Final Report

Project Title: Optimal Methods for Notifying Clinicians about Epilepsy Surgery Patients

Principal Investigator: Judith W Dexheimer, PhD
Team Members:
Benjamin Wissel, BS
Hans M. Greiner, MD
Katie D. Holland-Bouley, MD, PhD
Robert Faist, MS
Francesco T. Mangano, MD
Tracey Glauser, MD
Nanhua Zhang, PhD
Rhonda D. Szczesniak, PhD
John Pestian, PhD, MBA

Organization: **Cincinnati Children's Hospital Medical Center**
Dates: **September 30,2016 – September 29, 2018**

Project Officer: Steve Bernstein

Funding: This study was funded by a grant from the Agency for Healthcare Research and Quality (AHRQ 1 R21 HS024977-01).

Grant #: AHRQ R21 HS24977-01

1. Abstract

Purpose: To prospectively evaluate a machine learning algorithm in identifying epilepsy candidates for neurological surgery consultation using natural language processing.

Scope: A machine learning algorithm was built and trained to classify pediatric candidates for epilepsy surgery consultation using natural language processing of free-text clinical notes.

Methods: To determine prospective performance, the algorithm was integrated into an electronic health record (EHR) system and used to evaluate patients in real-time over the course of one year. Patients identified by the algorithm were screened by two epileptologists for surgical candidacy. The sensitivity, specificity, positive- and negative-predictive value, and F1 score of the algorithm, and the kappa coefficient of agreement between the epileptologists were calculated. Performance of the algorithm was determined by the number of missed surgical candidates it identified.

Results: Of 6,395 patients with epilepsy seen in an outpatient neurology clinic between 10/31/16 - 10/30/17, the algorithm identified 200 as potential surgical candidates. As of 9/24/18, 12 (6%) of these patients have been referred for presurgical evaluation, and 2 (1%) have received surgical treatment. Two epileptologists confirmed that the algorithm found an additional 42/200 (21%) patients who should be considered for surgery. The NLP system increased the number of identified surgical candidates by 43% (N = 98 vs. 140; p = 2 * 10-14). The number needed to treat to achieve seizure freedom in one patient was 4.8. An EHR-integrated machine learning algorithm can, in real-time, aid clinicians in identifying patients with epilepsy who could benefit from surgery.

Key Words: Epilepsy, Machine Learning, Artificial Intelligence, Medical Informatics

2. Purpose

To prospectively evaluate a machine learning algorithm in identifying epilepsy candidates for neurological surgery consultation using natural language processing. To determine prospective performance, the algorithm was integrated into an electronic health record (EHR) system and used to evaluate patients in real-time over the course of one year. In the second year, the algorithm results were provided to the treating neurologist randomized in one of three ways. The specific aims were as follows:

Specific Aim 1: Implement and prospectively evaluate the existing NLP system by integrating the system with the electronic health record for patients identified as potential surgical candidates
Hypothesis: The NLP system can be fully integrated with the EHR and will have equivalent prospective sensitivity and specificity when compared to the retrospective values.

Specific Aim 2: Perform a clinical pilot test to evaluate the effectiveness of electronic alerts, honest broker reminders, and no intervention (standard of care) for eligible patients.
Hypothesis: Health IT-based, electronic health record reminders will be more effective at prompting physicians to consider surgical evaluations compared with honest broker reminders and the current standard of care.

3. Scope

Background and Significance

More than 143,000 children have drug-resistant epilepsy. Cincinnati Children's Hospital Medical Center (CCHMC) is a leader in pediatric epilepsy care. Because there is no systematic process for reviewing candidacy for surgical evaluation, it is necessary to standardize who is a neurosurgery candidate. We developed a novel Natural Language Processing (NLP) algorithm to detect patients who may be eligible for epilepsy surgical consults. The NLP can be integrated into neurology practice. Once implemented, we will be able to extend the knowledge to additional institutions and standardize care.

Significance

The scientific premise of the proposed project is that integrating an existing NLP system directly into an EHR and clinical care can help improve patient outcomes and care. To measure the effectiveness of the system, we will use two different alerting methodologies. We will concurrently measure which of the two alerting methodologies is more likely to impact physician decisions.

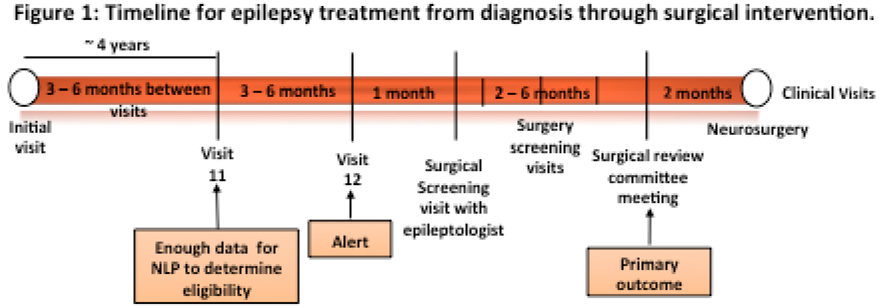
Epilepsy & Pediatrics

Epilepsy is a neurologic disorder characterized by seizures.16 It is one of the leading neurological disorders in the United States affecting more than 479,000 children and over 2 million adults.1,2 In 70% of the population with epilepsy, seizures are controlled with anti-epileptic medications.4 The other 30% are intractable who are not adequately controlled with medication.4 Of these, 55-59% of children may be seizure free after surgery and up to 77% may have improved quality of life with appropriate surgery.5-7 However, identifying surgical candidates is a laborious and complex process11 taking approximately 6 years from the date of epilepsy onset to surgery12. Early surgery has been shown to improve cognitive and seizure outcomes.17 Patient outcomes after surgery are good with only a 3% complication rate.10 While rules exist3 and early identification of pediatric patients is important;18 there is no streamlined process to identify patients meeting criteria for neurosurgical intervention.

Neurology Clinic

There are 29,000 patient encounters annually at the CCHMC neurology clinics. Epileptic patients are seen every 3-6 months in clinic. The neurologist modifies anti-epileptic medications as necessary if seizures continue. To be candidates for surgical evaluation, patients must have a diagnosis of intractable epilepsy and have at least 2 unsuccessful anti-epileptic drug trials. The general flow from initial visit to surgical intervention is shown in Figure 1. Currently CCHMC performs approximately 50 patient craniotomies annually.

Figure 1: Timeline for epilepsy treatment from diagnosis through surgical intervention.



