

# Large-scale identification of patients with cerebral aneurysms using natural language processing

Victor M. Castro, MS  
Dmitriy Dligach, PhD  
Sean Finan  
Sheng Yu, PhD  
Anil Can, MD  
Muhammad Abd-El-Barr,  
MD, PhD  
Vivian Gainer, MS  
Nancy A. Shadick, MD,  
MPH  
Shawn Murphy, MD,  
PhD  
Tianxi Cai, PhD  
Guergana Savova, PhD  
Scott T. Weiss, MD  
Rose Du, MD, PhD

Correspondence to  
Dr. Du:  
rdu@partners.org

## ABSTRACT

**Objective:** To use natural language processing (NLP) in conjunction with the electronic medical record (EMR) to accurately identify patients with cerebral aneurysms and their matched controls.

**Methods:** ICD-9 and Current Procedural Terminology codes were used to obtain an initial data mart of potential aneurysm patients from the EMR. NLP was then used to train a classification algorithm with .632 bootstrap cross-validation used for correction of overfitting bias. The classification rule was then applied to the full data mart. Additional validation was performed on 300 patients classified as having aneurysms. Controls were obtained by matching age, sex, race, and healthcare use.

**Results:** We identified 55,675 patients of 4.2 million patients with ICD-9 and Current Procedural Terminology codes consistent with cerebral aneurysms. Of those, 16,823 patients had the term aneurysm occur near relevant anatomic terms. After training, a final algorithm consisting of 8 coded and 14 NLP variables was selected, yielding an overall area under the receiver-operating characteristic curve of 0.95. After the final algorithm was applied, 5,589 patients were classified as having aneurysms, and 54,952 controls were matched to those patients. The positive predictive value based on a validation cohort of 300 patients was 0.86.

**Conclusions:** We harnessed the power of the EMR by applying NLP to obtain a large cohort of patients with intracranial aneurysms and their matched controls. Such algorithms can be generalized to other diseases for epidemiologic and genetic studies. *Neurology*® 2017;88:164-168

## GLOSSARY

**AUC** = area under the receiver-operating characteristic curve; **CPT** = Current Procedural Terminology; **EMR** = electronic medical record; **ICD-9-CM** = *International Classification of Diseases, 9th Revision, Clinical Modification*; **LASSO** = least absolute shrinkage and selection operator; **NLP** = natural language processing; **PPV** = positive predictive value; **UMLS** = Unified Medical Language System.

Cerebral aneurysm is a potentially devastating disorder that affects nearly 3% of the population.<sup>1</sup> Epidemiologic and genetics studies of aneurysm patients are often encumbered by insufficient power and typically require large, costly, multicenter studies. Our goal is to use electronic medical records (EMRs) to identify patients with cerebral aneurysms and to obtain a large cohort within a single health system.

The simplest means by which patients can be extracted from EMRs is via codified billing data. Unruptured aneurysms are coded with the ICD-9-CM diagnosis code of 497.3, but many such codings may be inaccurate. Ruptured aneurysms, on the other hand, are not coded directly with ICD-9 but may be inferred from the presence of subarachnoid hemorrhage or procedural codes for treatment of aneurysms. Therefore, it is difficult to use codified diagnosis and procedure data alone to accurately identify all patients with aneurysms.

To address this, we used codified data along with natural language processing (NLP) from the EMR with the goal of accurately identifying the subset of patients with true aneurysms. The

Supplemental data  
at *Neurology.org*

From Research Information Systems and Computing (V.M.C., V.G., S.M.), Partners Healthcare; Boston Children's Hospital Informatics Program (D.D., S.F., G.S.); Harvard Medical School (D.D., S.Y., A.C., M.A.-E.-B., N.A.S., S.M., S.T.W., R.D.); Department of Medicine (S.Y., S.T.W.), Department of Neurosurgery (A.C., M.A.-E.-B., R.D.), Division of Rheumatology, Immunology and Allergy (N.A.S.), and Channing Division of Network Medicine (S.T.W., R.D.), Brigham and Women's Hospital, Boston, MA; Center for Statistical Science (S.Y.), Tsinghua University, Beijing, China; Department of Neurology (S.M.), Massachusetts General Hospital; and Biostatistics (T.C.), Harvard School of Public Health, Boston, MA.

