node is connected to a node of a second class, the graph is bipartite (72). Networks can be partitioned into communities of like nodes (also called clusters) (73, 74). Barabási (75) defines a community as "a locally dense connected subgraph in a network (page 325)," and that "… we expect nodes that belong to a community to have a higher probability of linking to other members of that community than to nodes that do not belong to the same community …." Some of the partitioning algorithms depend upon the maximization of modularity which measures how well each community is separated from other communities.

Network analysis has proven useful in visualizing complex relationships between the phenotypes, genes, proteins, and metabolic pathways that underlie human diseases (76–81). Network analysis has provided important insights in into brain connectivity, and neuroimaging (82, 83). Network analysis has identified potential genetic causes of autism (84) and has clustered autism subjects by phenotype (85, 86). Network analysis has been used to identify genes that govern MS susceptibility (87–89), proteins implicated in the etiology of MS (90), as well as brain areas that undergo disconnection in MS (45, 91–93).

Proposed approach

The review of prior work suggested that there is a gap in identifying subtypes of MS based on signs and symptoms. Our goal was to identify clinical subtypes of RRMS using network analysis after feature reduction. We found 244 unique neurologic signs and symptoms in a cohort of 113 subjects with relapsing-remitting MS, mapped them to classes in a neuro-ontology, and then collapsed the classes into sixteen superclasses (**Figures 1A,B**). For each subject, the count of signs and symptoms in each superclass was normalized. A bipartite graph was created using NetworkX, with each subject node connected to one of sixteen superclass nodes by an edge proportional to the normalized count of signs and symptoms. Distances between subjects were calculated by the cosine similarity of their signs and symptoms. A unipartite graph was created in NetworkX where the nodes were subjects, and the edges were inter-subject distances. The unipartite and bipartite graphs were visualized in Gephi and partitioned into communities based on the Louvain algorithm (94). Modularity scores were used to evaluate the quality of the partitions. We used feature maps to characterize the communities. This approach could lead to classifying MS patients by clinical phenotype and supplement the phenotyping of MS subjects by disease course.

Methods

Subjects

One hundred and twenty MS subjects followed at the University of Illinois-Neuroscience Center were enrolled in the University of Illinois at Chicago (UIC) Neuroimmunology Biobank between August 2018 and March 2020 (mean age 42.7 $\pm$ 12.8 years, 73% female, 27% male, 58% Black, 42% White). The Biobank is approved by the Institutional Review Board (IRB) of the University of Illinois College of Medicine. All subjects provided informed written consent at enrollment. Subjects were between 18-80 years old and had a diagnosis of RRMS based on the 2017 McDonald criteria (95). Subjects had been recruited for a study of blood biomarkers in MS where RRMS was an inclusion criterion and progressive MS was an exclusion criterion. Seven subjects with normal neurological examinations were excluded from the analysis leaving a final study sample of 113 subjects.

Neuro-phenotyping

The neurological progress notes from the electronic health record of all subjects were reviewed, and neurological signs and symptoms were recorded (11). The cumulative signs and symptoms (both active and resolved) of each subject were recorded and mapped to concepts in a neuro-ontology with 1,600 possible concepts (96). Subjects had 13.2 $\mp$ 9.2 signs and symptoms (mean $\mp$ standard deviation). The 113 subjects had 1,453 total signs and symptoms (244 unique signs and symptoms). Subsumption (97) was used to collapse the signs and symptoms (**Figure 1A**) into 16 superclasses (**Figure 1B**) that included *behavior*, *cognitive*, *cranial nerve*,

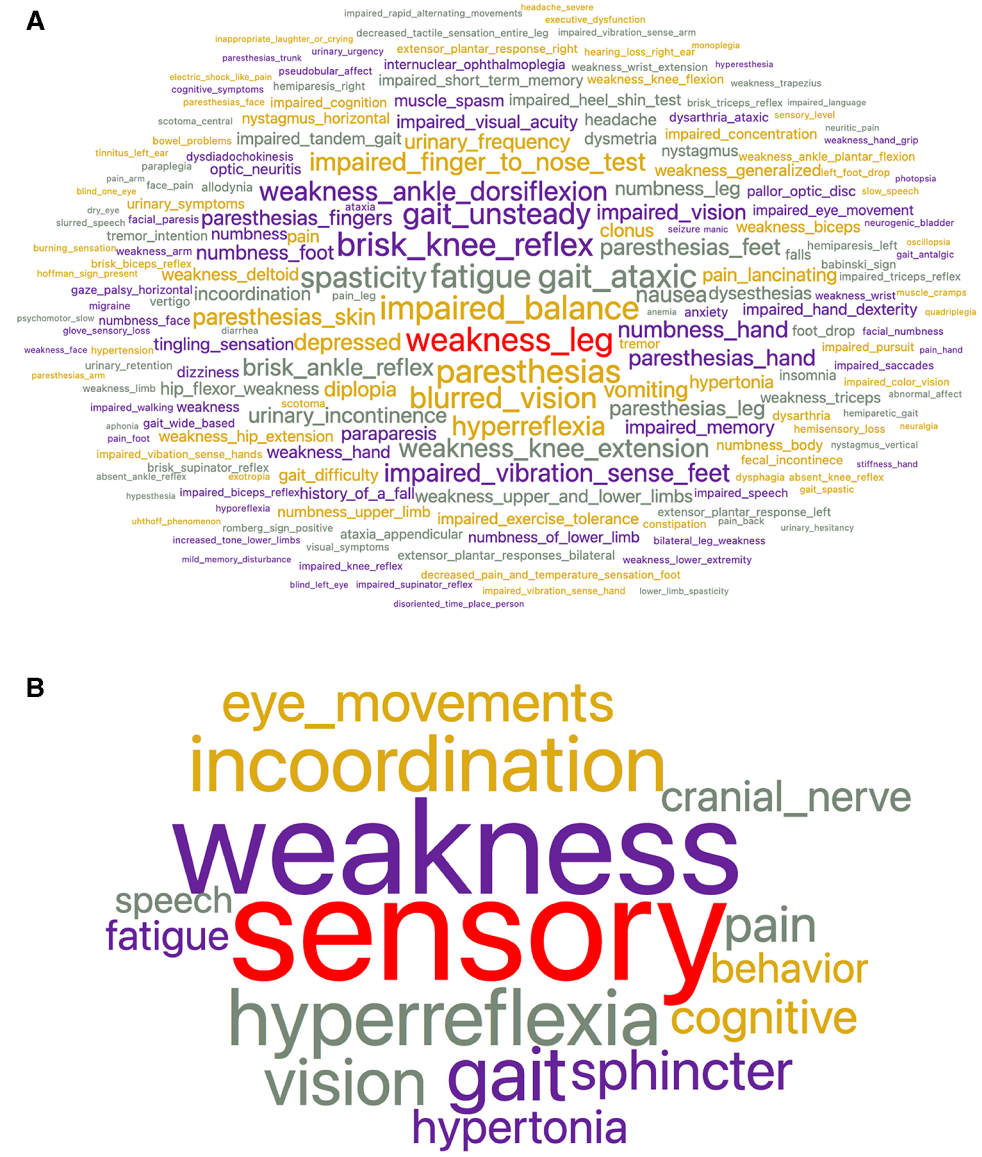


FIGURE 1
(**A**) Word cloud representing the frequency of signs and symptoms in the entire MS cohort before subsumption. Word size is proportional to frequency. There were 244 unique signs and symptoms. The most frequent signs and symptoms were leg weakness, impaired balance, fatigue, and paresthesias. Supporting files available on the project's GitHub site. (**B**) Word cloud representing the frequency of signs and symptoms in the entire MS cohort after subsumption into 16 superclasses. Word size is proportional to frequency. The largest superclasses are sensory, weakness, hyperreflexia, and incoordination. Supporting files available on the project GitHub site.

eye movement, *fatigue*, *gait*, *hyperreflexia*, *hypertonia*, *incoordination*, *pain*, *sensory*, *speech*, *sphincter*, *tremor*, *vision*, and *weakness*. The largest superclasses were weakness, sensory, incoordination, and hyperreflexia. Each subject was represented as a 17-dimension vector where the first element of the vector was the case identification label, and the subsequent sixteen elements were the count for each of the sixteen superclasses (**Figure 2A**). Counts were normalized over the interval [0, 1] using the *continuize* widget in Orange 3.32.0 (98) (**Figure 2B**). We chose to normalize counts because counts varied significantly between superclasses. For supporting data, see the project GitHub site.

Network analysis, distance metrics, feature maps

Network analyses were performed on normalized 113 × 17 data arrays (89, 98, 99). NetworkX (100) converted the data arrays to GraphML files compatible with Gephi. Bipartite

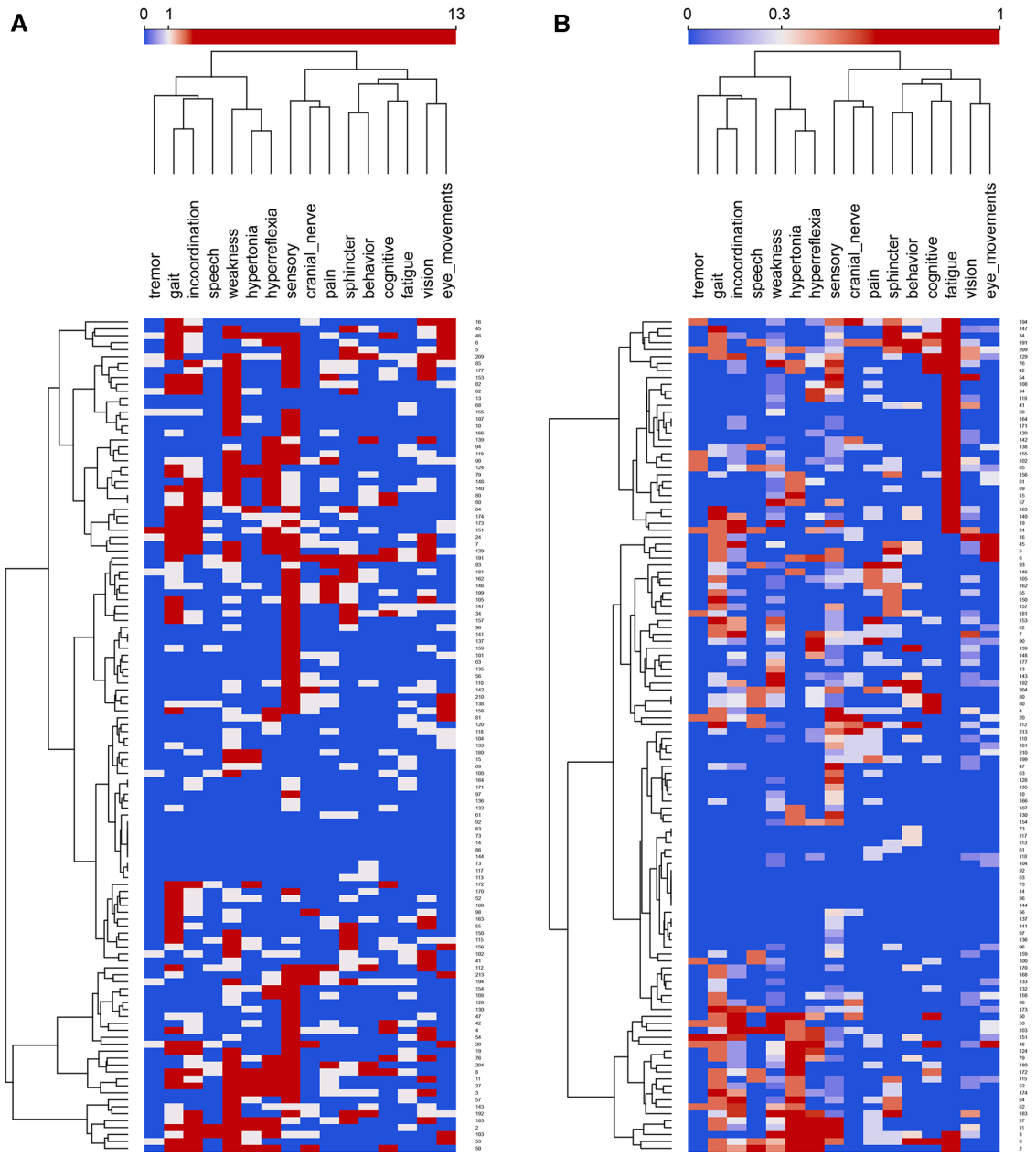


FIGURE 2
(A) Feature map of the entire cohort of MS patients before normalization. Rows are subjects, and columns are superclasses. Normalized feature counts in the columns range between 0 to 13 and the color scale is centered on 3 features. Rows and columns are clustered hierarchically with Ward linkage. Column distances by Pearson correlation coefficient; row distances are Euclidean. (B) Feature map of the entire cohort of MS patients after normalization. Rows are subjects, and columns are superclasses. Normalized feature counts in the columns range between 0 to 1 and the color scale is centered on 0.3 features. Rows and columns are clustered hierarchically with Ward linkage. Column distances by Pearson correlation coefficient; row distances are Euclidean.

networks were visualized in Gephi 0.9.7 using a variety of layouts, with the final analysis using the Force Atlas layout with a repulsion force of 10,000. Visual inspection showed Force Atlas to have the optimal spacing of nodes and clarity of visualization. The bipartite network contained nodes of subjects and features (signs and symptoms) as nodes with a magnitude of the edges connecting subjects to features equal to the normalized feature score for each subject. In the bipartite networks, there were no direct subject-subject or feature-feature edges. Node sizes were proportional to the average weighted degree of each node. Communities were named based on their predominant features. Nodes were colored by their community membership, and colors were used consistently across graphs based on feature

predominance. Edge widths were proportional to edge weight for the bipartite graphs. The unipartite networks were based on distances between subjects derived from the feature vectors for each subject. Distances were calculated in Orange using the *distances* widget for Pearson, Euclidean, and cosine distances. Visual inspection of the network graphs showed that the cosine-based graphs were superior to those based on the Pearson or Euclidean distances. Only the cosine distances were retained for further analysis (101). For the unipartite graphs, all nodes were subjects, and the edges were subject similarity based on the cosine distances. Node size was proportional to the degree (number of edges for each node). The edge width was fixed. Gephi was used to partition the unipartite and bipartite networks into communities based on the Louvain algorithm (94). The Louvain algorithm maximizes modularity (a measure of community separation). Modularity rises from 0.0 as the number of intra-community edges increases relative to inter-community edges. Larger values of modularity reflect a more robust separation of the communities. The degree, average degree, and modularity class for each node were calculated by Gephi. Modularity resolution was set to 1.0 for the unipartite graph and 1.15 for the bipartite graph. For the unipartite graph, two subjects were excluded from the final analysis as they formed communities with only one node. For the normalized unipartite graph, a cosine distance threshold of 0.4 was used to exclude weak edges. Feature means for each community were calculated by SPSS 28 (IBM, Chicago, IL). Differences between community feature means were tested by one-way ANOVA (SPSS). Feature maps were created with the *heat map* widget from Orange. The word cloud was created with the *word cloud* widget from Orange. The concordance for set membership between communities was measured by the Jaccard Index (102) where J is the Jaccard Index, and A and B are the set memberships of two communities:

$$J = \frac{A \cap B}{A \cup B}.$$

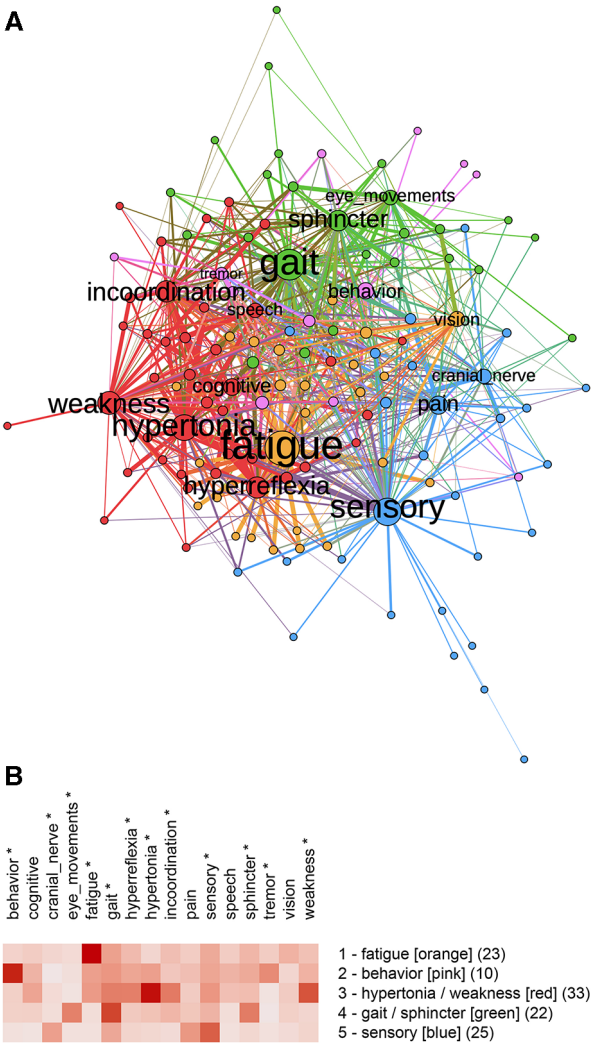


FIGURE 3
(A) Bipartite graph of normalized features. Labeled nodes are features; unlabeled nodes are subjects. Weighted connections (edges) are between features and subjects. (B) Feature map of the five communities identified by partitioning the bipartite graph with unnormalized features. Modularity analysis of the bipartite graph of normalized data showed five communities. Community 1 was predominantly fatigue, Community 2 was predominantly behavioral, Community 3 was weakness and hypertonia, Community 4 was gait and sphincter, and Community 5 was predominantly sensory features indicating features that differed significantly by the community (One-way ANOVA, $p < 0.05$, $df = 4$).

Results

The largest superclasses of signs and symptoms in this cohort of MS subjects were sensory, weakness, incoordination, and hyperreflexia (**Figure 1B**). To prevent the superclasses of weakness and sensory from dominating the network analysis, the signs and symptoms were normalized on the interval [0, 1] before network analysis and partitioning. Visual inspection of the feature map of the MS cohort suggested some clustering of subjects on signs and symptoms (**Figures 2A,B**) and that a network analysis to identify distinct communities would be fruitful.

The bipartite graph was partitioned into five communities (**Figure 3A**) with a modularity score of 0.25. Communities were named and color-coded by the one or two features with the highest community means as **fatigue** ($n = 23$), **behavior** ($n = 10$), **hypertonia/weakness** ($n = 33$), **gait/sphincter** ($n = 22$), and **sensory** ($n = 25$) (**Figure 3B**). ANOVA showed significant differences between communities for behavior ($p < .001$), cranial nerve ($p = .008$), eye movements ($p < .001$), fatigue ($p < .001$), gait ($p = .029$), hyperreflexia

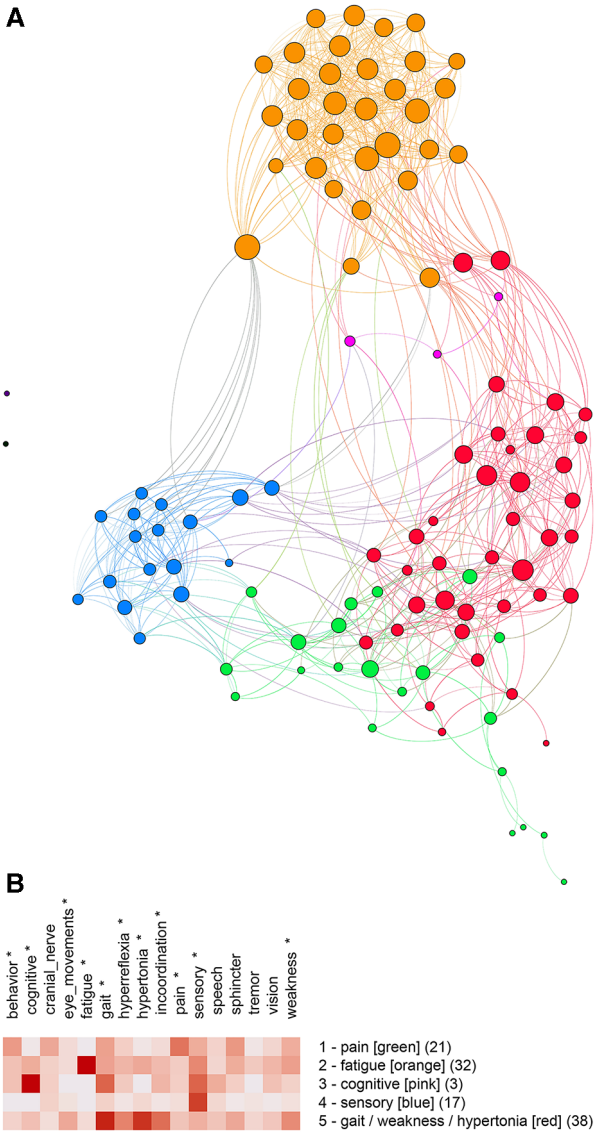


FIGURE 4
(**A**) Unipartite graph based on normalized features. Nodes are subjects, and node size is proportional to the number of edges. The largest communities are gait/weakness/hypertonia (red, $n = 38$) and fatigue (orange, $n = 32$). Note the small cognitive community (pink, $n = 3$). (**B**) Feature map of the five communities identified by partitioning the unipartite network graph based on normalized features. Asterisks indicate features that differed significantly by the community (One-way ANOVA, $p < 0.05$, $df = 4$).

($p = .031$), hypertonia ($p < .001$), incoordination ($p = .002$), sensory ($p = .018$), sphincter ($p = .021$), tremor ($p = .006$), and weakness ($p = .001$).

The unipartite graph was partitioned into five communities (Figure 4A) with a modularity score of 0.49. Communities were named by their predominant features: **pain**, **fatigue**, **cognitive**, **sensory**, and **weakness/gait/hypertonia** (Figure 4B). ANOVA analysis showed significant differences between communities for behavior ($p = .033$), cognitive ($p < .001$), eye movements ($p = .047$), fatigue ($p < .001$), gait ($p < .001$), hyperreflexia ($p = .014$), hypertonia ($p < .001$), incoordination ($p < .001$), pain ($p < .001$), sensory ($p < .001$), and weakness ($p = .037$).

Although partitioning the bipartite and unipartite graphs produced somewhat different communities, similarities between community membership and graphs were notable. We used the Jaccard Index (a set similarity measure) to assess the similarity between communities. Membership for the **fatigue** ($J = 0.72$) and **sensory** ($J = 0.56$) communities was similar for the unipartite and bipartite graphs. The unipartite graph community **gait/weakness/hypertonia** showed similarity to the bipartite graph communities **hypertonia/weakness** ($J = 0.36$) and **gait/sphincter** ($J = 0.36$). A complete table of Jaccard Index values is available on the project's GitHub site.

Discussion

Multiple sclerosis can present as sensory loss, weakness, incoordination, sphincter disturbance, diplopia, visual loss, cognitive impairment, fatigue, or even pain. We have used network analysis to identify distinct clinical subtypes of multiple sclerosis based on signs and symptoms. We first mapped the signs and symptoms of a cohort of multiple sclerosis subjects to concepts from neuro-ontology. We then created a bipartite graph, where subjects and their signs and symptoms were nodes in a graph (Figure 3A). In a bipartite graph, subjects are connected to signs and symptoms and not to other subjects. When the signs and symptoms of a subject are converted to vectors, distances between subjects can be calculated so that subject nodes can be connected to other subjects to form a unipartite graph (Figure 4A). Network analysis allowed us to identify communities of multiple sclerosis subjects who shared signs and symptoms in common. Partitioning of the unipartite and bipartite graphs based on modularity score identified communities with strong fatigue and sensory feature predominance. Both partitions had communities characterized by weakness combined with hypertonia or gait findings. Partitioning of the bipartite graph produced a small community with behavioral changes (depression, anxiety, etc.) and a gait/sphincter community. Partitioning of the unipartite graph produced a small community with cognitive findings and a medium-sized community with pain (Figure 4B).

Partitions of the unipartite graph yielded higher modularity scores than the bipartite graph, suggesting that the partitioning of the unipartite graph was more robust. The named communities for Figure 4B (**pain**, **fatigue**, **cognitive**, **sensory**, and **gait/weakness/hypertonia**) deserve special consideration as potentially identifiable multiple sclerosis subtypes. We found a strong overlap between the fatigue and sensory communities across both graphs as measured by the Jaccard Index. Significant overlap between the **gait/weakness/hypertonia** community from the unipartite graph with the **gait/sphincter** and **hypertonia/**

weakness communities from the bipartite graph was noted. Although the partitioning of networks based on features suggests that identifiable MS subtypes may exist, variability across partitions does not permit a definitive characterization of subtypes.

Although we did not correlate community features with MRI findings, the communities detected may reflect the anatomic location of MS lesions (103, 104). Of particular interest is the tendency for relapses to occur at the sites of previous MS attacks (57–59). Recurring relapses at the same anatomic site could lead to increased symptoms in certain domains (e.g., weakness, incoordination, sphincter, visual, etc.) and make subtypes of MS more discernible. On the other hand, "pure" subtypes of MS (i.e., pure motor, pure sensory, pure cognitive) are uncommon; nearly all MS patients in our cohort have signs and symptoms in multiple symptomatic domains (see, for example, **Figures 2A,B**).

Two strengths of this study should be mentioned. First, community detection was done by network analysis which offers an alternative to unsupervised machine learning algorithms based on cluster analysis (105, 106). Second, we used subsumption and the hierarchical organization of signs and symptoms in an ontology to reduce the number of features used in the analysis (97). The current study demonstrates that subsumption can successfully group signs and symptoms of MS subjects into superclasses (**Figures 1A, B**). These superclasses can be used to characterize the clinical features of communities identified by network analysis.

The current study has several limitations. The sample size was small ($N = 113$). The small sample size could cause a selection bias that influenced the communities found by network analysis. Network analysis of larger sample sizes may detect more robust communities with a different profile of predominant features. In particular, we did not identify communities of MS subjects with predominant vision, cranial nerve, or incoordination signs and symptoms, although such communities likely exist (15, 20–22). Another limitation was that we evaluated only one partitioning algorithm (Louvain). A limitation of the Louvain algorithm is that it does not exhaustively examine all possible partitions, so partitioning is non-deterministic, and partitions may change with each run (73, 107, 108). Other partitioning algorithms are available and might yield different results. We used subsumption to reduce the number of clinical features from 244 to sixteen. Different subsumption strategies would likely yield different results. We calculated distances between subjects using the cosine distance metric; other distance metrics are available and may have resulted in different results. Although the modularity scores of the partitions are comparable to those obtained on standard datasets like the Karate Club (73, 108), they are still modest (0.25–0.49). Another limitation was that subjects in the study were diagnosed with the RRMS phenotype. Without further analysis, our data cannot be extrapolated to other disease course phenotypes. Our analysis did not consider the race or sex of the subjects, which could influence clinical subtype (60, 109, 110). Finally, we partitioned MS subjects based on their accumulated signs and symptoms. Examining networks based on signs and symptoms at a single time would be instructive.

Conclusions

MS phenotypes based on the clinical course are well-established. Clinical subtypes of MS based on clinical presentation are increasingly recognized. After mapping the signs and symptoms of a cohort of MS patients to classes in neuro-ontology and then collapsing these classes into sixteen superclasses, we used network analysis to identify clinical subtypes of MS based on signs and symptoms. Feature maps (**Figures 3B**, **4B**) suggest that identifiable subtypes of MS with predominant signs and symptoms related to weakness, sensation, behavior, cognition, pain, and fatigue deserve further investigation. The clinical subtyping of MS subjects could supplement phenotyping by disease course. Additional studies may reveal that MS subtypes correlate with epigenetic, radiological, immunologic, or protein biomarkers.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by The Biobank is approved by the Institutional Review Board (IRB) of the University of Illinois College of Medicine. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Design and concept by DBH and QHP. Data acquisition by MDC, DBH, QHP, and CO. Computations by DBH and QHP. Data analysis, interpretation, and writing by DBH, QHP, MDC, CO, and DCWII. Revisions and approval by DBH, QHP, MDC, CO, and DCWII. All authors contributed to the article and approved the submitted version.