Due to the risks of craniotomy, difficulty in clinically distinguishing epileptic syndromes in which surgery is indicated or contraindicated and rapidly evolving criteria for surgical candidacy, providers are cautious about recommending patients for surgical consults. Once recommended, the patient is sent for a visit with an epileptologist and, if indicated, surgical screening proceeds. Surgical screening takes 2-6 months. At the completion of screening, patient records are reviewed by the entire surgical review committee including neurologists, epileptologists, neurosurgeon, radiologists, and pathologists (see letter of support) before being recommended as a surgical candidate (Figure 1). The surgical complication rate is low at 3%, and earlier identification could offer significant quality of life improvement to eligible patients and their caregivers. 5-7,9,10

Current Methods for Determining Need for Surgical Evaluation Eligibility

Patients are seen as needed in the neurology outpatient clinics with visits approximately 3-6 months apart (figure 1). A neurologist is the primary physician treating the patient's epilepsy. Patients are trialed on anti-epileptic medications and adjustments to medications and dose are performed as needed. During a clinic visit, patients can be referred by neurology to an epileptologist for consideration of phase 1 pre-surgical evaluation. Patients who are eligible will have an appropriate noninvasive evaluation, and may go on to invasive evaluation with surgically implanted brain electrodes and/or ultimately craniotomy for surgical removal of seizure focus.

The clinic neurologist is the physician who will ultimately order the epilepsy consult. The NLP algorithm is a clinical decision support tool that will aid in the decision making process. Due to the sensitive nature of the surgical procedure that may result, we will confirm NLP-identified eligibility through validation, the oversight committee, and expert opinion that the patients we will be alerting on are potential surgical candidates.

Natural Language Processing

Natural Language Processing (NLP) is a theoretical technique for analyzing free text.19 NLP has a rich history and an extensive research base. The electronic health record (EHR) contains data relating to a patient's visit; as much as 30-50% of these data are only available in free text.20 Clinical care and research can benefit from using the unstructured text information.21,22

NLP is used for surveillance, adverse event detection,23-27 to identify patient medications,28 and to extract data from radiology reports.29-31 NLP can be applied to evaluate clinical notes and provide recommendations,13 but techniques are frequently experimental26 and not integrated into practice. In the case of intractable epilepsy patients, the NLP may help make surgical consults and evaluations sooner.12 There are limited studies illustrating the direct application of NLP to clinical practice;32 and NLP can help clinical decision support.26

Clinical Decision Support

Fifty-five percent of US medical institutions have EHRs;33 decision support is part of implementation and is "any program designed to help…make clinical decisions."34 It covers many aspects of care including patient-specific recommendations,35 information management,36 and guidelines.37 Decision support should be provided at the right place, the right person, and the right time38 and should be a tool not a hindrance.39 Effective decision support should follow recommendations40-42 and can improve compliance and patient outcomes.43

Support for the Use of the Decision Support in Epilepsy

The divisions of Neurology and Neurosurgery enthusiastically support the implementation and study of reminder methods to alert providers that patients may be eligible for surgical consults. (see Neurology Letter of Support) Translational research relating to epilepsy is ongoing. Comparing computers to physicians for selection of medications shows that algorithms can aid in drug selection.44,45 Research suggests that early identification of patients who may be eligible for surgical evaluation is important.18 And an informatics infrastructure that could be used for patient classification is vital46 but does not yet exist. Once created, the infrastructure can be applied across pediatric hospitals.

Epilepsy progress notes can be classified across hospitals,15 and the system can be implemented at additional pediatric institutions. This research will inform future Health IT research and drive successful implementations of NLP and ideal alerting mechanisms in neurological care. By creating a successful NLP-EHR integration coupled with the ideal alert, the entire system can be applied to additional NLP systems and conditions.

Decrease time to surgical evaluation

This project provides an innovative integration of NLP into the EHR and a novel comparison of an honest-broker alerting method compared to a fully computerized reminder. By providing a recommendation to suggest a surgical evaluation to patients identified most at risk earlier in the treatment process, we have the potential to decrease time from diagnosis to surgical evaluation, resulting in decreased suffering, improvement in quality of life for both the patients and caregivers, and a decreased treatment cost.52 The resulting integrated system can be generalizable across hospitals15 and used to leverage tertiary clinical expertise across large-scale healthcare networks.

Participants

Physicians: All neurologists treating patients during the study period will be eligible. Physicians will be recruited for participation. Prior to implementation, we will inform all attending physicians in the neurology clinic about the ongoing study through emails and education at faculty and staff meetings. These educational sessions will allow all providers to ask questions and address any concerns that may arise.

Patients: All patients identified by the NLP system and with a clinic visit during the study period will be eligible for inclusion. Patients will be randomized to the honest broker, EHR alert, or no intervention. Patients will be excluded if the oversight committee determines they are not eligible.

References

1. National Center for Chronic Disease Prevention and Health Promotion. CDC - Chronic Disease - Epilepsy - At A Glance. Available at: http://www.cdc.gov/chronicdisease/resources/publications/aag/epilepsy.htm. Accessed April 29, 2014.
2. National Survey of Children's Health. NSCH 2011/12. *Data query from the Child and Adolescent Health Measurement Initiative, Data Resourse Center for Child and Adolescent Health website* Available at: http://www.childhealthdata.org/browse/survey/results?q=2557&r=1. Accessed April 29, 2014.
3. Kwan P, Arzimanoglou A, Berg AT, et al. Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies. *Epilepsia.* 2010;51(6):1069-1077. PMID: 19889013.
4. Kwan P, Brodie MJ. Early identification of refractory epilepsy. *N Engl J Med.* 2000;342(5):314-319. PMID: 10660394.
5. Moseley BD, Nickels K, Wirrell EC. Surgical outcomes for intractable epilepsy in children with epileptic spasms. *J Child Neurol.* 2012;27(6):713-720. PMID: 22123426.
6. Paolicchi JM, Jayakar P, Dean P, et al. Predictors of outcome in pediatric epilepsy surgery. *Neurology.* 2000;54(3):642-647. PMID: 10680797.
7. Wyllie E, Comair YG, Kotagal P, Bulacio J, Bingaman W, Ruggieri P. Seizure outcome after epilepsy surgery in children and adolescents. *Ann Neurol.* 1998;44(5):740-748. PMID: 9818929.
8. Berg AT, Vickrey BG, Langfitt JT, et al. The multicenter study of epilepsy surgery: recruitment and selection for surgery. *Epilepsia.* 2003;44(11):1425-1433. PMID: 14636351.
9. Loddenkemper T, Holland KD, Stanford LD, Kotagal P, Bingaman W, Wyllie E. Developmental outcome after epilepsy surgery in infancy. *Pediatrics.* 2007;119(5):930-935. PMID: 17473093.
10. Greiner HM, Horn PS, Arya R, et al. Acute postoperative seizures and long-term outcome following pediatric epilepsy surgery [ePub ahead of print]. *Seizure.* Feb 28, 2014. PMID: 24680553.
11. Ryvlin P, Rheims S. Epilepsy surgery: eligibility criteria and presurgical evaluation. *Dialogues Clin Neurosci.* 2008;10(1):91-103. PMID: 18472487.
12. Matykiewicz P, Cohen KB, Holland KD, et al. Earlier Identification of Epilepsy Surgery Candidates Using Natural Language Processing. *ACL 2013.* 2013:1.
13. Deleger L, Brodzinski H, Zhai H, et al. Developing and evaluating an automated appendicitis risk stratification algorithm for pediatric patients in the emergency department. *J Am Med Inform Assoc.* 2013;20(e2):e212-220. PMID: 24130231.
14. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform.* 2009;42(5):760-772. PMID: 19683066.
15. Connolly B, Matykiewicz P, Bretonnel Cohen K, et al. Assessing the similarity of surface linguistic features related to epilepsy across pediatric hospitals. *J Am Med Inform Assoc.* 2014. PMID: 24692393.
16. Chang BS, Lowenstein DH. Epilepsy. *N Engl J Med.* 2003;349(13):1257-1266. PMID: 14507951.
17. Skirrow C, Cross JH, Cormack F, Harkness W, Vargha-Khadem F, Baldeweg T. Long-term intellectual outcome after temporal lobe surgery in childhood. *Neurology.* 2011;76(15):1330-1337. PMID: 21482948.

