

Henry Ford Health

Henry Ford Health Scholarly Commons

Neurology Articles

Neurology

7-20-2022

Development of a natural language processing algorithm to extract seizure types and frequencies from the electronic health record

Barbara M. Decker

Alexandra Turco

Jian Xu

Samuel W. Terman

Nikitha Kosaraju

See next page for additional authors

Follow this and additional works at: https://scholarlycommons.henryford.com/neurology_articles

Authors

Barbara M. Decker, Alexandra Turco, Jian Xu, Samuel W. Terman, Nikitha Kosaraju, Alisha Jamil, Kathryn A. Davis, Brian Litt, Colin A. Ellis, Pouya Khankhanian, and Chloe E. Hill



Contents lists available at ScienceDirect

Seizure: European Journal of Epilepsy

journal homepage: www.elsevier.com/locate/seizure

Short communication

Development of a natural language processing algorithm to extract seizure types and frequencies from the electronic health record



Barbara M. Decker^{a,b,*}, Alexandra Turco^a, Jian Xu^c, Samuel W. Terman^d, Nikitha Kosaraju^a, Alisha Jamil^a, Kathryn A. Davis^a, Brian Litt^a, Colin A. Ellis^a, Pouya Khankhanian^e, Chloe E. Hill^{d,#}

^a Department of Neurology, University of Pennsylvania, Philadelphia, PA, United States

^b Department of Neurological Sciences, University of Vermont Medical Center, Burlington, VT, United States

^c Department of Neurology, Henry Ford Health System, Detroit, MI, United States

^d Department of Neurology, University of Michigan, Ann Arbor, MI, United States

^e Kaiser Permanente, Oakland, CA, United States

ARTICLE INFO

Keywords:

Epilepsy

Natural language processing

Seizure frequency

Electronic health record

Automated extraction

ABSTRACT

Objective: To develop a natural language processing (NLP) algorithm to abstract seizure types and frequencies from electronic health records (EHR).

Background: Seizure frequency measurement is an epilepsy quality metric. Yet, abstraction of seizure frequency from the EHR is laborious. We present an NLP algorithm to extract seizure data from unstructured text of clinic notes. Algorithm performance was assessed at two epilepsy centers.

Methods: We developed a rules-based NLP algorithm to recognize terms related to seizures and frequency within the text of an outpatient encounter. Algorithm output (e.g. number of seizures of a particular type within a time interval) was compared to seizure data manually annotated by two expert reviewers (“gold standard”). The algorithm was developed from 150 clinic notes from institution #1 (development set), then tested on a separate set of 219 notes from institution #1 (internal test set) with 248 unique seizure frequency elements. The algorithm was separately applied to 100 notes from institution #2 (external test set) with 124 unique seizure frequency elements. Algorithm performance was measured by recall (sensitivity), precision (positive predictive value), and F1 score (geometric mean of precision and recall).

Results: In the internal test set, the algorithm demonstrated 70% recall (173/248), 95% precision (173/182), and 0.82 F1 score compared to manual review. Algorithm performance in the external test set was lower with 22% recall (27/124), 73% precision (27/37), and 0.40 F1 score.

Conclusions: These results suggest NLP extraction of seizure types and frequencies is feasible, though not without challenges in generalizability for large-scale implementation.

1. Introduction

Documentation of seizure frequencies is a key American Academy of Neurology (AAN) epilepsy quality measure to improve care of patients with epilepsy [1,2]. Seizure frequency is a critical indicator of disease severity, the metric for treatment titration, and the most common primary outcome of clinical studies.

Reliable extraction of documented seizure frequency from electronic

health records (EHR) remains an important challenge - laborious manual review is required due to a lack of “smart” fields for automated abstraction and poor standardization of semiology/frequency descriptors. The rapidly evolving field of natural language processing (NLP) uses computational techniques to efficiently mine structured and/or unstructured EHR text. Such techniques leveraging big data could greatly expand clinical research by allowing large-scale automated chart reviews.

* Corresponding author at: Department of Neurological Sciences, University of Vermont Medical Center, 111 Colchester Ave, Burlington, VT 05401, United States.
E-mail address: Barbara.decker@uvmhealth.org (B.M. Decker).

Senior Author: Chloe E. Hill, MD MS, Department of Neurology, University of Michigan, 1500 E. Medical Center Drive, 1914 Taubman Center SPC 5316, Ann Arbor, MI 48109.

<https://doi.org/10.1016/j.seizure.2022.07.010>

Received 22 June 2022; Received in revised form 16 July 2022; Accepted 18 July 2022

Available online 20 July 2022

1059-1311/© 2022 British Epilepsy Association. Published by Elsevier Ltd. All rights reserved.