Funding

MDC acknowledges research funding from the Department of Veterans Affairs (BLR&D Merit Award BX000467) and prior support from Biogen. DCW II acknowledges research funding

support from the Mary K. Finley Endowment at the Missouri University of Science and Technology.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Kerr DA. The lumping, splitting of inflammatory CNS diseases. *Neurology*. (2006) 66(10):1466–7. doi: 10.1212/01.wnl.0000221747.37657.c6

2. Lublin FD, Reingold SC, SC for the National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. Defining the clinical course of multiple sclerosis: results of an international survey. *Neurology*. (1996) 46(4):907–11. doi: 10.1212/WNL.46.4.907

3. Marcus JF, Waubant EL. Updates on clinically isolated syndrome, diagnostic criteria for multiple sclerosis. *Neurohospitalist*. (2013) 3(2):65–80. doi: 10.1177/1941874412457183

4. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sørensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology*. (2014) 83(3):278–86. doi: 10.1212/WNL.0000000000000560

5. Lublin FD. New multiple sclerosis phenotypic classification. *Eur Neurol*. (2014) 72(1):Suppl. 1–5. doi: 10.1159/000367614

6. Engelhard J, Oleske DM, Schmitting S, Wells KE, Talapala S, Barbato LM. Multiple sclerosis by phenotype in Germany. *Mult Scler Relat Disord*. (2022) 57:103326. doi: 10.1016/j.msard.2021.103326

7. Kantarci OH. Phases and phenotypes of multiple sclerosis. *Continuum*. (2019) 25(3):636–54. doi: 10.1212/CON.0000000000000737

8. Oh J, Vidal-Jordana A, Montalban X. Multiple sclerosis: clinical aspects. *Curr Opin Neurol*. (2018) 31(6):752–9. doi: 10.1097/WCO.0000000000000622

9. Robinson PN. Deep phenotyping for precision medicine. *Hum Mutat*. (2012) 33(5):777–80. doi: 10.1002/humu.22080

10. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping, genomic data. *Nature*. (2018) 562 (7726):203–9. doi: 10.1038/s41586-018-0579-z

11. Hier DB, Yelugam R, Azizi S, Wunsch III DC. A focused review of deep phenotyping with examples from neurology. *Eur Sci J*. (2022) 9:4–19. doi: /10.19044/esj.2022.v18n4p4

12. Bove R, Chitnis T, Cree BA, Tintoré M, Naegelin Y, Uitdehaag BM, et al. SUMMIT (serially unified multicenter multiple sclerosis investigation): creating a repository of deeply phenotyped contemporary multiple sclerosis cohorts. *Mult Scler J*. (2018) 24(11):1485–98. doi: 10.1177/1352458517726657

13. Delude CM. Deep phenotyping: the details of disease. *Nature*. (2015) 527 (7576):S14–5. doi: 10.1038/527S14a

14. Poser C. Onset symptoms of multiple sclerosis. *J Neurol Neurosurg Psychiatr*. (1995) 58(2):253. doi: 10.1136/jnnp.58.2.253-a

15. Ford H. Clinical presentation, diagnosis of multiple sclerosis. *Clin Med*. (2020) 20(4):380. doi: 10.7861/clinmed.2020-0292

16. Loonstra FC, De Ruiter LR, Doesburg D, Lam KH, Van Lierop ZY, Moraal B, et al. Project Y: the search for clues explaining phenotype variability in MS. *Mult Scler Relat Disord*. (2022) 57:103337. doi: 10.1016/j.msard.2021.103337

17. Naismith RT, Trinkaus K, Cross A. Phenotype and prognosis in African-Americans with multiple sclerosis: a retrospective chart review. *Mult Scler J*. (2006) 12(6):775–81. doi: 10.1177/1352458506070923

18. Cree BA, Reich DE, Khan O, De Jager PL, Nakashima I, Takahashi T, et al. Modification of multiple sclerosis phenotypes by African ancestry at HLA. *Arch Neurol*. (2009) 66(2):226–33. doi: 10.1001/archneurol.2008.541

19. Ayache SS, Chalah MA, Al-Ani T, Farhat WH, Zouari HG, Créange A, et al. Tremor in multiple sclerosis: the intriguing role of the cerebellum. *J Neurol Sci*. (2015) 358(1-2):351–6. doi: 10.1016/j.jns.2015.09.360

20. Mills RJ, Yap L, Young CA. Treatment for ataxia in multiple sclerosis. *Cochrane Database Syst Rev*. (2007) 1. doi: 10.1002/14651858.CD005029.pub2

21. Jacobs DA, Galetta SL. Multiple sclerosis, the visual system. *Ophthalmol Clin North Am*. (2004) 17(3):265–73. doi: 10.1016/j.ohc.2004.05.011

22. Costello F. Vision disturbances in multiple sclerosis. *Semin Neurol*. (2016) 36 (02):185–95. doi: 10.1055/s-0036-1579692

23. Rae-Grant AD, Eckert NJ, Bartz S, Reed JF. Sensory symptoms of multiple sclerosis: a hidden reservoir of morbidity. *Mult Scler J*. (1999) 5(3):179–83. doi: 10.1177/135245859900500307

24. O'Connor AB, Schwid SR, Herrmann DN, Markman JD, Dworkin RH. Pain associated with multiple sclerosis: systematic review, proposed classification. *Pain*. (2008) 137(1):96–111. doi: 10.1016/j.pain.2007.08.024

25. Kratz AL, Whibley D, Alschuler KN, Ehde DM, Williams DA, Clauw DJ, et al. Characterizing chronic pain phenotypes in multiple sclerosis: a nationwide survey study. *Pain*. (2021) 162(5):1426. doi: 10.1097/j.pain.0000000000002136

26. Sanders E, Arts R. Paraesthesiae in multiple sclerosis. *J Neurol Sci*. (1986) 74 (2-3):297–305. doi: 10.1016/0022-510X(86)90115-2

27. Rizzo M, Hadjimichael O, Preiningerova J, Vollmer T. Prevalence and treatment of spasticity reported by multiple sclerosis patients. *Mult Scler J*. (2004) 10(5):589–95. doi: 10.1191/1352458504ms1085oa

28. Hoang PD, Gandevia SC, Herbert RD. Prevalence of joint contractures and muscle weakness in people with multiple sclerosis. *Disabil Rehabil*. (2014) 36 (19):1588–93. doi: 10.3109/09638288.2013.854841

29. Cordani C, Hidalgo de la Cruz M, Meani A, Valsasina P, Esposito F, Pagani E, et al. MRI correlates of clinical disability and hand-motor performance in multiple sclerosis phenotypes. *Mult Scler J*. (2021) 27(8):1205–21. doi: 10.1177/1352458520958356

30. Nociti V, Cianfoni A, Mirabella M, Caggiula M, Frisullo G, Patanella AK, et al. Clinical characteristics, course and prognosis of spinal multiple sclerosis. *Spinal Cord*. (2005) 43(12):731–4. doi: 10.1038/sj.sc.3101798

31. Wiesel PH, Norton C, Glickman S, Kamm MA. Pathophysiology and management of bowel dysfunction in multiple sclerosis. *Eur J Gastroenterol Hepatol*. (2001) 13(4):441–8. doi: 10.1097/00042737-200104000-00025

32. Amato MP, Zipoli V, Portaccio E. Cognitive changes in multiple sclerosis. *Expert Rev Neurother*. (2008) 8(10):1585–96. doi: 10.1586/14737175.8.10.1585

33. De Meo E, Portaccio E, Giorgio A, Ruano L, Goretti B, Niccolai C, et al. Identifying the distinct cognitive phenotypes in multiple sclerosis. *JAMA Neurol*. (2021) 78(4):414–25. doi: 10.1001/jamaneurol.2020.4920

34. Noffs G, Perera T, Kolbe SC, Shanahan CJ, Boonstra FM, Evans A, et al. What speech can tell us: a systematic review of dysarthria characteristics in Multiple Sclerosis. *Autoimmun Rev*. (2018) 17(12):1202–9. doi: 10.1016/j.autrev.2018.06.010

35. Miglis MG, Muppidi S. Autonomic dysfunction in multiple sclerosis and other updates on recent autonomic research. *Clin Auton Res*. (2018) 28 (4):391–3. doi: 10.1007/s10286-018-0548-5

36. Sá MJ. Psychological aspects of multiple sclerosis. *Clin Neurol Neurosurg*. (2008) 110(9):868–77. doi: 10.1016/j.clineuro.2007.10.001

37. Soyuer F, Mirza M, Erkorkmaz Ü. Balance performance in three forms of multiple sclerosis. *Neurol Res*. (2006) 28(5):555–62. doi: 10.1179/016164105X49373

38. Pop R, Kipfer S. Paroxysmal kinesigenic dyskinesia–like phenotype in multiple sclerosis. *Mult Scler J*. (2017) 23(13):1795–7. doi: 10.1177/1352458517702535

39. Induruwa I, Constantinescu CS, Gran B. Fatigue in multiple sclerosis—a brief review. *J Neurol Sci*. (2012) 323(1–2):9–15. doi: 10.1016/j.jns.2012.08.007

40. Kos D, Kerckhofs E, Nagels G, D'hooghe M, Ilsbroukx S. Origin of fatigue in multiple sclerosis: review of the literature. *Neurorehabil Neural Repair*. (2008) 22 (1):91–100. doi: 10.1177/1545968306298934

41. Amato MP, Fratiglioni L, Groppi C, Siracusa G, Amaducci L. Interrater reliability in assessing functional systems and disability on the Kurtzke scale in multiple sclerosis. *Arch Neurol.* (1988) 45(7):746–8. doi: 10.1001/archneur.1988.00520310052017

42. Ó Donnchadha S, Burke T, Bramham J, O'Brien MC, Whelan R, Reilly R, et al. Symptom overlap in anxiety and multiple sclerosis. *Mult Scler J.* (2013) 19 (10):1349–54. doi: 10.1177/1352458513476742

43. Koch MW, Patten S, Berzins S, Zhornitsky S, Greenfield J, Wall W, et al. Depression in multiple sclerosis: a long-term longitudinal study. *Mult Scler J.* (2015) 21(1):76–82. doi: 10.1177/1352458514536086

44. Ganesvaran G, Greer J, Pender M. Prominent brainstem and cerebellar involvement in multiple sclerosis with psoriasis. *Mult Scler J.* (2009) 15 (6):763–6. doi: 10.1177/1352458509103612

45. Rocca MA, Valsasina P, Meani A, Falini A, Comi G, Filippi M. Impaired functional integration in multiple sclerosis: a graph theory study. *Brain Struct Funct.* (2016) 221(1):115–31. doi: 10.1007/s00429-014-0896-4

46. Hancock LM, Galioto R, Samsonov A, Busch RM, Hermann B, Matias-Guiu JA. A proposed new taxonomy of cognitive phenotypes in multiple sclerosis: the international classification of cognitive disorders in MS (IC-CoDiMS). *Mult Scler J.* (2022):13524585221127941.

47. Staff NP, Lucchinetti CF, Keegan BM. Multiple sclerosis with predominant, severe cognitive impairment. *Arch Neurol.* (2009) 66(9):1139–43. doi: 10.1001/archneurol.2009.190

48. Leavitt VM, Tosto G, Riley CS. Cognitive phenotypes in multiple sclerosis. *J Neurol.* (2018) 265(3):562–6. doi: 10.1007/s00415-018-8747-5

49. Cabeça HLS, Rocha LC, Sabbá AF, Tomás AM, Bento-Torres NVO, Anthony DC, et al. The subtleties of cognitive decline in multiple sclerosis: an exploratory study using hierarchichal cluster analysis of CANTAB results. *BMC Neurol.* (2018) 18(1):140. doi: 10.1186/s12883-018-1141-1

50. Zurawski J, Healy B, Ratajska A, Barker L, Glanz B, Houtchens M. Identification of a predominant cognitive phenotype in patients with multiple sclerosis. *Eur J Neurol.* (2020) 27(6):1083–8. doi: 10.1111/ene.14186

51. Beckerman H, Eijssen IC, van Meeteren J, Verhulsdonck MC, de Groot V. Fatigue profiles in patients with multiple sclerosis are based on severity of fatigue and not on dimensions of fatigue. *Sci Rep.* (2020) 10(1):1–10. doi: 10.1038/s41598-020-61076-1

52. Borsboom D, Deserno MK, Rhemtulla M, Epskamp S, Fried EI, McNally RJ, et al. Network analysis of multivariate data in psychological science. *Nat Rev Dis Primers.* (2021) 1(1):1–18.

53. Kalincik T, Buzzard K, Jokubaitis V, Trojano M, Duquette P, Izquierdo G, et al. Risk of relapse phenotype recurrence in multiple sclerosis. *Mult Scler J.* (2014) 20(11):1511–22. doi: 10.1177/1352458514528762

54. Stewart T, Spelman T, Havrdova E, Horakova D, Trojano M, Izquierdo G, et al. Contribution of different relapse phenotypes to disability in multiple sclerosis. *Mult Scler J.* (2017) 23(2):266–76. doi: 10.1177/1352458516643392

55. Scott T, Wang P, You X, Mann M, Sperling B. Relationship between sustained disability progression and functional system scores in relapsing-remitting multiple sclerosis: analysis of placebo data from four randomized clinical trials. *Neuroepidemiology.* (2015) 44(1):16–23. doi: 10.1159/000369621

56. Rivel M, Achiron A, Dolev M, Stern Y, Zeilig G, Defrin R. Unique features of central neuropathic pain in multiple sclerosis: results of a cluster analysis. *Eur J Pain.* (2022) 26(5):1107–22. doi: 10.1002/ejp.1934

57. Tsantes E, Leone MA, Curti E, Cantello R, Vecchio D, Granella F. Location of first attack predicts the site of subsequent relapses in multiple sclerosis. *J Clin Neurosci.* (2020) 74:175–9. doi: 10.1016/j.jocn.2020.02.017

58. Mowry EM, Deen S, Malikova I, Pelletier J, Bacchetti P, Waubant E. The onset location of multiple sclerosis predicts the location of subsequent relapses. *J Neurol Neurosurg Psychiatry.* (2009) 80(4):400–3. doi: 10.1136/jnnp.2008.157305

59. Deen S, Bacchetti P, High A, Waubant E. Predictors of the location of multiple sclerosis relapse. *J Neurol Neurosurg Psychiatry.* (2008) 79(10):1190–3. doi: 10.1136/jnnp.2007.136440

60. Cree B, Khan O, Bourdette D, Goodin D, Cohen J, Marrie R, et al. Clinical characteristics of African Americans vs Caucasian Americans with multiple sclerosis. *Neurology.* (2004) 63(11):2039–45. doi: 10.1212/01.WNL.0000145762.60562.5D

61. Gerbis N, Parratt J. Severe unilateral optic neuritis in multiple sclerosis. *J Neurol Neurosurg Psychiatry.* (2018) 89:A41. doi: 10.1136/jnnp-2018-ANZAN.103

62. Diaz-Olavarrieta C, Cummings JL, Velazquez J, Garcia de al Cadena C. Neuropsychiatric manifestations of multiple sclerosis. *J Neuropsychiatry Clin Neurosci.* (1999) 11(1):51–7. doi: 10.1176/jnp.11.1.51

63. Zhang Y, Taylor BV, Simpson Jr S, Blizzard L, Campbell JA, Palmer AJ, et al. Feelings of depression, pain, and walking difficulties have the largest impact on the quality of life of people with multiple sclerosis, irrespective of clinical phenotype. *Mult Scler J.* (2021) 27(8):1262–75. doi: 10.1177/1352458520958369

64. Tahedl M, Levine SM, Greenlee MW, Weissert R, Schwarzbach JV. Functional connectivity in multiple sclerosis: recent findings and future directions. *Front Neurol.* (2018) 9:828. doi: 10.3389/fneur.2018.00828

65. Schee JP, Viswanathan S. Pure spinal multiple sclerosis: a possible novel entity within the multiple sclerosis disease spectrum. *Mult Scler J.* (2019) 25 (8):1189–95. doi: 10.1177/1352458518775912

66. Kira Ji. Neuromyelitis optica and opticospinal multiple sclerosis: mechanisms and pathogenesis. *Pathophysiology.* (2011) 18(1):69–79. doi: 10.1016/j.pathophys.2010.04.008

67. Takeuchi W, Fujimori J, Nakashima I. Multiple sclerosis limited to spinal cord lesions. *Clin Exp Neuroimmunol.* (2021) 12(2):111–5. doi: 10.1111/cen3.12635

68. Noseworthy J, Vandervoort M, Wong C, Ebers G. Interrater variability with the Expanded Disability Status Scale (EDSS) and Functional Systems (FS) in a multiple sclerosis clinical trial. *Neurology.* (1990) 40(6):971. doi: 10.1212/WNL.40.6.971

69. Yang Z, Pou-Prom C, Jones A, Banning M, Dai D, Mamdani M, et al. Assessment of natural language processing methods for ascertaining the expanded disability status scale score from the electronic health records of patients with multiple sclerosis: algorithm development and validation study. *JMIR Med Inform.* (2022) 10(1):e25157. doi: 10.2196/25157

70. Dahlke F, Arnold DL, Aarden P, Ganjgahi H, Häring DA, Čuklina J, et al. Characterisation of MS phenotypes across the age span using a novel data set integrating 34 clinical trials (NO, MS cohort): age is a key contributor to presentation. *Mult Scler J.* (2021) 27(13):2062–76. doi: 10.1177/1352458520988637

71. Bergamaschi R, Beghi E, Bosetti C, Ponzio M, Santucci C, Lepore V, et al. Do patients' and referral centers' characteristics influence multiple sclerosis phenotypes? Results from the Italian multiple sclerosis and related disorders register. *Neurol Sci.* (2022):1–11. doi: 10.1007/s10072-022-06169-7

72. Pavlopoulos GA, Kontou PI, Pavlopoulou A, Bouyioukos C, Markou E, Bagos PG. Bipartite graphs in systems biology and medicine: a survey of methods and applications. *GigaScience.* (2018) 7(4):giy014. doi: 10.1093/gigascience/giy014

73. Newman ME. Modularity and community structure in networks. *Proc Natl Acad Sci.* (2006) 103(23):8577–82. doi: 10.1073/pnas.0601602103

74. Kramer J, Boone L, Clifford T, Bruce J, Matta J. Analysis of medical data using community detection on inferred networks. *IEEE J Biomed Health Inform.* (2020) 24(11):3136–43. doi: 10.1109/JBHI.2020.3003827

75. Barabási AL. *Network science.* Cambridge, UK: Cambridge University Press (2016).

76. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet.* (2011) 12(1):56–68. doi: 10.1038/nrg2918

77. Carter H, Hofree M, Ideker T. Genotype to phenotype via network analysis. *Curr Opin Genet Dev.* (2013) 23(6):611–21. doi: 10.1016/j.gde.2013.10.003

78. Bertolero MA, Yeo BT, D'Esposito M. The diverse club. *Nat Commun.* (2017) 8(1):1–11. doi: 10.1038/s41467-017-01189-w

79. Chen Y, Zhang X, Zhang GQ, Xu R. Comparative analysis of a novel disease phenotype network based on clinical manifestations. *J Biomed Inform.* (2015) 53:113–20. doi: 10.1016/j.jbi.2014.09.007

80. Gosak M, Markovič R, Dolenšek J, Rupnik MS, Marhl M, Stožer A, et al. Network science of biological systems at different scales: a review. *Phys Life Rev.* (2018) 24:118–35. doi: 10.1016/j.plrev.2017.11.003

81. Ren X, Wang S, Huang T. Decipher the connections between proteins and phenotypes. *Biochim Biophys Acta Proteins Proteom.* (2020) 1868(11):140503. doi: 10.1016/j.bbapap.2020.140503

82. Betzel RF, Bertolero MA, Gordon EM, Gratton C, Dosenbach NU, Bassett DS. The community structure of functional brain networks exhibits scale-specific patterns of inter-and intra-subject variability. *Neuroimage.* (2019) 202:115990. doi: 10.1016/j.neuroimage.2019.07.003

83. Bassett DS, Zurn P, Gold JI. On the nature and use of models in network neuroscience. *Nat Rev Neurosci.* (2018) 19(9):566–78. doi: 10.1038/s41583-018-0038-8

84. Emberti Gialloreti L, Enea R, Di Micco V, Di Giovanni D, Curatolo P. Clustering analysis supports the detection of biological processes related to autism spectrum disorder. *Genes.* (2020) 11(12):1476. doi: 10.3390/genes11121476

85. Matta J, Zhao J, Ercal G, Obafemi-Ajayi T. Applications of node-based resilience graph theoretic framework to clustering autism spectrum disorders phenotypes. *Appl Netw Sci.* (2018) 3(1):1–22. doi: 10.1007/s41109-018-0093-0

86. Matta J, Dobrino D, Yeboah D, Howard S, Yasser EM, Obafemi-Ajayi T. Connecting phenotype to genotype: PheWAS-inspired analysis of autism spectrum disorder. *Front Hum Neurosci*. (2022) 16:1–16. doi: 10.3389/fnhum.2022.960991

87. Baranzini SE, Khankhanian P, Patsopoulos NA, Li M, Stankovich J, Cotsapas C, et al. Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls. *Am J Hum Genet*. (2013) 92(6):854–65. doi: 10.1016/j.ajhg.2013.04.019

88. Slim L, Chatelain C, Foucauld Hd, Azencott CA. A systematic analysis of gene–gene interaction in multiple sclerosis. *BMC Med Genomics*. (2022) 15 (1):1–14. doi: 10.1186/s12920-022-01247-3

89. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, function using NetworkX. In: Varoquaux G, Vaught T, Millman J, editors. *Proceedings of the 7th Python in Science*; Pasadena, CA, USA. Los Alamos, NM: Los Alamos National Laboratory (LANL); 2008.

90. Cervantes-Gracia K, Husi H. Integrative analysis of multiple sclerosis using a systems biology approach. *Sci Rep*. (2018) 8(1):1–14. doi: 10.1038/s41598-018-24032-8

91. Schiavi S, Azzari A, Mensi A, Graziano N, Daducci A, Bicego M, et al. Classification of multiple sclerosis patients based on structural disconnection: a robust feature selection approach. *J Neuroimaging*. (2022) 32:647–55. doi: 10.1111/jon.12991

92. Schoonheim MM, Meijer KA, Geurts JJ. Network collapse, cognitive impairment in multiple sclerosis. *Front Neurol*. (2015) 6:82. doi: 10.3389/fneur.2015.00082

93. Fleischer V, Radetz A, Ciolac D, Muthuraman M, Gonzalez-Escamilla G, Zipp F, et al. Graph theoretical framework of brain networks in multiple sclerosis: a review of concepts. *Neuroscience*. (2019) 403:35–53. doi: 10.1016/j.neuroscience.2017.10.033

94. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech: Theory Exp*. (2008) 2008(10): P10008. doi: 10.1088/1742-5468/2008/10/P10008

95. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol*. (2018) 17(2):162–73. doi: 10.1016/S1474-4422(17)30470-2

96. Hier DB, Brint SU. A Neuro-ontology for the neurological examination. *BMC Med Inform Decis Mak*. (2020) 20(1):1–9. doi: 10.1186/s12911-020-1066-7

97. Wunsch III DC, Hier DB. Subsumption is a novel feature reduction strategy for high dimensionality datasets. *Eur Sci J*. 2022;18:20–36. Accessed August 12, 2022. doi: 10.190444/esj.2022.v18n4p20

98. Demsar J, Zupan B. Orange: data mining fruitful and fun-A historical perspective. *Informatica*. (2013) 37(1):55.

99. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 3. Burnaby, BC Canada: Public Knowledge Project (2009). p. 361–362. doi: 10.1609/icwsm.v3i1.13937

100. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using networkX. In: Varoquaux G, Vaught T, Millman J, editors. *Proceedings of the 7th Python in Science Conference*; Pasadena, CA, USA; scipy.org 2008. p. 11–15.

101. Hier DB, Kopel J, Brint SU, Wunsch DC, Olbricht GR, Azizi S, et al. Evaluation of standard and semantically-augmented distance metrics for neurology patients. *BMC Med Inform Decis Mak*. (2020) 20(1):1–15. doi: 10.1186/s12911-020-01217-8

102. Jaccard P. Distribution de la flore alpine dans le Bassin des Drouces et dans quelques régions voisines. *Bull Soc Vaud Sci Nat*. (1901) 37(140):241–72.

103. Huber SJ, Paulson GW, Chakeres D, Pakalnis A, Brogan M, Phillips BL, et al. Magnetic resonance imaging and clinical correlations in multiple sclerosis. *J Neurol Sci*. (1988) 86(1):1–12. doi: 10.1016/0022-510X(88)90002-0

104. Stevens JC, Farlow MR, Edwards MK, Yu Pl. Magnetic resonance imaging: clinical correlation in 64 patients with multiple sclerosis. *Arch Neurol*. (1986) 43 (11):1145–8. doi: 10.1001/archneur.1986.00520110039011

105. Xu R, Wunsch D. Survey of clustering algorithms. *IEEE Trans Neural Netw*. (2005) 16(3):645–78. doi: 10.1109/TNN.2005.845141

106. Xu R, Wunsch DC. Clustering algorithms in biomedical research: a review. *IEEE Rev Biomed Eng*. (2010) 3:120–54. doi: 10.1109/RBME.2010.2083647

107. Fortunato S, Hric D. Community detection in networks: a user guide. *Phys Rep*. (2016) 659:1–44. doi: 10.1016/j.physrep.2016.09.002

108. Motschnig N, Ramharter A, Schweiger O, Zabka P, Foerster KT. On comparing and enhancing common approaches to network community detection [Preprint] (2021). Available at: https://doi.org/10.48550/arXiv.2108.13482.

109. Weinstock-Guttman B, Jacobs L, Brownscheidle C, Baier M, Rea D, Apatoff B, et al. Multiple sclerosis characteristics in African American patients in the New York State Multiple Sclerosis Consortium. *Mult Scler J*. (2003) 9(3):293–8. doi: 10.1191/1352458503ms909oa

110. Caldito NG, Saidha S, Sotirchos ES, Dewey BE, Cowley NJ, Glaister J, et al. Brain and retinal atrophy in African-Americans versus Caucasian-Americans with multiple sclerosis: a longitudinal study. *Brain*. (2018) 141(11):3115–29. doi: 10.1093/brain/awy245

frontiers | Frontiers in Digital Health

TYPE Brief Research Report
PUBLISHED 27 January 2023
DOI 10.3389/fdgth.2023.1064936

OPEN ACCESS

EDITED BY
Elena Casiraghi,
Università degli Studi di Milano, Italy

REVIEWED BY
Manisha Mantri,
Center for Development of Advanced Computing (C-DAC), India,
Hui Zheng,
Zhejiang Normal University, China

*CORRESPONDENCE
Daniel B. Hier
✉ hierd@mst.edu

SPECIALTY SECTION
This article was submitted to Health Informatics, a section of the journal Frontiers in Digital Health

RECEIVED 08 October 2022
ACCEPTED 10 January 2023
PUBLISHED 27 January 2023

CITATION
Hier DB, Yelugam R, Carrithers MD and Wunsch III DC (2023) The visualization of Orphadata neurology phenotypes. Front. Digit. Health 5:1064936. doi: 10.3389/fdgth.2023.1064936

The visualization of Orphadata neurology phenotypes

Daniel B. Hier[1,2]*, Raghu Yelugam[1], Michael D. Carrithers[2] and Donald C. Wunsch III[3]

[1]Applied Computational Intelligence Laboratory, Department of Electrical & Computer Engineering, Missouri University of Science & Technology, Rolla, MO, United States, [2]Department of Neurology and Rehabilitation, University of Illinois at Chicago, Chicago, IL, United States, [3]National Institute of Diabetes and Digestive and Kidney Diseases, Liver Diseases Branch, Bethesda, MD, United States

Disease phenotypes are characterized by signs (what a physician observes during the examination of a patient) and symptoms (the complaints of a patient to a physician). Large repositories of disease phenotypes are accessible through the Online Mendelian Inheritance of Man, Human Phenotype Ontology, and Orphadata initiatives. Many of the diseases in these datasets are neurologic. For each repository, the phenotype of neurologic disease is represented as a list of concepts of variable length where the concepts are selected from a restricted ontology. Visualizations of these concept lists are not provided. We address this limitation by using subsumption to reduce the number of descriptive features from 2,946 classes into thirty superclasses. Phenotype feature lists of variable lengths were converted into fixed-length vectors. Phenotype vectors were aggregated into matrices and visualized as heat maps that allowed side-by-side disease comparisons. Individual diseases (representing a row in the matrix) were visualized as word clouds. We illustrate the utility of this approach by visualizing the neuro-phenotypes of 32 dystonic diseases from Orphadata. Subsumption can collapse phenotype features into superclasses, phenotype lists can be vectorized, and phenotypes vectors can be visualized as heat maps and word clouds.

KEYWORDS
neurology, phenotyping, subsumption, ontology, visualization, heat maps, feature reduction

Introduction

The signs and symptoms of a disease characterize its phenotype. In addition to signs (what a physician observes in a patient) and symptoms (the complaints of a patient), a clinical phenotype can include the age at the onset of a disease, its mode of onset, its rate of progression, its mode of inheritance, and its response to treatment. Some researchers include biochemical, radiological, electrophysiological, and biosensor findings as part of the disease phenotype (1–5). Large phenotype repositories are available on the internet. The On-Line Mendelian Inheritance in Man (OMIM) has over 9,500 disease profiles (6) and Orphadata has phenotype profiles of 4,245 rare diseases (7). The Human Phenotype Ontology (HPO) draws phenotype profiles from Orphadata and OMIM so that some genetic diseases have alternative profiles from each registry (8,9). All three repositories have sophisticated search engines that retrieve phenotype features by disease or gene (1). Phenotypic features are recorded as concepts (terms) from restricted vocabularies such as the Human Phenotype Ontology (20,246 terms) (10), or the Online Mendelian Inheritance of Man ontology (99,165 terms) (11).

Neuro-phenotypes

The June 2022 release of Orphadata lists 7,261 rare diseases, with 1,740 classified as rare neurological diseases (https://www.orphadata.com/linearisation/). Orphadata provides phenotype profiles on 1,184 rare neurologic diseases (https://www.orphadata.com/phenotypes/). Neuro-phenotyping is the deep phenotyping of neurological disease (1). We have suggested that most neuro-phenotyping can be done with a restricted vocabulary of about 1,600 concepts (12). Although lists of phenotypic features for neurological diagnoses can be retrieved from Orphadata, OMIM, or HPO, these lists are difficult to visualize.

Visualizations of disease phenotypes have limitations

OMIM, Orphanet, and HPO yield lists of phenotype features of variable length, sorted by alphabetical order, feature frequency, or body system. For example, the Orphadata annotations for Dystonia Type 13 (DYT13) are:

Very frequent
- stereotypy
- torsion dystonia
- torticollis

Frequent
- limb dystonia
- dystonia
- craniofacial dystonia
- jerky head movements

Occasional
- postural tremor
- action tremor
- focal dystonia

Rare
- Generalized dystonia
- Hoarse voice

Although useful, these lists have limitations. The lists may be long. In the Orphanet dataset, 25% of the lists are more than 34 features in length. Many of these lengths are beyond the length of 7 ± 2 that is easily comprehended (13). Side-by-side comparisons of lists are difficult (**Table 1**). Lists of signs and symptoms from Orphadata may contain pathologies (e.g., gliosis, Lewy bodies), radiological findings (e.g., abnormal PET FDG), biochemical findings, electrophysiological findings, and modes of inheritance. Although terms in Orphadata are from the HPO-controlled vocabulary (20,246 classes) (10), redundancies, near-synonyms, hypernyms, and hyponyms populate the lists (e.g., dysarthria and slow slurred speech; bradykinesia and hypokinesia; masked facies and hypomimia, etc.) Furthermore, OMIM, Orphadata, and HPO do not provide native methods for visualization of phenotype.

Prior work

Limited work has been done on visualizing phenotype lists retrieved from HPO, OMIM, or Orphadata. Xu et al. (14) visualized the distances between genetic diseases and their underlying phenotypes using t-SNE (stochastic neighborhood embedding) maps. The phenotype features from the OMIM dataset were used to calculate distances between genetic diseases. The t-SNE maps are a 2-dimensional representation of the distances between genetic diseases derived from multi-dimensional data. Although these t-SNE maps provide instructive information about the distances between genetic diseases, they do not reveal the details of the underlying phenotypes. Network analysis and network graphs have been used to visualize the distances between diseases based on their phenotype (15–17). However, these network diagrams do not elucidate the underlying phenotypic differences between the diseases. Several methods have been proposed to visualize disease-phenotype relationships, including radar graphs (18), co-occurrence charts (19), and sunburst diagrams (20). Cao et al. have developed visualization techniques called DICON, FacetAtlas, and SolarMap that show promise for visualizing phenotype features by disease (21–24).

An additional barrier to visualizing neurology phenotype profiles is the large number of terms in the HPO ($N = 20,390$), making the number of columns in heat maps or tables impractical. A feature reduction strategy that chunks phenotype features into a more manageable number of superclasses is needed. For example, Hier and Pearson (25) have suggested chunking problems in the electronic health record by body system to increase the readability of the problem list. Both OMIM and HPO chunk phenotype features by body system. Orphanet chunks phenotype features by feature frequency (common to rare). Yauy et al. (26) have chunked 16,600 phenotypic traits into 390 interacting symptom groups. However, the chunking of phenotype features by body system is unlikely to yield useful visualizations because dissimilar phenotypic features are grouped together. For example, chunking concepts by a nervous system category would put the unlike concepts of hypertonia, hypotonia, hyperreflexia, and hyporeflexia into the same category, a grouping of little diagnostic value. Although the chunking of phenotype concepts by body system or other schemes helps organize phenotype features, it does not reduce the number of features. Since the HPO is a hierarchical containment ontology, we have suggested that subsumption can create superclasses of phenotypic features and reduce the number of features (27,28).

Proposed approach and use case

We propose to improve the visualization of neurology phenotypes in the Orphdata dataset utilizing a combination of subsumption, vectorization, heat maps, and word clouds.

As proof of concept, we illustrate the utility of this approach with a use case that visualizes the phenotype lists of 32 dystonic diseases from Orphadata. In 1911 Oppenheim described the disease *dystonia musculorum deformans* and coined the term dystonia (29). Albanese et al. (30) defined dystonia as "a rare movement disorder characterized by sustained or intermittent muscle contractions causing abnormal, often repetitive movements, postures, or both." Since the description of dystonia by Oppenheim, many forms of dystonia have been described. Dystonia is classified along two axes: clinical and etiologic (30).

TABLE 1 The upper half of the Table shows lists of signs and symptoms for each dystonic disease from Orphadata.

Dystonias →	DYT16	DYT6	DYT27	DYT4
List of classes↓	Dystonia 16	Dystonia 6	Dystonia 27	Dystonia 4
	Limb dystonia	Dystonia	Oral dystonia	Blepharospasm
	Torticollis	Generalized dystonia	Laryngeal dystonia	Dementia
	Dysarthria	Dysarthria	Action tremor	Dysphagia
	Parkinsonism	Torticollis	Writer's cramp	Dysphonia
	Hyperreflexia	Blepharospasm	Limb dystonia	Generalized dystonia
	Dysphonia	Laryngeal dystonia	Axial dystonia	Laryngeal dystonia
	Dysphagia	Craniofacial dystonia	Focal dystonia	Abnl tongue movement
	Bradykinesia	Lingual dystonia	Postural tremor	Open mouth
	Postural tremor	Limb dystonia		Torticollis
	Orofacial dyskinesia			Gait disturbance
	Unsteady gait			Eunuchoid habitus
	Pyramidal sign			Sunken cheeks
	Lower limb pain			Involuntary movements
	Motor delay			Kyphoscoliosis
	Intellectual disability			Dysdiadochokinesis
				Respiratory distress
				Postural Tremor
Vector of superclasses ↓	Subsumption and vectorization ↓	Subsumption and vectorization ↓	Subsumption and vectorization ↓	Subsumption and vectorization ↓
Ataxia	0	0	0	1
Cognitive	1	0	0	1
Cranial nerve	1	0	0	2
Gait	1	0	0	1
Hyperkinesia	3	8	7	6
Hyperreflexia	1	0	0	0
Hypokinesia	2	0	0	0
Miscellaneous	2	0	0	4
Pain	1	0	0	0
Speech	2	1	0	1
Tremor	1	0	1	1

In the lower half of the Table, lists of classes have been converted to vectors of superclasses using subsumption governed by a lookup table. Counts are the number of times each class occurs in the superclass and is the input for the row values for the heat maps. Columns from the top half are variable length lists; columns from the bottom half are fixed length vectors.

Clinical classification is by age at onset, body distribution, the temporal pattern of symptoms, and associated phenotype features. Etiologic classification is by genetic versus non-genetic causation. Dystonia is one of the hyperkinetic movement disorders which also encompasses chorea, athetosis, hemiballismus, tics, tremors, stereotypy, myoclonus, and dyskinesia (31). Although all diseases labeled dystonia have a core symptom of dystonia, there is considerable variability in the clinical presentation (signs and symptoms) of the dystonias (29,32,33), making it an excellent use case for phenotype visualization. Furthermore, better characterization and classification of the dystonias is a major initiative of the European Reference Network for Rare Diseases, and Orphadata (34,35).