18. Glauser TA, Loddenkemper T. Management of childhood epilepsy. *Continuum (Minneap Minn).* 2013;19(3 Epilepsy):656-681. PMID: 23739103.
19. Pestian J, Deleger L, Savova G, Dexheimer JW, Solti I. Natural Language Processing - The Basics In: Hutton JJ, ed. *Pediatric biomedical informatics computer: applications in pediatric research*. New York: Springer,; 2012:149-172.
20. Hicks J, Rand Graduate School. *The potential of claims data to support the measurement of health care quality.* Santa Monica, CA.: RAND; 2003.
21. Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc.* 2005;12(4):448-457. PMID: 15802475.
22. Murff HJ, FitzHenry F, Matheny ME, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA.* 2011;306(8):848-855. PMID: 21862746.
23. Bates DW, Evans RS, Murff H, Stetson PD, Pizziferri L, Hripcsak G. Detecting adverse events using information technology. *J Am Med Inform Assoc.* 2003;10(2):115-128. PMID: 12595401.
24. Petratos GN, Kim Y, Evans RS, Williams SD, Gardner RM. Comparing the effectiveness of computerized adverse drug event monitoring systems to enhance clinical decision support for hospitalized patients. *Appl Clin Inform.* 2010;1(3):293-303. PMID: 23616843.
25. Tinoco A, Evans RS, Staes CJ, Lloyd JF, Rothschild JM, Haug PJ. Comparison of computerized surveillance and manual chart review for adverse events. *J Am Med Inform Assoc.* 2011;18(4):491-497. PMID: 21672911.
26. Ye Y, Tsui FR, Wagner M, Espino JU, Li Q. Influenza detection from emergency department reports using natural language processing and Bayesian network classifiers. *J Am Med Inform Assoc.* 2014. PMID: 24406261.
27. Conway M, Dowling JN, Chapman WW. Using chief complaints for syndromic surveillance: a review of chief complaint based classifiers in North America. *J Biomed Inform.* 2013;46(4):734-743. PMID: 23602781.
28. Savova GK, Olson JE, Murphy SP, et al. Automated discovery of drug treatment patterns for endocrine therapy of breast cancer within an electronic medical record. *J Am Med Inform Assoc.* 2012;19(e1):e83-89. PMID: 22140207.
29. Dublin S, Baldwin E, Walker RL, et al. Natural Language Processing to identify pneumonia from radiology reports. *Pharmacoepidemiol Drug Saf.* 2013;22(8):834-841. PMID: 23554109.
30. Elkin PL, Froehling D, Wahner-Roedler D, et al. NLP-based identification of pneumonia cases from free-text radiological reports. *AMIA Annu Symp Proc.* 2008:172-176. PMID: 18998791.
31. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc.* 1994;1(2):161-174. PMID: 7719797.
32. Fiszman M, Haug PJ. Using medical language processing to support real-time evaluation of pneumonia guidelines. *Proc AMIA Symp.* 2000:235-239. PMID: 11079880.
33. Jamoom E, Beatty P, Bercovitz A, Woodwell D, Palso K, Rechtsteiner E. Physician adoption of electronic health record systems: United States, 2011. *NCHS Data Brief.* 2012(98):1-8. PMID: 23050588.
34. Musen MA, Shakar Y, Shortliffe EH. Clinical Decision-Support Systems In: Shortliffe EH, Cimino JJ, eds. *Biomedical informatics : computer applications in health care and biomedicine*. 3rd ed. New York: Springer; 2006:698-736.
35. Slagle JM, Gordon JS, Harris CE, et al. MyMediHealth - designing a next generation system for child-centered medication management. *J Biomed Inform.* 2010;43(5 Suppl):S27-31. PMID: 20937481.
36. Chute CG, Beck SA, Fisk TB, Mohr DN. The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. *J Am Med Inform Assoc.* 2010;17(2):131-135. PMID: 20190054.
37. Bell LM, Grundmeier R, Localio R, et al. Electronic health record-based decision support to improve asthma care: a cluster-randomized trial. *Pediatrics.* 2010;125(4):e770-777. PMID: 20231191.
38. Sirajuddin AM, Osheroff JA, Sittig DF, Chuo J, Velasco F, Collins DA. Implementation pearls from a new guidebook on improving medication use and outcomes with clinical decision support. Effective CDS is essential for addressing healthcare performance improvement imperatives. *J Healthc Inf Manag.* 2009;23(4):38-45. PMID: 19894486.
39. Friedman CP. A "fundamental theorem" of biomedical informatics. *J Am Med Inform Assoc.* 2009;16(2):169-170. PMID: 19074294.

40. Garg AX, Adhikari NK, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA.* 2005;293(10):1223-1238. PMID: 15755945.
41. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ.* 2005;330(7494):765. PMID: 15767266.
42. Sittig DF, Wright A, Osheroff JA, et al. Grand challenges in clinical decision support. *J Biomed Inform.* 2008;41(2):387-392. PMID: 18029232.
43. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. *Ann Intern Med.* 2014;160(1):48-54. PMID: 24573664.
44. Pestian J, Matykiewicz P, Holland-Bouley K, Standridge S, Spencer M, Glauser T. Selecting anti-epileptic drugs: a pediatric epileptologist's view, a computer's view. *Acta Neurol Scand.* 2013;127(3):208-215. PMID: 22998126.
45. Pestian J, Spencer M, Matykiewicz P, Zhang K, Vinks AA, Glauser T. Personalizing Drug Selection Using Advanced Clinical Decision Support. *Biomed Inform Insights.* 2009;2:19-29. PMID: 19898682.