Prior NLP algorithms to address seizure frequency have been limited [2–5]. One particular challenge is extracting seizure frequencies for patients with multiple seizure types. In this study, we developed a rules-based NLP algorithm to identify seizure frequencies from the EHR unstructured narrative free text of a clinical encounter and to report numeric frequency descriptors for unique semiologies. We applied this algorithm to data from two comprehensive epilepsy centers to evaluate performance and generalizability.

2. Methods

2.1. Standard protocol approvals, registrations, and patient consents

The University of Pennsylvania (Institution #1) and University of Michigan (Institution #2) Institutional Review Boards approved this study.

2.2. Datasets

Institution #1: 417 adult patients seen for an epilepsy diagnosis from 2010-2018 were sampled from a cohort of records previously annotated for seizure frequency in a study of clobazam efficacy, thus a ready dataset for algorithm development [6]. This cohort was randomly sampled for the development dataset and a separate portion was randomly sampled for the internal test set. Clinic notes were written by 50 unique providers.

Institution #2: 100 adult patients with an epilepsy diagnosis were randomly sampled from a data pull of outpatient neurology visits for epilepsy from 2015-2019. This dataset (“external test set”) was used to measure the accuracy of the algorithm previously developed with Institution #1 data. Clinic notes were written by 29 unique providers.

2.3. Data collection

For each patient, the most recent clinic note was evaluated for seizure frequency documentation. Preceding clinic visit notes were also reviewed, when necessary, to interpret current seizure frequency in relation to a prior visit. Notes were excluded if seizure frequency was ambiguous or not documented, yielding totals of 150 clinical notes for the algorithm development set, 219 notes for the internal test set, and 96 notes for the external test set.

A “seizure frequency element” was defined as a phrase that conveys the frequency of a particular seizure type (e.g., “he had three generalized tonic-clonic seizures [GTCs] in the last month”). Each seizure frequency element communicates a rate (typically with a numeric term [e.g. “three”] and a temporal term [e.g. “last month”], though sometimes with a single word [e.g. “daily”]) and a seizure descriptor term (“GTCs”). A given note may contain one or more seizure frequency elements (e.g., “daily auras and one GTC this month”) or statements of seizure freedom (e.g., “seizure free for two years”).

For the development set, notes were independently annotated by two expert reviewers (AJ, AT). The two independent reviewers had an initial agreement rate of 73% within the development set. In cases of disagreement, notes were discussed between reviewers and, when necessary, by the larger group of authors. If unresolvable, the note was removed from the final analysis.

2.4. Algorithm overview

A rules-based NLP algorithm was developed from the 150-note development set using R programming software (Supplement) [6]. Briefly, the algorithm used pattern matching and regular expressions to search for seizure frequency elements within a note and extract seizure type and the quantitative frequency (e.g., number of seizures per year). The range of formats (terms and patterns) for seizure frequency elements are shown in Table 1.

Table 1
Terms and patterns of a seizure frequency element.

TERM	EXAMPLES
SEIZURE TYPE	Seizure(s), aura(s), sz, convulsion(s), drop attack(s), myoclonic jerk(s), spell(s), grand mal(s), petit mal(s), episode(s), GTC(s), BTC(s), CPS(s), SPS(s), head dip(s), head drop(s), absence(s), shaking, complex partial seizure(s), complex partial(s), simple partial seizure(s), simple partial(s), cluster(s), staring, staring spell(s), rolling eyes, FAS, FIAS
TIME (NOUN)	Day(s), week(s), month(s), monthly, year (s),
TIME (ADVERB)	Daily, weekly, monthly, yearly, annually
NUMERIC (QUALITATIVE)	Few, many, several, multiple, frequent, infrequent, periodic, occasional, rare
NUMERIC (QUANTITATIVE)	Zero, one, once, two, twice, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen, twenty, thirty, forty, fifty, sixty, seventy, eighty, ninety, hundred, 1-100
PREPOSITION	Per, /, in a, in one, in the past, a, each, every, times per, times, times a, throughout the, this, in
PATTERN	EXAMPLE
NUMERIC – SEIZURE TYPE – PREPOSITION – TIME	“5 GTCs per day”
SEIZURE TYPE – NUMERIC – PREPOSITION – TIME	“drop attacks 5 times daily”
NUMERIC – SEIZURE TYPE – TIME	“5 FAS daily”
SEIZURE TYPE – TIME	“auras daily”
SEIZURE TYPE – PREPOSITION – NUMERIC – TIME	“shaking every 5 days”
NUMERIC – SEIZURE TYPE – PREPOSITION – NUMERIC – TIME	“5 spells in the past 5 days”
TIME – SEIZURE TYPE	“daily head dips”

Algorithmic abstraction was run on the most current encounter note and the preceding encounter note for each patient. When the algorithm extracted both current and older information (i.e. data copied forward from the previous encounter) for a given seizure type for a patient, only the most recent or novel seizure frequency elements were reported. When the algorithm extracted multiple current phrases regarding the same seizure type in the same note, the phrase with the higher seizure frequency was reported.