Go to *Neurology.org* for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

results were then validated against a well-phenotyped cohort. Although NLP has been previously applied to other disease processes, including asthma and rheumatoid arthritis,<sup>2-4</sup> the current state of the art calls for customization to each disease entity. The distinguishing feature of cerebral aneurysms compared to other entities such as asthma is the reliance on imaging studies and the nuances present in radiology reports and clinical notes.

**METHODS** We used the Partners Healthcare Research Patients Data Registry as the source of the EMRs from which data are extracted (figure).<sup>5,6</sup> The Research Patients Data Registry encompasses 4.2 million patients who have received care at the Brigham and Woman's Hospital or Massachusetts General Hospital over the past 20 years. It includes both codified data (e.g., ICD-9 codes, Current Procedural Terminology [CPT] codes) and unstructured data such as radiology reports, outpatient clinic notes, inpatient discharge summaries, and operative reports.

Patients with intracranial aneurysms were initially identified with the use of broad inclusion criteria to maximize the sensitivity. Search criteria used were ICD-9 codes for unruptured aneurysms (437.3), berry aneurysm of the anterior communicating artery (747.81), ruptured syphilitic aneurysm (094.87), subarachnoid hemorrhage (430), intracerebral hemorrhage (431), and cerebral vasospasm (435.9) and procedure codes for angioplasty for vasospasm (39.50), intra-arterial injection for treatment of vasospasm (99.29), cerebral angiogram (88.41), clipping of aneurysm (39.51), and coiling of aneurysm (38.82, 39.52, 39.71, 39.72, 39.73, 39.74, 39.75, 39.76, 39.77, 39.78, 39.79). CPT codes for endovascular treatment of aneurysm (61708, 61710, 61623, 61624, 61626) and surgery for aneurysm (61680,

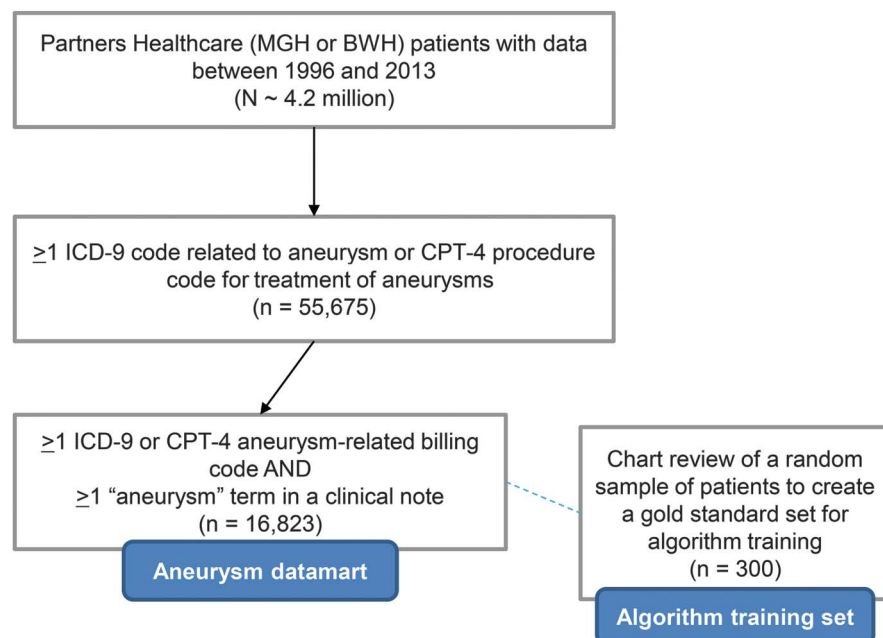
61682, 61684, 61686, 61690, 61692, 61697, 61698, 61700, 61702, 61703, 61705, 61711) were also included. In addition, patients with a mention of the term aneurysm appearing near relevant words such as cerebral, cranial, head, brain, and locations of aneurysms (table e-1 at Neurology.org) were included. We excluded notes with the word aneurysm near abdominal, aortic, or iliac and disregarded radiology reports that were not related to the head or brain. A data mart containing all clinical data from the refined population was generated with the i2b2 server software (i2b2 version 1.7, i2b2, Boston, MA).<sup>7</sup> Fifty patients from this refined data mart were then used to determine whether the data mart has a sufficiently high prevalence of aneurysms and to determine the number of patients needed to form the training set.