4. Methods

There are approximately 29,000 neurology patient encounters, including 12,000 epilepsy encounters from 6,500 patients, and 100 patients evaluated for epilepsy surgery at Cincinnati Children's Hospital Medical Center (CCHMC) annually. Appropriate surgical referrals for patients with epilepsy are made by neurologists according to International League Against Epilepsy (ILAE) criteria and American Academy of Neurology (AAN) guidelines. Once recommended, patients visit an epileptologist and, if indicated, presurgical evaluation proceeds. A surgical review committee, which includes neurologists, epileptologists, neurosurgeons, radiologists, pathologists, and psychologists, reviews patient records before recommending patients for surgery. This process generally takes 2-6 months.

Study Design

The design will include a prospective evaluation of the NLP system, then two different alerting methods integrated into practice compared to standard of care. The design will include the following steps: evaluation of patients currently identified by the NLP system as surgical-consult eligible, a weekly evaluation of identified patients with upcoming appointments; then providing these recommendations to the neurologists. We will integrate the recommendations from the NLP system with the EHR. The EHR is a hospital-wide information system, Epic (Verona, WI), that has been in place since 2008.

<u>Phase 1: Prospective Evaluation of the NLP-identified Patient List</u>

This phase will be devoted to the integration and validation of the NLP-identified patients who may be eligible for epilepsy neurosurgical consults. This aim will provide a gold standard set of patients. Specific Aim 1: Implement and prospectively evaluate the existing NLP system by integrating the system with the electronic health record for patients identified as potential surgical candidates Hypothesis: The NLP system can be fully integrated with the EHR, and will have equivalent prospective sensitivity and specificity when compared to the retrospective values.

<u>Oversight Committee Participants</u>

The oversight committee consists of 2 attending neurologists (HG, KB), 1 neurosurgeon (FM), 1 NLP expert (JP), and 1 informaticist (JD). This group will meet monthly to track study progress, ensure safety and appropriateness of recommendations, and combat any potential issues that may arise. The physician members of the oversight committee will participate in patient verification from the NLP-system in Aim 1 and will weekly evaluate responses to recommendations in Aim 2.

<u>Justification and Feasibility</u>

While the system was developed and evaluated retrospectively, it has not been integrated with the EHR or prospectively evaluated. These steps are necessary to provide alerts and recommendations. Our retrospective data presents a sensitivity = 98%, specificity = 79%, positive predictive value = 96%, and negative predictive value = 90%.12

<u>Study Setting</u>

This study will occur in conjunction with the neurology and neurosurgery divisions at CCHMC.

<u>Research Design</u>

6,163 chronically ill patients have enough data to have reliable NLP recommendation results. 593 of these have been identified as "unknown," these will be evaluate for eligibility. Of these, 89 have an upcoming visit scheduled with their neurologist.

<u>Integration</u>

To implement the NLP system, we will prospectively evaluate the NLP recommendations and then integrate the output with the EHR. First we will evaluate a sample of patients identified by the NLP system focusing initially on those with upcoming appointments. We will provide all relevant clinical information to the two neurologists including the notes, medications, and medical record number. The neurologists will assign the international league against epilepsy (ILAE) criteria.3 Five percent of these data will be randomly selected and used for assessing inter-rater reliability. Discrepancies will be reviewed and reconciled by an adjudicating neurosurgeon and oversight committee. In each review, the neurologists will separately assess charts and make a surgical consult determination. The order in which the neurologists review the charts, both in the inter-rater-reliability phase and the complete-review phase, will be randomly assigned. From previous work, it takes 1 hour to review each chart. Once all charts are reviewed, we will convene the oversight committee for a review of each of the patients to discuss disagreements. Disagreements will be resolved by consensus.

The NLP system will be fully integrated with the Epic EHR. The NLP system will be run asynchronously to patient care on a weekly basis; it will not be run on-demand. Patients with a neurology visit in the upcoming week will be eligible for randomization for Aim 2. We expect approximately 1-2 patients/week. The NLP system integration will be through a web-based API (Bridges). The data will be sent to the EHR automatically and stored in secure biomedical informatics servers. The NLP system will be run weekly with recommendations provided to the neurologists on Sunday evenings when a patient has an appointment that week.

Measures

The expected outcomes from Aim 1 are the informatics infrastructure to support integration, the prospective evaluation of the NLP system, and a set of patients who are potentially eligible for epilepsy surgical evaluation. These patients will be the sample for aim 2. We will collect the inter-rater agreement; we will follow if any of the identified patients had a neurosurgical consult and/or resulting neurosurgery. As a primary outcome, we will calculate the sensitivity, specificity, positive and negative predictive values of the NLP-system and F1 measures as verified by the expert clinicians and compare these prospective operational characteristics to the same retrospective characteristics. Secondary outcomes will include the inter-rater agreement of the two neurologists and demographic information for patients identified by the NLP system.

Statistical Analysis

The analysis for this aim will occur in two stages. In the first stage, we will measure the inter-rater reliability between the two neurologists using Cohen's kappa statistic.67 In the second stage, we will compute the sensitivity and specificity for the NLP system to identify patients with respect to the neurologist identification. The value of F1 will be calculated by taking the product of sensitivity and specificity. The 95% CI for each estimate will be obtained using bootstrap methods.

We estimated the precision with which we can assess the sensitivity in our prospective study, given the findings from our previously published results. In that preliminary study, we had a sensitivity (95% CI) of .0.963 (0.957 to 0.970). We use the width of this 95% z-interval to indicate the precision of the sensitivity estimate. By enrolling a minimum of 300 patients, we will have sufficient precision (0.010) with which we can estimate the ability of the NLP to correctly identify referral patients.

Phase 2: Implementation and Evaluation of Alerts in the Neurology Clinic

This phase will be devoted to the integration, implementation and evaluation of three different interventions.

Justification and Feasibility

Electronic alerts have been shown to improve practitioner performance. They are a common method of providing recommendations to clinicians to perform tasks and aid in complex disease management. There are currently 593 patients the NLP system identified as eligible for a surgical consult and 89 of them have upcoming visits. As the study progresses, we expect more patients to schedule visits. The typical time between neurology visits for a patient with intractable epilepsy is 3-6 months.

We will evaluate several methods to determine the ideal way to alert providers using the honest broker approach including methods that were successful in the literature such as emails,74 telephone calls,75 and others suggested by neurology faculty. The honest broker reminder will provide the same information found in the electronic alert. While these studies are not offering reminders for the same gravity of situation, using an effective reminder approach helps to ensure a more successful intervention.