We downloaded the most recent Orphadata file with phenotype annotations of 4,254 rare diseases, including 1,184 rare neurological diseases. We identified 2,946 unique HPO terms used to characterize the signs and symptoms of rare neurological diseases and created a lookup table to map each term to one of 30 superclasses based on subsumption and expert opinion. The lists of phenotypic features for 32 dystonic diseases from Orphadata were converted into 31-element vectors, with the first element of the vector being the disease name and the next 30 elements being the count of features (signs and symptoms) for each superclass. The full 32-row × 31-column matrix of the dystonic diseases can be visualized as a feature map (**Figure 2**); individual rows can be visualized as word clouds (**Figure 3B**).

Methods

Phenotype feature lists by disease (data acquisition)

An XML file with 4,254 rare disease disorders and 112,256 phenotypic annotations was downloaded (June 2022 release of Orphadata: (https://www.orphadata.com/phenotypes/). Phenotype features are coded using the HPO ontology. Orphadata defines a rare disease as affecting less than 1 in 2,000 individuals in Europe and classifies 1,184 of the diseases as rare neurological diseases. We used python to parse the XML file and create a variable-length list of phenotypic features for each disease. We retained phenotypic annotations that were clinical signs or symptoms and filtered out phenotypic annotations related to disease course (progressive, static, etc.), mode of inheritance (recessive, dominant, etc.), biochemical abnormality, radiological abnormality, pathological abnormality, or electrophysiological abnormality. Based on published literature, Orphadata classifies the frequency of each phenotypic feature from rare (1–4%) to always present (100%). We retained phenotypic features classified as occasional or higher (5–100%).

Lookup table to convert phenotype classes to superclasses (subsumption)

The HPO (10) is organized as a hierarchical subsumption ontology so that more-specific concepts in the ontology are subsumed by more general concepts (28). We identified 2,946 unique concepts that Orphadata used to phenotype neurological diseases. We collapsed these concepts into 30 superclasses using subsumption and domain expert opinion. Example class memberships and class counts are shown for each superclass below.

1. alertness (53 terms) delirium, drowsy, somnolence
2. ataxia (62 terms) asynergia, clumsiness, dystaxia
3. atrophy (69 terms) muscle atrophy, atrophy, limb fasciculations
4. behavior (238 terms) apathy, anxiety, delusions
5. cognitive (202 terms) agnosia, apraxia, forgetfulness
6. cranial nerve (203 terms) ageusia, hyperacusis, facial diplegia
7. dysautonomia (35 terms) hypohidrosis, orthostatic syncope, dysautonomia
8. eye movements (272 terms) upgaze palsy, nystagmus, hypometric saccades
9. fatigue (26 terms) muscle fatigue, fatigable weakness, fatigue
10. gait (110 terms) ataxic gait, falls, unsteady gait
11. head (263 terms) microcephaly, macrocephaly, increased head size
12. hyperkinesia (157 terms) dyskinesia, dystonia, hyperkinesia
13. hyperreflexia (58 terms) increased reflexes, clonus, hyperreflexia
14. hypertonia (58 terms) increased muscle tone, rigidity, spasticity
15. hypokinesia (66 terms) bradykinesia, akinesia, hypomimia
16. hyporeflexia (43 terms) areflexia, hyporeflexia, absent ankle reflex
17. hypotonia (19 terms) decreased tone, muscle flaccidity, limb hypotonia
18. other muscle (119 terms) myokymia, muscle hypertrophy, myotonia
19. neck (48 terms) stiff neck, neck rigidity, meningismus
20. pain (145 terms) pain, arm pain, allodynia
21. seizure (358 terms) seizure, tonic-clonic seizure, febrile seizure
22. sensory (192 terms) hyperesthesia, dysesthesia, hypesthesia
23. skin (194 terms) cafe au lait spots, petechiae, rash
24. sleep (48 terms) cataplexy, narcolepsy, hypersomnia
25. speech_language (116 terms) dysarthria, aphasia, echolalia
26. sphincter (67 terms) urinary incontinence, constipation, enuresis
27. tremor (48 terms) tremor, resting tremor, action tremor
28. vision (450 terms) achromatopsia, scotoma, optic atrophy
29. weakness (159 terms) proximal weakness, foot drop, triceps weakness
30. miscellaneous (618 terms) nausea, vomiting, bradycardia

We used python to assign each phenotypic feature (sign or symptom) to one of the thirty superclasses based on the lookup table (see **Table 1** for an illustration of how individual phenotype features were mapped to superclasses). The lookup table is available in the **Supplementary Materials**.

Vectorization (conversion of phenotype lists to phenotype vectors)

Variable-length lists of phenotypic features were converted into vectors of fixed length 31 elements. The first element of the list was the disease label, and the following 30 elements were the counts of features in each of the 30 superclasses based on the lookup table. When the phenotype is represented as a vector, phenotypes can be compared by distance metrics. Furthermore, the magnitude of each element in the phenotype vector carries additional information that allows comparisons between diseases. For example, one disease with hyperkinetic features dystonia, chorea, and athetosis would have a hyperkinesia superclass value of $n = 3$, whereas a disease with only dystonia would have a hyperkinesia superclass value of $n = 1$. Such weightings could be useful in distinguishing between phenotypes of similar diseases.

Visualization (creation of heat maps and word clouds

Heat maps and word clouds were based on the phenotype vectors generated by python. Heat maps were created using the *heat map widget* from Orange (36). The score mapped for each superclass was the count of the phenotype features subsumed by that class. When a superclass had no features assigned to it, that superclass was dropped from the heat map. Word clouds were produced using the *word cloud widget* from Orange. Word size in the word cloud reflected the frequency of phenotypic features for a group of diseases (**Figure 1B**) or a single disease (**Figure 3B**).

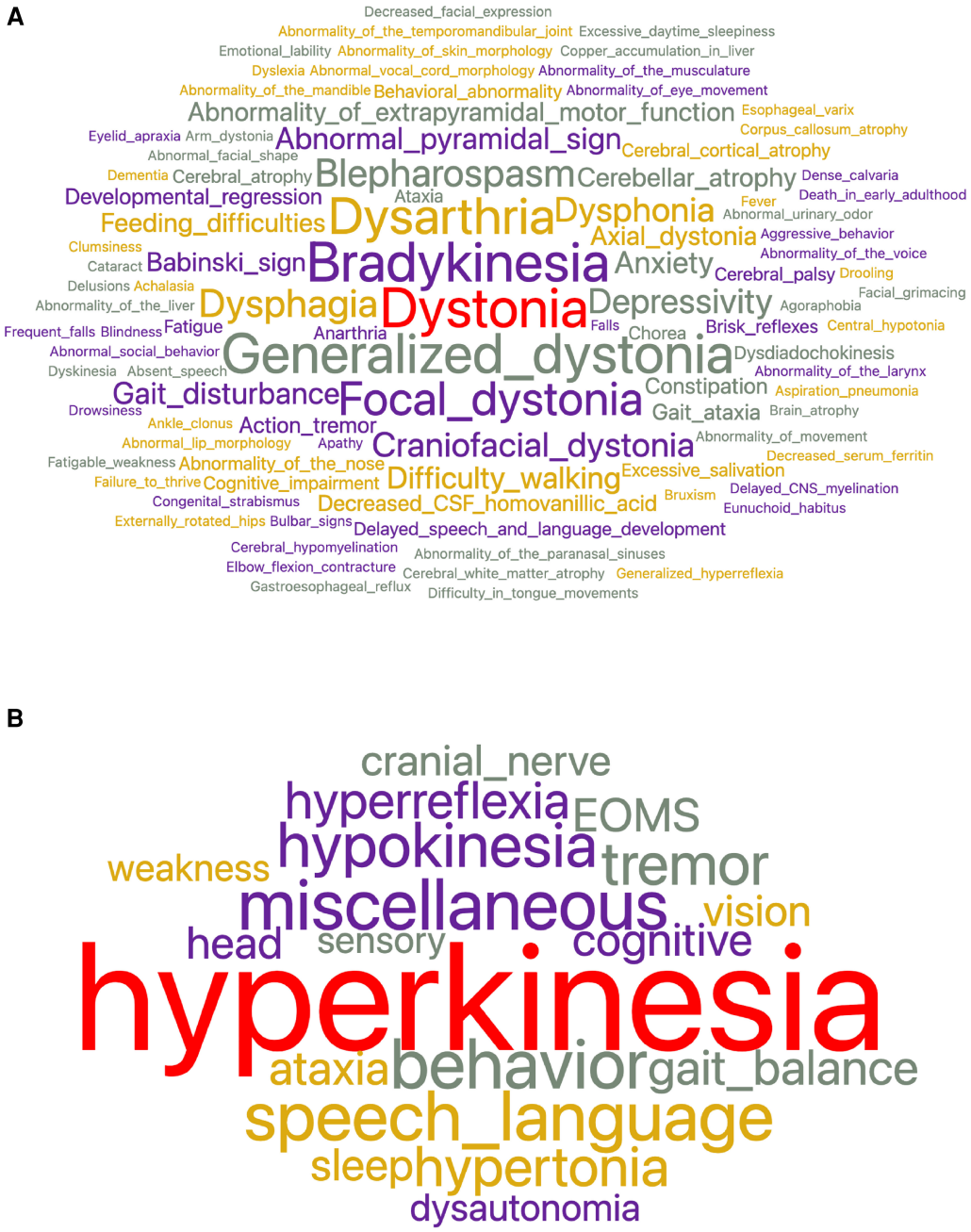


FIGURE 1
(A) To characterize the 32 dystonic diseases, 528 total concepts and 252 unique concepts were used. The most frequent concepts used were dystonia, bradykinesia, generalized dystonia, dysarthria, and focal dystonia. (B) After feature reduction by subsumption, the number of superclasses needed to characterize dystonia diseases was reduced to nineteen. The largest superclass is hyperkinesia which encompasses dystonia, generalized dystonia, focal dystonia, blepharospasm, craniofacial dystonia, and others.

Results

As our use case, we examined the phenotype profiles of 32 disease variants of dystonia in Orphadata. Phenotype profiles were lists of features (see **Table 1** for examples of DYT4, DYT6, DYT16, and DYT27). Feature lists ranged from 5 to 48 elements, with a mean of 18.4 features $\pm$10.5. The 252 unique features in the phenotype lists were reduced by subsumption into one of the 19 available 30 superclasses (**Table 1** and **Figure 1A,B**). This allowed visualization of the entire dystonia disease set of 32 variants as a heat map (**Figure 2**). This heat map allows an easy distinction of pure dystonia (e.g., DYT25 and DYT26) from dystonias with sensory loss (e.g., autosomal dominant dopa-responsive dystonia), cognitive impairment (e.g., DYT4) and hypokinesia (e.g., adult-onset dystonia-parkinsonism). Individual rows in the heat map (**Figure 3A** can be further visualized with word clouds which emphasize phenotypic differences between the dystonia variants (see **Figure 3B** for word clouds of DYT4, DY6, DYT16, and DYT 27.)

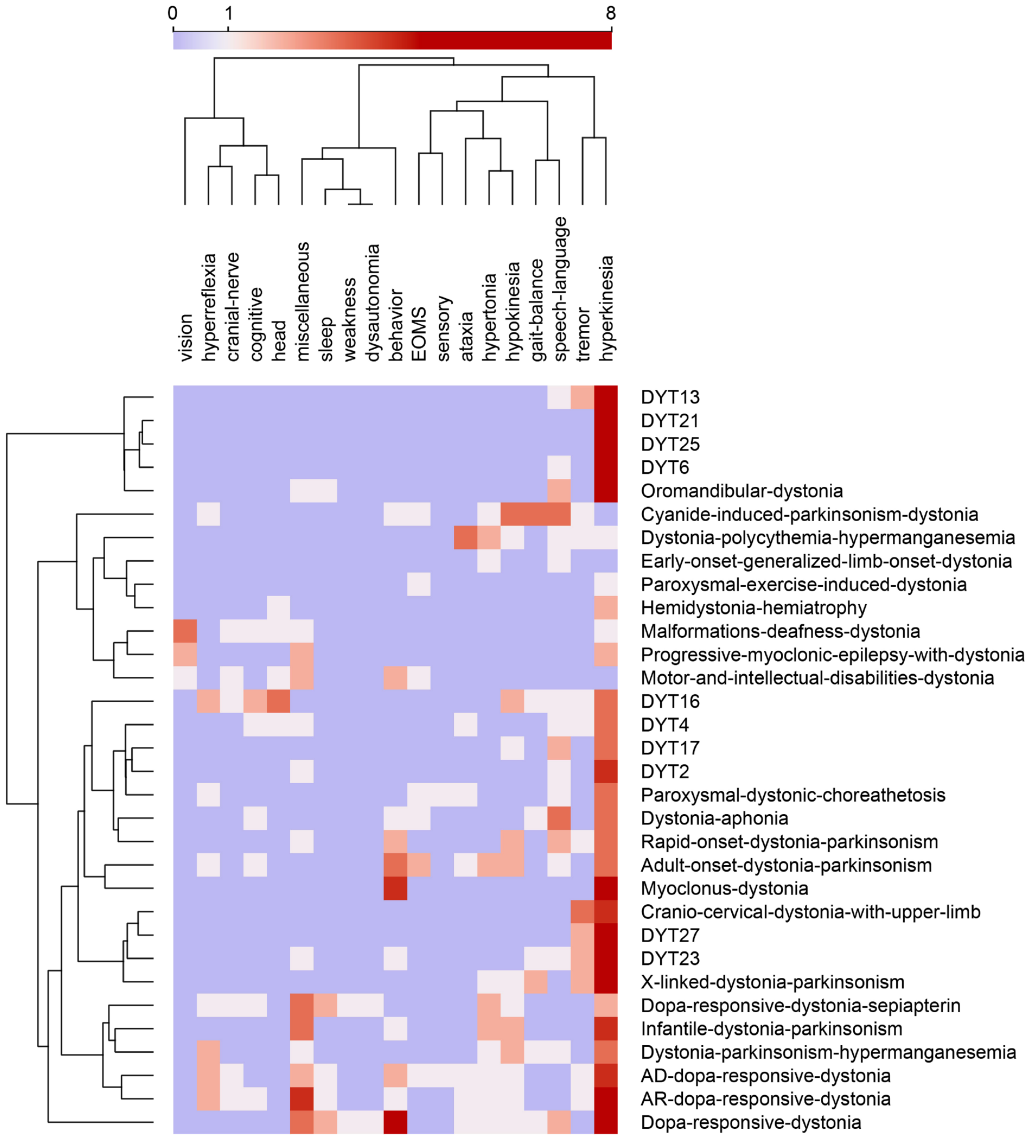


FIGURE 2
Feature map of 32 dystonias from Orphadata. Each row is a different variant of dystonia. Each column is one of 19 phenotype superclasses. Counts in columns range from 0 to 8. The color scale is centered at 1. Rows and columns are clustered by hierarchical clustering with Ward linkage. Distances between columns are by Pearson correlation coefficient. Distances between rows are by Euclidean distance. Hyperkinesia is the most frequent feature, followed by tremor, behavior, hypokinesia, speech_language, and miscellaneous (See word cloud in Figure 1B). Data underlying this table is available in the Supplementary Materials.

Discussion

Rich and detailed information on the phenotypes of neurological diseases is held in online repositories such as OMIM, HPO, and Orphadata. Detailed phenotypic data is available for download and can be used to gain insights into the inter-relationships between genes, disease, and phenotypes. Nonetheless, the visualization of the phenotypes retrieved as lists remains problematic. We identified several limitations to the visualization of disease phenotypes that included:

1. Phenotype feature lists are long.
2. Too many of the phenotype features are near synonyms, hyponyms, or hypernyms.
3. The number of unique features is large.
4. Side-by-side comparisons of phenotypes are difficult.
5. Phenotype lists of signs and symptoms are co-mingled with radiological, pathological, biochemical, and electrophysiological findings.

To address these limitations, we proposed restricting our attention to visualizing the phenotypes of rare neurological diseases in Orphadata ($N = 1,184$). We mapped each of the 4,505 unique features used to describe signs and symptoms in Orphadata into one of 30 superclasses (see list in the Methods section). This allowed us to convert phenotype lists of variable length to vectors of fixed length (31 elements), in which the first element of the vector was the disease label and the next 30 elements were the count of features for each of the 30 superclasses. This process of converting a list to a vector is illustrated in Table 1 for DYT4,

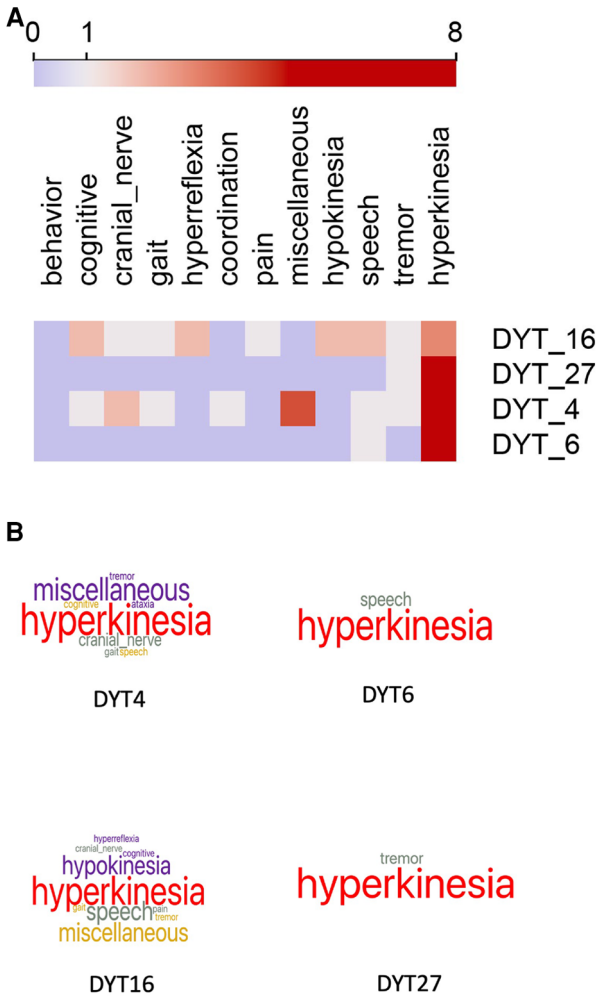


FIGURE 3
(A) Heat map of four selected cases of dystonia. Columns are feature superclasses, and rows are diseases. heat maps and word clouds are based on Table 1. Each row in the heat map represents a column of signs and symptoms from Table 1. Feature scores range from 0 to 8, with the color scale centered at 1. Word cloud visualizations of each row are below. Compare to Table 1 for comprehensibility. (B) Word clouds for the four forms of dystonia represented in the heat maps above and Table 1. Word size reflects the feature count in each superclass. DYT6 and DYT27 are pure dystonia, whereas DYT4 and DYT16 have other non-dystonic features. Compare to Table 1 for comprehensibility. Underlying data available in Supplementary Materials.

DYT6, DYT16, and DYT27. Only 11 of the 30 superclasses were needed to represent these four dystonias. Once phenotype lists are converted to vectors, a group of diseases can be represented as a matrix. For example, 32 dystonic diseases from Orphadata can be converted to a matrix with 32 rows (each row a disease) and 20 columns (each column a superclass of phenotypic features plus one column for the disease label) and then visualized as a heat map (**Figure 2**). For easy readability, individual rows (diseases) in the heat maps can be converted to word clouds to visualize better the phenotype (**Figure 3B**).

We have addressed limitation (1) (long feature lists) by using subsumption to collapse 4,505 phenotypic classes into 30 neurological superclasses. This subsumption of numerous phenotypic features into 30 superclasses also addressed limitation (2) (too many near-synonyms) and limitation (3) (too many unique features). Once phenotype lists of variable length are converted to vectors of fixed length, side-by-side comparisons of diseases become feasible through the use of heat maps and word clouds (**Figures 3A,B**); addressing limitation (4). Another advantage of vectorization is that it allows the calculation of distances between phenotypes using standard distance metrics such as cosine and Euclidean. **Figure 2** demonstrates the clustering of rows (dystonic diseases) using the Euclidean distance. We filtered out biochemical, radiological, electrophysiological, and pathological features to address limitation (5) (thus, limiting the phenotype to signs and symptoms.)

This work has some significant limitations. First, collapsing granular phenotype features into superclasses by subsumption involves information loss. The superclasses retain no laterality information (left-sided versus right-sided weakness, etc.) The superclasses retain no topographical information (proximal versus distal weakness, etc.) The high information value of some granular phenotype features, such as impaired vertical gaze (a sign of progressive supranuclear palsy) or internuclear ophthalmoplegia (a sign of multiple sclerosis), is lost when the granular features are collapsed into the superclass of abnormal eye movements. Second, our current process of collapsing phenotype concepts into superclasses requires a manually constructed lookup table that assigns each concept to a superclass. Errors can be made in assigning concepts to superclasses. We are looking at ways to improve the subsumption process that collapses ontology concepts into superclasses. Third, heat map scales are non-linear. For each superclass score, we counted the number of features in that superclass. For example, a disease phenotype with the term *hemiparesis* would have a superclass score of 1 for weakness. In contrast, a disease phenotype with terms *arm weakness* and *leg weakness* would have a superclass score of 2. Furthermore, we did not weight phenotype features by importance. In building the features maps, a more general concept like *hyperreflexia* carries the same weight as a more limited concept such as *increased biceps reflex*. We are exploring whether normalization or other transformations of the underlying data would improve the utility of the heat maps. Fourth, the size and granularity of the superclasses were not uniform. For example, the vision superclass subsumed 450 concepts and had many different types of visual impairment, whereas the fatigue superclass subsumed only 26 concepts and reflected the concept of fatigue alone. Fifth, our selection of thirty superclasses was somewhat arbitrary and subject to modification. Although the selection of the thirty superclasses reflected domain expert opinion and the underlying structure of the ontologies, other useful partitions of the ontology into superclasses are possible. For example, chorea or dystonia could have been distinct superclasses instead of subsumed into hyperkinesia. Speech (e.g., dysarthria) and language disorders (e.g., aphasia) could have been separate superclasses. Sixth, the superclasses were restricted to neurological terms and neurological diseases. As a result, the heat maps will not be useful in visualizing the phenotypes of non-neurological diseases. Furthermore, the heat maps will not adequately visualize important non-neurological signs and symptoms of diagnostic value (such as Kayser-Fleisher rings for Wilson's disease (37)). Although true pathognomonic signs and

symptoms are rare in neurology (1,38–40), the heat maps lack the granularity to show pathognomonic signs. Furthermore, the current heat maps do not support a *drill down* to the underlying granular phenotype features. Although we used Orange to create the heat maps, suitable heat maps are also available in python, and R. Other heat map color schemes are available and may give better visualizations. The Orphadata phenotype datasets are undergoing revisions and improvements. Some diseases are phenotyped more completely than others. Although the dataset is curated, omissions, errors, and discrepancies can still occur. Finally, a similar analysis could have been done with phenotypic annotations from the OMIM or HPO datasets.

Despite these limitations, combining feature reduction by subsumption with vectorization of phenotype lists followed by visualization by heat maps and word clouds offers a robust method to explore neurology phenotypes. Subsumption permits the reduction of thousands of ontological concepts into a reduced number of phenotype superclasses. Vectorization allows the conversion of variable-length phenotype feature lists into superclass vectors of fixed length. Matrices of superclass vectors allow the side-by-side comparison of disease phenotypes as heat maps. Individual rows in the heat maps can be visualized with word clouds, providing an easy-to-grasp representation of a disease phenotype.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: www.orphadata.com/phenotypes/ and in Supplementary Materials.

Author contributions

Concept by DBH and RY. Data analysis by DBH and RY. Data interpretation by DBH, MDC, RY, and DCW III. Writing, revision, and approval by DBH, MDC, RY, and DCW III. All authors contributed to the article and approved the submitted version.

Funding

MDC received financial support from the Veterans Administration and Biogen.

Acknowledgments

We acknowledge helpful discussions with our colleagues Chelsea Oommen, Quentin Howlett-Prieto, and Fahime Shojaei.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2023.1064936/full#supplementary-material.

References

1. Hier D, Yelugam R, Azizi S, Wunsch III D. A focused review of deep phenotyping with examples from neurology. *Eur Sci J*. (2022) 18:4–19. doi: 10.19044/esj.2022.v18n4p4

2. Gupta AS. Digital phenotyping in clinical neurology. *Semin Neurol*. (2022) 42:48–59. doi: 10.1055/s-0041-1741495

3. Delude CM. Deep phenotyping: the details of disease. *Nature*. (2015) 527:S14–5. doi: 10.1038/527S14a

4. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping, genomic data. *Nature*. (2018) 562:203–9. doi: 10.1038/s41586-018-0579-z

5. Robinson PN. Deep phenotyping for precision medicine. *Hum Mutat*. (2012) 33:777–80. doi: 10.1002/humu.22080

6. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: online mendelian inheritance in man (OMIM®), an online catalog of human genes, genetic disorders. *Nucleic Acids Res*. (2015) 43:D789–98. doi: 10.1093/nar/gku1205

7. Maiella S, Rath A, Angin C, Mousson F, Kremp O. Orphanet and its consortium: where to find expert-validated information on rare diseases. *Rev Neurol*. (2013) 169 (Suppl 1):S3–S8. doi: 10.1016/s0035-3787(13)70052-3

8. Groza T, Köhler S, Moldenhauer D, Vasilevsky N, Baynam G, Zemojtel T, et al. The human phenotype ontology: semantic unification of common and rare disease. *Am J Hum Genet*. (2015) 97:111–24. doi: 10.1016/j.ajhg.2015.05.020

9. Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, et al. The human phenotype ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res*. (2014) 42:D966–74. doi: 10.1093/nar/gkt1026

10. Kohler S, Robinson P. Human Phenotype Ontology (2022). Available from: https://bioportal.bioontology.org/ontologies/HP

11. McKusick-Nathans Institute for Genetic Medicine. Online Mendelian Inheritance in Man (2022). Available from: https://bioportal.bioontology.org/ontologies/OMIM

12. Hier DB, Brint SU. A neuro-ontology for the neurological examination. *BMC Med Inform Decis Mak*. (2020) 20:1–9. doi: 10.1186/s12911-020-1066-7

13. Miller GA. The magic number seven plus or minus two: some limits on our capacity for processing information. *Psychol Rev*. (1956) 63:91–7. doi: 10.1037/h0043158

14. Xu W, Jiang X, Hu X, Li G. Visualization of genetic disease-phenotype similarities by multiple maps t-SNE with Laplacian regularization. *BMC Med Genomics*. (2014) 7:1–9. doi: 10.1186/1755-8794-7-S2-S1

15. Emmert-Streib F, Tripathi S, Simoes RDM, Hawwa AF, Dehmer M. The human disease network: opportunities for classification, diagnosis, and prediction of disorders and disease genes. *Syst Biomed*. (2013) 1:20–8. doi: 10.4161/sysb.22816

16. Wei DH, Kang T, Pincus HA, Weng C. Construction of disease similarity networks using concept embedding and ontology. *Stud Health Technol Inform*. (2019) 264:442. doi: 10.3233/SHTI190260

17. Köhler S, Doelken SC, Rath A, Aymé S, Robinson PN. Ontological phenotype standards for neurogenetics. *Hum Mutat*. (2012) 33:1333–9. doi: 10.1002/humu.22112

18. Clementz BA, Trotti RL, Pearlson GD, Keshavan MS, Gershon ES, Keedy SK, et al. Testing psychosis phenotypes from bipolar–schizophrenia network for intermediate phenotypes for clinical application: biotype characteristics and targets. *Biol Psychiatry*. (2020) 5:808–18. doi: 10.1016/j.bpsc.2020.03.011

19. Glueck M, Gvozdik A, Chevalier F, Khan A, Brudno M, Wigdor D. Phenostacks: cross-sectional cohort phenotype comparison visualizations. *IEEE Trans Vis Comput Graph*. (2016) 23:191–200. doi: 10.1109/TVCG.2016.2598469

20. Glueck M, Hamilton P, Chevalier F, Breslav S, Khan A, Wigdor D, et al. Phenoblocks: phenotype comparison visualizations. *IEEE Trans Vis Comput Graph*. (2015) 22:101–10. doi: 10.1109/TVCG.2015.2467733

21. Cao N, Sun J, Lin YR, Gotz D, Liu S, Qu H. Facetatlas: multifaceted visualization for rich text corpora. *IEEE Trans Vis Comput Graph*. (2010) 16:1172–81. doi: 10.1109/TVCG.2010.154

22. Cao N, Gotz D, Sun J, Lin YR, Qu H. SolarMap: multifaceted visual analytics for topic exploration. *2011 IEEE 11th International Conference on Data Mining*. IEEE (2011). p. 101–10. Available from: https://doi.org/10.1109/ICDM.2011.135

23. Cao N, Gotz D, Sun J, Qu H. DICON: interactive visual analysis of multidimensional clusters. *IEEE Trans Vis Comput Graph*. (2011) 17:2581–90. doi: 10.1109/TVCG.2011.188

24. Gotz D, Sun J, Cao N. Multifaceted visual analytics for healthcare applications. *IBM J Res Dev*. (2012) 56:6–1. doi: 10.1147/JRD.2012.2199170

25. Hier DB, Pearson J. Two algorithms for the reorganisation of the problem list by organ system. *BMJ Health Care Inform*. (2019) 26:e100024. doi: 10.1136/bmjhci-2019-100024

26. Yauy K, Duforet-Frebourg N, Testard Q, Beaumeunier S, Audoux J, Simard B, et al. Learning phenotypic patterns in genetic disease by symptom interaction modeling. *medRxiv* (2022). Available from: https://doi.org/10.1101/2022.07.29.22278181

27. Wunsch III DC, Hier DB. Subsumption is a novel feature reduction strategy for high dimensionality datasets. *Eur Sci J*. (2022) 18:20–33. doi: 10.19044/esj.2022.v18n4p20

28. Wunsch DC, Hier DB. Subsumption reduces dataset dimensionality without decreasing performance of a machine learning classifier. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE (2021). p. 1618–21. Available from: htts://doi.org/10.1109/EMBC46164.2021.9629897

29. Grütz K, Klein C. Dystonia updates: definition, nomenclature, clinical classification, and etiology. *J Neural Transm*. (2021) 128:395–404. doi: 10.1007/s00702-021-02314-2

30. Albanese A, Bhatia K, Bressman SB, DeLong MR, Fahn S, Fung VS, et al. Phenomenology, classification of dystonia: a consensus update. *Mov Disord*. (2013) 28:863–73. doi: 10.1002/mds.25475

31. Jankovic J. Treatment of hyperkinetic movement disorders. *Lancet Neurol*. (2009) 8:844–56. doi: 10.1016/s1474-4422(09)70183-8

32. Lange LM, Junker J, Loens S, Baumann H, Olschewski L, Schaake S, et al. Genotype–phenotype relations for isolated dystonia genes: MDSgene systematic review. *Mov Disord*. (2021) 36:1086–103. doi: 10.1002/mds.28485

33. di Biase L, Di Santo A, Caminiti ML, Pecoraro PM, Di Lazzaro V. Classification of dystonia. *Life*. (2022) 12:206. doi: 10.3390/life12020206

34. Centen LM, Pinter D, van Egmond ME, Graessner H, Kovacs N, Koy A, et al. Dystonia management across Europe within ERN-RND: current state, future challenges. *J Neurol*. (2022) 1–13. doi: 10.1007/s00415-022-11412-4. [Epub ahead of print]

35. Graessner H, Brunelle A, Reinhard C, Hermanns S, Post A. European reference network for rare neurological diseases–ERN-RND. *Information Brochure* (2020). Available from: https://www.ern–rnd.eu/

36. Demšar J, Curk T, Erjavec A, Č G, Hočevar T, Milutinovič M, et al. Orange: data mining toolbox in python. *J Mach Learn Res*. (2013) 14:2349–53. doi: 10.5555/2567709.2567736

37. Finelli PF. Kayser-Fleischer ring: hepatolenticular degeneration (Wilson's disease). *Neurology*. (1995) 45:1261–2. doi: 10.1212/wnl.45.7.1261

38. Janeway EG. Limitations of pathognomonic signs and symptoms. *J Am Med Assoc*. (1884) 3:116–20. doi: 10.1001/jama.1884.02390540004001a

39. Barrows HS, Bennett K. The diagnostic (problem solving) skill of the neurologist: experimental studies and their implications for neurological training. *Arch Neurol*. (1972) 26:273–7. doi: 10.1001/archneur.1972.00490090099009

40. Chimowitz MI, Logigian EL, Caplan LR. The accuracy of bedside neurological diagnoses. *Ann Neurol*. (1990) 28:78–85. doi: 10.1002/ana.410280114

frontiers | Frontiers in Digital Health

TYPE Perspective
PUBLISHED 22 March 2023
DOI 10.3389/fdgth.2023.1149154

OPEN ACCESS

EDITED BY
Daniel B. Hier,
Missouri University of Science and Technology, United States

REVIEWED BY
Jorge Cancela,
Roche, Switzerland
John McNeil,
Monash University, Australia

*CORRESPONDENCE
Allan D. Wu
allan.wu@northwestern.edu

SPECIALTY SECTION
This article was submitted to Health Informatics, a section of the journal Frontiers in Digital Health

RECEIVED 21 January 2023
ACCEPTED 01 March 2023
PUBLISHED 22 March 2023

CITATION
Wu AD and Wilson AM (2023) Parkinson's disease population-wide registries in the United States: Current and future opportunities. Front. Digit. Health 5:1149154. doi: 10.3389/fdgth.2023.1149154

Parkinson's disease population-wide registries in the United States: Current and future opportunities

Allan D. Wu[1,2,3]* and Andrew M. Wilson[2,4]

[1]Division of Movement Disorders, Department of Neurology, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States, [2]Department of Neurology, David Geffen School of Medicine, University of California Los Angeles (UCLA), Los Angeles, CA, United States, [3]Stanley Manne Children's Research Institute, Ann & Robert H. Lurie Children's Hospital, Chicago, IL, United States, [4]Department of Neurology, Greater Los Angeles VA, Los Angeles, CA, United States

Parkinson's disease (PD) is a neurodegenerative disease with both genetic and environmental risk factors. Efforts to understand the growing incidence and prevalence of PD have led to several state PD registry initiatives in the United States. The California PD Registry (CPDR) is the largest state-wide PD registry and requires electronic reporting of all eligible cases by all medical providers. We borrow from our experience with the CPDR to highlight 4 gaps to population-based PD registries. Specifically we address (1) who should be included in PD registries; (2) what data should be collected in PD case reports; (3) how to ensure the validity of case reports; and (4) how can state PD registries exchange and aggregate information. We propose a set of recommendations that addresses these and other gaps toward achieving a promise of a practical, interoperable, and scalable PD registry in the U.S., which can serve as a key health information resource to support epidemiology, health equity, quality improvement, and research.