A total of 300 patients, including the initial 50 patients, were randomly selected from the data mart as the gold standard training set. Detailed medical record review was performed for these 300 patients by 2 neurosurgeons (R.D. and M.A.-E.-B.). The presence of aneurysm was categorized as definite aneurysm, definite no aneurysm, possible aneurysm, and unknown/insufficient information. Interobserver agreement was assessed by the Cohen  $\kappa$  coefficient. Any discrepancy was resolved by consensus.

**NLP and training algorithm.** A custom dictionary of terms relevant to cerebral aneurysms was created with the use of the Unified Medical Language System (UMLS)<sup>8</sup> terms by querying unique concept identifiers related to the locations of aneurysms (e.g., middle cerebral artery, anterior communicating artery), other clinical phenotypes related to cerebral aneurysms (e.g., saccular aneurysm, subarachnoid hemorrhage), associated conditions (e.g., polycystic kidney disease), and competing diagnoses (e.g., arteriovenous malformation). Unique concept identifiers and their terms were added to the dictionary manually if they were not included in the UMLS (table e-2).

Clinical notes, discharge summaries, operative reports, and radiology reports were then processed with the clinical Text

**Figure** Training algorithm for aneurysm cohort selection



BMW = Brigham and Women's Hospital; CPT-4 = Current Procedural Terminology, Fourth Edition; ICD-9 = *International Classification of Diseases, Ninth Revision*; MGH = Massachusetts General Hospital.

**Table 1** Features selected for final cerebral aneurysm algorithm (coding + natural language processing, specificity = 0.95)

Features	$\beta$ Coefficient
Patient's age	0.0157
Patient's sex	-0.5141
Log-adjusted number of all notes in patient record	-0.1804
Log-adjusted count of aneurysm ICD-9 diagnosis codes	1.7621
Log-adjusted count of cerebral hemorrhage ICD-9 diagnosis codes	0.1585
Log-adjusted count of hypertension ICD-9 diagnosis codes	-0.1750
Log-adjusted count of aneurysm treatments (CPT-4)	0.2386
Log-adjusted count of brain imaging procedure codes (CPT-4)	-0.3628
Log-adjusted count of cTAKES NLP term Acoa	0.0583
Log-adjusted count of cTAKES NLP term A1	0.0544
Log-adjusted count of cTAKES NLP term carotid artery	0.0879
Log-adjusted count of cTAKES NLP term carotid terminus rad	0.0002
Log-adjusted count of cTAKES NLP term paraclinoid rad	0.0002
Log-adjusted count of cTAKES NLP term supraclinoid rad	0.0307
Log-adjusted count of cTAKES NLP term cavernous	0.0481
Log-adjusted count of cTAKES NLP term location on vessel rad	-0.0383
Log-adjusted count of cTAKES NLP term location on vessel	-0.0454
Log-adjusted count of cTAKES NLP term outpouching	-0.1980
Log-adjusted count of cTAKES NLP term aneurysm rad	0.0277
Log-adjusted count of cTAKES NLP term aneurysm	0.4162
Log-adjusted count of cTAKES NLP term exclusion criteria	-0.2633
Log-adjusted count of cTAKES NLP term intracerebral aneurysm	0.1909

Abbreviations: Acoa = anterior communicating artery; CPT-4 = Current Procedural Terminology, Fourth Edition; cTAKES = clinical Text Analysis and Knowledge Extraction System; ICD-9 = International Classification of Diseases, Ninth Revision; NLP = natural language processing.