Study Setting
This study will occur in conjunction with the neurology and neurosurgery divisions at CCHMC

Research Design
We propose to study the effectiveness of two different alerting reminders compared to no intervention: an honest broker-delivered alert and an EHR alert. Given the sensitive nature of recommending a patient for epilepsy surgical evaluation and evaluation we will examine an honest broker alert method. The honest-broker is commonly used in biobanks.76 We will use a third party to deliver the NLP recommendations to the neurologists. We will trial the two methods to see which is more effective. Effectiveness of the alerting method is defined as a patient being sent for surgical consult or an appropriate refusal by the treating neurologist.

The oversight committee will oversee implementation and prospective evaluation (Aim 1). During phase 2, we will use data visualization to evaluate how well each intervention is progressing.77 The data visualization method requires definitions of behavior to passive, negative and positive to ensure comparable interventions.

Intervention Design
We will design the two interventions with the participation and feedback of the neurology clinic physicians. The EHR alert will be designed using principles of human factors engineering through user-centered design (UCD) process, which will include iterative design workshops and usability testing.78 Low-fidelity mock-ups will be used to explore usability, functionality, and context of alerts as well as envisioning alternative interactions during design workshops. The design workshops may be augmented by individual cognitive walkthrough sessions with practitioners. During these walkthroughs, practitioners will be shown interim mock-ups and asked to describe what each element on the interface does, where they would see this alert occur and concerns or issues that they could see. This phase will culminate in working prototype alerts that will be evaluated through formal usability testing prior to implementation.79,80 In usability tests, we will invite clinicians to evaluate the prototypes, assess system functionality, and uncover use errors, design flaws, and potential barriers to effective use. We will limit the alert adaptations to constraints within the Epic EHR, but generalized to other commercial EHRs. We will electronically survey neurologists to determine their opinions on providing recommendations, when they would like to receive them, and how they would prefer to be informed. By participating in faculty meetings, individual interviews, design, and clinic observation, we will present the most informative and least intrusive recommendations to providers.

Intervention 1: Honest broker: We will trial several honest broker delivery methods depending on input from neurology. Our honest broker trial may include an automated email with contact information for an epilepsy expert, an EHR in-basket message, or a telephone call. Each of these three methods will be vetted in faculty meetings and through emails to ensure that a) the information the neurologist needs to make a decision is available; and b) the information to empower the neurologists to place trust in the NLP recommendation and validation is available. We will ask for clinician feedback within one week after the eligible patient's visit to assess opinions about the honest broker recommendation and iteratively make improvements.

Intervention 2: EHR alert: We will do semi-structured interviews with a convenience sample of neurologists and observations to determine EHR interaction during patient care. We will design an EHR embedded alert tied to a patient chart and turned on one week prior to the scheduled visit. The alert will contain the existing refusal reasons collected as epilepsy quality measures. It will display once per patient visit to minimize alert fatigue.81-83 As patients are seen once every 3-6 months, we estimate 2 alerts per provider per week. We will ask for clinician feedback one week after the patient's visit to assess opinions.

Prior to implementation, we will attend faculty and staff meetings and send out informational emails starting one month prior to implementation to educate the physicians about the upcoming study. Upon each encounter, the patient will be randomly assigned for their treating physician to receive one of the interventions.

Data Collection

Data will be collected from the NLP system, and the EHR through chart reviews. A clinical research associate will abstract patient information using a standardized data collection tool. We will collect surgical evaluation refusal reasons already present in the EHR for all patients presenting to the neurology clinic. We will collect demographic data including any missed eligible by the NLP system as those patients who have a surgical consult but were not identified by NLP. All patients will be followed throughout the study period.

Measures

The expected outcomes from aim 2 are responses to alerts measured through patients sent for surgical consult and refusal reasons from the two alerts. We will compare the intervention groups with our primary outcome of the patient receiving a surgical consult or appropriate recommendation decline as defined by the oversight committee. Secondary outcomes will include surgical consults patients not identified by the system, time from diagnosis to surgical consult, refusal reasons by the providers, and demographic data.

Statistical Analysis

To determine the most effective alerting method, we will estimate the probability of receiving surgical consult for the three intervention groups using generalized estimating equations (GEE) with a logit link function.84 The GEE approach will be used to account for the clustered nature of the data, due to some of the physicians participating in more than one encounter. An odds ratio (OR) and 95% CI will be calculated for each pairwise comparison of interventions. The intervention with the highest likelihood for the receipt of surgical consults will be considered the most effective. For our secondary outcomes, we will use a logistic regression model to estimate the probability of surgical patient not being identified by the system. We will use survival analysis to estimate the median time from diagnosis until surgical evaluation. We will use a descriptive analysis to examine the frequency and percentage of refusal reasons. Members of the research team will conduct semi-structured interviews with physicians to understand 1) perspectives on the information to identify appropriate candidates for research, 2) feedback on the alerting mechanisms and 3) recommendations for improvement.

Power

With an expected 80% of existing identified patients to be eligible for surgical consult, we expect approximately 136 eligible patients with roughly 45 in each group. We calculated power across a range of sample sizes for the primary outcome (whether the subject received a surgical consult). Power was calculated assuming that the OR that corresponds to the greatest difference between interventions is 2.70 and using the Chi-square test for differences between two proportions. To adjust for the three comparisons, power estimates are based on an alpha level of 0.017. If we enroll 136 patients, we will have sufficient power (86.5%) to detect the OR corresponding to a medium effect size (0.30) for the maximum difference between interventions.

5. Results

Phase 1 Results

There were 27,769 ambulatory neurology visits over the one-year study period. Of these, 12,019 were epilepsy-related from 6,395 individual patients. The NLP system identified 200 patients as potential surgical candidates. Patients were 2 to 48 years of age. Demographics are shown in Table 1. There was a positive correlation between the number of neurology office visits and the proportion of patients recommended by the NLP system ($p = 0.04$) (Figure 2A). The majority of patients who were recommended by the NLP system had between 5 and 22 neurology visits (173/200; 77%) (Figure 2B). The NLP system's classifications were more accurate when given notes from more office visits ($p = 4 * 10^{-4}$) (Figure 2C).

Table 1. Demographics of patients identified as surgical candidates by the machine learning algorithm.