KEYWORDS
Parkinson's disease, parkinsonism, registries, epidemiology, population informatics

Introduction

Parkinson's disease (PD) is the most rapidly growing neurodegenerative disease across the globe (1). Epidemiology studies, using claims datasets, have estimated prevalence and incidence of PD (2, 3), and observational cohort studies have identified both environmental and genetic risk factors for the development of PD (4–6). To expand upon this work, true population-wide PD registries, leveraging real-time electronic health record (EHR) data associated with clinical care, hold promise to address more comprehensive questions about epidemiological risk factors, treatment, healthcare utilization, and outcomes across the wide diversity of people and community settings.

In the United States, statewide PD surveillance registries are growing in momentum to assess the prevalence, incidence, and distribution of cases and to support public health education, outreach, and research. Nebraska was the first statewide PD registry (1997), requiring reporting of new PD cases (7). The California Parkinson's Disease Registry (CPDR) was established in 2005 to determine the incidence and prevalence of PD in California, to examine disparities in PD risk, and to conduct demographic and epidemiological research. The CPDR started requiring mandatory reporting of all PD cases in 2018 (8, 9). Multiple states have smaller registries or legislation pending for PD

or neurodegenerative registries (10, 11). In parallel, Congress authorized the Centers for Disease Prevention and Control (CDC) to develop a National Neurologic Conditions Surveillance System (NNCSS), with initial conditions being PD and Multiple Sclerosis (12). To date, efforts to align state PD registries to form an effective network of statewide PD registries are limited.

In this *Perspective*, we discuss the CPDR as an all-electronic, near real-time PD registry and the largest current example of a state PD registry. The CPDR requires all providers to report encounters where PD is treated or diagnosed, regardless of encounter type or specialty—a particularly broad set of criteria. Cases can be reported in real-time using electronic health record (EHR) case reports (or near real-time in quarterly batches); an online portal is used for manual reporting of individual cases. As of 2021, the CPDR has received 534,583 reports from 550 reporting entities across most counties, covering 93,928 unique PD patients (13). Reporting from California practices is not yet considered complete and no prevalence estimates have been released. For researchers, a data disclosure policy and procedure was released in 2021.

The CPDR also exemplifies many challenges and gaps faced by population-wide PD registry design, implementation, and usability. To help address these gaps, the Michael J. Fox Foundation for Parkinson's Research recently supported an independent project at the University of California, Los Angeles (the UCLA-CPDR-EHR PD UCE-PD project). The UCE-PD project aims were to assess the accuracy and completeness of data collected by automated means at a single large academic site and to develop, implement, and demonstrate a framework of tools to improve upon CPDR accuracy and completeness. The UCE-PD project was led by a multidisciplinary team including movement disorders specialists, general neurologists, and primary care physicians with expertise representing clinical practice, epidemiology, clinical informatics, and health services research.

We borrow from our experience with the CPDR and UCE-PD project to highlight 4 gaps in population-based PD registries. Specifically we address (1) who should be included in PD registries; (2) what data should be collected in case reports; (3) how to ensure the validity of case reports; and (4) how can PD registries exchange information? We conclude by presenting a list of recommendations to consider as next steps toward realizing a population-wide PD registry.

Gap 1: who should be included in a PD registry?

The clinical diagnosis of PD can be challenging as there is no confirmatory test or biomarker. Current diagnostic criteria for PD rely on clinical expertise and factors that are uncommonly coded reliably or accurately in EHRs (14, 15). There are circumstances when the diagnosis of PD cannot be made with confidence (16), particularly early cases of parkinsonism or those confounded by alternate causes (e.g., drug-related or vascular). The diagnosis of PD is also confounded by related, though distinct, neurodegenerative parkinsonism syndromes (NPS) such as progressive supranuclear palsy (PSP), corticobasal syndrome (CBS), multiple systems atrophy (MSA), or dementia with Lewy bodies (DLB), which may only become clinically clear after years of being diagnosed with PD. Notably, persons with NPS are of interest to PD registries because they may share epidemiological risks and have similar health resource needs as persons with PD (PwPD). Excluding NPS risks reducing potential value in a PD registry.

There is also no consensus EHR phenotype for PD. A combination of diagnosis codes, medications, provider specialties, and lookback intervals have been used by many published algorithms for detecting PD from EHR data and support the ability to detect PD or NPS (17–21). Unfortunately, estimates are that only 75%–82% of cases of PD detectable by codes are actually PD (21–23). Such issues, well-known among neurologists, researchers, and clinicians, contribute to some skepticism for cases included (or not included) in registries. Additionally, performance of algorithms are challenging to interpret because of variation in whether the focus is on the detection of PD itself, PD with NPS, or parkinsonism in general (17). Further, algorithms developed in one system have rarely been tested using data across differing systems, and consensus algorithms have not yet emerged (24).

The CPDR relies only on ICD10 diagnosis codes (G20 and G90.3) to trigger encounters to report. While G90.3 is intended to identify MSA, other NPS were excluded. The G20 code represents PD, but is also used when the clinician codes for less certain parkinsonism "not otherwise specified." When the UCE-PD team reviewed a sample of 456 patients identified using six parkinsonism codes, we found that the two code CPDR combination had a lower positive predictive value for PD than G20 alone or the broader set of parkinsonism codes (**Figure 1A**).

Gap 2: what data should be collected in PD data reports?

Currently, individual statewide PD registries separately develop data specifications, which limits harmonization of collected data and reduces the potential benefit of such registries. Consensus standards for data elements recommended for population-wide PD registries have yet to be established.

The CPDR experience is illustrative of the challenge of determining a minimum required specification across a diverse healthcare system. Initial CPDR proposed specifications included both administrative data (reporting entity, patient demographics, provider information) and required available clinical data elements, such as PD symptoms, medications and comorbidities (25). Due to non-standard and variable nature of how clinical data elements are documented among different hospital and clinic settings, advocacy groups responded with concerns about reporting feasibility and burden. The final CPDR specification places nearly all clinical data elements into an optional category, with exceptions being encounter diagnosis codes and the date of diagnosis [(9), **Supplementary Material S1**]. This oversimplification of required PD data elements limited clinical

A

Diagnosis codes	PD code	CPDR codes	Parkinsonism codes
ICD10 codes included:	G20	G20, G90.3	G20, G90.3, G23.1, G23.2, G31.83, G31.85
PPV[cs] for Probable PD	347/420 (83%)	348/433 (80%)	348/456 (76%)
PPV[cs] for Probable or Possible PD	416/420 (99%)	422/433 (97%)	445/456 (98%)

B

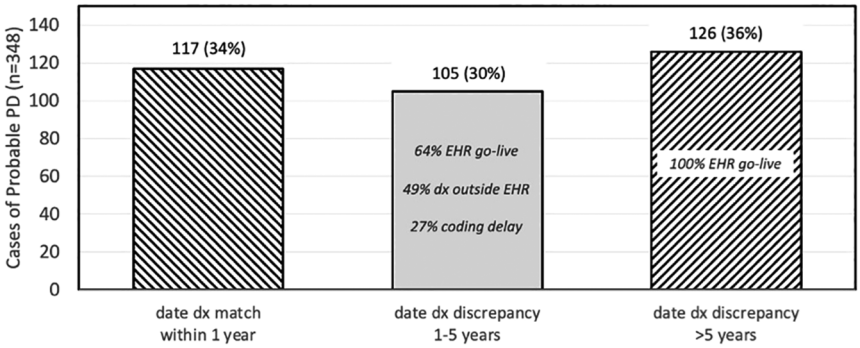


FIGURE 1
Performance of CPDR diagnosis codes and date of diagnosis. (A) Effect of different ICD10 trigger codes on case identification of Probable or Possible PD in a cross-sectional (CS) sample of cases. Probable and Possible PD assignment was made by a neurologist after manual chart review. The lowest positive predictive value (PPV) for Probable or Possible PD was with CPDR codes. (B) For cases of Probable PD ($n = 348$), the date of diagnosis discrepancy between the CPDR reported date and the date from chart review was within 1 year in 34% of cases. The percentage of cases with discrepancies less than or more than 5 years (middle and right bar, respectively), and the reasons for those discrepancies, are shown. Some cases had more than one discrepancy reason. EHR go-live 5 years prior to data collection resulted in a floor for automated CPDR-reported dates.

utility and reduced some ability to de-duplicate or validate cases reported.

Some data elements that may be considered for PD registries may not be easily, reliably, or accurately captured. For example, the CPDR specification requires that each report include a date of diagnosis. Even mature cancer registries, where this date is a key data element and anchored by pathological confirmation, struggle to obtain this information from oncology specialty practices (26). To ensure success of initial CPDR implementation, the registry allows the earliest date of a trigger diagnosis on the Problem List or the earliest encounter date to be used as the date of diagnosis (9). This data definition favors completeness of data reported with potential risk to accuracy. When reviewed in a sample of 348 Probable PD cases, the UCE-PD team found that the date of diagnosis reported by CPDR specifications was accurate within 1 year in 34% of cases when compared to that by gold-standard manual chart review (**Figure 1B**). This work emphasizes that certain data elements being considered as specifications for a PD registry should be assessed for quantifiable risk of accuracy or completeness.

Gap 3: how should PD case reports and registry data be validated?

Quality assurance (QA) and case validation ensure that registries are capturing data that are complete and high quality. QA is arguably even more important in real-world, EHR-driven, population-wide registries, where data are created as a byproduct of clinical care (27). Because PD and NPS do not have a definitive biomarker or standard nomenclature (unlike cancer staging), the need for a transparent and robust validation process to build confidence among stakeholders is important to include in PD registry design.

Cancer registries, consolidated under the CDC National Program of Cancer Registries (NPCR), cover 97% of the US population and collect timely data on incidence, treatments, and outcomes (28, 29). To achieve this, resources are available to train and certify cancer registry abstractors, usually near the point of care. As such, population cancer registries focus abstractor efforts on the collection and submission of high quality data at the source. Many existing Parkinson disease registry efforts similarly collect high quality data, requiring considerable resources, from selected movement disorder specialty sites (30).

This point of care approach is not practical for population-scale PD registries where cases are reported across a wide spectrum of medical practices. Validation within large registries typically sample cases and compare them against gold-standard neurologic assessment or manual chart review. The outcome is to assess whether variations in PD data in the registry between sites are due to differences in coding practices (17), distribution of care (3), or represent actual differences in incidence or prevalence. The CPDR provides completeness data to reporting sites, but has not yet adopted guidance for validating cases reported to the registry.

For the UCE-PD project, we developed a proof of principle validation workflow for CPDR-eligible cases. The strategy used was to select cases for review, have trained abstractors manually review charts using a standardized abstraction tool, and have experienced adjudicators confirm the PD diagnosis classification by reviewing summary information from the abstraction tool. QA tools were developed including abstractor training modules, feedback sessions, and inter-rater dual abstraction reliability checks. Challenging cases were escalated for further chart review to neurologists to finalize an adjudicated classification. Each final case classification would represent the gold-standard for validation purposes.

This validation process is theoretically scalable because of its potential federated approach. The validation abstraction process and associated QA checks can be conducted within each local site. Importantly, clinician review would not be required for most adjudications. Applied across all sites reporting PD cases, this process can provide standardized validation information that can help enhance the trust of patients, clinicians, and researchers participating in PD registries.

Gap 4: how can PD registries exchange and aggregate information?

It is unrealistic to think that any one singular registry can house the requisite information to address the epidemiological, clinical, and health services questions of the future. A successful population-wide PD registry will require an interoperability infrastructure that supports data exchange among registries.

Interoperability requires a common standard of codes that represent data elements captured from all EHRs used. To the extent possible, registries will specify mappings of required data elements to standard terminology code sets, such as ICD10 for diagnoses, RxNorm for medications, CPT for procedures, etc. (31). However, some concepts important for PD registries may not yet have a standard code. As an example, movement disorders, as a neurological subspecialty, is not represented in standard specialty taxonomy code sets (32). In circumstances where gaps exist in standard code sets, an interim step can be to partner with health information exchanges (HIE) that could support non-standard data elements of importance.

To illustrate, the UCE-PD team developed a focused data dictionary of symptoms that are commonly encountered in PD. We worked with an EHR vendor (Epic Systems, Verona WI) to create common PD registry data elements within the default EHR system. These symptom data elements are now automatically available and semantically interoperable for all customers within the vendor-specific HIE (**Supplementary Material S2**).

The technical tools and trust framework of sharing PD registry data are an area for innovation and ongoing evolution. Health Level 7 International (HL7) sets widely used standards for the exchange, integration, sharing, and retrieval of electronic health information. An electronic case report (eCR) standard was released by HL7 in 2017 with data elements that represent a consensus "minimum necessary" for public health case reports (33). Prior to 2019, implementation of eCR by entities reporting to the CPDR was low. As eCR was widely promoted during the Covid pandemic (34), there was a significant increase in entities that supported eCR infrastructure. The CPDR saw an increase in eCR format reports from none in 2019 to 67% of reported cases in 2021 (personal communication, CPDR). Unfortunately, piecemeal adoption of eCR formats by individual county and state public health departments, rather than broad adoption, remains a barrier to full interoperability.

Discussion

We briefly outlined some of the current state and challenges of developing population-wide PD registries. We discussed CPDR as an example of a statewide PD registry implementation, recognizing the growing momentum toward additional statewide registries in the near future. With this context, we propose a set of recommendations that addresses these and other gaps toward achieving a promise of practical, interoperable, and scalable population-wide PD registries. While this *Perspective* focuses on aspects of a U.S. state implementation, how to adapt these recommendations to international sites must also be considered. The state-by-state (e.g., California, Nebraska) approach that characterizes U.S. public health presents challenges may be less prevalent in centralized healthcare systems. Our vision is that, while the initial population-wide registries will first support use cases of public health surveillance, epidemiology, and assessment of health care utilization, the maturation of broad interoperability frameworks will enable development of these PD registries as key population-wide health information resources. For example, when eventually linkable to other patient outcomes, clinical trial data, quality registries, genomics, and biorepository resources, discoveries can be inferred, developed, and applied at scale as public health interventions to advance access and health equity outcomes, quality improvement initiatives, and research efforts.

Overview recommendations:

1. Propose a series of symposia or workshops to develop consensus around a core set of infrastructure decisions to support population-wide PD registries. Participation should include subject matter experts, patient advocacy groups, specialty societies, health system informaticists, state public health departments, and CDC NNCSS to develop broad stakeholder engagement.
2. Develop a vision and mission statement about the role of population-wide and state PD registries. This statement should reflect direct goals supporting public health surveillance, health services equity, and epidemiology research as well as longer-term goals to support efforts in public health intervention, quality improvement, and research. Endorsement of FAIR (Findable, Accessible, Interoperable, and Reusable) principles for data management should be encouraged (35).

3. Develop, publish, and maintain a repository of standards and guidelines. As registries are repositories of systematically collected data, a central set of consensus documentation around use cases, implementation strategies, data collection standard operating procedures, and data dictionaries, as further discussed in points below, will be needed.
4. Develop guidelines at state and federal levels to address common issues for large public health registries including, but not limited to patient confidentiality, the balance of public health vs. right to privacy, ownership of data, reusability of data, and return of benefits of the registry to stakeholders.
5. Evaluate, develop, and share models for financial and resource sustainability for individual state PD registries, exploring partnerships with academia, third-party vendors, federal regulatory agencies, or other solutions. To date, uncertainties in state budgets have adversely affected operations in state-funded registries (i.e., CPDR, Nebraska PD Registry). As an exemplar, a combination of federal, state, and private funds have helped sustain cancer registries in the US (36).

Scientific considerations:

6. Encourage inclusive PD/parkinsonism registries that will encompass both PD and NPS. This recommendation is supported by clinical overlap, challenges in detecting early possible PD or NPS, and need to understand scientific, clinical, and health care delivery similarities and differences with both PD and NPS.
   a. Initial CPDR criteria of six ICD10 codes for parkinsonism are a good starting point, but further scientific consensus on reporting criteria is recommended.
7. Support development of a practical intermediate classification system for labelling each case reported in a broadly inclusive PD/parkinsonism registry. The classification should be granular enough to reflect real-world uncertainties in PD diagnosis, yet high-level and discrete enough to facilitate automatic interoperable mapping between state registries.
   a. The UCE-PD team developed a consensus diagnostic classification scheme to account for the variations in data quality and diagnostic uncertainty commonly encountered when validating cases of potential PD. **Figure 2A** (and **Supplementary Material S3**) illustrates the UCE-PD consensus nomenclature and conceptual definitions for labelling case reports.
8. Develop a consensus data dictionary of elements recommended for reporting to population-wide PD registries, prioritizing those elements that are readily available (feasible and accurate) and are

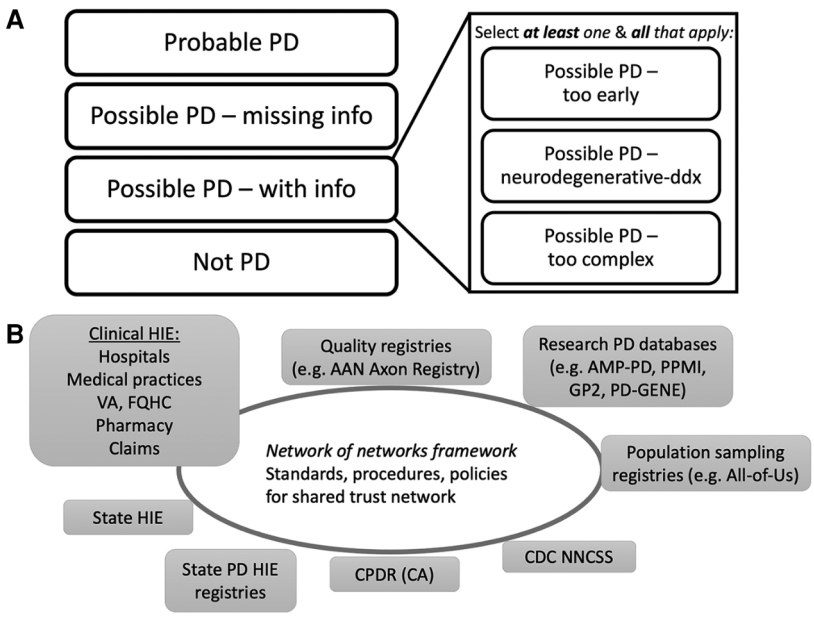


FIGURE 2
Proposed EHR PD classifications and future interoperability network with PD registries. (A) Overview of UCE-PD classification for each reported case at a point in time in PD/parkinsonism registries. Each case has a unique classification (left) with non-exclusive subclassifications for Possible PD (right) (Supplementary Table S2 for details). (B) Schematic of a future interoperability network of networks to support electronic health information exchange (HIE) relevant for PD registries. A national trust framework (Trusted Exchange Framework Common Agreement, TEFCA) will facilitate exchanges among PD registries and balance public health mandatory reporting with the sharing of clinical, quality, or research information within appropriate legal, compliance, confidentiality, and privacy policies. Individual state PD registries (CPDR for example) could connect to the framework directly, could form interstate PD-specific HIEs, or connect indirectly *via* other clinical HIEs. Other specific PD registries (quality, research), if part of the trust framework, could request information from population-wide PD registries for relevant context, and vice versa. EHR, electronic health record; VA, Veteran's Administration; FQHC, federally qualified health center; AAN, American Academy of Neurology; AMP-PD, Accelerating Medicines Partnership Parkinson's Disease; PPMI, Parkinson's Progression Markers Initiative; GP2, Global Parkinson's Genetics Program; PD-GENE, PD GENEration; CDC NNCSS, Centers for Disease Control National Neurologic Conditions Surveillance System; CPDR, California Parkinson's Disease Registry; CA, California. EHR, electronic health record; VA, Veteran's Administration; FQHC, federally qualified health center; AAN, American Academy of Neurology; AMP-PD, Accelerating Medicines Partnership Parkinson's Disease; PPMI, Parkinson's Progression Markers Initiative; GP2, Global Parkinson's Genetics Program; PD-GENE, PD GENEration; CDC NNCSS, Centers for Disease Control National Neurologic Conditions Surveillance System; CPDR, California Parkinson's Disease Registry; CA, California.

essential to case de-duplication and validation. Such a data dictionary will ensure a common base of terminology for interoperable data exchange among PD registries.
   a. Elements range from demographic elements (most feasible), administrative data elements (e.g., encounter dates and types, specialties, medications; feasible, some variability), to clinical symptom, diagnostic certainty, and disease severity elements (most challenging to standardize within real-time workflows).
9. Recommend a basic, minimum dataset standard for mandatory reporting across all sites reporting cases. While this dataset emphasizes feasibility for automated reporting, a minimum data standard will reduce risk of oversimplification of specifications.
   a. Additional data dictionary specifications as distinct data modules can be considered or added as population-wide PD registries mature. Such an approach addresses the problem of missing data when desired data elements are specified as "required if available" or optional. Sites with sufficient reporting capabilities, resources, or interest (e.g., neurology or movement disorder practices) may be incentivized to report on additional specified data elements.
10. Assess and ensure that recommended data elements are represented and mapped to standard concept codes. Gaps identified should be addressed with a strategy to develop, test, and create appropriate codes with appropriate standard development organizations.
   a. As an example, a consensus strategy to update the current ICD10 code for PD (G20) can be considered to separate out alternate diagnoses of a nonspecific parkinsonism or an uncertain early PD.
11. Support evaluation of data elements that are considered for population-wide PD registries, but will be more challenging to collect. Data elements can be proposed as provisional and tested before being approved within either a basic or higher tier data specification.
12. Support, develop, and incentivize a systematic and scalable validation process for population-based PD registries. As a starting point for discussion, the UCE-PD team has developed proof-of-principle processes and tools to support abstraction and case adjudication for PD registries.

Registry implementation:

13. Recommend that each state PD/parkinsonism registry maintain a standing scientific and patient advisory committee to ensure stakeholder engagement and alignment with consensus guidelines. A forum should be available where state PD/parkinsonism registries and the CDC NNCSS can communicate, share strategies, and align on national goals.
14. Prioritize automated reporting through certified EHR mechanisms. As eCR is now a mandatory component of the 2023 Medicare Promoting Interoperability incentive payment system (37), we recommend that an eCR specification be used as a preferred public health report system for PD registries.
15. Monitor and evaluate technologies and policies covering interoperability solutions as relevant to the development of a network of interoperable state and population-wide PD registries (**Figure 2B**). Examples include:
   a. Alignment with the United States Core Data for Interoperability (USCDI), the federally required set of data elements that certified EHR systems must support for interoperability. USCDI+ was recently announced as a possible domain-specific extension for which PD registries could be an ideal domain use case (38).
   b. Fast Healthcare Interoperability Resources (FHIR) standards hold promise for enabling interoperability between population-based registries and can support domain-specific data dictionaries (39).
   c. The Trusted Exchange Framework and Common Agreement (TEFCA) is a set of principles, technical requirements, and policies that support a nationwide system for securely sharing interoperable electronic health information (40). PD registries may be an ideal public health use case for the TEFCA network.

Conclusion

With the advent of statewide PD registries, we believe now is the time to (re)address scope, design, implementation, validation, and interoperability issues. We call on PD registry owners and stakeholders to consider these gaps and recommendations as we work toward a feasible framework for a truly inclusive population-wide PD registry that serves as a trusted resource for public health, clinical care, and research.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by UCLA IRB. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

ADW and AMW conceived of the project, contributed to summary of work presented, and finalized manuscript. All authors contributed to the article and approved the submitted version.

Funding

Supported by The Michael J. Fox Foundation for Parkinson's Research (MJFF-010561) to ADW and AMW; Centers for Disease Control (CDC) Intergovernmental Personnel agreements (IPA) for ADW and AMW. ADW has received honoraria from the American

Academy of Neurology. The opinions expressed here are those of the authors based on their experiences with clinical registries.

Acknowledgments

We thank the UCE-PD advisory committee including Beate Ritz, Caroline Tanner, Lawrence Dardick, Keith Garb, Michael and Nessa Weinman, Kristin Mendenhall, Mark Siegel. We acknowledge the UCLA UCE-PD team for work summarized in this manuscript. We thank the California Parkinson's Disease Registry (CPDR) for helpful discussions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2023.1149154/full#supplementary-material.

References

1. Ray Dorsey E, Elbaz A, Nichols E, Abd-Allah F, Abdelalim A, Adsuar JC, et al. Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the global burden of disease study 2016. *Lancet Neurol.* (2018) 17(11):939–53. doi: 10.1016/S1474-4422(18)30295-3

2. Marras C, Beck JC, Bower JH, Roberts E, Ritz B, Ross GW, et al. Prevalence of Parkinson's disease across North America. *NPJ Parkinsons Dis.* (2018) 4(1):1–7. doi: 10.1038/s41531-018-0058-0

3. Mantri S, Fullard ME, Beck J, Willis AW. State-level prevalence, health service use, and spending vary widely among medicare beneficiaries with Parkinson disease. *NPJ Parkinsons Dis.* (2019) 5(1):1–9. doi: 10.1038/s41531-019-0074-8

4. Goldman SM. Environmental toxins and Parkinson's disease. *Annu Rev Pharmacol Toxicol.* (2014) 54(1):141–64. doi: 10.1146/annurev-pharmtox-011613-135937

5. Ascherio A, Schwarzschild MA. The epidemiology of Parkinson's disease: risk factors and prevention Vol. 15 (2016). Available at: https://www.thelancet.com/journals/laneur/home.

6. Chuang YH, Paul KC, Sinsheimer JS, Bronstein JM, Bordelon YM, Ritz B. Genetic variants in nicotinic receptors and smoking cessation in Parkinson's disease. *Park Relat Disord.* (2019) 62:57–61. doi: 10.1016/j.parkreldis.2019.01.031

7. Strickland D, Bertoni JM. Parkinson's prevalence estimated by a state registry. *Mov Disord.* (2004) 19(3):318–23. doi: 10.1002/mds.10619

8. Hampton T. Parkinson disease registry launched. *J Am Med Assoc.* (2005) 293 (2):149. doi: 10.1001/jama.293.2.149

9. California Department of Public Health. The California Parkinson's disease registry implementation guide version 3.0. California Department of Public Health; 2018 (Accessed October 13, 2018).

10. Bertoni JM, Sprenkle PM, Strickland D, Noedel N. Evaluation of Parkinson's disease in entrants on the Nebraska state Parkinson's disease registry. *Mov Disord.* (2006) 21(10):1623–6. doi: 10.1002/mds.21026

11. Kim HM, Leverenz JB, Burdick DJ, Srivatsal S, Pate J, Hu S-C, et al. Diagnostic validation for participants in the Washington state Parkinson disease registry. *Parkinsons Dis.* (2018) 2018:1–6. doi: 10.1155/2018/3719578

12. Centers for Disease Control. Interim report to congress on national neurological conditions surveillance system (2020). Available at: https://www.cdc.gov/surveillance/neurology/congress_report.html (Accessed January 18, 2023).

13. California Department of Public Health. California Parkinson's disease registry program summary (2022). Available at: www.cdph.ca.gov/parkinsons (Accessed October 5, 2022).

14. Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord.* (2015) 30 (12):1591–601. doi: 10.1002/mds.26424

15. Berg D, Postuma RB, Adler CH, Bloem BR, Chan P, Dubois B, et al. MDS research criteria for prodromal Parkinson's disease. *Mov Disord.* (2015) 30:1600–11. doi: 10.1002/mds.26431

16. Armstrong MJ, Okun MS. Diagnosis and treatment of Parkinson disease: a review. *J Am Med Assoc.* (2020) 323(6):548–60. doi: 10.1001/jama.2019.22360

17. Harding Z, Wilkinson T, Stevenson A, Horrocks S, Ly A, Schnier C, et al. Identifying Parkinson's disease and parkinsonism cases using routinely collected healthcare data: a systematic review. *PLoS One.* (2019) 14:1–19. doi: 10.1371/journal.pone.0198736

18. Butt DA, Tu K, Young J, Green D, Wang M, Ivers N, et al. A validation study of administrative data algorithms to identify patients with parkinsonism with prevalence and incidence trends. *Neuroepidemiology.* (2014) 43(1):28–37. doi: 10.1159/000365590

19. Swarztrauber K, Anau J, Peters D. Identifying and distinguishing cases of parkinsonism and Parkinson's disease using ICD-9 CM codes and pharmacy data. *Mov Disord.* (2005) 20:964–70. doi: 10.1002/mds.20479

20. Szumski NR, Cheng EM. Optimizing algorithms to identify Parkinson's disease cases within an administrative database. *Mov Disord.* (2009) 24(1):51–6. doi: 10.1002/mds.22283

21. Wei WQ, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. *J Am Med Informatics Assoc.* (2016) 23 (e1):20–7. doi: 10.1093/jamia/ocv130

22. Wermuth L, Funch Lassen C, Himmerslev L, Olsen J, Ritz B. Validation of hospital register-based diagnosis of Parkinson's disease. Available at: https://ugeskriftet.dk/dmj.

23. Wermuth L, Cui X, Greene N, Schernhammer E, Ritz B. Medical record review to differentiate between idiopathic Parkinson's disease and parkinsonism: a danish record linkage study with 10 years of follow-up. *Parkinsons Dis.* (2015) 2015:1–9. doi: 10.1155/2015/781479

24. Brandt PS, Kho A, Luo Y, Pacheco JA, Walunas TL, Hakonarson H, et al. Characterizing variability of electronic health record-driven phenotype definitions. *J Am Med Informatics Assoc.* (2022) 30(3):427–37. doi: 10.1093/jamia/ocac235

25. California Department of Public Health. The California Parkinson's disease registry implementation guide version 1.1 (Accessed June 18, 2018).

26. Porter KR, Chao C, Quinn VP, Hsu J-WY, Jacobsen SJ. Variability in date of prostate cancer diagnosis: a comparison of cancer registry, pathology report, and electronic health data sources. *Ann Epidemiol.* (2014) 24(11):855–60. doi: 10.1016/j.annepidem.2014.09.004

27. Franklin PD, Lurie J, Tosteson TD, Tosteson ANA. Integration of registries with EHRs to accelerate generation of real-world evidence for clinical practice and learning health systems research: recommendations from a workshop on registry best practices. *J Bone Joint Surg Am.* (2020) 102(19):e110. doi: 10.2106/JBJS.19.01464

28. German RR, Wike JM, Bauer KR, Fleming ST, Trentham-Dietz A, Namiak M, et al. Quality of cancer registry data: findings from CDC-NPCR's breast and prostate cancer data quality and patterns of care study. *J Regist Manag.* (2011) 2(38):75–86. PMID: 22096878.