Analysis and Knowledge Extraction System (<http://ctakes.apache.org>).<sup>9</sup> For each concept, the number of its positive mention was used as the NLP feature of that concept.

A classification algorithm was trained with the use of the gold standard labeled training set via a penalized logistic regression model with adaptive least absolute shrinkage and selection operator (LASSO) penalty that accounts for model complexity.<sup>10,11</sup> The penalty parameter was selected to optimize the bayesian information criterion. The final algorithm estimates the probability that a patient has an aneurysm given his/her feature information. We evaluated 2 different classifications of patients for the purpose of selecting the best algorithm and classification system: no/possible/unknown aneurysm vs definite aneurysm and no/unknown aneurysm vs possible/definite aneurysm. To evaluate algorithm performance, we computed the area under the receiver-operating characteristic curve (AUC), sensitivity, positive predictive value (PPV), and negative predictive value corresponding to a specificity level of 95%. The .632 bootstrap cross-validation was used to correct for overfitting bias.<sup>12,13</sup> The classification rule associated with 95% specificity for distinguishing definite aneurysm from no/possible/unknown aneurysm was then applied to the full data mart to identify aneurysm cases. From the patients who were classified as having aneurysms and were not part of the gold standard training set, an additional 300 patients were randomly selected for additional validation. Detailed medical record

review was performed by 2 physicians (R.D. and A.C.), and the PPV was determined.

**Controls.** To obtain the control group, we used a previously validated method that matches the healthcare use of the cases and controls to compensate for differences in data collection practices in electronic health records.<sup>14</sup> A pool of patients with at least one visit to either Massachusetts General Hospital or Brigham and Women's Hospital, no mention of the term aneurysm in a clinical note or aneurysm-related billing code, and no inpatient admissions lasting >2 weeks served as potential control patients ( $n = 1,436,010$ ). We included only patients with no mention of the term aneurysm in a note or aneurysm-related billing code to ensure that there are no cases in the control group. The exclusion of patients with inpatient admission lasting >2 weeks was imposed to avoid overly ill patients who could skew matching based on the number of recorded events. Ten control patients were matched to each selected patient on the basis of age, sex, race, number of recorded events (diagnosis, procedures, medications, and vital signs), and earliest and latest visits. The number of recorded events is representative of the healthcare use of each individual patient.

**Standard protocol approvals, registrations, and patient consents.** This study was approved by the Partners Institutional Review Board.

**RESULTS** Using ICD-9 and CPT codes alone, we initially identified 55,675 patients with diagnosis or procedure codes indicating the presence of an aneurysm. A total of 16,823 patients had both diagnosis and procedure codes for aneurysm and the term aneurysm that occurred near (within 4 words) terms describing brain anatomy, including cerebral, cranial, head, and brain (table e-1); those with aneurysm near abdominal, iliac, and aorta were excluded.

The training set consists of 115 (38%) with definite aneurysm, 133 (44.3%) with no aneurysm, 28 (9%) with possible aneurysm, and 24 (8%) unknown. Interrater agreement was high ( $\kappa = 0.94$ , 95% confidence interval 0.92–0.97). The reasons for the interrater differences were small outpouchings ( $n = 9$ ), fusiform dilatation of/dysplastic/irregular vessel ( $n = 4$ ), error in the clinical note by nonneurosurgeon/nonneurologist ( $n = 2$ ), low-quality magnetic resonance angiography ( $n = 1$ ), presence in only one clinic note ( $n = 1$ ), and twist in vessel initially thought to be an aneurysm ( $n = 1$ ).

A dictionary of features to be used for training an algorithm was derived with the use of domain expertise and ontological references in the UMLS. The feature set included 15 coded variables (i.e., diagnosis, imaging procedures, and surgical procedures) and 127 NLP variables. Of the NLP variables, 29 had been mentioned for at least 10% of the patients and were selected as candidate features for algorithm training.