	Patients found by NLP (N = 200)			**All epilepsy patients (N = 6,395)**			
	N	**%**	**Mean ± SD**	**N**	**%**	**Mean ± SD**	**p-value**
Age (yrs)			15.0 ± 8.4			13.0 ± 7.3	<0.001
Gender							0.318
Female	84			2,914			
Male	116			3,481			
Race							0.299
White	169			5,153			
Black or African-American	22			712			
Asian	3			99			
Other	3			107			
Multi-racial	1			230			
Unknown	2			94			
Number of Neurology Office Visits	200	100	14.1 ± 7.5	6,395	10	9.2 ± 7.5	<0.001

N: number of patients; SD: standard deviation

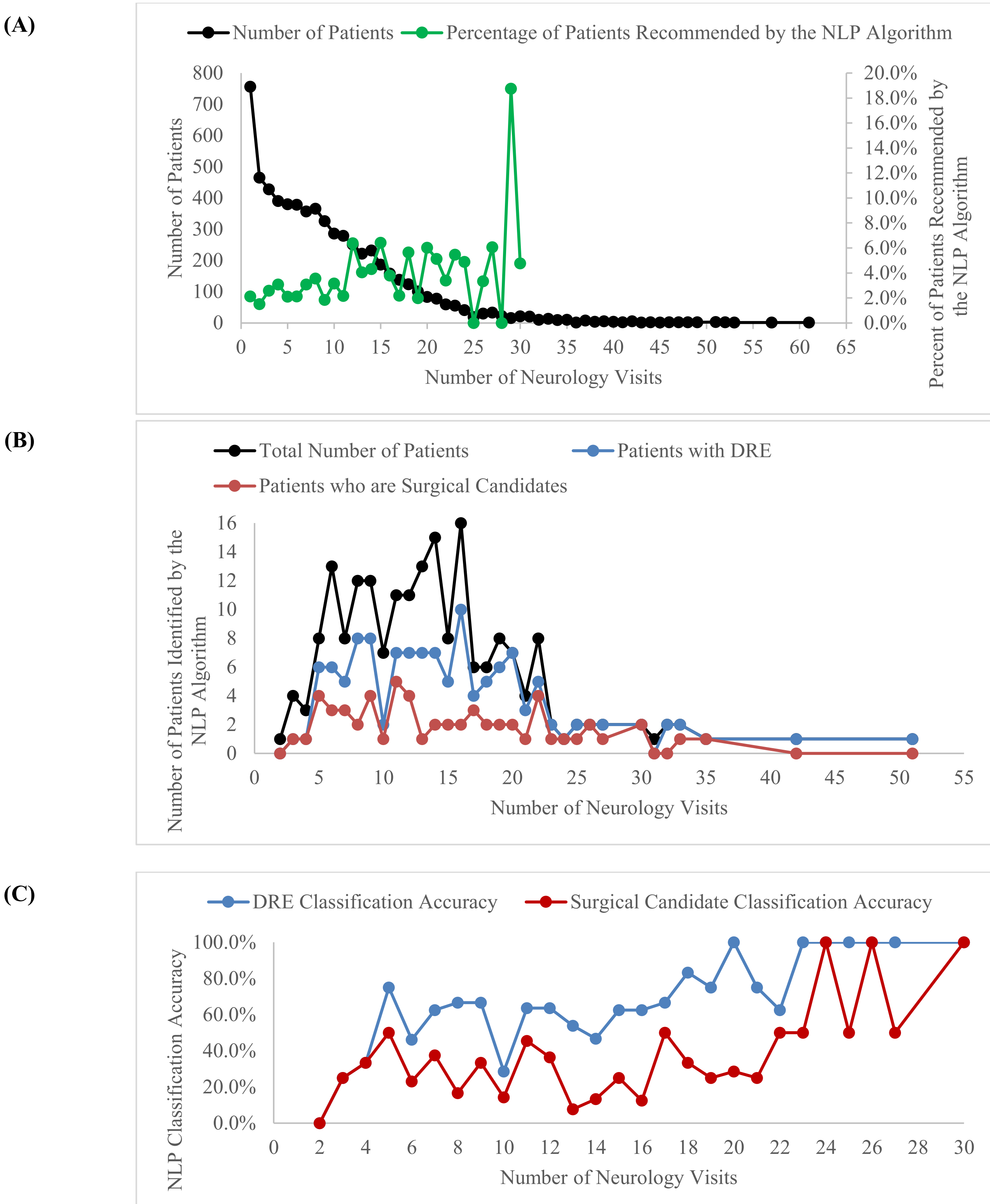


Figure 2: (A) Total number of patients with epilepsy at CCHMC and the proportion of patients recommended by the NLP system by the number of outpatient neurology visits. (B) Number of patients recommended by the NLP system by the number of outpatient neurology visits. (C) Percentage of patients recommended by the NLP system who had drug-resistant epilepsy or were surgical candidates by the number of outpatient neurology visits. Outliers who had more than 3X office visits (8/200 [4%]) were excluded from (A) and (C).

Of the 200 patients identified, 12 (6%) were referred for presurgical evaluation and 2 (1%) have received resective surgical treatment (not including neurostimulator implantation), as of 9/24/18, 23 months since the study began. NLP system identification antedated referrals for presurgical evaluation in 11/12 (92%) cases. A description of these 12 patients is provided in Table 2.

A summary of patient flow through the NLP system is shown in Figure 3. The NLP system identified a total of 54 potential candidates (27% of patients recommended by the NLP system). Of the surgical candidates, 42/54 (79%) have not been referred for a presurgical evaluation by their providers. Of all patients identified by the NLP system, 128/200 (64%) had DRE. Of patients recommended by the NLP system who had DRE, and additional 74/128 (58%) would have been considered surgical candidates if not for the following reasons: patient was developmentally devastated and surgery would not substantially increase quality of life (N = 28), though diagnosed with DRE, it was unclear from the EHR if the patient had seizures in the previous year (N = 17), patient had a medical contraindications to epilepsy surgery (N = 12), patient had strong clinical, MRI, and/or EEG evidence of a multi-focal or generalized epilepsy syndrome (N = 10), patient was already evaluated for and/or underwent surgical treatment (N = 6), or the patient/family was comfortable with current level of seizure control (N = 1). The remaining 72/200 (36%) were true misclassifications. The NLP system increased the number of identified surgical candidates by 43% (N = 98 vs. 140; p = 2 * 10-14). The NNT to achieve seizure freedom in one patient was 4.8.