29. CDC/NCCDPHP. National program of cancer registries (NPCR). Available at: https://health.gov/healthypeople/objectives-and-data/data-sources-and-methods/data-sources/national-program-cancer-registries-npcr (Accessed January 18, 2023).

30. Parashos SA, Bloem BR, Browner NM, Giladi N, Gurevich T, Hausdorff JM, et al. What predicts falls in Parkinson disease?: observations from the Parkinson's foundation registry. *Neurol Clin Pract.* (2018) 8(3):214–22. doi: 10.1212/CPJ.0000000000000461

31. The office of the national coordinator for health information technology. 2022 Interoperability standards advisory reference edition. Available from: https://www.healthit.gov/isa/sites/isa/files/inline-files/2022-ISA-Reference-Edition.pdf (Accessed January 17, 2023).

32. National uniform claim committee - provider taxonomy. Available at: https://www.nucc.org/index.php/code-sets-mainmenu-41/provider-taxonomy-mainmenu-40 (Accessed January 17, 2023).

33. HL7 CDA® R2 Implementation guide: public health case report, release 2 STU release 1.1-US realm the electronic initial case report (eICR). 2017. Available at: http://www.ama-assn.org/ama/pub/physician-.

34. CDC. eCR now: COVID-19 electronic case reporting for healthcare providers (2019). Available at: https://www.cdc.gov/coronavirus/2019-ncov/hcp/electronic-case-reporting.html.

35. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. Comment: the FAIR guiding principles for scientific data management and stewardship. *Sci Data.* (2016) 3:160018. doi: 10.1038/sdata.2016.18

36. Tangka FKL, Subramanian S, Beebe MC, Weir HK, Trebino D, Babcock F, et al. Cost of operating central cancer registries and factors that affect cost: findings from an economic evaluation of centers for disease control and prevention national program of cancer registries. *J Public Heal Manag Pract.* (2016) 22(5):452–60. doi: 10.0.4.73/phh.0000000000000349

37. 2023 Promoting interoperability performance. Available at: https://qpp.cms.gov/mips/promoting-interoperability?py=2023 (Accsseed January 20, 2023).

38. HealthIT.gov. USCDI+. Available at: https://www.healthit.gov/topic/interoperability/uscdi-plus (Accessed January 20, 2023).

39. Nicholson N, Perego A. Interoperability of population-based patient registries. *J Biomed Inform.* (2020) 112:100074. doi: 10.1016/j.yjbinx.2020.100074

40. Office of the National Coordinator for Health Information Technology. The trusted exchange framework (TEF): principles for trusted exchange (2022). Available at: https://www.healthit.gov/sites/default/files/draft-trusted-exchange-framework.pdf.

frontiers | Frontiers in Digital Health

TYPE Original Research
PUBLISHED 08 June 2023
DOI 10.3389/fdgth.2023.1091508

OPEN ACCESS

EDITED BY
Benjamin Kummer,
Icahn School of Medicine at Mount Sinai, United States

REVIEWED BY
Satya S. Sahoo,
Case Western Reserve University, United States
Michael Purcaro,
Worcester Foundation for Biomedical Research, United States

*CORRESPONDENCE
Biswajit Maharathi
bmahar2@uic.edu

RECEIVED 07 November 2022
ACCEPTED 23 May 2023
PUBLISHED 08 June 2023

CITATION
Maharathi B, Mir F, Hosur K and Loeb JA (2023) INTUITION: a data platform to integrate human epilepsy clinical care and support for discovery. Front. Digit. Health 5:1091508. doi: 10.3389/fdgth.2023.1091508

INTUITION: a data platform to integrate human epilepsy clinical care and support for discovery

Biswajit Maharathi[1,2]*, Fozia Mir[1], Karthik Hosur[1] and Jeffrey A. Loeb[1,2]

[1]Department of Neurology and Rehabilitation, University of Illinois, Chicago, IL, United States, [2]Center for Clinical and Translational Science, University of Illinois at Chicago, Chicago, IL, United States

To make appropriate clinical decisions, clinicians consider many types of data from multiple sources to arrive at a diagnosis and plan. However, the current health systems have siloed data, making it challenging to develop information platforms that integrate this process into a single place for comprehensive clinical evaluation and research. INTUITION is a human brain integrative data system that facilitates multimodal data integration, unified storage, cohort selection, and analysis of multidisciplinary datasets. In this article, we describe the use of INTUITION to include electronic health records together with co-registered neuroimaging and EEG from patients who undergo invasive brain surgery for epilepsy. In addition to providing clinically useful visualizations and analytics to help guide surgical planning, INTUITION also links a bank of human brain epileptic tissues from specific brain locations to quantitative EEG, imaging, histology, and omics studies in a unique, completely integrated informatics platform. Having a clinically useful platform for integrating multimodal datasets can not only aid in clinical management decisions but also in creating a unique resource for research and discovery when linked to spatially mapped tissue samples.

KEYWORDS

epilepsy, systems biology, integrated informatics, neurology, intuition

Introduction

Epilepsy: a challenging case with enormous potential

Approximately one-third of all epilepsy patients are resistant to anti-epileptic drugs (1). Some of these patients benefit from surgical resection of the epileptic brain tissue to become seizure-free. The decision to surgically resect brain tissue requires accurate localization of seizure onset. The tissue localization process requires integrated evaluation of multimodal data derived from Electronic Health Record (EHR), spatially co-registered electrodes to the brain surface, evaluation of intracranial Electroencephalographs (EEGs), imaging, neuropsychiatric tests, and lab testing. The tissue is then removed following detailed group meeting discussions to assess outcome (seizure freedom or reduction). Tissue that is precisely mapped to the underlying electrical signals offers a unique opportunity to explore the causes of epilepsy and develop new treatments (2). Until recently, major limitations in maximizing the research utility of tissue removed from epilepsy surgery patients have been finding ways to link different data modalities, establishing streamlined data processing pipelines, and enabling integrated informatics. The situation is compounded since multidisciplinary data types needed for clinical care and interdisciplinary research are siloed on different computer systems. This requires data aggregation, curation, quality control, inventory management for different data coming

from various sources through different data collection protocols, and robust governance that takes care of security, compliance, and data access based on the specific needs of researchers.

The removal of human brain tissue to treat or cure epileptic disorders offers an exceptional research opportunity that is not possible in most other human brain disorders. Critical to the success of this research is precise co-registration of tissue with both the location of intracranial recording electrodes and multimodal imaging as described in a recent review (2). This has allowed us to study the cellular (histological) and omics (genomic, proteomic, metabolomics) correlates of specific physiological and anatomical measures from EEGs and imaging studies of multiple epileptic brain regions. Without this localization, tissue removed has limited value since there is significant heterogeneity in the electrical signals in different brain regions determined from intracranial recordings. The integrated understanding of electrophysiology, neuroimaging, histology, omics information, along with patient history, can elucidate the complex mechanism of epilepsy and significantly advance the field.

Creating a multidimensional database of the human epileptic brain

Several investigator-led initiatives have created comprehensive neurological disease-related databases for the past few decades to enhance research and knowledge discovery. Such databases store large, curated datasets, including EEG, imaging, genomics, and clinical details. Epilepsy is a common neurological condition of recurrent seizures, where many diverse types of data are used to evaluate and surgically treat patients who fail to respond to medical management. Platforms such as IEEG.ORG, EPILEPSIAE (3), and Temple EEG Database (4) store EEG datasets along with clinical and imaging metadata specifically for epilepsy. There are also databases that store more specific EEG datasets like the neonatal EEG database (5). On the other hand, databases like EpimiRBase (6) store epilepsy related microRNA datasets, and several imaging databases store imaging modality specific high resolution brain scans for humans (7) and animals along with genomic information (8). There are additional disease and condition specific initiatives such as LONI (9) which also hold both EEG and imaging information along with clinical details for traumatic brain injury and epilepsy, and our own PTRD database (10) that holds clinical and preclinical data on subarachnoid hemorrhage that includes electronic health records, imaging, EEG, and derived research information along with intuitive visualizations on patients and animal models.

To date, there has not been a comprehensive system that collects all the raw data and expands the scope of the system to accommodate datasets generated through basic, clinical, and translational research. Given the wealth of data and tissues produced during the clinical workup for epilepsy surgery, we have developed a system that collects clinical epilepsy data and spatially registers all EEG data onto brain imaging, allowing for precise spatial mapping of resected tissue samples at electrode positions (**Figure 1**). This data platform we call 'INTUTION' has enabled multimodal research studies on the tissues and integrated datasets, to develop a better understanding of the underlying disease and drive better treatment plans (2, 11, 12).

Methods

INTUITION enables interdisciplinary collaborative studies. Along with the raw datasets from several clinical systems, INTUITION supports asynchronous data processing tools that help to de-identify EEG and brain imaging, post-process EEG for epileptic spike-seizure detection, create 3D brain models from imaging data, and facilitate genomics analysis. The system enables semi-automatic electrode co-registration of EEG electrode positions on 3D brain models and registers resected brain tissue samples to the 3D surface. INTUITION also supports inventory management for omics samples and maintains the scanned histology images at different scanning resolutions. Such a well-curated and controlled data set not only helps in surgical planning but also helps researchers to understand and develop new treatments for epilepsy (**Figure 1**). Here we describe our integrated informatics approach enabled by INTUITION that orchestrates several data pipelines to optimize clinical decision making in epilepsy and supports translational research.

Patient population

INTUITION manages multimodal information from over two hundred epilepsy patients who have undergone two-stage epilepsy surgery in the past 20 years. Following patient informed consent, INTUITION initially collects all raw data, including identifiable features. Data de-identification modules remove all identifiable features when the data is moved from the clinical systems to the research environment. While all the patients had surgery for epilepsy, many of the patients had additional conditions, including polymicrogyria, tuberous sclerosis, focal cortical dysplasia, hippocampal sclerosis, brain injuries, and tumors, further enhancing the value of the tissue/data collection. We further link this patient data to a tissue bank containing over one thousand pieces of human brain tissues precisely mapped to each recording electrode. The patient population ranges from pediatric cases (as early as six months old) to older adults (above 50 years). With ongoing data collection and plans to use the platform at multiple sites, the platform aims to collect richer datasets from an even more diverse patient population.

Patient data collection

At the University of Illinois at Chicago, the clinical care of epilepsy patients generates a significant amount of data, a copy of which is stored in the clinical research data warehouse (CRDW). INTUITION has an established data request with the CRDW to prospectively extract new epilepsy patient data to a REDCap project dedicated to INTUITION. INTUITON, through

10.3389/fdgth.2023.1091508

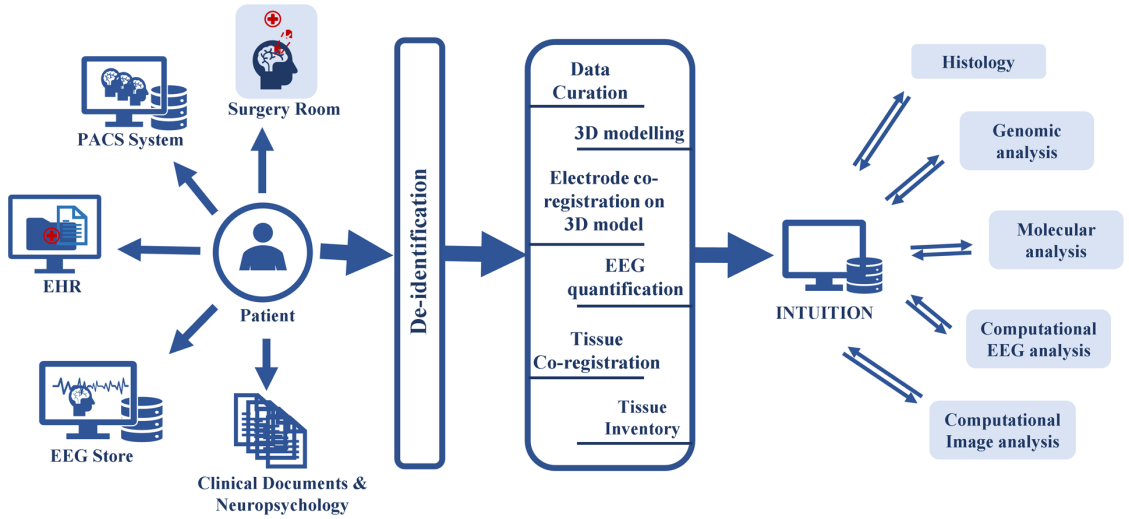


FIGURE 1
INTUITION workflow depicting upstream and downstream data flow. INTUITION has two distinct data flows. In the upstream data flow, INTUITION receives patient demographic information (EHR), brain imaging (PACS system), EEG and other electrophysiology data (EEG Store), tissue samples, intraoperative photographs (surgery room), other clinical notes, reports, neuropsychology notes, surgical evaluations, and physician observations (clinical documents not covered anywhere else). It then de-identifies the information and passes to an intermediate stage, where the data is curated, and tissue samples are co-registered, tagged, and updated in the inventory. All verified data is then entered into the database for storage. This dataset is further filtered depending on the scientific question and cohort selection and downloaded for offline processing. The offline dataset is used to evaluate the cellular structure, genomic profiling, quantitative image analysis, and quantitative EEG analysis. Single modality studies or multimodal integrative studies are performed depending on the scientific question. The processed data is then uploaded back to the platform for future use.

its automated data pipelines, downloads, transforms, de-identifies, and loads patient data to the database. Research users download the radiology images manually through a local PACS reviewer workstation and EEG data from the EEG lab in the hospital. The data details are as below.

Clinical information

We populate de-identified clinical information, including demographics, clinical diagnoses, procedures, unstructured clinical notes (blobs), and reports of imaging and neuropsychological evaluation, into the system through REDCap APIs. Research coordinators manually enter the surgical outcomes, seizure observations, intraoperative surgery data, and copies of the original report files. The research team and physician always validate the information. Based on the different study designs and needs, we expand the scope of clinical information with additional information.

EEG datasets

Every patient in INTUITION had scalp EEG recordings and the implantation of intracranial electrodes at precise brain regions using long-term video-EEG recording sessions. While storing the raw EEG data sets, the data size can expand to approximately 20 GB for one-day high-density EEG recording (124 channel, 1000 Hz sampling rate). If the data includes video files, it can expand to 40–60 GB per patient daily for video-EEG. During the entire recording period, this data can be up to several terabytes (TBs) (Table 1). For each patient, at least 3×10 min of EEG segments are collected from the intracranial EEG, which includes interictal activity. Location and EEG of seizures are also collected. The data is always reviewed and extracted under physician supervision to maintain the data quality. For each patient, we use a variety of algorithms to measure interictal epileptic waveforms and seizures and link these to precise locations using offline processing tools as described below.

Multimodal imaging

Imaging modalities are critical for localizing the seizure foci and related lesions. When combined with the intracranial EEG studies, they provide the spatial framework for designing therapeutic brain resections. Brain imaging includes multiple magnetic resonance imaging (MRI) sequences, computed tomography, x-Rays, positron emission tomography, single photon emission ictal scans, magnetoencephalography, and event-related optical imaging. While not all the recordings or scans are performed on each patient, usually one or more tests and scans are performed based on the need. For each patient, pre-electrode implantation, post-electrode implantation and post-operative imaging scans are collected, which are needed for the co-registration of EEG electrodes on a multitude of imaging studies. INTUITION de-identifies and stores raw images used to understand brain structure, volumetric information, structural connectivity, tissue co-registration, identify brain structures, highlight abnormalities and perform computational image analysis (13–15).

TABLE 1 Data elements with approximate data size per patient.

Data domain	Data module	Data type	Data records	Data records
Clinical	Demographics	tabular (numeric/string)	1	KB
	Diagnosis	tabular(string)	~10–100s	KB
	Procedures	tabular(string)	~10–100s	KB
	Medications	tabular(numeric/string)	~100s	KB
	Labs	tabular(numeric/string)	~1000s	KB
	Notes	tabular(text)	~100s	MB
	Ontology	tabular(string)	1	KB
Neuropsychology	All tests	JSON	~1–2	KB
Outcome	Outcome	tabular(string)	1	Bytes
EEG	Reports	tabular(text)	~1–10s	MB
	EEG files + meta info	file (.eeg/.edf) + tabular(string)	~3–10s	1 GB—2TBs
Radiology	Reports	tabular(text)	~1–30s	MB
	Imaging Files + meta info	file(DICOM RAW, NIFTI, DCM) + tabular(string)	~1–30s	~GBs
	3D electrode coregistration	file(DFS, OBJ) + tabular(string,C SV)	~1–3s	<1GB
Surgery	Surgical Data	(JPG) + tabular	~10s	~100MB
Tissue	Images andinventory	(JPG) + tabular	~1–10s	~100MB
Histology	Images	(jpeg2000/TIFF) + tabular	~1–10s	~GB—TBs
Studies/Omics	Tissue usage & results	Files, hyperlinks	~1–10000s	~TB—PB

Tissue samples

Tissues removed as part of the surgical procedure and not needed for diagnosis are used for research purposes and stored in our NeuroRepository following informed consent. These samples post-surgery is immediately mapped to the precise location of the brain using pictures in the surgery room and brain arterial patterns. Further, these samples are co-registered with EEG and MRI following standardized protocol (2). These stored samples are used for histological analysis, staining, genomics, proteomics, and metabolomics. In this way, each brain location with specific imaging features and corresponding electrical properties recorded from *in vivo* (e.g., spikes or seizures) can be linked to tissue histology and molecular/genetic attributes. INTUITION provides an inventory of the entire dataset, including tissue stored in refrigerators, sectioned tissue slides, RNA, DNA, and protein inventory (quantity remaining and storage location), and the links to EEG electrodes, EEG quantified results (spikes, seizure onset), MRI co-ordinates (for comparison with brain lesion locations). This information is used for cohort selection and research.

System design, infrastructure and management

INTUITION is developed on Django's model-view-template (MVT) architecture with PostgreSQL database, Python middleware, Django web framework, JINJA template engine, and HTML, CSS, and JavaScript for additional frontend work. An outline of the system architecture is shown in **Figure 2**. The system has a base application along with several service apps to support tools, viewers, and data operation (**Figure 2**). Some packages used for the application are Django Object Relationship Mapper (ORM) to interface with the database, Django-Migrate for database migrations, Django-REST for building RESTful APIs, Django-Cryptography for hashing utilities, Django-SimpleJWT to enable JSON Web Tokens (JWT) for API and other security features, Django-CORS for handling Cross Origin Resource Sharing (CORS), Allauth for handling authentication and authorization. We used the Requests library for making HTTP requests, Pandas for data transformations, Plotly for visualization, and CRISPY Forms for creating forms in the templates, which provides advantages of implementing bootstrap v5 forms while enforcing form functionalities like marking fields as required, hidden, dropdowns and multi-select checkbox. We secure the forms from cross-site request forgery using Django inbuilt functionality of Django CSRF, handling file uploads and JSONminify to reduce data transmission load between the client and server. We use the inbuilt Django-Storage library in the backend to create the file upload field and data storage management within the media files, which improves the efficiency of reading and writing files compared to using file storage in a database. We used Jinja2 templates, HTML (Bootstrap 5), and JavaScript to make the user interface seamless. We have followed Jakob Nielsen's ten general heuristics for interaction design to ensure our user interactions are seamless and on par with the global UI/UX standards (16). We also have an asset caching system that allows us to reduce latency between request and response by caching static resources in the user's browser. We also included features to bulk import and export CSV data from the system. We also use the PyDICOM library to process DICOM files and manipulate them to reduce payload for transmitting it to the client and pylibjpeg to manage image modifications and functionalities like contrast, brightness, and rotate images. We use the Django Simple History library to log all the data creation and modifications performed on Intuition. The Django-Authentication module manages the user access control. The library manages user authentication, session management, password hashing, password validation, and handling of password reset requests.

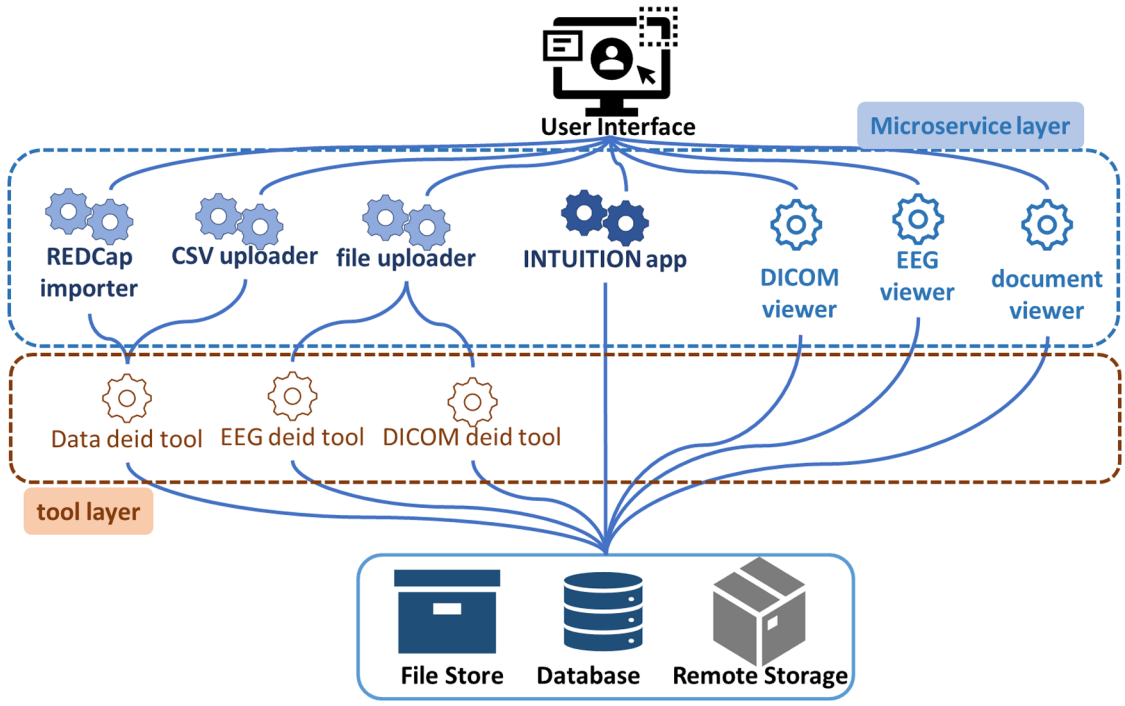


FIGURE 2
INTUITION system design.

We used PostgreSQL (v13) as our database because it is robust for online transaction processing (OLTP) and a competent analytical database that integrates well with Django. PostgreSQL also can manage high volumes of scientific data, is scalable, can be easily extended, and is freely available. The files for EEG, imaging, and histology module, along with any document, were in the Windows Encrypting File System (EFS); the index of these files was stored in a relational table with entity-attribute-value (EAV) data model. In the case of the EEG and Imaging dataset, where the uploaded file count for each patient is quite variable, and each entry can have associated support files (specifically for EEG), we have used a simple JSON field in corresponding tables to store names and associated files with file tags, file types, extensions, and associated user-created files.

The system's interactive search lets users search through any information available on INTUITION. The basic search enables users to enter keyword-based search terms parsed by the system to display a list of available patients. Additionally, an advanced search lets users select predefined search boxes for more complex patient selections.

INTUITION has a concept of PROJECT, which aggregates a set of patients for a particular research study and users who have access to that project enabling investigators to create study cohorts within the platform and share it with collaborators.

INTUITION is hosted on a Windows server (IIS version 10.0 on Windows 2022). INTUITION maintains end-to-end TLS using HTTPS protocol. Currently, the INTUITION application is deployed within the University of Illinois Chicago Private network for an added layer of security.

Offline data processing in INTUITION

Data de-identification

We have developed a set tool to de-identify each patient's imaging and EEG data.

DICOM de-identification: We have developed a web application tool that uses Pydicom to perform image modifications. The program works in two steps. First, it reads all the attributes of the DICOM header of the uploaded series. It deletes the header elements that are private and unknown attributes (VR or UN tag) as per the DICOM standard. It uses three sets of predefined attributes to perform de-identification. The patient name is updated with user-entered shifted data and patient code for the attributes marked as "Update Fields," such as study date. It will assign random string values to these attributes if these fields are not provided. Attributes defined as "Keep Fields" are kept as it is. These attributes are key to understanding the scan details, parameters, data orientation, and the data matrix. The "Remove Fields" attributes must be deleted if found. These fields correspond to accession numbers, machine serial numbers, hospital addresses, and provider details. In the second step, the tool displays the DICOM pixel data to the user to select regions needing removal from the images. The users can select one or more rectangular regions and whether they need to be deleted from one slice or all slices of the DICOM series. Once the selection is made, the tool removes the pixel values and regenerates the DICOM images in a de-identified format. Further, the data can be stored on INTUITION, or the user can download it to their local environment.

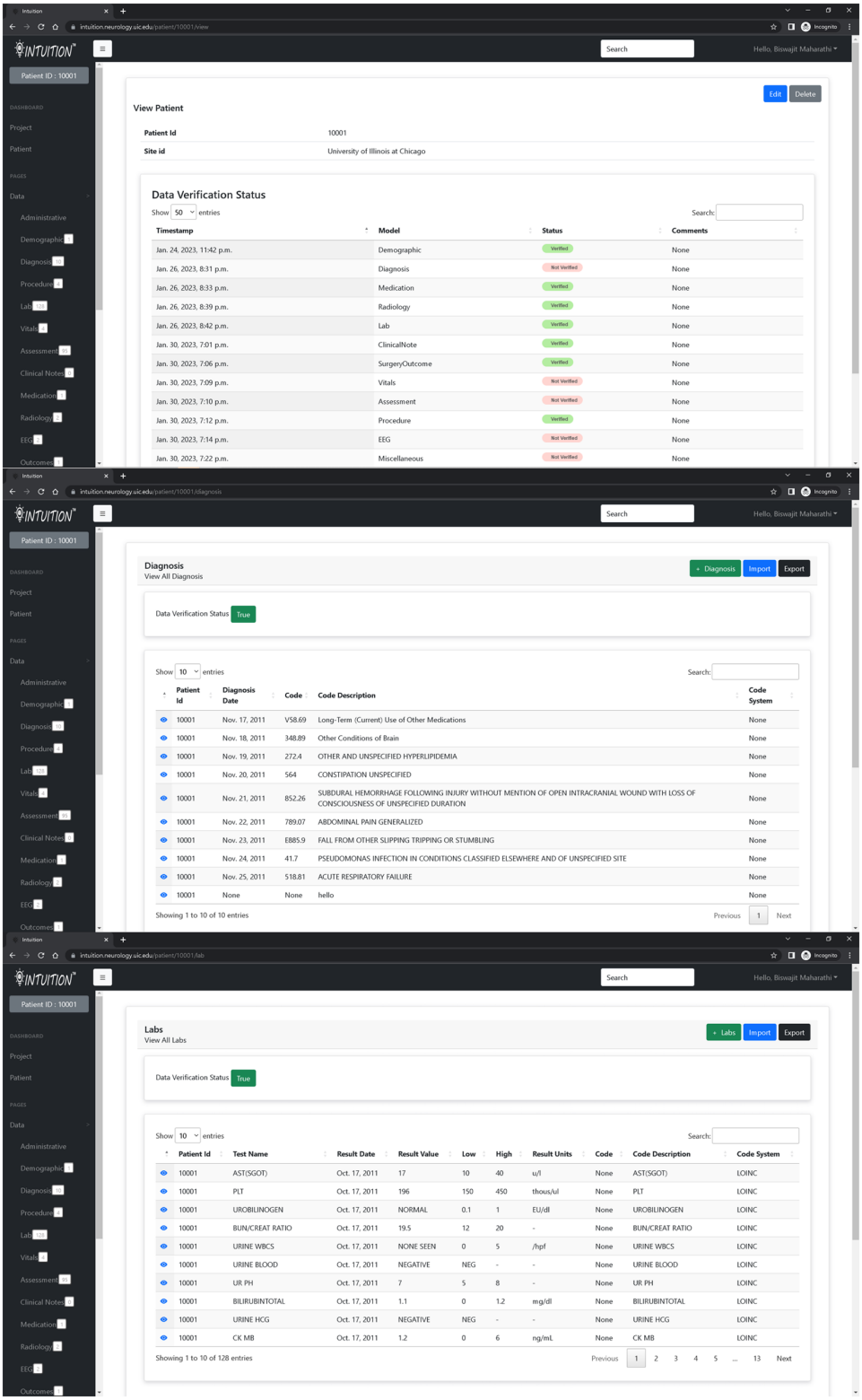


FIGURE 3
INTUITION pages displaying information extracted from electronic health records. The top figure shows the overall status of all modules whereas the sidebar status shows the records available within each data module. The middle and lower image shows sample of diagnosis and lab measurements for patient 10001.

10.3389/fdgth.2023.1091508

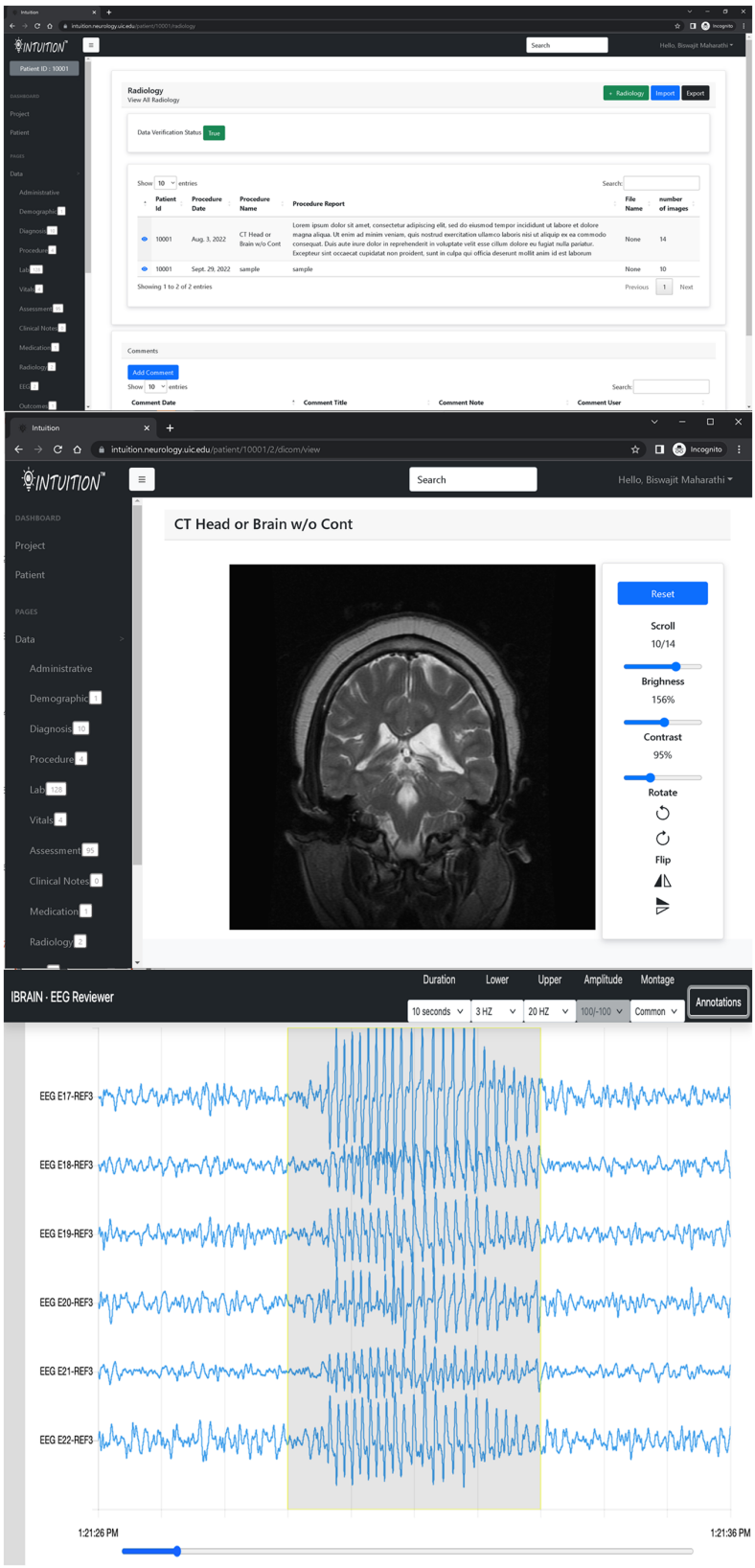


FIGURE 4
INTUITION imaging and EEG visualization tool. Top two panels display an MRI record with capability to review radiology notes and the DICOM image series on an inbuilt browser. The bottom panel shows the EEG browser with sample EEG data with a seizure event annotated.