After the LASSO procedure was applied to train a model to classify cases as definite aneurysms vs possible/no aneurysms and the .632 bootstrap cross-validation was applied, 8 coded variables and 14

**Table 2** Accuracy parameters of aneurysm algorithms

Algorithm	Specificity = 0.95			
	AUC (SE)	Sensitivity (SE)	PPV (SE)	NPV (SE)
COD + NLP	0.946 (0.010)	0.78 (0.04)	0.91 (0.01)	0.99 (0.03)
NLP	0.904 (0.015)	0.62 (0.06)	0.89 (0.02)	0.99 (0.03)
COD	0.912 (0.015)	0.59 (0.06)	0.88 (0.02)	0.99 (0.03)
COD + NLP vs NLP	0.042 (0.013)	0.16 (0.06)	0.02 (0.01)	0.00 (0.03)
COD + NLP vs COD	0.034 (0.012)	0.19 (0.06)	0.03 (0.01)	0.00 (0.03)

Abbreviations: AUC = area under the receiver-operating characteristic curve; COD = algorithm based on electronic medical record billing codes only; COD + NLP = algorithm relying on electronic medical record billing codes and natural language processing features; EMR = electronic medical record; NLP = natural language processing algorithms only; NPV = negative predictive value; PPV = positive predictive value.

The last 2 rows show the difference in the accuracy parameters between the COD + NLP algorithm and the NLP algorithm and COD algorithm.

NLP variables were selected in the final algorithm (table 1), yielding an overall AUC of 0.946. Models developed with only coded variables or only NLP variables yielded substantially lower AUCs (0.912 and 0.904, respectively; table 2). For the final model, we chose a cutoff of 95% specificity with an estimated sensitivity of 0.78, a PPV of 0.91, and a negative predictive value of 0.99. Of 16,823 patients, 5,589 patients were classified as having definite aneurysms and 11,298 were classified as having no/possible aneurysms after the final trained model was applied. With the use of the case-control matching algorithm

described above, 54,952 controls were matched to the algorithm-selected cases at a 10:1 ratio. Demographic comparisons of selected cases and controls are reported in table 3.

Detailed medical record review of 300 patients classified as having definite aneurysm resulted in 258 patients with definite aneurysms not associated with other vascular malformations such as arteriovenous malformations. This results in a PPV of 0.86.

**DISCUSSION** The EMR has become an integral part of modern healthcare. The wealth of electronic health data is a potential resource both for clinical research and for evaluating hospital quality measures. However, the ability to take advantage of the large volume of electronic data is encumbered by the mostly unstructured nature of EMRs. In this study, we demonstrate the utility of NLP in extracting structured data and accurately identifying a small subset of patients with an uncommon disorder, cerebral aneurysms, from a large database consisting of 4.2 million patients.

Using codified data alone, we would have >55,000 patients who satisfy our search criteria for aneurysms. This would have been an infeasible number of records to review manually. Even after narrowing the search to those with diagnosis/procedure codes for aneurysm and those with the term aneurysm in a clinical note and excluding those with the term aneurysm near other anatomic parts of the body, we were left with nearly 17,000 records. On the basis of the subset of 300 gold standard patients who subsequently underwent detailed manual review of the medical records by clinical experts, we found the prevalence of this set of 16,823 patients to be 38%. Without further refinement, this would have resulted in the chart review of >10,000 negative records, which would be both immensely time-consuming and unfruitful. With the algorithm-selected cases, the prevalence of true cases (PPV) was increased dramatically to 91% in the training set with bootstrap cross-validation and 86% in the validation set. The slight difference is likely due to variations in the sample and in the labeling process (e.g., a possible infundibulum may be called an aneurysm by a radiologist). The combined model using both codified data and NLP allows us to identify >30% additional true cases with a high level of specificity, which is significant in the context of a rare phenotype and when patient recruitment is difficult. Nevertheless, with this model, ≈450 of 5,589 patients would be misclassified as having true aneurysms. The misclassification has a number of implications, depending on the type of downstream study involved. For genomic studies and epidemiologic studies, the misclassification can introduce bias and reduce the power of the study.