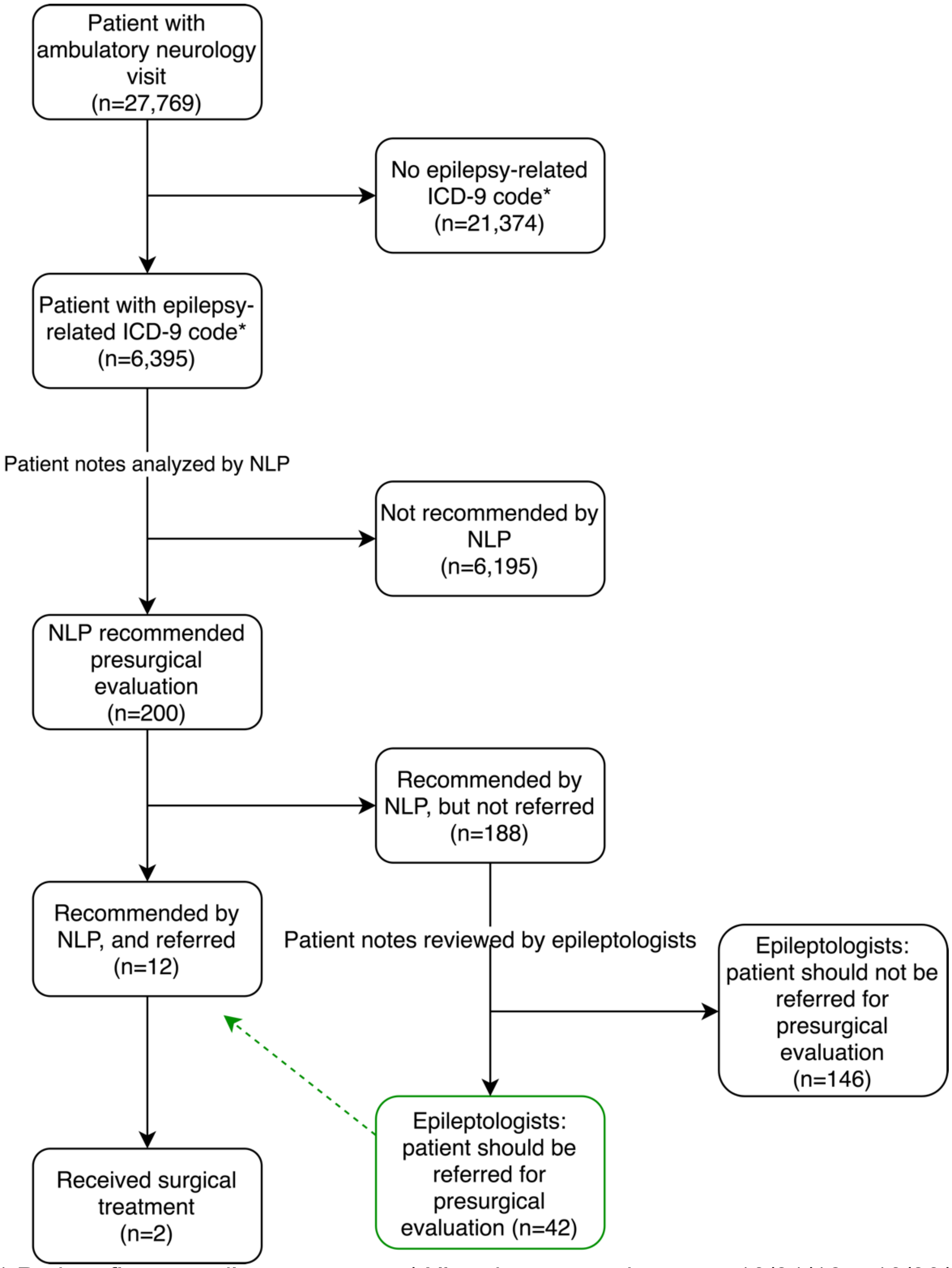


Figure 3: Phase 1 Patient flow to epilepsy surgery. *All patients seen between 10/31/16 – 10/30/17 at Cincinnati Children's Hospital Medical Center (CCHMC) with ICD-9 codes for epilepsy or convulsions (345.*, 780.3*, and 779.0) were eligible. Of 6,395 patients with epilepsy with outpatient neurology visits during the study period, clinicians referred 98 for presurgical evaluation and 39 received surgery.

During the study period, there were 98 referred for presurgical evaluation and 46 patients who received epilepsy surgery. The NLP system identified 1/46 (2%) patients who received epilepsy surgery and 12/98 (12%) patients who were referred for presurgical evaluation. In these 12 cases, the NLP system recommendations antedated patient referral and surgery in all cases, except for Patient #7 in Table 2. The NLP system identified Patient #7 9 days after his/her referral.

Table 2. List of patients identified as surgical candidates by the NLP system who were referred for presurgical evaluation. NLP alerts antedated referral for presurgical evaluation in 11/12 (92%) cases.

Patient ID	NLP ID Date	Referral Date	Time Elapsed (days)	Surgery Date	Reason for No Surgery
1	10/31/16	7/25/17, 3/13/2018	498	N/A	VNS on 5/31/18
2	12/5/16	10/11/17	310	N/A	Social
3	1/8/17	1/13/17	5	7/11/17	N/A
4	2/5/17	2/10/17	5	4/10/18	N/A
5	2/12/17	2/17/17	5	N/A	Inconclusive workup
6	3/19/17	3/23/17	4	N/A	Social
7	3/19/17	3/10/17	-9	N/A	Poor surgical candidate
8	5/14/17	5/18/17	4	N/A	Poor surgical candidate
9	7/9/17	7/13/17	4	N/A	Social
10	7/16/17	9/15/17	61	N/A	Medical
11	8/6/17	1/3/18	150	N/A	Poor surgical candidate; considering VNS
12	8/6/17	3/30/18	236	N/A	Poor surgical candidate; considering VNS

ID: identification; VNS: vagal nerve stimulator; N/A: not applicable

Cohen's kappa statistic for the epileptologists' inter-rater reliability was moderate, at 0.61 (95% C.I.: 0.49 – 0.73; p = 0.027).34 NLP system's sensitivity was 0.39, specificity was 0.98, positive predictive value was 0.27, and negative predictive value was 0.99. The resulting F1 score was 0.38.

Phase 2 Preliminary Results

There were approximately 28,000 ambulatory neurology visits during the study period. Of these, 11,937 were epilepsy-related from 6,418 individual patients. The CDSS ran weekly on Sunday nights and identified 176 patients as potentially eligible for a presurgical evaluation. Patients were 2 to 36 years of age. Demographics are shown in Table 2.

Table 2: Demographics of patients identified as surgical candidates by the natural language processing system.

	Patients found by NLP (N = 176)			All epilepsy patients (N = 6,418)			
	N	**%**	**Mean ± SD**	**N**	**%**	**Mean ± SD**	**p-value**
Age (yrs)			13.8 ± 8.0			12.7 ± 7.3	0.119
Gender							
Female	46	42		2,920	45		0.493
Male	63	58		3,498	55		
Race							
White	91	83		5,203	81		0.523
Black or African-American	13	12		700	11		0.735
Asian	2	2		109	2		0.913
Hispanic/Latino	1	1		35	1		0.603
Multi-racial	2	2		212	3		0.378
Other	0	0		50	1		0.355
Unknown	0	0		119	2		0.151
Number of Neurology Office Visits							

NLP: Natural language processing system; N: number of patients; SD: standard deviation

The NLP system generated a cumulative total of 121 physician alerts and 61 null alerts. A summary of these alerts and the actions taken by physicians is shown in Table 3. Honest broker email and EHR alerts prompted similar provider responses. They consistently responded to both alert types (55/60 ([90%] and 60/61 [95%], respectively; p = 0.80) and elected to be re-alerted (55/60 [90%] and 60/61 [95%], respectively; p = 0.80). The odds ratio of patient referral after electronic alert compared to no alert was 2.5 (95% CI: 0.8-5.0; p = 0.21), and 2.8 (95% CI: 0.9-5.1; p = 0.20) compared to patients not identified by the CDSS.

Table 3: Provider responses to natural language processing system alerts.

Intervention	**Email**	**EHR alert**	**None**	**P-value**
N	60	61	61	
Received alert feedback	58	60	n/a	
Referrals	2	3	1	
Response:				
Epilepsy surgery discussed but patient declined procedure	5	4	n/a	
Patient was previously referred to epilepsy surgery program	1	4	n/a	
Patient is currently being referred to the epilepsy surgery program	1	0	n/a	
Patient has medical contraindications to surgery	2	2	n/a	
Patients epilepsy does not fill ILAE's criteria for intractability	25	29	n/a	
Referral considered but deferred at this time	23	22	n/a	
Re-prompt in 6 months? (yes)	51	55	n/a	

*According to International League Against Epilepsy (ILAE) criteria.[3] EHR: electronic health record; N: number of patients

Discussion

In our study, a machine learning algorithm successfully interpreted free-text and structured data from EHR notes to identify patients with epilepsy who were strong candidates for surgical treatment in real time. The NLP system identified patients earlier than clinician 94% of the time and increased the number of identified surgical candidates by 43%. This exemplifies how machine learning can be used to identify patients with complex, chronic disorders and recommend them to their providers for advanced surgical care. NLP is an emerging screening tool in epilepsy which can save providers and patients time and improve effectiveness of visits. In order to come to the same conclusion as the NLP system, a provider would have to read all previous office notes and review EEG and MRI information during a follow-up visit. In our experience, this requires over an hour for the average patient with at least a 2-year history of epilepsy care. Given the time constraints enforced upon providers by the current healthcare environment, this is not feasible for most visits. NLP decision support makes consideration of epilepsy surgery feasible for every follow-up visit in every epilepsy patient. This improves standardization of care in epilepsy and could result in decreased time to referral for surgical evaluation, in addition to identifying patients who would not otherwise be referred at all.

This is the first report of an EHR surveillance system (EHRSS) that provides surgical treatment decision support in real time. However, other automated systems have been developed for identification, health care quality improvement, and clinical decision support. Clinical decision support systems have been utilized to provide diagnostic support19 and predict neurosurgical outcomes. In other studies, surveillance systems have been used to identify rare syndromes and disease outbreaks. The degree to which these applications have been implemented and prospectively validated varies widely.

Although effective, our NLP algorithm was not perfectly accurate. NLP, in general, has limitations in its performance interpreting shorthand, abbreviations, and idioms in free-text notes. However, clinicians have the ability to disregard these misclassifications. When designing this NLP algorithm, we considered it more valuable to maximize the catchment of missed surgical candidates. This is explained in previous work and demonstrated here in the results. Even without increasing the overall accuracy, the NLP system’s threshold for surgery recommendation can be tuned according to the clinician’s preference for high sensitivity vs. specificity. Interestingly, the NLP system did not identify any patients who already received surgical treatment during the study period. These patients were not explicitly excluded from being evaluated by the NLP system, but the NLP did not duplicate the clinicians’ recommendations. We speculate that the language used in these patients’ EHR notes changed after they were identified as surgical candidates by their providers. The NLP system was trained on patients who had undergone epilepsy surgery, but only on their notes prior to being referred for presurgical consultation. Language in the EHR notes are likely to change once these patients were referred, explaining the why algorithm did not recommend 97/98 (99%) patients who were already referred in this prospective study. Additionally, we suspect that more than 12/54 (22%) NLP-identified surgical candidates would have been referred for presurgical consultation if the NLP system’s recommendations were forwarded to providers. The study team designed this study to validate the prospective performance of the NLP system to ensure its recommendations to providers are meaningful.

Some factors that limit this study, and NLP broadly, should be considered. First, misclassifications limited the F1 score of the NLP system. However, many patients considered here as misclassifications indeed met ILAE criteria for intractable epilepsy,12 and were poor surgical candidates for others reasons. Since the NLP system automatically re-trains itself on these and all new patients, its accuracy is likely to increase over time. Additional features can be incorporated into the algorithm as patterns in incorrectly classified patients reveal themselves, such as excluding patients who are developmentally devastated from an invasive brain tumor, for example. Efforts on this front could yield the ability to predict prognosis37 on top of its current binary surgical candidacy classifications. Also, although there were a large number of patients screened by the NLP system, the system’s classifications could be biased toward the population served by CCHMC.

We are performing additional analysis for phase 2 and additional chart review to verify why patients were missed from the alerts.

6. List of Publications and Products

Presentation

1. Dexheimer JW, Greiner H, Holland-Bouley K, Faist R, Mangano F, Pestian J. Prospective Evaluation of a Natural Language Processing System for Epilepsy Identification. Council on Clinical Information Technology: American Academy of Pediatrics. Orlando, FL. 2018. (attached)
2. Artificial Intelligence and Machine Learning in Biomedical Research. NIH Workshop: Harnessing Artificial Intelligence and Machine Learning to Advance Biomedical Research. Bethesda, MD. July 23, 2018.

Under Review

3. Dexheimer JW, Wissel B, Greiner HM, Holland-Bouley K, Faist R, Mangano F, Pestian J. Identifying epilepsy surgery candidates with machine learning. (attached)

In process (analysis underway)

4. Dexheimer JW, Wissel B, Greiner HM, Holland-Bouley K, Faist R, Mangano F, Pestian J. Natural Language Processing for Identifying Epilepsy Surgery Candidates: A Randomized, Controlled Pilot Trial

Press

5. 10/26/16 - New Grant Funds System to Help Physicians Identify Epilepsy Surgery Candidates https://www.cincinnatichildrens.org/research/divisions/b/bmi/news/2016/2016-10-17-new-grant-funds-system-to-help-physicians-identify-epilepsy-surgery-candidates
6. 9/11/18 - Novel Algorithm Identifies Epilepsy Surgical Candidates Sooner https://www.liebertpub.com/doi/abs/10.1089/clinomi.05.05.16?journalCode=clinomi