EEG de-identification: The EEG is always extracted in European Data Format (EDF). The EEG de-identification tool reads the EEG file headers, removes, or modifies the fields that correspond to name and data attributes and rewrites the data back to EEG file in a fashion like the image de-identification. Further, we also remove all annotations from the file and either store it on INTUITION or let the user download them.

OMICS data de-identification: OMICS data de-identification is a complex process that involves ethical, legal, and technical challenges. To de-identify OMICS data, we follow several steps that involve technical, data governance, and sharing policy changes. The details of these steps are as follows. First, we remove personally identifiable information from the standardized omics files, such as patient names, IDs, age, gender, clinical conditions, and tissue location. Second, we manage the coded OMICS data on a separate server with limited connectivity to the INTUITION system. Access to this server requires additional privileges. Third, INTUITION holds derived information from genomic data, which has a minimal scope of re-identification. The genomic store is also encrypted to ensure added security. Fourth, complete genome records are never shared without appropriate request and approval from the data governance committee.

Fifth, data access is limited to requests with valid scientific purposes, and minimum data is shared based on the type and scope of the data request. Finally, upon a suitable agreement, we share an aggregated and coded dataset. The process of re-coding and aggregation reduces the risk of patient re-identification.

Overall, these steps aim to ensure that OMICS data is de-identified and shared in a secure and responsible manner.

EEG and MRI analysis

Our primary goal is to analyze EEG data to identify epileptic events such as interictal spikes and epileptic seizures, which are the electrophysiological biomarkers of epilepsy; co-register these event's occurrence, location, and propagation patterns (13, 17) across brain locations over 100 electrodes on a 3D brain model. Once the tissue is resected, perform additional co-registration of their locations on the same 3D model. In addition, we can examine many other EEG features from the raw data not limited to periodic discharges, frequency band-specific epilepsy signatures, and high-frequency oscillations. The computational work is performed offline using in-house algorithms and scripts developed in MATLAB and Python, and the processed information is passed back to INTUITION. While the initial design used MATLAB as the offline tool for signal processing, the data can be connected to other analytical platforms through appropriate APIs. In relation to the EEG work, we have developed an EEG de-identification tool, an EEG visualization tool.

For brain images, we use established tools like BrainSuite or *freesurfer* (18) for offline 3D re-constructions of the brain, cortical thickness, and geodesic distance measures. We use an in-house developed algorithm to use MRI and CT to co-register the electrode location and match it with intra-operative images for accuracy. This method provides an accurate localization of the electrodes and corresponding geodesic distance and cortical thickness. Furthermore, since many patients with epilepsy have brain lesions such as developmental abnormalities or tumors, our multimodal analytical framework also localizes these lesions in relation to the electrode locations. This helps us understand lesional and non-lesional epilepsy in a detailed manner. To manage brain imaging data, we have a DICOM-de-identification tool, DICOM visualization tool, and 3D model rendering tool that are web-based and integrated into INTUITION.

Histological and molecular/-omics analysis and image data storage

A unique characteristic of our approach is that each piece of resected brain tissue is precisely mapped to a specific brain electrode location. This enables a direct link between brain structure and electrophysiology at that specific region to the histology (or cell structure) and molecular features of the underlying tissue. Our previous work outlines the meticulous way we subdivide each block of tissue underlying a specific brain region for these analyses. High-resolution digital images of stained sections from each brain region, genomics (19–21), proteomics (21), and metabolomic (22) are stored within INTUITION and readily available for focused, discovery-based projects to understand and develop better treatments for epilepsy. While INTUITION stores the results of omics analysis on-premises, the raw files, often large (**Table 1**), are stored on the data server, and their links are stored on the INTUITION inventory module.

Results

The INTUITION application currently holds two hundred epilepsy patient data totaling 2 TB of on-premises storage and several terabytes of omics and imaging data on remote servers. The system supports a status dashboard that displays the status of data entries and whether a data reviewer has verified them. The navigation sidebar additionally provides information regarding the record counts for investigators to know how much data has been collected for each patient (**Figure 3**).

In addition, INTUITION has a patient timeline and data dashboard view (**Figure 4**) that has been tested for a previously published expansion on traumatic brain injury and epilepsy (10).

Search and cohort discovery

INTUITION provides a simple search interface that accepts text-based search criteria for demographics, diagnosis, and medications; and returns the qualifying list of patients. This list can be further aggregated, stored as a project, and can be shared with other investigators or reused at a later point. INTUITON also provides advanced search features that help in cohort discovery. The advanced search page takes information such as demographic filter (age, gender, date ranges), presence of any

diagnosis code (ICD codes or descriptive text), prescribed medication, procedures conducted, presence of specific lab measurements, and the range of values for that measurement, availability of the specific type of medical imaging scan along with the number of times the scans was performed (e.g., Pre and post-surgery MRI scans), availability of EEG, tissue inventory, and keyword-based search through available unstructured clinical notes.

Challenges ahead

We have created a multimodal database platform for uploading, storing, searching, and analyzing system biology datasets collected from two-stage epilepsy brain surgery patients. The platform INTUITION, named to mimic what a physician requires to make a diagnosis after absorbing many types of data, currently holds two hundred well-curated patient datasets along with location co-registered tissue samples used for histology and omics analysis of thousands of brain regions. While this form of 'Big Data' is not large compared to other datasets, the comprehensive, multimodal nature allows discoveries not possible with other, larger, but less integrated datasets. This is achieved through direct linkages between clinical, electrical, imaging, histological, and molecular data. Having both a clinical and research environment protects the patient's identity and enables the connection of all deidentified data modalities for the same patient. INTUITION and the datasets within INTUITION have led to important discoveries summarized in a recent review article (2). These discoveries range from improving our fundamental understanding of epilepsy to new therapeutics, new diagnostic approaches, and discoveries about what makes the human brain unique. The INTUITION system has fostered several patent applications, new drugs, and brain imaging strategies that would not have been possible without a highly curated, integrated, multimodal dataset of the human brain.

Moving forward, this platform has multiple challenges that are currently being addressed through further development of the INTUITION platform: (1) Finding ways to integrate the heterogeneous datasets for each patient and across all patients at multiple sites/surgical programs to create a secure, federated platform of deidentified data; (2) The search for epileptogenic biomarkers presents specific challenges on combined electrophysiology and imaging data and histology-omics data requiring new, offline workflow processes to create metadata sets that reduce the complexity for meaningful insight; (3) The platform is highly labor-intensive and requires a significant amount of manual data processing and entry; (4) Most datasets that feed into INTUITION reside on different, siloed servers that do not link to one another.

Conclusion

INTUITION is a unique data platform that brings together multiple data types within a focused human disease. Building this for patients who undergo epilepsy surgery provides some of the most detailed multimodal data that exists on the human brain linked to fresh human brain tissue samples. While not a large database, the carefully curated metadata for each patient offers an unparalleled opportunity to understand and develop novel diagnostic and treatment approaches for patients with epilepsy. Advances made in the building, expansion, and automation of INTUITION will further advance its utility and allow discoveries in many other human disorders.

Data availability statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author/s.

Ethics statement

Ethical review and approval was not required for this study in accordance with the local legislation and institutional requirements.

Author contributions

BM, JL, FM, and KS: contributed to the conception and design of the study. BM and KS: organized the database, developed the platform, and maintained it. FM: maintains the data. BM: wrote the first draft of the manuscript. JL, FM, and KS: wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The work is currently supported by grants from the NIH (Center for Clinical Translational Science) UL1TR002003, NIH R01 NS109515 (JAL) and Discovery Partners Institute (JAL).

Acknowledgment

We thank the many physicians, computer scientists, engineers, graduates and medical students at both Wayne State University and the University of Illinois at Chicago for their varied contributions. We also thank all our patients and their families, who have undergone invasive epilepsy surgeries, for allowing us to use their information and tissues to build INTUITION.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Picot M-C, Baldy-Moulinier M, Daurs J-P, Dujols P, Crespel A. The prevalence of epilepsy and pharmacoresistant epilepsy in adults: a population-based study in a western European country. *Epilepsia*. (2008) 49:1230–8. doi: 10.1111/j.1528-1167.2008.01579.x

2. Kirchner A, Dachet F, Loeb JA. Identifying targets for preventing epilepsy using systems biology of the human brain. *Neuropharmacology*. (2019):107757. doi: 10.1016/j.neuropharm.2019.107757

3. Ihle M, Feldwisch-Drentrup H, Teixeira CA, Witon A, Schelter B, Timmer J, et al. EPILEPSIAE—a European epilepsy database. *Comput Methods Programs Biomed*. (2012) 106:127–38. doi: 10.1016/j.cmpb.2010.08.011

4. Obeid I, Picone J. The temple university hospital EEG data corpus. *Front Neurosci*. (2016) 10:196. doi: 10.3389/fnins.2016.00196

5. Stevenson NJ, Tapani K, Lauronen L, Vanhatalo S. A dataset of neonatal EEG recordings with seizure annotations. *Sci Data*. (2019) 6:190039. doi: 10.1038/sdata.2019.39

6. Mooney C, Becker BA, Raoof R, Henshall DC. EpimiRBase: a comprehensive database of microRNA-epilepsy associations. *Bioinformatics*. (2016) 32:1436–8. doi: 10.1093/bioinformatics/btw008

7. Lüsebrink F, Sciarra A, Mattern H, Yakupov R, Speck O. T1-weighted in vivo human whole brain MRI dataset with an ultrahigh isotropic resolution of 250 µm. *Sci Data*. (2017) 4:170032.

8. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature*. (2007) 445:168–76. doi: 10.1038/nature05453

9. Duncan D, Vespa P, Pitkänen A, Braimah A, Lapinlampi N, Toga AW. Big data sharing and analysis to advance research in post-traumatic epilepsy. *Neurobiol Dis*. (2019) 123:127–36. doi: 10.1016/j.nbd.2018.05.026

10. Maharathi B, Wong J, Geraghty JR, Serafini A, Davis JM, Butler M, et al. Multi-modal data integration platform combining clinical and preclinical models of post subarachnoid hemorrhage epilepsy. *Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Int Conf*. (2022) 2022:3459–63. doi: 10.1109/EMBC48229.2022.9871864

11. Loeb JA. A human systems biology approach to discover new drug targets in epilepsy. *Epilepsia*. (2010) 51:171–7. doi: 10.1111/j.1528-1167.2010.02635.x

12. Loeb JA. Identifying targets for preventing epilepsy using systems biology. *Neurosci Lett*. (2011) 497:205–12. doi: 10.1016/j.neulet.2011.02.041

13. Maharathi B, Wlodarski R, Bagla S, Asano E, Hua J, Patton J, et al. Interictal spike connectivity in human epileptic neocortex. *Clin Neurophysiol*. (2018) 130:270–9. doi: 10.1016/j.clinph.2018.11.025

14. Maharathi B, Loeb JAJA, Patton J. *Estimation of resting state effective connectivity in epilepsy using direct-directed transfer function. in 2016 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Vols 2016-octob* 716–9 (IEEE, 2016).

15. Maharathi B, Loeb JA, Patton J. *Central sulcus is a barrier to causal propagation in epileptic networks. in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 2555–9 (IEEE, 2019). doi: 10.1109/EMBC.2019.8857401

16. Nielsen J. *Enhancing the explanatory power of usability heuristics. in Proceedings of the SIGCHI conference on human factors in computing systems celebrating interdependence—cHI '94* 152–8 (ACM Press, 1994). doi: 10.1145/191666.191729

17. Maharathi B, Patton J, Serafini A, Slavin K, Loeb JA. Highly consistent temporal lobe interictal spike networks revealed from foramen ovale electrodes. *Clin Neurophysiol*. (2021) 132:2065–74. doi: 10.1016/j.clinph.2021.06.013

18. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*. (1999) 9:179–94. doi: 10.1006/nimg.1998.0395

19. Beaumont TL, Yao B, Shah A, Kapatos G, Loeb JA. Layer-specific CREB target gene induction in human neocortical epilepsy. *J Neurosci*. (2012) 32:14389–14401a. doi: 10.1523/JNEUROSCI.3408-12.2012

20. Lipovich L, Dachet F, Cai J, Bagla S, Balan K, Jia H, et al. Activity-dependent human brain coding/noncoding gene regulatory networks. *Genetics*. (2012) 192:1133–48. doi: 10.1534/genetics.112.145128

21. Dachet F, Bagla S, Keren-Aviram G, Morton A, Balan K, Saadat L, et al. Predicting novel histopathological microlesions in human epileptic brain through transcriptional clustering. *Brain*. (2015) 138:356–70. doi: 10.1093/brain/awu350

22. Wu HC, Dachet F, Ghoddoussi F, Bagla S, Fuerst D, Stanley JA, et al. Altered metabolomic-genomic signature: a potential noninvasive biomarker of epilepsy. *Epilepsia*. (2017) 58:1626–36. doi: 10.1111/epi.13848

frontiers | Frontiers in Digital Health

TYPE Brief Research Report
PUBLISHED 13 June 2023
DOI 10.3389/fdgth.2023.1075771

OPEN ACCESS

EDITED BY
Lina F. Soualmia,
Université de Rouen, France

REVIEWED BY
Xia Jing,
Clemson University, United States
Alec Chapman,
The University of Utah, United States

*CORRESPONDENCE
Daniel B. Hier
✉ hierd@mst.edu

RECEIVED 20 October 2022
ACCEPTED 26 May 2023
PUBLISHED 13 June 2023

CITATION
Oommen C, Howlett-Prieto Q, Carrithers MD and Hier DB (2023) Inter-rater agreement for the annotation of neurologic signs and symptoms in electronic health records. Front. Digit. Health 5:1075771. doi: 10.3389/fdgth.2023.1075771

Inter-rater agreement for the annotation of neurologic signs and symptoms in electronic health records

Chelsea Oommen[1], Quentin Howlett-Prieto[1], Michael D. Carrithers[1] and Daniel B. Hier[1,2]*

[1]Department of Neurology and Rehabilitation, University of Illinois at Chicago, Chicago, IL, United States, [2]Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO, United States

The extraction of patient signs and symptoms recorded as free text in electronic health records is critical for precision medicine. Once extracted, signs and symptoms can be made computable by mapping to signs and symptoms in an ontology. Extracting signs and symptoms from free text is tedious and time-consuming. Prior studies have suggested that inter-rater agreement for clinical concept extraction is low. We have examined inter-rater agreement for annotating neurologic concepts in clinical notes from electronic health records. After training on the annotation process, the annotation tool, and the supporting neuro-ontology, three raters annotated 15 clinical notes in three rounds. Inter-rater agreement between the three annotators was high for text span and category label. A machine annotator based on a convolutional neural network had a high level of agreement with the human annotators but one that was lower than human inter-rater agreement. We conclude that high levels of agreement between human annotators are possible with appropriate training and annotation tools. Furthermore, more training examples combined with improvements in neural networks and natural language processing should make machine annotators capable of high throughput automated clinical concept extraction with high levels of agreement with human annotators.

KEYWORDS

natural language processing, annotation, electronic health records, phenotype, clinical concept extraction, inter-rater agreement, neural networks, signs and symptoms

Introduction

Extracting medical concepts from electronic health records is key to precision medicine (1). The signs and symptoms of patients (part of the patient phenotype) are generally recorded as free text in progress notes, admission notes, and discharge summaries (2). Clinical phenotyping of patients involves the mapping of free text to defined terms that are concepts in an ontology (3,4). This is a two-step process that involves identifying appropriate text spans in narratives and then converting the text spans to target concepts in an ontology (5,6). The process of mapping free text to defined classes in an

ontology, illustrated in (1) and (2), has been termed **normalization** (7,8).

$$\text{patient movements were } \textbf{ataxic} \Rightarrow \textbf{ataxia} \Rightarrow \text{UMLS CUI}: \textbf{C0004134} \quad (1)$$

$$\text{freetext} \Rightarrow \text{clinical concept} \Rightarrow \text{machine readable code} \quad (2)$$

In this example 1, an annotator highlights the term ataxic, then it is mapped to the concept ataxia, and the UMLS code CUI C0004134 is retrieved (9). This is a slow and error-prone process for human annotators. Agreement between human raters for annotation of clinical text is often low. A study on the agreement for SNOMED CT codes between coders from three professional coding companies yielded about 50 percent agreement for exact matches with slightly higher agreement when adjusted for near matches (10). Another study of SNOMED CT coding of ophthalmology notes yielded low levels of inter-rater agreement ranging from 33 to 64% (11). Identified sources of disagreement between coders included human errors (lack of applicable medical knowledge, lack of recognition of abbreviations for concepts, and general carelessness), annotation guideline flaws (under specified and unclear guidelines), ontology flaws (polysemy of coded concepts), interface term issues (inconsistent categorization of clinical jargon), and language issues (interpretation difficulties due to use of ellipsis, anaphora, paraphrasing, and other linguistic concepts) (12).

The goal of high throughput phenotyping is to use natural language processing (NLP) to automate the annotation process (13). Approaches to high throughput clinical concept extraction have included rule-based systems, traditional machine learning algorithms, deep learning algorithms, and hybrid methods that combine algorithms (6). Tools for concept extraction based on rules, linguistic analysis, and statistical models, such as cTAKES and MetaMap, generally have accuracy and recall between 0.38 and 0.66 (5,14,15). Neural networks are being used for concept recognition with increasing success. Arbabi et al. developed a convolutional neural network that matches input phrases to concepts in the Human Phenotype Ontology with high accuracy (16). Other deep learning approaches, including neural networks based on bidirectional encoder representations from transformers (BERT), show promise for automated clinical concept extraction (5,6,17,18).

In this paper, we examine inter-rater agreement for text-span identification of neurological concepts in notes from electronic health records. In addition to the agreement between human annotators, we examine the agreement between human annotators and a machine annotator based on a convolutional neural network.

Methods

Annotation tool

Prodigy (Explosion AI, Berlin, Germany) was used to annotate neurologic concepts in the EHR physician notes. Prodigy runs under python in the terminal mode of macOS, Windows, or Linux. It creates a web interface locally (**Figures 1A,B**). As input, Prodigy requires free text to be converted to JSON format.

$$\{\text{``text''}: \text{``The patient had } \textbf{weakness} \text{ and} \textbf{sensoryloss}''\} \quad (3)$$

Each line of text from a JSON file 3, appears as a separate screen for annotation by Prodigy (**Figures 1A,B**). Annotations are stored in an SQLite database and are exportable with annotations and text spans as a JSON file. Prodigy is integrated with the *spaCy* natural language processing toolkit (Explosion AI) and can train neural networks for named entity recognition and text classification.

The Kappa statistic was used to assess agreement between the three annotators and the neural network. The Kappa statistic corrects observed rater agreement for chance rater agreement. It ranges from 0 to 1, where 1 is complete agreement, 0 is a chance agreement. Values of Kappa of 0.6 to 0.79 are considered substantial agreement, values between 0.8 and 0.90 are considered strong agreement, and values over 0.90 are considered near perfect agreement (19,20). For each line of text that had one or more annotations (3), the agreement was rated 1 for the annotations if both annotators agreed and rated 0 if the annotators disagreed. A line of text with no annotations (null_annotations) by either annotator was scored 1 for agreement. The total number of annotations considered by the Kappa statistic for two raters **A** and **B** was $(A \cup B + \text{null\_annotations})$.

Rater training and instructions

Three annotators participated in the research. Annotator 1 (A1) was a senior neurologist, Annotator 2 (A2) was a pre-medical student majoring in neuroscience, and Annotator 3 (A3) was a third-year medical student. Raters first reviewed neurologic signs and symptoms in the neuro-ontology of neurological concepts (21) and then were instructed to find all neurological concepts in the neurology notes. Signs and symptoms (ataxia, fatigue, weakness, memory loss, etc.) were annotated but not disease entities (Alzheimer's disease, multiple sclerosis, etc.) Raters annotated the neurologic concepts and ignored laterality and other modifiers (e.g., *arm pain* for *right arm pain*, *back pain* for *severe back pain*, etc.) In addition, annotators tagged each text span with an category label (see **Figures 1A,B**). Category labels included *unigrams* (one-word concepts such as ataxia), *bigrams* (two-word concepts such as double vision), *trigrams* (three-word concepts such as low back pain), *tetragrams* (four-word concepts such as relative afferent pupil defect), *extended* (text span annotations longer than four words), *compound* (multiple concepts in one text span such as brisk ankle and knee reflex), and *tabular* (concepts represented in tabular or columnar format, usually showed right and left body sides). Our motivation for tagging signs and symptoms by the length and type of the text span was a hypothesis that neural networks trained to recognize signs and symptoms in medical text would

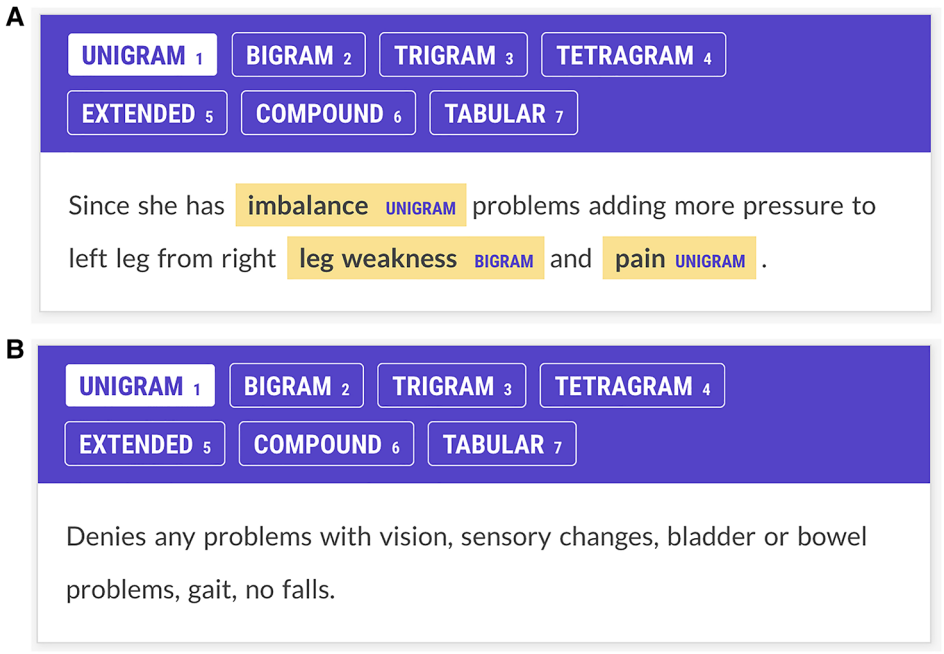


FIGURE 1
(A) Annotator screen for a patient with multiple sclerosis. The patient complains of imbalance, leg weakness, and pain, and these concepts have been annotated. Imbalance and pain are labeled as unigrams; leg weakness is labeled a bigram. Annotators were trained to ignore laterality (e.g., right leg weakness.) Each Prodigy screen reflects one line of text from the JSON input file. This screen has three potential items to contribute to the Kappa statistic: imbalance, leg weakness, and pain. (B) Annotator screen for neurological concepts for a patient with multiple sclerosis. The patient denies problems with vision, sensation, bladder, bowel, gait, or falls. The annotators are trained not to annotate negated concepts. The NN had no specific negation rule but learned not to tag negated concepts through training examples. Since there are no signs and symptoms in this screen, if both annotators show no annotations, a score of 1 is assigned to the Kappa statistic for agreement on this screen. If one annotator shows no annotations and another shows annotations on this screen, annotator disagreement is scored.

exhibit lower accuracies with longer text spans. This hypothesis was confirmed by a recent study from our group (18).

The machine annotator

The machine annotator (NN) was a neural network that was trained to recognize text spans containing neurology concepts in the electronic health record physician notes. The NN was the default spaCy named entity recognition model based on a four-layer convolutional neural network (CNN) that looked at four words on either side of each token using *tok2vec* with an initial learning rate 1×10^{-3}. The default parameters provided by Prodigy were used for training. NN was trained on 11,000 manually annotated sentences derived from neurology textbooks, online neurological disease descriptions, and electronic health record notes. Further details on training the NN are available in (18).

Annotations

Five patient EHR notes were annotated for each of the three rounds. The annotation of EHR clinical notes for research purposes was approved by the Institutional Review Board of the University of Illinois (UIC Neuroimmunology Biobank 2017-0520Z). Informed patient consent for use of clinical notes was obtained from all subjects through the UIC Biobank Project. Three human annotators (A1, A2, and A3) and the machine annotator (NN) annotated each note. After each round, the annotators met and reviewed any annotation disagreements. The annotations of each annotator were stored in an SQLite database and exported as a JSON file for scoring for inter-rater agreement in Python. Text spans were mapped to concepts in the neuro-ontology (21) utilizing a lookup table with 3,500 target phrases and the similarity method from spaCy (22) (pp. 152–54). Univariate analysis of variance and Cohen's Kappa statistic were calculated with SPSS (IBM, version 28).

Results

Annotators identified neurological signs and symptoms in physician notes from electronic health records. Each annotator identified the text span associated with each sign and symptom and assigned a category label to each annotation (e.g., unigram, bigram, trigram, etc.) Inter-rater agreement (adjusted and unadjusted) was calculated between the three human annotators and the machine annotator (NN).

Although five EHR notes were annotated for each round, the notes varied in length. Each line in the EHR note was converted to a single line in the JSON file and generated one annotation screen in the Prodigy annotator. Round 1 had 625 annotation

screens with 139 signs and symptoms to annotate, Round 2 had 674 annotation screens with 205 signs and symptoms to annotate, and Round 3 had 523 annotation screens with 138 signs and symptoms to annotate. Since the number of signs and symptoms was less than the number of annotation screens, many annotation screens had no signs or symptoms to annotate (null screens). When both annotators agreed that the annotation screen had no signs or symptoms, this was scored as annotator agreement for both the adjusted and unadjusted metrics (Kappa and concordance).

Concordance (unadjusted agreement) on the text span task was 88.9% ± 3.2 (mean ± SD) between the human annotators and was 83.9% ± 4.6 (mean ± SD) between the human annotators and the machine annotator (human-human mean was higher, one-way ANOVA, df = 1, $p = 0.016$). Concordance (unadjusted agreement) on the category label task was 87.7% ± 4.4 (mean ± SD) between human annotators and was 84.6% ± 5.5 (mean ± SD) between the human annotators and the machine annotator (means did not differ, one-way ANOVA, df = 1, $p = 0.212$).

Cohen's Kappa statistic (κ) was high for both the text span task (0.715 to 0.893) and the category label task (0.72 to 0.89) (**Figures 2A,B**). On the text span identification task (**Figure 3A**) κ was higher for the human-human pairs (0.85 ± 0.05 mean ± SD) than the human-machine pairs (0.76 ± 0.06). On the category label task, κ (**Figure 3B**) was similar between the human-human pairs (0.83 ± 0.05 mean ± SD) and the human-machine pairs (0.82 ± 0.06). κ for the text span task and the category label task did not differ by round (for p values and means see **Figures 4A,B**).

Discussion

Signs and symptoms are an important component of a patient's phenotype. Extracting these phenotypic features from electronic health records and converting them to machine-readable codes makes them computable (23). These computable phenotypes are

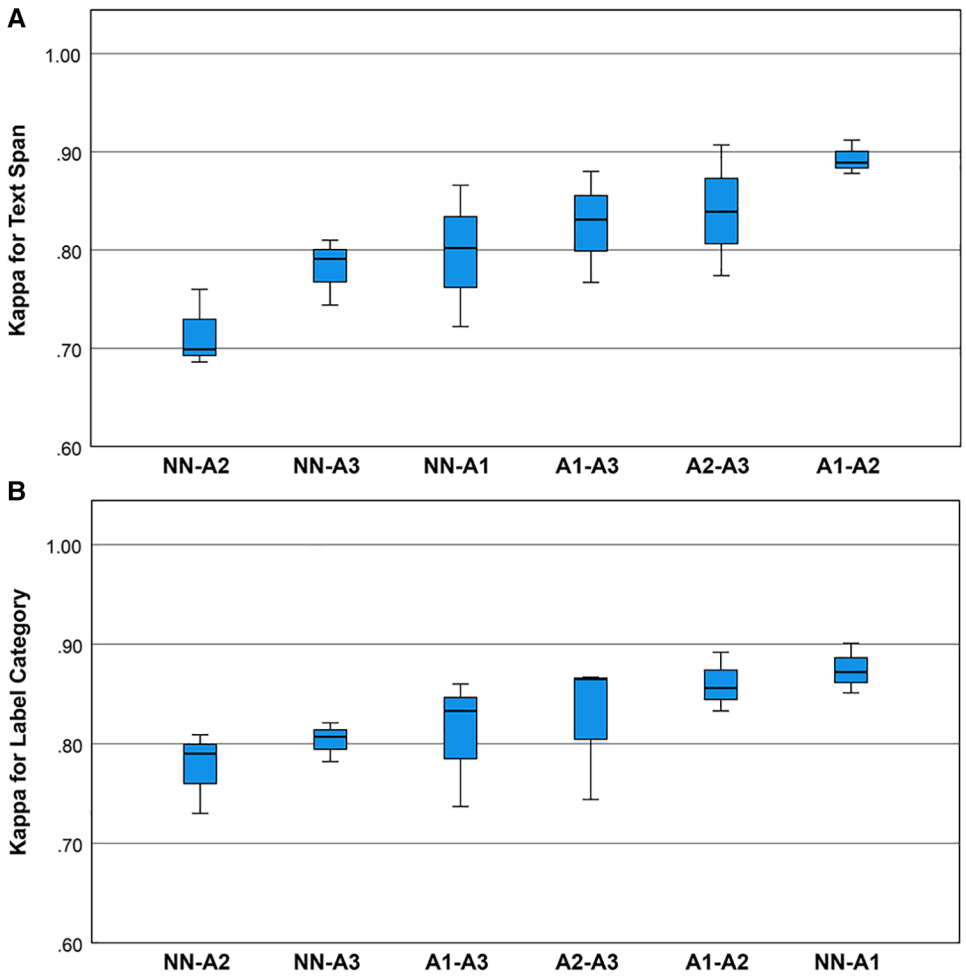


FIGURE 2
(A) Boxplots for the Kappa statistic for inter-rater agreement for text spans for the neurological concepts. Univariate analysis of variance showed that mean inter-rater agreement differed by rating pair (one-way ANOVA, df = 5, $p = 0.021$). Post hoc comparisons by the Bonferroni method showed that pair A1-A2 outperformed pair NN-A2. (B) Boxplots for the Kappa statistic for inter-rater agreement for category labels for the neurological concepts. Univariate analysis of variance showed that mean Kappa for category label agreement did not differ by rating pair (one-way ANOVA, $p = 0.165$, df = 5).