**Table 3** Demographics of the final aneurysm cohort and matched controls

	Aneurysm cohort	Controls
n	5,589	55,026
Age, mean ± SD, y	62.0 ± 13.7	61.7 ± 13.9
Female, %	76	76
Race, %		
White	80	79
Black	6	6
Hispanic	6	6
Asian	3	3
Other/unknown	6	7
No. of recorded events, median (IQR)	281 (123–626)	287 (124–584)
Year of earliest visit, median (IQR)	2003 (1998–2007)	2002 (1997–2007)
Year of latest visit, median (IQR)	2012 (2009–2013)	2012 (2010–2013)
Public insurance, %	38	28
Past/current smoker, %	36	17
BMI, maximum, mean ± SD, kg/m <sup>2</sup>	29.6 ± 6.6	30.2 ± 7.6
History of hypertension, %	55	42
History of dyslipidemia, %	39	41

Abbreviations: BMI = body mass index; IQR = interquartile range.



To account for this, one can build the expected misclassification rate in the cases into the analysis, as per Sinnott et al.<sup>15</sup> For other uses such as recruitment for a prospective clinical trial, a lower specificity may be tolerable because the patients will be screened further. Even in the most stringent scenario in which the medical records of all predicted cases are reviewed, the significant reduction in negative cases would render such reviews much more manageable.

We also demonstrated the utility of using codified data in identifying a matching control group. The use of healthcare use in selecting the control group through electronic health records has recently been validated by Castro et al.<sup>14</sup> Hospital-based controls selected from a tertiary care institution can be heterogeneous in terms of their exposure to various diseases, unlike the control population selected for clinical trials. This use-based method has been shown to be superior to the use of hospital-based controls matched on age/sex/race alone. The ability to identify both the cases and matching controls from whom discarded tissue samples can be obtained will serve as a powerful means of future epidemiologic and genetic studies.

Overall, we have demonstrated the power of using NLP in conjunction with the EMR in accurately identifying a large cohort of patients with intracranial aneurysms and their matched controls. The use of NLP renders an otherwise formidable task achievable. The accurate identification of large cohorts of patients with specific disease processes depends on the clinical understanding of the disease process and will be important in future epidemiologic and genetic studies.

#### AUTHOR CONTRIBUTIONS

Victor M. Castro, Dmitriy Dligach, Sean Finan, and Sheng Yu: study concept and design, acquisition of data, analysis and interpretation of data, critical revision of manuscript for intellectual content. Anil Can and Muhammad Abd-El-Barr: acquisition of data, critical revision of manuscript for intellectual content. Vivian Gainer, Nancy A. Shadick, and Shawn Murphy: study concept and design, critical revision of manuscript for intellectual content. Tianxi Cai and Guergana Savova: study concept and design, analysis and interpretation of data, critical revision of manuscript for intellectual content. Scott T. Weiss: study concept and design, critical revision of manuscript for intellectual content. Rose Du: study concept and design, acquisition of data, analysis and interpretation of data, critical revision of manuscript for intellectual content.

#### STUDY FUNDING

This study was supported by Partners Personalized Medicine.

#### DISCLOSURE

The authors report no disclosures relevant to the manuscript. Go to [Neurology.org](http://Neurology.org) for full disclosures.

Received June 22, 2016. Accepted in final form September 28, 2016.