2.5. Algorithm performance

Algorithm output was compared to the “gold standard” of manual annotation by two independent expert reviewers. Algorithm performance was measured by recall, precision, and F1 score. Recall (or sensitivity) was measured as correct algorithm-reported seizure frequency elements divided by total reviewer-annotated elements. Precision (or positive predictive value) was calculated as the correct algorithm-reported seizure frequency elements divided by total algorithm-reported elements. F1 score was calculated as the geometric mean of precision and recall (for this statistic, a value of 1 indicates perfect accuracy).

3. Results

3.1. Development set

A total of 194 unique seizure frequency elements were annotated by expert reviewers from 150 clinical encounter notes (Table 2). The algorithm reported 169 of these, of which 161 were correct. As such, the development set demonstrated a recall of 83% (161/194), precision of 95% (161/169), and an F1 score of 0.89.

Table 2
Algorithm performance.

	Development Set (Institution 1)	Internal test set (Institution 1)	External test set (Institution 2)
Total clinic encounter notes reviewed	417 (from which 48 were excluded due to ambiguous or absent seizure frequency)		100 (from which 4 were excluded due to ambiguous or absent seizure frequency)
Total clinic encounter notes included	150	219	96
Total number of unique seizure frequency elements by <u>reviewer annotation</u> , "gold standard"	194	248	124
Total number of unique seizure frequency elements by <u>algorithm report</u>	169	182	37
Number of correct <u>algorithm-reported</u> elements	161	173	27
Recall (%)	161/194 (83%)	173/248 (70%)	27/124 (22%)
Precision (%)	161/169 (95%)	173/182 (95%)	27/37 (73%)
F1 score	0.89	0.82	0.40

3.2. Internal test set

Of 219 clinical encounter notes, experts annotated 248 unique seizure frequency elements. The algorithm reported 182 total elements, of which 173 were correct. Application of the algorithm to this test set demonstrated 70% recall (173/248), 95% precision (173/182), and 0.82 F1 score.

3.3. External test set

Of the 96 total clinical encounter notes within the external test set, reviewers annotated 124 unique seizure frequency elements. The algorithm reported a total of 33 elements, of which 27 were correct. In the external test set, algorithm performance was lower with 22% recall (27/124), 73% precision (27/37), and an F1 score of 0.40. Notably, in the external test set, the algorithm performed particularly poorly with statements of seizure freedom (precision 0/24).

4. Discussion

In this study, we developed an NLP algorithm to report seizure frequencies from unstructured EHR text, and we assessed performance across institutions. Our algorithm is novel in that it interprets text to answer multiple seizure type frequencies within one note [2–5]. This methodology is well aligned with recent AAN epilepsy quality measures to document in greater clinical detail seizure types, frequencies, and time since last seizure [7]. The algorithm demonstrated good accuracy for notes within the same institution; however, performance was poorer with application to an outside institution.

Variation in language, content organization, or template use within the clinic note could account for discrepant performances between institutions. In the external set, the algorithm often erroneously identified old seizure frequencies. Additionally, the development dataset over-sampled patients with active epilepsy; we found the algorithm was not well trained to identify seizure freedom in the external set. Despite the well-established, updated International League Against Epilepsy operational classifications of seizure types (e.g., focal impaired awareness seizures, focal-to-bilateral tonic-clonic seizures, etc.), notes commonly

included non-standardized descriptions [8]. Additionally, there was great variation in how frequency was documented. While greater standardization and/or rigor in describing seizures could be helpful, it may be more practical to plan for broad inclusion of possible descriptive terms and language patterns in future seizure algorithm development. Future studies should focus on generalizability across institutions, a challenge highlighted by this multi-center study.

Limitations of this study include the exclusion of notes with ambiguous statements of seizure frequency; this was necessary for our study design but will be an important consideration for algorithm application in practice. Discriminating between seizures and non-epileptic spells was outside the scope of this study; our goal was to tabulate frequencies of unique semiologies. Notably, minor discrepancies tallied as algorithm errors may be clinically insignificant, and it is worth considering what resolution of seizure frequency reporting would be necessary to inform clinical care or to perform research studies [9]. There are certainly some scenarios in which a less meticulous estimate of seizure frequency may suffice, for example, examining the presence of seizures vs. seizure-freedom or alternatively if seizure frequency was described in broader buckets (e.g. daily vs. weekly vs. monthly).

5. Conclusions

Automated, accurate seizure frequency extraction would be beneficial to epilepsy patient care and epilepsy research. Our rules-based NLP tool for the extraction of seizure frequencies from clinical notes showed acceptable performance within the development institution but did not generalize well. These results suggest NLP text extraction and reporting of seizure frequency by type is feasible, and highlight the important challenge of generalizability across institutions for large-scale implementation.

Study funding

Barbara Decker is supported by NIH T32-NS-061779. Samuel Terman is supported by the Susan S Spencer Clinical Research Training Scholarship and the Michigan Institute for Clinical and Health Research J Award UL1TR002240. Kathryn Davis is supported by the Thornton Foundation, NINDS R01NS116504, NINDS R56NS099348. Brian Litt is supported by Pioneer Award Number DP1NS122038. Colin A Ellis is supported by the National Institute Of Neurological Disorders And Stroke of the National Institutes of Health Award Number K23NS121520; by the American Academy of Neurology Susan S Spencer Clinical Research Training Scholarship; by the Thomas B and Jeannette E Laws McCabe Fund at the University of Pennsylvania; and by the Mirowski Family Foundation. Pouya Khankhanian is supported by NIH T32-NS-091008. Chloe Hill is supported by NIH KL2TR002241.

Author Contributions

BD, SW, KD, BL, CE, PK and CH - Drafting/revision of the manuscript for content, including medical writing for content; major role in data acquisition; study concept or study design; analysis and interpretation of data.

AT, JX, NK and AJ - Drafting/revision of the manuscript for content, including medical writing for content; major role in data acquisition.

Declaration of Competing Interest

None.

Acknowledgements

Mirowski Family Foundation, Bonnie and Jonathan Rothberg.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.seizure.2022.07.010](https://doi.org/10.1016/j.seizure.2022.07.010).

References

- [1] Patel AD, Baca C, Franklin G, Herman ST. Quality improvement in neurology Epilepsy Quality Measurement Set 2017 update. Published online 2018:829–837. [10.1212/WNL.0000000000006425](https://doi.org/10.1212/WNL.0000000000006425).
- [2] Xie K, Litt B, Roth D, Ellis CA. Quantifying clinical outcome measures in patients with epilepsy using the electronic health record. In: Proceedings of the 21st workshop on biomedical language processing; 2022. p. 369–75. Published online, <https://aclanthology.org/2022.bionlp-1.36>.
- [3] Xie K, Gallagher RS, Conrad EC, et al. Extracting seizure frequency from epilepsy clinic notes: a machine reading approach to natural language processing. *J Am Med Inform Assoc* 2022;29(5):873–81. <https://doi.org/10.1093/jamia/ocac018>.
- [4] Decker BM, Hill E, Baldassano SN, Khankhanian P. Seizure : European journal of epilepsy can antiepileptic efficacy and epilepsy variables be studied from electronic health records ? A review of current approaches. 2021;85(September 2020):138–144. [10.1016/j.seizure.2020.11.011](https://doi.org/10.1016/j.seizure.2020.11.011).
- [5] Fonferko-Shadrach B, Lacey AS, Roberts A, et al. Using natural language processing to extract structured epilepsy data from unstructured clinic letters: development and validation of the ExECT (extraction of epilepsy clinical text) system. *BMJ Open* 2019;9(4):e023232. <https://doi.org/10.1136/bmjopen-2018-023232>.
- [6] Jamil A, Levinson N, Gelfand M, Hill CE, Khankhanian P, Davis KA. Efficacy and tolerability of clobazam in adults with drug-refractory epilepsy. *Neurol Clin Pract* 2020. <https://doi.org/10.1212/cpj.0000000000000992>. Published online.
- [7] Clary HM, Josephson SA, Franklin G, et al. Seizure frequency process and outcome quality measures: quality improvement in neurology. *Neurology* 2022;98(14):583–90. <https://doi.org/10.1212/WNL.0000000000200239>.
- [8] Fisher RS, Cross JH, Souza CD, et al. Instruction manual for the ILAE 2017 operational classification of seizure types. Published online 2017:531–542. [10.1111/epi.13671](https://doi.org/10.1111/epi.13671).
- [9] Choi H, Hamberger MJ, Munger Clary H, et al. Seizure frequency and patient-centered outcome assessment in epilepsy. *Epilepsia* 2014;55(8):1205–12. <https://doi.org/10.1111/EPI.12672/SUPPINFO>.