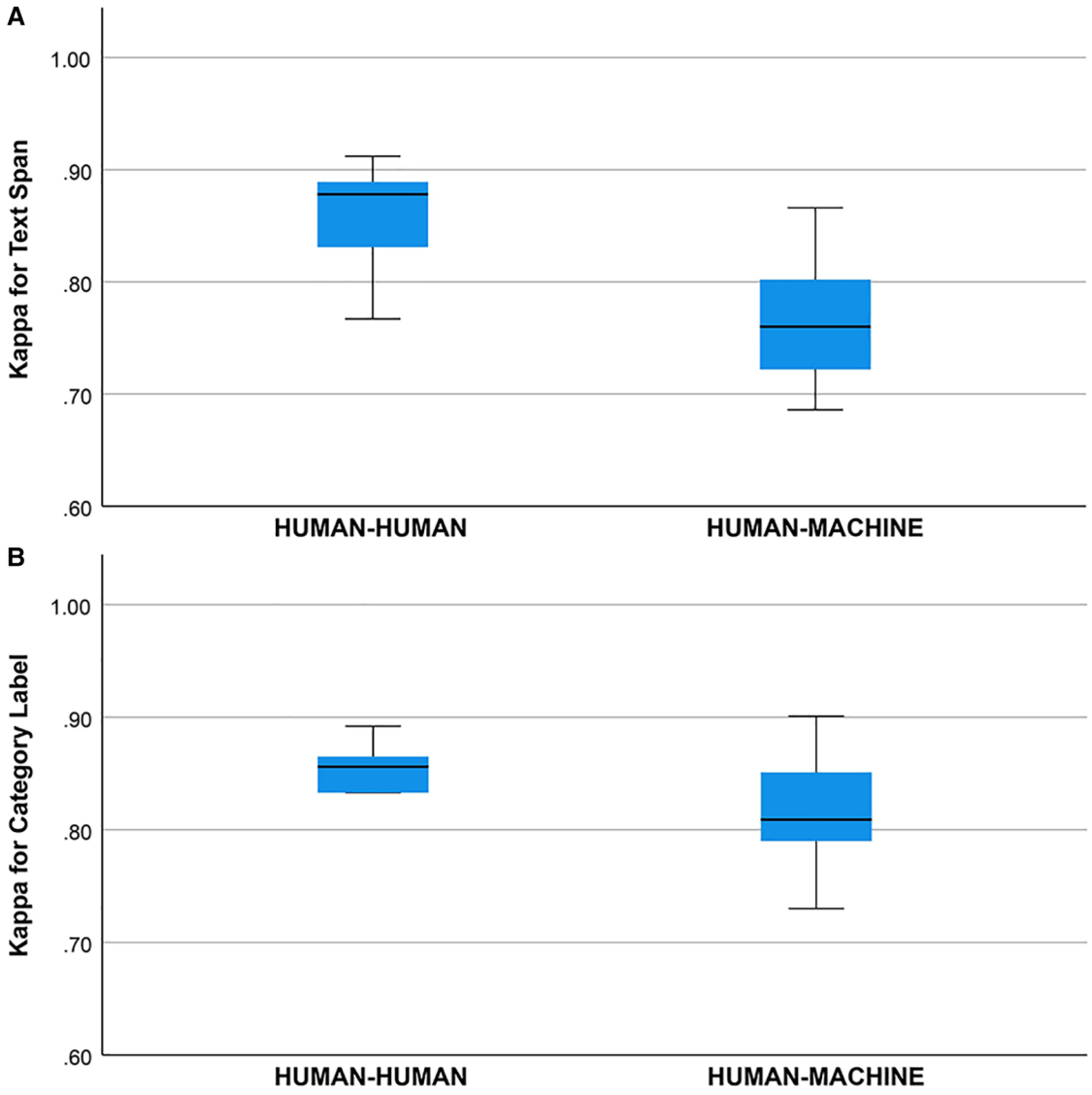


FIGURE 3
(A) Kappa statistic for agreement between human-human and human-machine raters for text span. Groups differed, one-way ANOVA, df = 1, $p = 0.004$. (B) Kappa statistic for agreement between human-human and human-machine raters for category label. Groups did not differ, one way ANOVA, df = 1, $p = 0.589$.

critical to precision medicine initiatives (24–26). Agrawal et al. (5) have conceptualized clinical entity extraction as a two-step process of text span recognition followed by clinical entity normalization. Text span recognition is the identification of signs and symptoms in the free text; entity normalization is the mapping of this text to canonical signs and symptoms in an ontology such as UMLS (9). We have focused on an inter-rater agreement for text span annotation. For entity normalization, we depended on a look-up table that mapped text spans to concepts in neuro-ontology. We found high inter-rater concordance (unadjusted agreement) among the human annotators (approximately 89%) with a lower concordance (unadjusted) agreement between the human annotators and the machine annotator (approximately 84%).

The concordance (unadjusted agreement) for category labels was lower than the inter-rater agreement for text spans which may have been due to factors such as the use of hyphens in the free text of the EHR notes and annotator uncertainty about which types of text spans required the tabular label. The Kappa statistic (adjusted agreement) for human-human raters was between 0.77 and 0.91, and the Kappa statistic for the human-machine agreement was between 0.69 and 0.87 (**Figure 3A**). We consider the inter-rater adjusted agreement between the human raters (0.77 to 0.91) good, especially when contrasted with the inter-rater adjusted agreement between trained neurologists eliciting patient signs and symptoms (27,28). For trained neurologists eliciting signs and symptoms such as weakness, sensory loss, ataxia, aphasia, dysarthria, and drowsiness, the κ statistic ranges from 0.40 to 0.70 (27,28).

The higher levels of agreement in this study may reflect that eliciting a sign or symptom from a patient is more difficult than

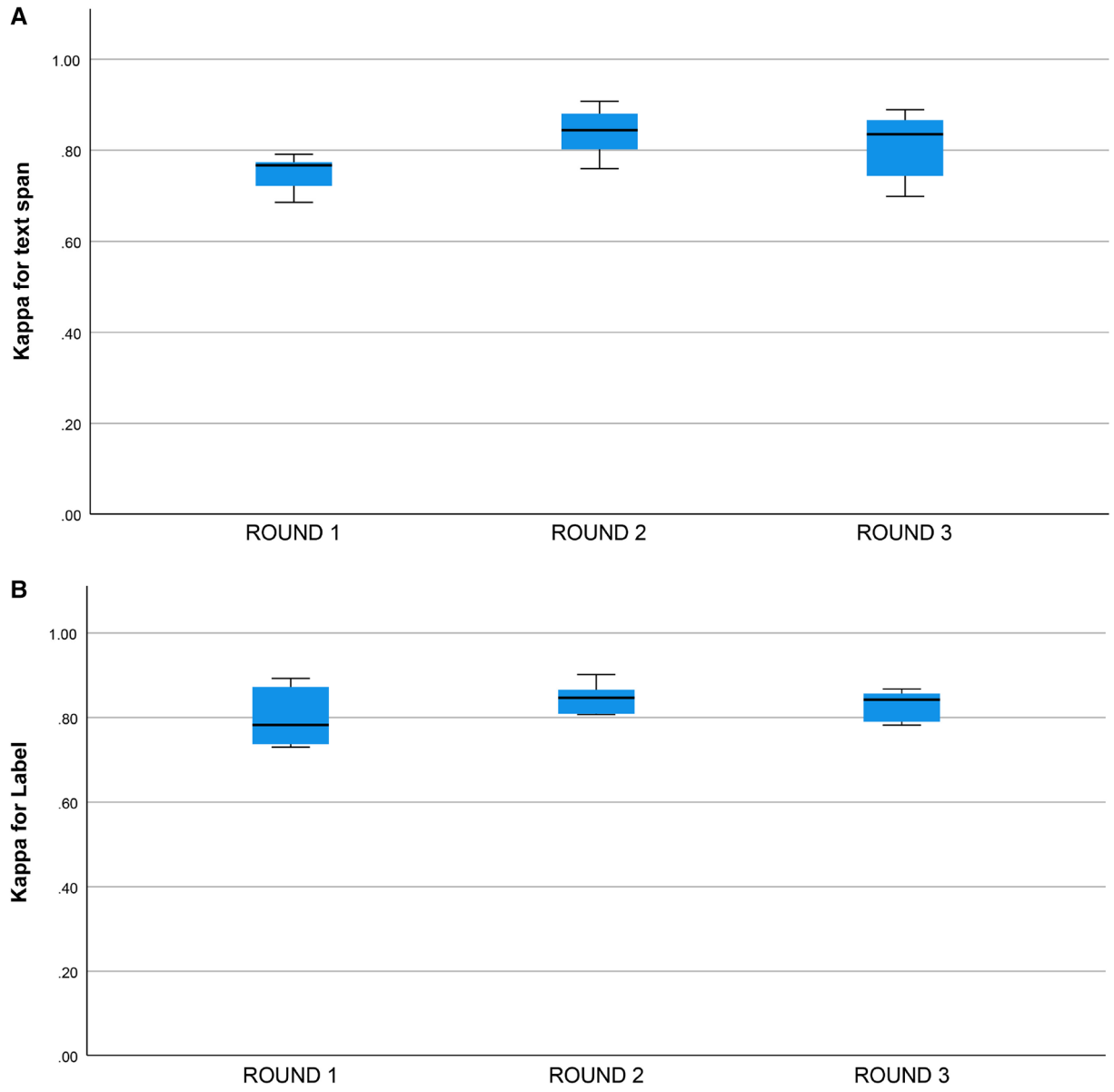


FIGURE 4
(A) Kappa statistic for inter-rater agreement for text span by round. Round 1: 0.78 ± 0.03 (mean ± SE), Round 2: 0.84 ± 0.03, Round 3: 0.81 ± 0.03, groups do not differ, one-way ANOVA, df = 2, $p = 0.310$. (B) Kappa statistic for inter-rater agreement for category label by round. Round 1: 0.80 ± 0.21 (mean ± SE). Round 2: 0.85 ± 0.21, Round 3: 0.83 ± 0.21, groups do not differ, one-way ANOVA, df = 2, $p = 0.306$.

annotating a sign or symptom in an EHR. Nonetheless, the adjusted agreement (κ) was higher in this study than in prior annotation studies (10,11), possibly reflecting the training of the annotators, the use of a neuro-ontology, the decision not to code severity or laterality of the symptoms, and the use of a sophisticated annotation tool.

We did not find a training effect for the human annotators across rounds (**Figures 4A,B**). Although the annotators met after each round and discussed discrepancies in their annotations, inter-rater adjusted and unadjusted agreement did not improve significantly between rounds. This suggests that there may be a ceiling for inter-rater agreement for text span annotation with a Kappa of 0.80 to 0.90 and that higher levels of agreement may not be possible due to the complexity of the task and random factors that are not addressable with additional training or experience. This ceiling effect for the human inter-rater agreement has implications for the potential for higher rates of inter-rater agreement between humans and machines (**Figure 3B**). Mean inter-rater adjusted agreement for text span was higher for the human-human pairs ($\kappa = 0.85$) than the human-machine pairs ($\kappa = 0.76$). Additional training examples would likely improve the performance of the machine annotator on the text span and category label tasks. Furthermore, other neural networks are likely to outperform the convolutional neural network (CNN), which is the baseline for Prodigy. We have found that a neural network based on bidirectional encoder

representations from transformers (BERT) can improve performance on the text span task by 5 to 10% (18). Others have found that deep learning approaches based on BERT outperform approaches based on CNN for concept identification and extraction tasks (17). A ceiling effect for inter-rater agreement for annotating signs and symptoms, whether human-human or human-machine, near a κ of 0.90 is likely.

Given the heavy documentation burden on physicians and physician burn-out attributed to electronic health records, physician documentation of signs and symptoms will likely continue as free text. Structured documentation of signs and symptoms as an alternative to free text is too burdensome in the current environment (29–34). A medium-sized medical center with a daily inpatient census of 300 and a daily outpatient census of 2,000 generates at least 5,000 clinical notes daily or over 1.5 million notes annually (unpublished estimates based on two academic medical centers). The sheer volume of clinical notes in electronic health records makes the manual annotation of signs and symptoms impractical. Extracting signs and symptoms for precision medicine initiatives will depend on advances in natural language processing and natural language understanding.

Although high throughput phenotyping of electronic health records by manual methods is impractical (13), the manual annotation of free text in electronic health records can be used to train neural networks for phenotyping. Neural networks can also speed up the manual annotation process. The annotator Prodigy (35,36) has an annotation mode called *ner.correct*, which uses a trained neural network to accelerate the manual annotation of signs and symptoms.

With suitable training and guidelines, high levels of inter-rater agreement between human annotators for signs and symptoms are feasible. Restricting the annotation to a limited domain (e.g., neurological signs and symptoms) and restricted ontology (e.g., neuro-ontology) simplifies manual annotation. Although the inter-rater agreement between human and machine annotators was lower than between human annotators, advances in natural language processing should bring inter-rater agreement between machines and humans closer and make high throughput phenotyping of electronic health records feasible.

This work has limitations. The sample of clinical notes was small (five patient notes per annotation round). A larger sample of notes would have been desirable. The annotation process was restricted to neurological signs and symptoms in neurology notes. The target ontology was a limited neuro-ontology with 1600 concepts (21). We evaluated only one machine annotator based on a convolutional neural network. Other neural networks are likely to perform better. Our results on an inter-rater agreement might not generalize to other medical domains and ontologies. Although we had three raters for this study, we did not designate any of them as the "gold standard," and we elected to calculate inter-rater agreement for each pair of raters separately. In our opinion, unadjusted agreement at the 90% level between human raters should be considered high. Likewise, machine annotators that can reach 90% unadjusted agreement with human annotators should be considered accurate. Because we lacked a gold standard, we chose to measure the performance of the machine annotator as concordance (unadjusted agreement) and Kappa statistic (adjusted agreement) rather than as accuracy, precision, and recall. Although we used ANOVA to assess the significance of differences in the means for adjusted and unadjusted agreement, we cannot be certain that all assumptions underlying ANOVA were met in our samples, including normality, homogeneity of variance, and independence.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board of the University of Illinois at Chicago. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Concept and design by DH. Data collection by DH, CO, and QH-P. Data analysis by CO and DH. Data interpretation by DH, MC, QH-P, and CO. Initial draft by DH and CO. Revisions, re-writing, and final approval by DH, CO, QH-P, and MC. All authors contributed to the article and approved the submitted version.

Funding

MC acknowledges research funding from the Department of Veterans Affairs (BLR&D Merit Award BX000467).

Conflict of interest

MC acknowledges prior support from Biogen.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Hebbring SJ, Rastegar-Mojarad M, Ye Z, Mayer J, Jacobson C, Lin S. Application of clinical text data for phenome-wide association studies (PheWASs). *Bioinformatics*. (2015) 31:1981–7. doi: 10.1093/bioinformatics/btv076

2. Kimia AA, Savova G, Landschaft A, Harper MB. An introduction to natural language processing: how you can get more from those electronic notes you are generating. *Pediatr Emerg Care*. (2015) 31:536–41. doi: 10.1097/PEC.0000000000000484

3. Alzoubi H, Alzubi R, Ramzan N, West D, Al-Hadhrami T, Alazab M. A review of automatic phenotyping approaches using electronic health records. *Electronics*. (2019) 8:1235. doi: 10.3390/electronics8111235

4. Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. *J Am Med Inform Assoc*. (2014) 21:221–30. doi: 10.1136/amiajnl-2013-001935

5. Agrawal M, O'Connell C, Fatemi Y, Levy A, Sontag D. Robust benchmarking for machine learning of clinical entity extraction. *Machine Learning for Healthcare Conference*. PMLR (2020). p. 928–49.

6. Fu S, Chen D, He H, Liu S, Moon S, Peterson KJ, et al. Clinical concept extraction: a methodology review. *J Biomed Inform*. (2020) 109:103526. doi: 10.1016/j.jbi.2020.103526

7. Mamlin BW, Heinze DT, McDonald CJ. Automated extraction, normalization of findings from cancer-related free-text radiology reports. *AMIA Annual Symposium Proceedings*. Vol. 2003. American Medical Informatics Association (2003). p. 420.

8. Leaman R, Islamaj Doğan R, Lu Z. Dnorm: disease name normalization with pairwise learning to rank. *Bioinformatics*. (2013) 29:2909–17. doi: 10.1093/bioinformatics/btt474

9. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res*. (2004) 32:D267–70. doi: 10.1093/nar/gkh061

10. Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. *J Am Med Inform Assoc*. (2007) 14:497–506. doi: 10.1197/jamia.M2372

11. Hwang JC, Alexander CY, Casper DS, Starren J, Cimino JJ, Chiang MF. Representation of ophthalmology concepts by electronic systems: intercoder agreement among physicians using controlled terminologies. *Ophthalmology*. (2006) 113:511–9. doi: 10.1016/j.ophtha.2006.01.017

12. Miñarro-Giménez JA, Martínez-Costa C, Karlsson D, Schulz S, Gøeg KR. Qualitative analysis of manual annotations of clinical text with SNOMED CT. *PLoS ONE*. (2018) 13:e0209547. doi: 10.1371/journal.pone.0209547

13. Hier DB, Yelugam R, Azizi S, Wunsch II DC. A focused review of deep phenotyping with examples from neurology. *Eur Sci J*. (2022) 18:4–19 (Accessed August 12, 2022). doi: 10.19044/esj.2022.v18n4p4

14. Divita G, Zeng QT, Gundlapalli AV, Duvall S, Nebeker J, Samore MH. Sophia: a expedient UMLS concept extraction annotator. *AMIA Annual Symposium Proceedings*. Vol. 2014. American Medical Informatics Association (2014). p. 467.

15. Hier DB, Yelugam R, Azizi S, Carrithers MD, Wunsch II DC. High throughput neurological phenotyping with metamap. *Eur Sci J*. (2022) 18:37–49 (Accessed August 12, 2022). doi: 10.19044/esj.2022.v18n4p37

16. Arbabi A, Adams DR, Fidler S, Brudno M. Identifying clinical terms in medical text using ontology-guided machine learning. *JMIR Med Inform*. (2019) 7:e12596. doi: 10.2196/12596. PMID: 31094361; PMCID: PMC6533869.

17. Yang X, Bian J, Hogan WR, Wu Y. Clinical concept extraction using transformers. *J Am Med Inform Assoc*. (2020) 27:1935–42. doi: 10.1093/jamia/ocaa189

18. Azizi S, Hier D, Wunsch ID. Enhanced neurologic concept recognition using a named entity recognition model based on transformers. *Front Digit Health*. (2022) 4:1–8. doi: 10.3389/fdgth.2022.1065581

19. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. (2012) 22:276–82. doi: 10.11613/BM.2012.031

20. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. (1960) 20:37–46. doi: 10.1177/001316446002000104

21. Hier DB, Brint SU. A neuro-ontology for the neurological examination. *BMC Med Inform Decis Mak*. (2020) 20:1–9. doi: 10.1186/s12911-020-1066-7

22. Altinok D. *Mastering spaCy*. Birmingham, UK: Packt Publishing (2021).

23. Hier DB, Yelugam R, Azizi S, Wunsch DC. A focused review of deep phenotyping with examples from neurology. *Eur Sci J*. (2022) 18:4–19. doi: 10.19044/esj.2022.v18n4p4

24. Haendel MA, Chute CG, Robinson PN. Classification, ontology, and precision medicine. *N Engl J Med*. (2018) 379:1452–62. doi: 10.1056/NEJMra1615014

25. Robinson PN. Deep phenotyping for precision medicine. *Hum Mutat*. (2012) 33:777–80. doi: 10.1002/humu.22080

26. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. (2015) 372:793–5. doi: 10.1056/NEJMp1500523

27. Shinar D, Gross CR, Mohr JP, Caplan LR, Price TR, Wolf PA, et al. Interobserver variability in the assessment of neurologic history and examination in the Stroke Data Bank. *Arch Neurol*. (1985) 42:557–65. doi: 10.1001/archneur.1985.04060060059010

28. Goldstein LB, Bertels C, Davis JN. Interrater reliability of the NIH stroke scale. *Arch Neurol*. (1989) 46:660–2. doi: 10.1001/archneur.1989.00520420080026

29. Vuokko R, Mäkelä-Bengs P, Hyppönen H, Lindqvist M, Doupi P. Impacts of structuring the electronic health record: results of a systematic literature review from the perspective of secondary use of patient data. *Int J Med Inform*. (2017) 97:293–303. doi: 10.1016/j.ijmedinf.2016.10.004

30. Cohen GR, Friedman CP, Ryan AM, Richardson CR, Adler-Milstein J. Variation in physicians' electronic health record documentation and potential patient harm from that variation. *J Gen Intern Med*. (2019) 34:2355–67. doi: 10.1007/s11606-019-05025-3

31. Joukes E, Abu-Hanna A, Cornet R, de Keizer NF. Time spent on dedicated patient care and documentation tasks before and after the introduction of a structured and standardized electronic health record. *Appl Clin Inform*. (2018) 9:046–53. doi: 10.1055/s-0037-1615747

32. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *J Am Med Inform Assoc*. (2011) 18:181–6. doi: 10.1136/jamia.2010.007237

33. Moy AJ, Schwartz JM, Chen RJ, Sadri S, Lucas E, Cato KD, et al. Measurement of clinical documentation burden among physicians and nurses using electronic health records: a scoping review. *J Am Med Inform Assoc*. (2021) 28:998–1008. doi: 10.1093/jamia/ocaa325

34. Downing NL, Bates DW, Longhurst CA. Physician burnout in the electronic health record era: are we ignoring the real cause? *Ann Intern Med*. (2018) 169:50–1. doi: 10.7326/M18-0139

35. Musabeyezu F. *Comparative study of annotation tools and techniques* [master's thesis]. African University of Science and Technology (2019).

36. Neves M, Ševa J. An extensive review of tools for manual annotation of documents. *Brief Bioinform*. (2021) 22:146–63. doi: 10.1093/bib/bbz130

frontiers | Frontiers in Digital Health

TYPE Original Research
PUBLISHED 07 August 2023
DOI 10.3389/fdgth.2023.1035442

OPEN ACCESS

EDITED BY
Benjamin Kummer,
Icahn School of Medicine at Mount Sinai, United States

REVIEWED BY
Leah Blank,
Icahn School of Medicine at Mount Sinai, United States
Sanay Muhammad Umar Saeed,
University of Engineering and Technology, Pakistan

*CORRESPONDENCE
Jared Ward
ward-jared@cooperhealth.edu

RECEIVED 02 September 2022
ACCEPTED 17 July 2023
PUBLISHED 07 August 2023

CITATION
Ward J, Green A, Cole R, Zarbiv S, Dumond S, Clough J and Rincon F (2023) Implementation and impact of a point of care electroencephalography platform in a community hospital: a cohort study. Front. Digit. Health 5:1035442. doi: 10.3389/fdgth.2023.1035442

Implementation and impact of a point of care electroencephalography platform in a community hospital: a cohort study

Jared Ward[1]*, Adam Green[1], Robert Cole[1], Samson Zarbiv[1], Stanley Dumond[2], Jessica Clough[3] and Fred Rincon[4]

[1]Department of Medicine, Division of Critical Care Medicine, Cooper University Hospital, Cooper University Medical School of Rowan University, Camden, NJ, United States, [2]Department of Medicine, Critical Care Medicine Fellowship, Inspira Medical Center, Vineland, NJ, United States, [3]Cardiopulmonary Department, Inspira Health, Vineland, NJ, United States, [4]Department of Neurology, Cooper University Hospital, Cooper University Medical School of Rowan University, Camden, NJ, United States

Objective: To determine the clinical and financial feasibility of implementing a poc-EEG system in a community hospital.
Design: Data from a prospective cohort displaying abnormal mentation concerning for NCSE or rhythmic movements due to potential underlying seizure necessitating EEG was collected and compared to a control group containing patient data from 2020.
Setting: A teaching community hospital with limited EEG support.
Patients: The study group consisted of patients requiring emergent EEG during hours when conventional EEG was unavailable. Control group is made up of patients who were emergently transferred for EEG during the historical period.
Interventions: Application and interpretation of Ceribell®, a poc-EEG system.
Measurement and main results: 88 patients were eligible with indications for poc-EEG including hyperkinetic movements post-cardiac arrest (19%), abnormal mentation after possible seizure (46%), and unresponsive patients with concern for NCSE (35%). 21% had seizure burden on poc-EEG and 4.5% had seizure activity on follow-up EEG. A mean of 1.1 patients per month required transfer to a tertiary care center for continuous EEG. For the control period, a total of 22 patients or a mean of 2 patients per month were transferred for emergent EEG. Annually, we observed a decrease in the number of transferred patients in the post-implementation period by 10.8 (95% CI: −2.17–23.64, $p = 0.1$). Financial analysis of the control found the hospital system incurred a loss of $3,463.11 per patient transferred for an annual loss of $83,114.64. In the study group, this would compute to an annual loss of $45,713.05 for an overall decrease in amount lost of $37,401.59. We compared amount lost per patient between historical controls and study patients. Implementation of poc-EEG resulted in an overall decrease in annual amount lost of $37,401.59 by avoidance of transfer fees. We calculated the amount gained per patient in the study group to be $13,936.44. To cover the cost of the poc-EEG system, 8.59 patients would need to avoid transfer annually.

Conclusion: A poc-EEG system can be safely implemented in a community hospital leading to an absolute decrease in transfers to tertiary hospital. This decrease in patient transfers can cover the cost of implementing the poc-EEG system. The additional benefits from transfer avoidance include clinical benefits such as rapid appropriate treatment of seizures and avoidance of unnecessary treatment as well as negating transfer risk and keeping the patient at their local hospital.

KEYWORDS

status epilepticus, point-of-care EEG, transfers, seizures, community hospital, finances

Introduction

A significant proportion of comatose patients in the emergency department (ED) or the intensive care unit (ICU) are at risk of developing nonconvulsive status epilepticus (NCSE), which is defined as a state of continuous or repetitive seizures without convulsions for more than 5 min (1). The annual incidence of status epilepticus (SE) is estimated to be 9.9–41 per 100,000 hospital admissions, with roughly one third of those classified as NCSE (2). Of all patients undergoing EEG in the ICU, 19% have been found to have seizure activity (3) and 48% of patients after convulsive status have been shown to have NCSE (4). Without timely diagnosis, treatment, and extenuation of NCSE, patients are at increased risk of neurological injury and death (5).

Continuous electroencephalogram (cEEG) remains the method of choice to diagnose NCSE with current guidelines recommending initiation within one hour of status epilepticus (1). Unfortunately, EEG is not available at all centers despite its association with improved outcomes due to resources required for its implementation, maintenance, and use (6). This leads to unnecessary transfer to tertiary centers for patients without NCSE resulting in a delay in further evaluation and treatment as well as additional costs and risks to the patient. Even at centers with cEEG, there is frequently a delay in initiation that falls outside the current guidelines (7, 8) An easy to deploy EEG system that allows for rapid diagnosis would fill these voids.

A poc-EEG platform uses fewer EEG leads than traditional EEG but in theory can be applied rapidly and with minimal training. Poc-EEG integrates three main tools for EEG interpretation: (a) raw EEG data, (b) sonification of EEG patterns, and (c) artificial intelligence (AI), which provides a percentage that reflects probability of seizure, obviating the necessity for bedside neurology interpretation. When compared to traditional EEG, poc-EEG has demonstrated similar accuracy in diagnosing NCSE (9, 10). Poc-EEG has been successfully implemented in academic centers allowing for timely and accurate assessment of patients in the critical care setting (8). A case series of 10 patients has shown the successful use of poc-EEG to assist in timely diagnosis and treatment of suspected seizure in a community hospital (11).

One poc-EGG system, Ceribell® (Mountain View, CA), is an FDA-approved limited montage 10 electrode EEG system that can be rapidly applied to patients with a suspicion of seizure. To ensure appropriate connectivity, each electrode has a gel which is expelled and after twisting the external part of the electrode a green light is displayed. Once all ten electrodes have an adequate connection, the device bedside monitor begins recording. Ceribell's® software algorithm used by the Seizure Detection module identifies sections of EEG that may correspond to electrographic seizures via preprocessing and segmenting signals into smaller events and then evaluating those signals based on time, frequency and channel features over a moving 5-min window. If the seizure thresholds are reached the device produces an alarm. The algorithm generated a seizure alert with 100% sensitivity if burden >50%, 88% if >10% but more importantly showed a negative predictive value of 99% if no seizure burden was reported (12). A comparison of Ceribell and conventional EEG characteristics is provided in **Table 1** and **Figure 1**.

In this study, we investigated the feasibility of implementing Ceribell®, a poc-EEG system, in a community hospital. We analyzed the effect of poc-EEG on clinical and financial outcomes. Specifically, our objectives were to: (a) describe implementation of a poc-EEG system at a community hospital; (b) characterize a study cohort who would undergo poc-EEG; and (c) provide basic analysis of potential cost benefit from transfer avoidance and apply that analysis to the cost of the technology.

Methods

Study design

Patients were prospectively identified with concern for seizure activity necessitating EEG for management during hours when conventional EEG was not available. This included patients with abnormal mentation potentially due to NCSE or rhythmic movements with concern for underlying seizure activity. A retrospective cohort consisting of patients requiring transfer for emergent EEG were used to determine the baseline number of patients transferred monthly and the costs associated with the transfer.

TABLE 1 Ceribell vs. Conventional EEG characteristics.

	Conventional	Ceribell
Frequency		
Sampling rate (range)	200–1,000 Hz	250 Hz
Frequency response (range)	0.01–500 Hz	0.5–100 Hz
Channels	32	8
Number of electrodes	21	10

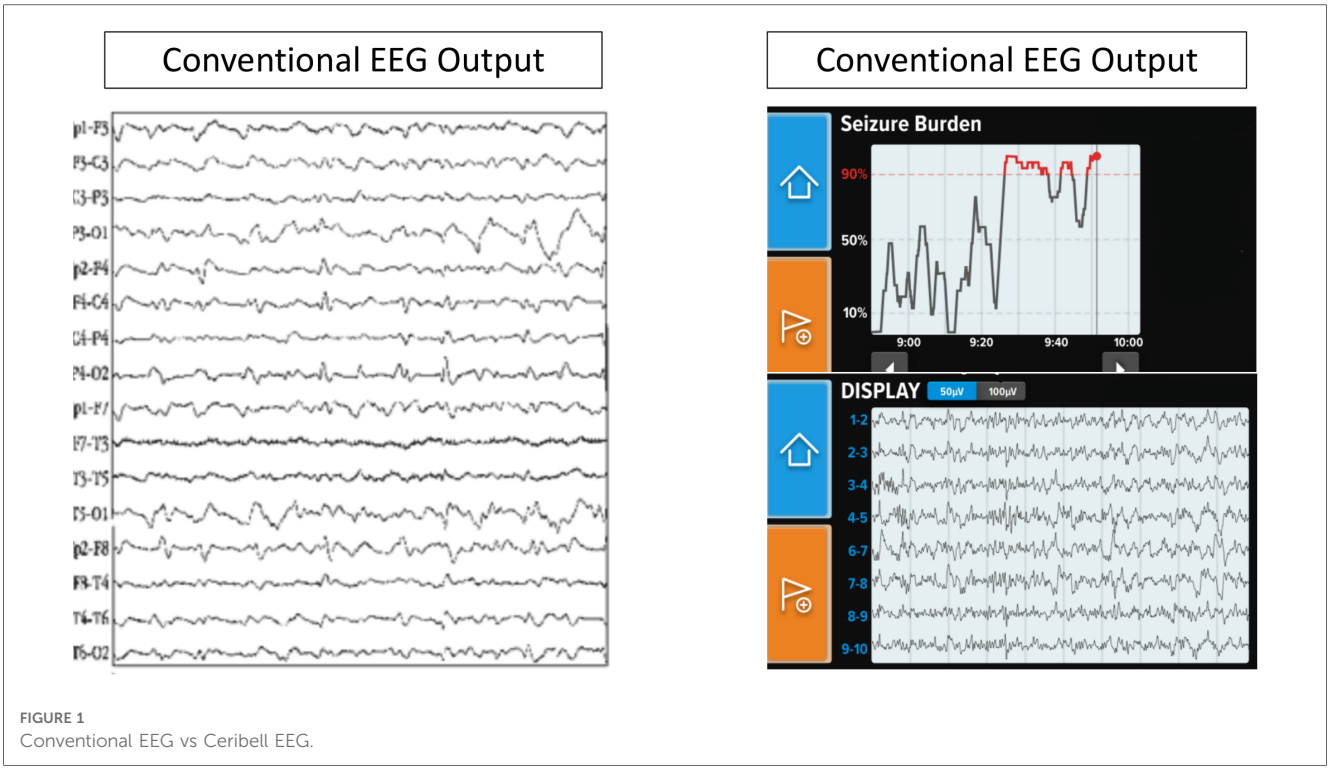


FIGURE 1
Conventional EEG vs Ceribell EEG.

Setting

Two community hospitals within the Inspira Health Network located in Vineland, NJ (262 beds) and Mullica Hill, NJ (210 beds) were identified. Neurology consultation was available at both hospitals as well as routine spot EEGs during regular business hours, specifically 9am–5pm with limited neurology coverage and no EEG technicians on weekend days. Twenty-four-hour critical care fellow support was available at both hospitals as well as intensivist daytime coverage and 24-hour availability. Data was collected between January and October 2021. Historical data was obtained from the calendar year 2020. Transfers during the month of December 2020 were excluded as this is when the poc-EEG device was piloted, which affected the number of transfers. The year preceding intervention was chosen as the historical cohort for several reasons. While practice patterns throughout the COVID-19 pandemic likely changed, the two years more closely resembled each other than if compared to a period not within the COVID-19 pandemic. Furthermore, prior to 2020 neurology consultation within the hospital system was even more sporadic as it relied on outpatient neurology coverage. Starting in 2020, a dedicated in-patient neurology consultant was hired to cover normal business hours. Prior to this year, patients were likely transferred for neurology consultation alone, making comparison impossible.

Intervention

Ten critical care fellows were trained in how to appropriately apply the Ceribell® headband (Figure 2) by the company's educator. Each fellow had a unique login for the mobile portal, allowing for further investigation. No training was needed by the bedside nurse, other than to notify the fellow if the device alarmed or connectivity failed (red light displayed).