#### REFERENCES

1. Vlak MH, Algra A, Brandenburg R, Rinkel GJ. Prevalence of unruptured intracranial aneurysms, with emphasis on sex, age, comorbidity, country, and time period: a systematic review and meta-analysis. *Lancet Neurol* 2011;10:626–636.
2. Castro V, Shen Y, Yu S, et al. Identification of subjects with polycystic ovary syndrome using electronic health records. *Reprod Biol Endocrinol* 2015;13:116.
3. Zeng QT, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. *BMC Med Inform Decis Mak* 2006;6:30.
4. Castro VM, Minnier J, Murphy SN, et al. Validation of electronic health record phenotyping of bipolar disorder cases and controls. *Am J Psychiatry* 2015;172:363–372.
5. Nalichowski R, Keogh D, Chueh HC, Murphy SN. Calculating the benefits of a research patient data repository. *AMIA Annu Symp Proc* 2006:1044.
6. Murphy SN, Morgan MM, Barnett GO, Chueh HC. Optimizing healthcare research data warehouse design through past COSTAR query analysis. *Proc AMIA Symp* 1999:892–896.
7. Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010;17:124–130.
8. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004;32:D267–D270.
9. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Am Med Inform Assoc* 2010;17:507–513.
10. Zou H. The adaptive lasso and its oracle properties. *J Am Stat Assoc* 2006;101:1418–1429.
11. Hastie T, Tibshirani R, Friedman J, Franklin J. The elements of statistical learning: data mining, inference and prediction. *Math Intell* 2005;27:83–85.
12. Efron B, Tibshirani R. Improvements on cross-validation: the .632+ bootstrap method. *J Am Stat Assoc* 1997;92:548–560.
13. Jiang W, Simon R. A comparison of bootstrap methods and an adjusted bootstrap approach for estimating the prediction error in microarray classification. *Stat Med* 2007;26:5320–5334.
14. Castro VM, Apperson WK, Gainer VS, et al. Evaluation of matched control algorithms in EHR-based phenotyping studies: a case study of inflammatory bowel disease comorbidities. *J Biomed Inform* 2014;52:105–111.
15. Sinnott JA, Dai W, Liao KP, et al. Improving the power of genetic association tests with imperfect phenotype derived from electronic medical records. *Hum Genet* 2014;133:1369–1382.

# Neurology®

## Large-scale identification of patients with cerebral aneurysms using natural language processing

Victor M. Castro, Dmitriy Dligach, Sean Finan, et al.

*Neurology* 2017;88:164-168 Published Online before print December 7, 2016

DOI 10.1212/WNL.0000000000003490

**This information is current as of December 7, 2016**

<b>Updated Information &amp; Services</b>	including high resolution figures, can be found at: <a href="http://www.neurology.org/content/88/2/164.full.html">http://www.neurology.org/content/88/2/164.full.html</a>
<b>Supplementary Material</b>	Supplementary material can be found at: <a href="http://www.neurology.org/content/suppl/2016/12/07/WNL.0000000000003490.DC1">http://www.neurology.org/content/suppl/2016/12/07/WNL.0000000000003490.DC1</a>
<b>References</b>	This article cites 13 articles, 3 of which you can access for free at: <a href="http://www.neurology.org/content/88/2/164.full.html##ref-list-1">http://www.neurology.org/content/88/2/164.full.html##ref-list-1</a>
<b>Subspecialty Collections</b>	This article, along with others on similar topics, appears in the following collection(s): <b>All Cerebrovascular disease/Stroke</b> <a href="http://www.neurology.org/cgi/collection/all_cerebrovascular_disease_stroke">http://www.neurology.org/cgi/collection/all_cerebrovascular_disease_stroke</a> <b>All epidemiology</b> <a href="http://www.neurology.org/cgi/collection/all_epidemiology">http://www.neurology.org/cgi/collection/all_epidemiology</a> <b>Subarachnoid hemorrhage</b> <a href="http://www.neurology.org/cgi/collection/subarachnoid_hemorrhage">http://www.neurology.org/cgi/collection/subarachnoid_hemorrhage</a>
<b>Permissions &amp; Licensing</b>	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: <a href="http://www.neurology.org/misc/about.xhtml#permissions">http://www.neurology.org/misc/about.xhtml#permissions</a>
<b>Reprints</b>	Information about ordering reprints can be found online: <a href="http://www.neurology.org/misc/addir.xhtml#reprintsus">http://www.neurology.org/misc/addir.xhtml#reprintsus</a>

*Neurology*® is the official journal of the American Academy of Neurology. Published continuously since 1951, it is now a weekly with 48 issues per year. Copyright © 2016 American Academy of Neurology. All rights reserved. Print ISSN: 0028-3878. Online ISSN: 1526-632X.