Patients in the ED or ICU with concerns for seizure activity were prospectively identified and Ceribell® headband device was applied. These patients demonstrated either abnormal mentation possibly due to NCSE or rhythmic movements potentially due to convulsive seizure activity. After application, the critical care fellow directly observed the patient for the first five minutes. The EEG headband was then allowed to record for a maximum of two hours. After removal of the headband, the critical care fellow would review the seizure threshold reached during the entire two hours via the online portal. A treatment algorithm was provided (Figure 3) that instructed the appropriate treatment intervention if the seizure threshold was <10%, 10%–70%, or >70%. Intensivist and neurology consultants were available if needed. Based on the findings and intervention required, a standardized note was documented (Figure 4). During the next regular business hours, a standard EEG was performed on all patients as well as a neurology consultation. In addition, the poc-EEG tracings were interpreted by the neurology consultant. Figure 5 describes the workflow.

Population

After adopting the use of poc-EEG, clinical criteria were established pertaining to appropriate use. Its use was restricted to patients requiring emergent diagnostic EEG during hours

FIGURE 2
Ceribell poc-EEG headband, recorder and portal.

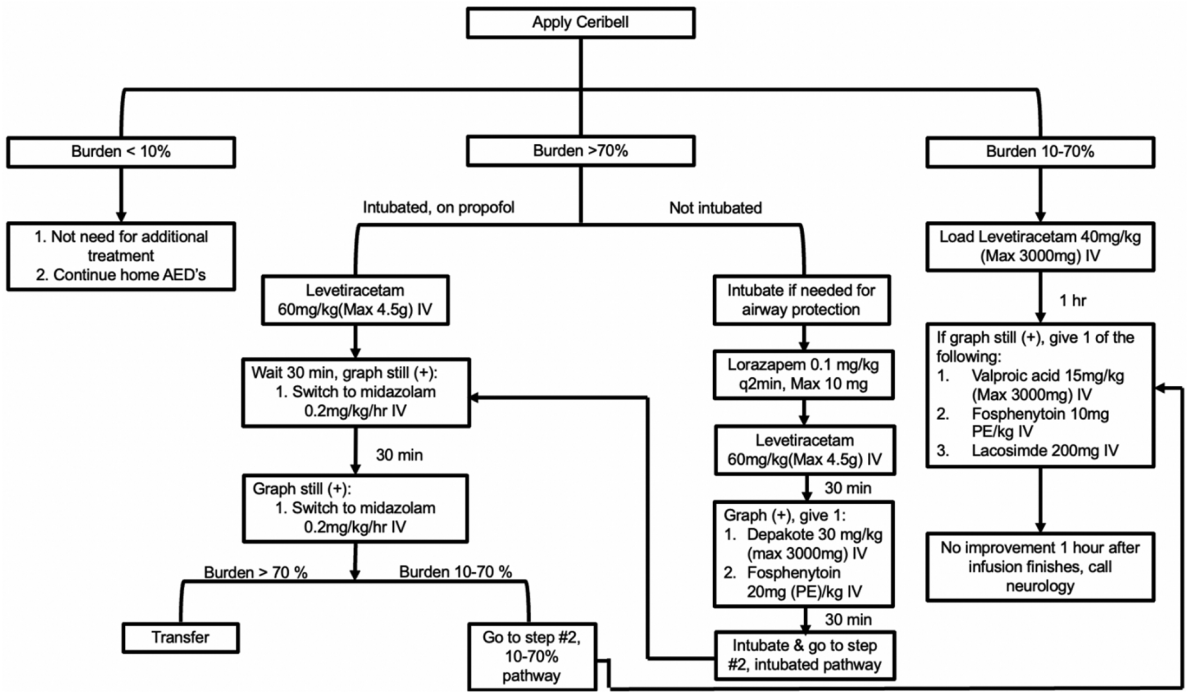


FIGURE 3
Treatment algorithm.

when conventional EEG was not available. The use of poc-EEG was left to the treating physician's preference and was not part of a study protocol. Common indications included patients with hyperkinetic movements post-cardiac arrest, patients with history of seizures and/or witnessed convulsive seizure activity without return to baseline mentation, and all other patients found unresponsive or stuporous upon admission with concern for NCSE (**Figure 6**). The study size was determined by the total number of patients receiving poc-EEG.

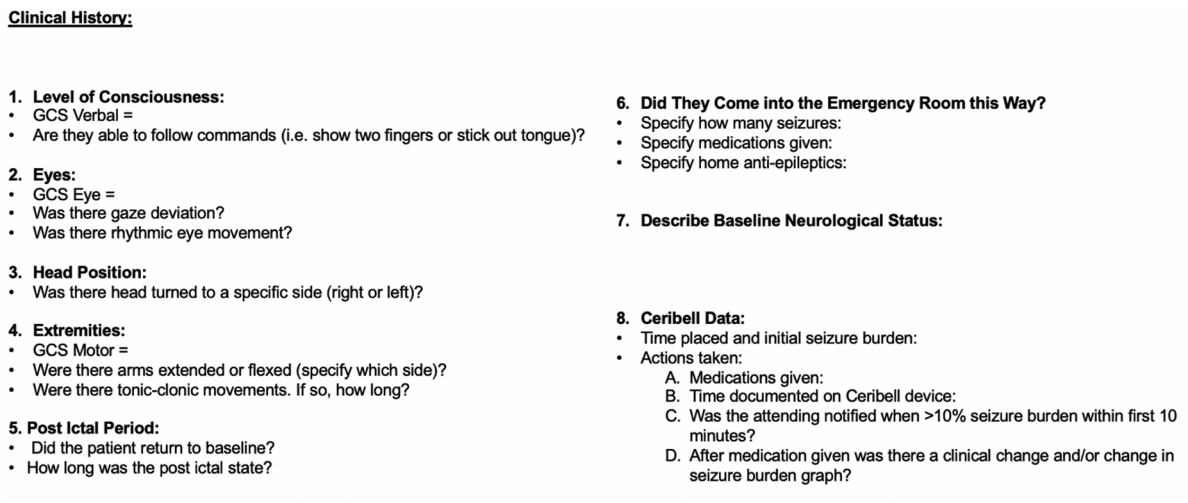


FIGURE 4
Poc-EEG event note.

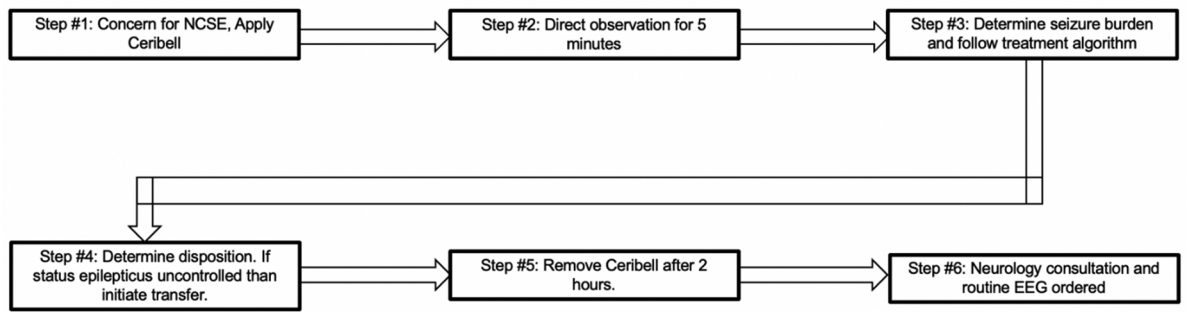


FIGURE 5
Poc-EEG workflow.

The control group consisted of patient transferred for emergent EEG during the historical period.

Data collection

Prior to data collection, the institutional review board of Inspira Medical Center approved the study (2022-02-001) and waived the need for informed consent. The standardized note was reviewed as well as the neurologists' dictation of the poc-EEG study and the corresponding standard EEG. Their official interpretation was used to determine electrographic seizure activity as well as response to treatment. Chart review was then completed to identify disposition outcome, mortality, and complications that occurred throughout the hospitalization.

After the patient was discharged and/or transferred, a coding summary was generated by the financial billings department and was included in the patient's medical record. Each patient's hospitalization record was reviewed, and the billing department provided the average net billed and accrued.

Financial analysis

Understanding the financial feasibility after implementation of a new product is important for its longevity. The focus of the analysis was to define the financial impact due to the absolute reduction in the number of transferred patients after implementation of poc-EEG. We chose to compare transfers to outside hospitals for emergent EEG between groups. While this may not capture the total financial impact of poc-EEG implementation, it can act as a surrogate reflecting the financial burden on the health care system. Transfers to outside hospitals have been linked to increasing health care costs, and, at least for NCSE, these costs have been shown to be decreased through the use of technology and AI (13, 14). Further analysis could be done to determine if the cost of the technology could be covered due to patient transfer avoidance.

For all analysis, we used the mean amount collected for each group. We did not have access to financial data at the level of individual patients. We used the amount collected as opposed to the amount charged as charged amounts are subject to pricing

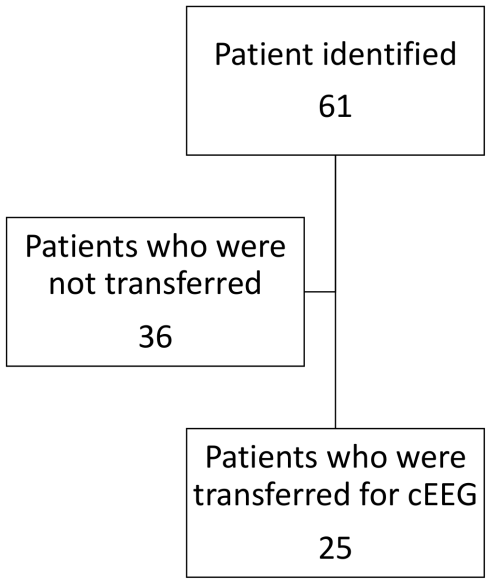


FIGURE 6
Poc-EEG cohort.

differences across institutions and therefore may limit generalizability of results.

To determine the financial impact, we did the following. First, we calculated the annual loss due to patients transferred for EEG. We took the control group and determined the mean collected per patient. We subtracted the expenses of transfer from this value. This represented the net loss per patient transferred. This could then be extrapolated to an annual loss based on total number of patients transferred during the control year. We assumed that there were no differential transfer costs between transferred patients in both groups. This allowed us to take the same amount lost per patient and apply it to the number of patients transferred in the treatment group providing an annual loss after implementation of poc-EEG. The difference between these two values represents the decrease in annual loss of patients requiring transfer for emergent EEG.

In order to determine the number of patients who avoided transfer needed to cover the fixed cost of the device, we did the following. We took the mean amount collected for the treatment group and subtracted the variable cost of the headband. We then subtracted the amount lost if the patient was transferred (calculated above). This represented the net earned by avoiding patient transfer. We then calculated the annual fixed cost of the Ceribell® system (monthly subscription fee × 12). By dividing, the annual cost of the technology by the amount earned per patient avoiding transfer, the number of transfers needed to cover the expense of the system could be determined.

Statistical analysis

The results are mostly descriptive in this study. Comparison of frequencies was preformed using the Chi Square tests. 95% confidence interval was used with statistical significance at $p < 0.05$. Quantitative data including reduction in transfers and financial outcomes did not require further analysis. Confidence intervals for financial analysis were unable to be calculated as only the mean for each group was collected. All data was accounted for.

Results

Clinical characteristics

From January through October 2021, we implemented and used poc-EEG in 88 subjects. Eligible subjects included patients with hyperkinetic movements post-cardiac arrest (19%, $n = 17/88$), patients with a history of seizures and/or with witnessed convulsive seizure activity and without return to baseline (46%, $n = 40/88$), and all other patients found unresponsive or stuporous upon admission to the hospital with concern for NCSE (35%, $n = 31/88$). Approximately 10% ($n = 9/88$) of the poc-EEG were applied in the emergency room, the rest of the patients were identified in the ICU (90%, $n = 79/88$). Of the 88 patients, 21% ($n = 19/88$) had significant electrographic seizure burden on poc-EEG and 4% ($n = 4/88$) had electrographic seizure activity confirmed on follow-up EEG. Another 5% ($n = 5/88$) were transferred immediately after poc-EEG identified high burden of electrographic seizure activity as an immediate need for continuous EEG was identified; therefore poc-EEG excluded and/or decreased ongoing concern for electrographic seizure activity in 78% ($n = 69/88$) of our inception cohort (**Figure 7**).

Demographic characteristics of the patients in this cohort showed a mean age of 57 years old (95% CI: 53.27–60.65), 52% ($n = 46/88$) were male, 46.5% ($n = 41/88$) female, and 1 person identified as transgender (gender not identified). Approximately 16% ($n = 14/88$) had a history of seizures on AEDs. Overall, the cohort was 64% Caucasian ($n = 56/88$) with 20% African American ($n = 18/88$), 11% Hispanic ($n = 10/88$), 0.3% Asian ($n = 3/88$), and 0.1% other ($n = 1/88$). All data was accounted for without any missing variables. (**Table 2**).

Only 2 patients where poc-EEG identified 0% electrographic seizure burden were found to have electrographic seizure activity on the follow-up standard EEG; thus only 2.4% of patients ($n = 2/83$) were found to have electrographic seizure activity on follow-up EEG despite a negative poc-EEG; 5 patients were transferred before follow-up EEG could be performed.

Transfer data

During the study period, eleven patients (mean of 1.1 per month) were transferred for emergent EEG. This constituted 13.4% ($n = 11/82$) of the total cohort. During 2020, 22 patients (mean of 2 per month) were transferred to a tertiary center for emergent EEG. The difference between these two values represents the decrease in the annualized number of patients requiring transfer for emergent EGG. This computed to an annual estimate of 10.8 patients (95% CI: −2.17–23.64, $p = 0.1$).

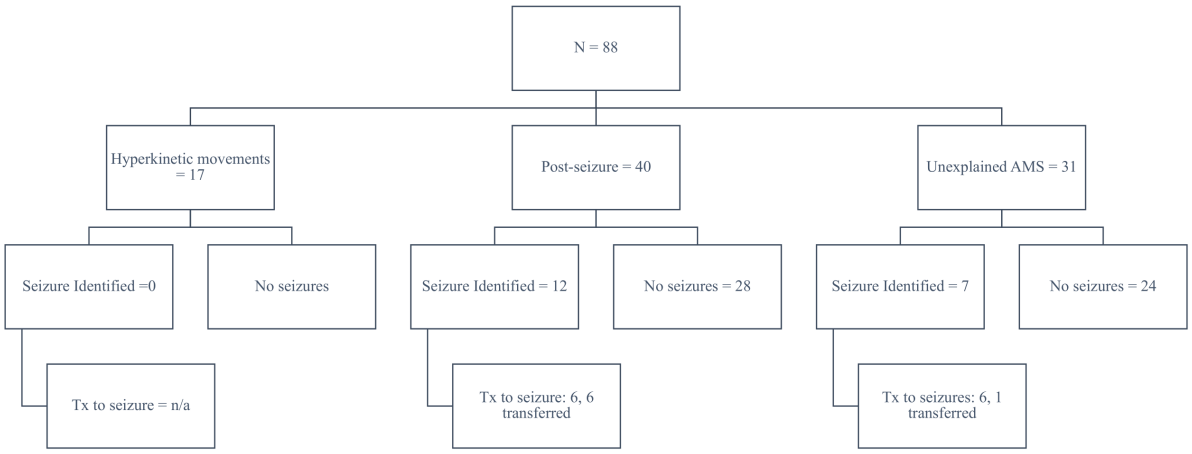


FIGURE 7
Poc-EEG cohort characteristics.

Financial analysis

The control group had a mean amount collected of $4,036.89 per patient. This is for the initial treatment and stabilization of the patient prior to transfer. For our hospital system, most patients requiring transfer go to one academic hospital that is roughly 40 miles away. Given the acuity of illness of these patients and need for ACLS trained nursing and appropriate monitoring capabilities, the transfer center reported a mean cost of $7,500 per patient billed to the sending facility. When adjusted for the amount collected per patient, this result in a mean loss of $3,463.11 per patient or an estimated annual loss of $83,114.64. This was calculated by multiplying $3,463.11 by 24, the number of patients transferred during the control period.

When the study period is analyzed, the eleven patients who required transfer would result in a loss of $38,094.21, $3,463.11 per patient, for the ten-month study period or estimated annual loss of $45,713.05. For this cohort after the implementation of poc-EEG, Inspira experienced an overall decrease in amount lost due to transferred patients of $37,401.59. This was calculated by subtracting $37,401.59 from $83,114,64.

The treatment cohort (those that received poc-EEG) had a mean collection of $11,161.33 per patient. As above, assuming each patient transferred incurs a loss of $3,463.11 per patient every patient kept because of poc-EEG would result in a net positive of $13,936.44 per patient after the cost of the poc-EEG headband was applied. The poc-EEG system has a monthly fixed cost of $9,975 for a multi-hospital system or $119,700 annual cost. To cover those costs, 8.59 patients per year (0.72 per month) would need to avoid transfer. We demonstrated a reduction in transfer of 0.9 patients per month (**Table 3**). It would take 9.5 months to recover upfront costs. The number needed to avoid transport to recuperate annual costs would be significantly lower if that patient required flight transport as these costs often exceed $40,000.00 and therefore would make avoidance of unnecessary transfer more important.

TABLE 2 Demographic characteristics.

	Cohort characteristics	Historical characteristics
Age (years)*	57 (53.27, 60.65)	58 (50.08, 67.84)
Gender		
Male	52.2%	60.9%
Female	46.6%	39.1%
Transgender	1.2%	0%
Race		
Caucasian	63.6%	56.5%
African American	20.5%	17.4%
Hispanic	11.3%	26.1%
Asian	3.3%	0%
Other	1.2%	0%

*Data presented as mean (25th percentile, 75th percentile).

TABLE 3 Poc-EEG net income.

	Revenue (per patient)		Variable expenses (per patient)	Net income (per patient)
	Billed	Collected		
Control	$11,361.30	$4,036.89	$7,500.00 (transfer cost)	Control Collected–Transfer cost = –$3,463.11
Ceribell	$28,585.49	$11,161.33	$688.00 (headband cost)	Ceribell collected–headband = $10,473.33
Savings by avoiding transfer	Ceribell net income – Control net income = $13,936.44			
Ceribell annual fixed cost	Ceribell monthly cost ($9,975.00) × 12 months = $119,700.00			
# of prevented transfers needed to cover costs	($119,700.00)/($13,936.44) = 8.59			

Discussion

Critically-ill neurological patients account for at least 10%–15% of admissions to intensive care units of which 8%–34% will experience seizure activity (15, 16). Approximately 3.3% of all critically ill patients experience seizures and a high index of suspicion needs to be had by providers especially in comatose patient or those without return to baseline mentation (17). Of seizures captured in one study, 34% were nonconvulsive seizures, and of these, 76% were NCSE (3). Emergent EEG has been noted to be of increasing importance in critical care but access to this diagnostic modality has remained limited. At one large US tertiary care medical center, where EEG availability and accessibility barriers should be minimal, the time to EEG in the ICU was 3.5 h (7). However, outside of these centers, even that time is unachievable as one study showed that in 286 emergent EEGs, the average interval from request to formal reporting was 1.13 days (18). A recent publication of the use of poc-EEG in COVID-19 patients showed that for 10 consecutive device applications, mean time to interpretation was 23.8 min compared to 126.5 min for routine 18-channel studies (19). Before the advent of poc-EEG, many smaller hospitals would often transfer patients for these services; one study conducted in 24 West Virginia hospitals found that the need for critical care and neurology services accounted for nearly 54% of all transfers during their study period (20). Thus, there is clearly a need for and adaptation of poc-EEGs aimed at reducing the overall time to EEG as well as expanding EEG availability outside of tertiary care centers but data on this is limited. Poc-EEG also has the added benefit of faster application and exposure to those applying the device to patients who may have communicable diseases.

Our experience provides a pragmatic framework on how to successfully implement this technology in a community setting with limited neurological coverage. The logistics regarding proper implementation and use of poc-EEG is often the largest obstacle to overcome. The stepwise approach provided here may provide guidance for other institutions with similar EEG availability and a means to fill that void. In addition, the data provided here demonstrates this can be done with a high concordance between poc-EEG and the following standard full montage EEG. This leads to improvement in care provided and a decrease in the absolute number of transfers to tertiary centers.

Avoidance of unnecessary transfers allows patients to be cared for in their own community. This decreases the burden of travel on the patients' family. It allows provides an opportunity for the patients' outpatient providers to continue to participate in the patients' care. This also eliminates risk associated with transfer. While the analysis did not show a statistically significant reduction in transfers, it did show an absolute reduction in number of transfers and a favorable financial analysis. The study was under powered and results could vary if examined on a larger scale. The analysis included provides financial justification for implementation of poc-EEG systems. Previous studies showed that transfers to referral centers are associated with higher costs to patients and often with no changes in treatment management (14). Transferred patients hospital cost were on average $9,600 higher compared with non-transfer patients and a recent US study showed transfer costs of $6,160 plus a $24.64 per mile charge for ground transport and $11,760 for air transfers, which excluded billing for other services (14, 21).

The major benefits of adapting poc-EEG include improved clinical care by addressing a diagnostic deficit (i.e., access to EEG during off hours). Other benefits, such as promoting patient satisfaction and minimizing transfer risk, are also notable. Financial analysis supported the cost of the implementation. This was done by examining per patient average collection as well as calculating the amount saved by minimizing transfer costs. While these savings did not cover the entire cost of the technology there is additional financial benefit from avoiding transfer. While this is challenging to calculate retrospectively as it is difficult to identify which patients would have been transferred if poc-EEG was not available. We were able to calculate on average how much each poc-EEG patient collected and from that determine how many patients needed to avoid transfer to cover the costs of the poc-EEG. Given the ease of use and absolute reduction in transfers between the two cohorts, poc-EEG will likely justify its associated cost and reduce out of hospital transfers.

Poc-EEG demonstrated a very low false negative rate for patients with minimal electrographic seizure activity on poc-EEG but confirmed electrographic seizure activity on standard EEG. Explanation for the false negatives could be attributed to the time between studies. False positives or those with electrographic seizure activity on poc-EEG but in fact negative standard EEG would be hard to identify in our study design. As all patients with positive poc-EEG would warrant anti-epileptic treatment and thus an explanation for the follow up negative standard EEG. There would need to be concurrent poc-EEG and standard EEG. An understanding of the outcomes of those transferred would add strength to our data.

Additionally, poc-EEG provide reliable data. As previously stated, a study showed 88% sensitivity for seizure burden >10% and 100% sensitivity if >50% but more importantly, a 99% negative predictive value if seizure activity was not identified by the device (12). Furthermore, when compared to conventional EEG, Ceribell® showed equivalent signal quality and durability (10). This previously published data mirrors our own experience of a low false negative rate (2.4% in our cohort) and leads to our determination that the poc-EEG device can be safely implemented in the community hospital setting.

Limitations

The retrospective nature of the historical cohort makes the data less granular than desired. Outcome measures, EEG findings, and reason for transfer can be hard to determine from chart review. We are confident that the main indication for transfer for both groups was need for emergent EEG, however there is always room for potential error. This limits comparison between the groups.

Given the retrospective design, we were limited in our ability to capture all drivers of transfer. An understanding of the treatment plan at the tertiary hospital and the outcomes of the EEGs at those hospitals is lacking from our data. The use of the poc-EEG was left to the treating physician's preference which introduces selection bias. While we believe the historical and study group have very similar patient demographics and severity of illness, it is possible there are seasonal or other changes that occur that are not recognized. A more robust data collection plan would have allowed for adjustment of both chronic conditions such as history of seizure or neurological injury as well as aspects of the acute illness that may have influenced the decision to transfer. This would have allowed for a more confident comparison between the groups. Both groups occurred during the COVID-19 pandemic, and this may have unknown effects. As the pandemic progressed, practice patterns changed which could influence decisions on transfer thus effecting our results.

The finances associated with treatment costs vary based on location, insurance, and many other factors. If more hospitals from diverse settings participated, the generalizability of the financial analysis would have increased. The cost cited in this study may not be replicated exactly by other institutions. In addition, more formal financial analysis could be implemented on future similar studies.

Conclusion

Our study highlights the continued importance for community medical centers to develop ways to provide rapid diagnosis and treatment for patients at risk of status epilepticus. This study is the largest study that shows the use of poc-EEG in a community setting and how it can lead to a decrease in unnecessary transfers with potential reduction in hospital costs.

Prior presentation

A portion of the work will be presented at SCCM Conference 2022; however, this manuscript has not been published elsewhere and is not under consideration by another journal.

Summary statement

Point-of-care EEG can be implemented in a community hospital with a high degree of diagnostic accuracy preventing transfers to tertiary centers with a very favorable financial profile.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Inspira Health Institutional Review Board File #2022-02-001. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Authors contributions

Contributed substantially to the conception and design of the study (JW, AG, RC, SZ, JC, FR), the acquisition of data (JW, JC, SD) and interpretation of data (JW, AG, FR). Drafted (JW, AG, FR) or provided critical revision of the article (all other authors). Provided the final approval of the version submitted for publication (all authors).

Acknowledgments

We would like to thank Phillip Dellinger, MD, MCCM, FCCP and Christa Schorr, DNP, MSN, RN, NEA-BC, FCCM for reviewing the contents of this manuscript.

Funding

Support was provided solely from institutional and/or departmental sources.

Conflict of interest

The authors declare that AG and FR receive consulting fees from Ceribell® Inc.; however, Ceribell® Inc. did not have any oversight for this manuscript.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Brophy GM, Bell R, Claassen J, Alldredge A, Bleck T, Glauser T, et al. Guidelines for the evaluation and management of status epilepticus. *Neurocrit Care*. (2012) 17 (1):3–23. doi: 10.1007/s12028-012-9695-z

2. Sanchez S, Rincon F. Status epilepticus: epidemiology and public health needs. *J Clin Med*. (2016) 5(8):71. doi: 10.3390/jcm5080071

3. Classen J, Mayer SA, Kowalski RG, Emerson RG, Hirsch LJ. Detection of electrographic seizures with continuous EEG monitoring in critically ill patients. *Neurology*. (2004) 62(10):1743–8. doi: 10.1212/01.WNL.0000125184.88797.62

4. DeLorenzo RJ, Waterhouse EJ, Town AR, Boggs JG, Ko D, DeLorenzo GA, et al. Persistent nonconvulsive status epilepticus after the control of convulsive status epilepticus. *Epilepsia*. (1998) 39(8):833–40. doi: 10.1111/j.1528-1157.1998.tb01177.x

5. Scholtes FB, Renier WO, Meinardi H. Generalized convulsive status epilepticus: causes, therapy, and outcome in 346 patients. *Epilepsia*. (1994) 35(5):1104–12. doi: 10.1111/j.1528-1157.1994.tb02562.x

6. Hill CE, Blank LJ, Thibault D, Davis K, Dahodwala N, Litt B, et al. Continuous EEG is associated with favorable hospitalization outcomes for critically ill patients. *Neurology*. (2019) 92(1):9–18. doi: 10.1212/WNL.0000000000006689

7. Gururangan K, Razavi B, Parvizi J. Utility of electroencephalography: experience from a U.S. Tertiary care medical center. *Clin Neurophysiol*. (2016) 127(10):3335–40. doi: 10.1016/j.clinph.2016.08.013

8. Vespa PM, Olson DM, John S, Hobbs K, Gururangan K, Nie K, et al. Evaluating the clinical impact of rapid response electroencephalography: the DECIDE multicenter prospective observational clinical study. *Crit Care Med*. (2020) 48 (9):1249–57. doi: 10.1097/CCM.0000000000004428

9. Rittenberger JC, Weissman A, Baldwin M, Flickinger K, Repine M, Guyette F, et al. Preliminary experience with point-of-care EEG in post-cardiac arrest patients. *Resuscitation*. (2019) 135:8–102. doi: 10.1016/j.resuscitation.2018.12.022

10. Kamousi B, Grant AM, Bachelder B, Jianchun Y, Hajinoroozi M, Woo R. Comparing the quality of signals recorded with a rapid response EEG and conventional clinical EEG systems. *Clin Neurophysiol Pract*. (2019) 4:69–75. doi: 10.1016/j.cnp.2019.02.002

11. Yazbeck M, Sra P, Parvizi J. Rapid response electroencephalography for urgent evaluation of patients in community hospital intensive care practice. *J Neurosci Nurs*. (2019) 51(6):308–12. doi: 10.1097/JNN.0000000000000476

12. Kamousi B, Karunakaran S, Gururangan K, Markert M, Decker B, Khankhanian P, et al. Monitoring the burden of seizures and highly epileptiform patterns in critical care with a novel machine learning method. *Neurocrit Care*. (2021) 34(3):908–17. doi: 10.1007/s12028-020-01120-0

13. Pelosi P, Ferguson ND, Frutos-Vivar F, Anzueto A, Putensen C, Raymondos K, et al. Management and outcome of mechanically ventilated neurologic patients. *Crit Care Med*. (2011) 39:1482–92. doi: 10.1097/CCM.0b013e31821209a8

14. Kuhn EN, Warmus BA, Davis MC, Oster RA, Guthrie BL. Identification and cost of potentially avoidable transfers to a tertiary care neurosurgery service: a pilot study. *Neurosurgery*. (2016) 79(4):541–8. doi: 10.1227/NEU.0000000000001378

15. Teng CY, Davis BS, Kahn JM, Rosengart MR, Brown JB. Factors associated with potentially avoidable interhospital transfers in emergency general surgery-A call for quality improvement efforts. *Surgery*. (2021) 170(5):1298–307. Erratum in: Surgery. 2022;172(2):779. doi: 10.1016/j.surg.2021.05.021

16. Westover MB, Shafi MM, Bianchi MT, Moura L, O'Rourke D, Rosenthal ES, et al. The probability of seizures during EEG monitoring in critically ill adults. *Clin Neurophysiol*. (2015) 126(3):463–71. doi: 10.1016/j.clinph.2014.05.037

17. Bleck TP, Smith MC, Pierre-Louis SJC, Jares JJ, Murray J, Hansen CA. Neurologic complications of critical medical illness. *Crit Care Med*. (1993) 21 (1):98–103. doi: 10.1097/00003246-199301000-00019

18. Firosh KS, Ashalatha R, Thomas SV, Sarma PS. Emergent EEG is helpful in neurology critical care practice. *Clin Neurophysiol*. (2005) 116(10):2454–9. doi: 10.1016/j.clinph.2005.06.024

19. LaMonte MP. Ceribell EEG shortens seizure diagnosis and workforce time and is useful for COVID isolation. *Epilepsia Open*. (2021) 6(2):331–8. doi: 10.1002/epi4.12474

20. Nair D, Gibbs MM. Inter-hospital transfers from rural hospitals to an academic medical center. *W V Med J*. (2013) 109(4):44–9. PMID: 23930562

21. Golestanian E, Scruggs JE, Gangnon RE, Mark RP, Wood KE. Effect of interhospital transfer on resource utilization and outcomes at a tertiary care referral center. *Crit Care Med*. (2007) 35(6):1470–76. doi: 10.1097/01.CCM.0000265741.16192.D9

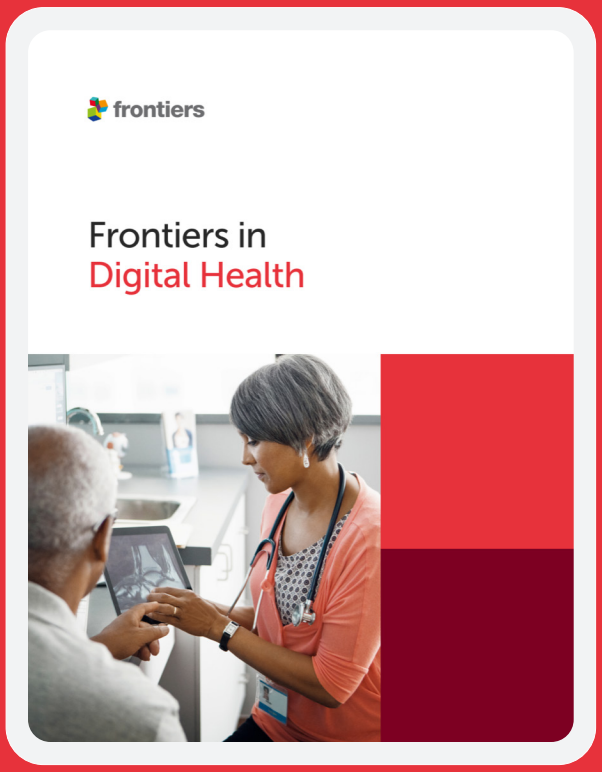
Frontiers in
Digital Health

Explores digital innovation to transform modern healthcare

A multidisciplinary journal that focuses on how we can transform healthcare with innovative digital tools. It provides a forum for an era of health service marked by increased prediction and prevention.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact