Epilepsy Among Elderly Medicare Beneficiaries

A Validated Approach to Identify Prevalent and Incident Epilepsy

Lidia M.V.R. Moura, MD, MPH,†‡ Jason R. Smith, BA,* Deborah Blacker, MD, ScD,†§ Christine Vogeli, PhD,|| Lee H. Schwamm, MD,*‡ Andrew J. Cole, MD,*‡ Sonia Hernandez-Diaz, MD, DPH,† and John Hsu, MD, MBA¶#*

Background: Uncertain validity of epilepsy diagnoses within health insurance claims and other large datasets have hindered efforts to study and monitor care at the population level.

Objectives: To develop and validate prediction models using longitudinal Medicare administrative data to identify patients with actual epilepsy among those with the diagnosis.

Research Design, Subjects, Measures: We used linked electronic health records and Medicare administrative data including claims to predict epilepsy status. A neurologist reviewed electronic health record data to assess epilepsy status in a stratified random sample of Medicare beneficiaries aged 65+ years between January 2012 and December 2014. We then reconstructed the full sample using inverse probability sampling weights. We developed prediction models using longitudinal Medicare data, then in a separate sample evaluated the predictive performance of each model, for example, area under the receiver operating characteristic curve (AUROC), sensitivity, and specificity.

Results: Of 20,945 patients in the reconstructed sample, 2.1% had confirmed epilepsy. The best-performing prediction model to identify prevalent epilepsy required epilepsy diagnoses with multiple claims at least 60 days apart, and epilepsy-specific drug claims: AUROC = 0.93 [95% confidence interval (CI), 0.90–0.96], and with an 80% diagnostic threshold, sensitivity = 87.8% (95% CI, 80.4%–93.2%), specificity = 98.4% (95% CI, 98.2%–98.5%). A similar model also performed well in predicting incident epilepsy ($k = 0.79$; 95% CI, 0.66–0.92).

Conclusions: Prediction models using longitudinal Medicare data perform well in predicting incident and prevalent epilepsy status accurately.

Key Words: epilepsy, epidemiology, elderly, claims data, algorithms

(*Med Care* 2019;00: 000–000)

From the *Department of Neurology, Massachusetts General Hospital; †Department of Epidemiology, Harvard T.H. Chan School of Public Health; ‡Department of Neurology, Harvard Medical School; Departments of §Psychiatry; ||Medicine; ¶Department of Medicine, Mongan Institute, Massachusetts General Hospital; and #Department of Health Care Policy, Harvard Medical School, Boston, MA.

Supported by the NIH (1K08AG053380-01A1) and the Harvard University Diversity Inclusion and Community Partnership program for protected research time to design and conduct all aspects of the study. L.M.V.R.M. receives funding from the NIH (1K08AG053380-01A1, 3P01AG032952-09S1, 1R01AG062282-01), the Harvard University Diversity Inclusion and Community Partnership Faculty Fellowship Award, the Epilepsy Foundation, and the Schwamm-Marriott Clinical Care Research Fellowship Award. D.B. receives funding from the NIH (P50 AG005134, 1P01AG036694, U01 AG042984, U01AG045390, 1U24NS100591-01, 1R01AG058063-01A1, 1R01AG062282-01, 2P01AG032952-09S1). S.H.-D. receives funding from the NIH (R21HD092879). J.H. receives funding from the NIH (1R01 CA164023-04, 2P01AG032952-06A1, R01 HD075121-04, R01 MH104560-02).

The authors declare no conflict of interest.

Reprints: Lidia M.V.R. Moura, MD, MPH, 55 Fruit Street, Wang ACC 739D, Boston, MA 02114. E-mail: lidia.moura@mgh.harvard.edu.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.lww-medicalcare.com.

ISSN: 0025-7079/19/000-000

Epilepsy is a life-threatening, often lifelong, disorder characterized by recurrent, spontaneous seizures.[1,2] Presentation of epilepsy can be varied, in part due to the range of etiologies. For example, epilepsy may arise from common structural pathologies including neurodegeneration, trauma, stroke, and tumors.[3,4] Epilepsy may present with subtle cognitive or behavioral symptoms, which could be mistaken for age-related cognitive impairment.[5,6] At times, both epilepsy and the seizures both go undetected.[3] Early detection and treatment, however, is critical as the risk of seizure recurrence after a first seizure can be as high as 80%.[7,8] Moreover, morbidity can be exacerbated by other comorbidities associated with advanced age.[9]

Despite the significant health burden posed by epilepsy, there are few national estimates of disease incidence or prevalence, though anecdotal estimates suggest that both have increased. Generating such national estimates and tracking care for patients across the country would require a well-validated approach to identify true epilepsy cases using data from large, national datasets, for example, Medicare administrative datasets including insurance claims. Relying only on diagnosis codes, however, is a potentially fraught process.[10–14] At the same time, many large datasets contain more information than only diagnosis codes, some of which could be useful for predicting disease status.

Furthermore, the diagnosis and treatment of epilepsy in the elderly requires nuances in clinical judgement and review of information that might not be captured in claims, for example, synthesis of nonspecific symptoms, detailed histories, and impressions from diagnostic tests. As a result, previous studies that ignore age and focus primarily on diagnostic and

treatment criteria have failed to produce accurate and validated claims-based measures to detect epilepsy among individuals aged 65+ years.[11,12,15–20] These methodological limitations contribute to the uncertainty about the true prevalence of epilepsy in the elderly population.

To address this gap in the literature, we used a population-based study design linking longitudinal Medicare administrative data (eg, claims) to electronic health records (EHR) data, combined with blinded expert review of the clinical data, to determine whether claims-based models can accurately predict which patients in the elderly Medicare population have true epilepsy disease.

METHODS

Data Sources and Sampling Approach

We used 3 longitudinal datasets: (1) The Partners Healthcare System EHRs; (2) The Partners Accountable Care Organization (ACO) claims data. We used Medicare claims from Medicare Parts A/B (hospital and physician services) and D (prescription drugs). The Partners Healthcare Institutional Review Board approved the study protocol.

We used a 3-year observation period (January 2012–December 2014), which is a reasonable length to capture an established diagnosis of epilepsy.[2,11,14,16] We then used the following eligibility criteria: (1) aged 65+ years by January 2012; (2) enrollment in both Medicare Parts A and B; (3) continuous alignment to the ACO between January 2012 and December 2014 or until death; (4) Medicare original reason for entitlement code of age or disability; and (5) community dwelling (not institutionalized) at the time of ACO entry.[21] To improve the efficiency of the chart reviews, we then performed a structured sampling on the full sample (n = 20,945) based on claims coded under the International Classification of Diseases, ninth revision (ICD-9) (Fig. 1) and Table 1 provides descriptive characteristics of the 3 samples (total n = 1534): (A) at least 1 claim potentially indicative of epilepsy (codes 345.xx; 780.39); (B) no claim for epilepsy but seen by a neurologist; and (C) no claims for epilepsy and not seen by a neurologist. Our rationale was 3-fold: (a) it increased the power for analysis involving between-group comparisons because analytical groups were of reasonably comparable sizes (ie, true epilepsy cases vs. not, as well as false-positives and false-negative cases); (b) it increased the feasibility of the study given finite resources; and (c) it enabled us to reconstruct the target population rates using the inverse of the probability of being sampled weights.

EHR-based Diagnosis (Reference-Standard)

An experienced neurologist reviewed and abstracted clinical data from all 1534 EHRs (including physician notes, emergency department visit records, and diagnostic tests such as electroencephalography files or brain imaging), without information on which of the 3 samples the patient was a member or on any of the claims information. Using the current International League Against Epilepsy (ILAE) guidelines (Text, Supplemental Digital Content 1, http://links.lww.com/MLR/B716 and Table, Supplemental Digital Content 2, http://links.lww.com/MLR/B717), the neurologist determined prevalent epilepsy disease status and epilepsy classification.[22,23]

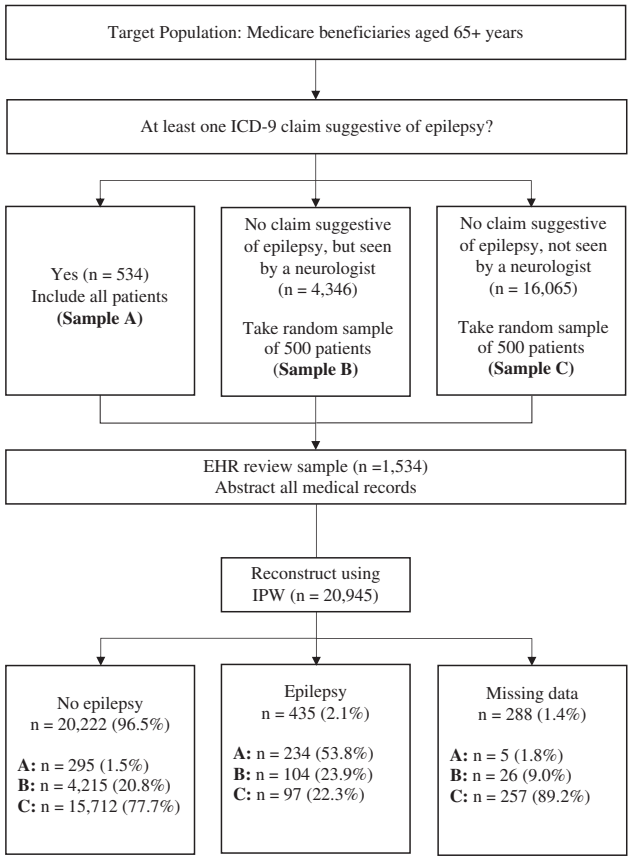

FIGURE 1. Sampling strategy and patients with adjudicated epilepsy diagnosis. Sampling strategy of eligible patients that resulted in a study sample of 1534 patients, broken up into 3 subsamples. ICD-9 claims related to epilepsy, the basis of stratification, were restricted to the time frame of January 2012–December 2014. EHR indicates electronic health record; ICD-9, International Classification of Diseases, Ninth Revision; IPW, inverse probability weighting.

Within the sample of prevalent cases, the neurologist then determined which of these first occurred between 2012 and 2014, that is, potential incident cases. The neurologist obtained the year of first epilepsy diagnosis (ie, the index event) and the year of the first documented seizure. The neurologist exclusively designated potential incident epilepsy cases after careful review of physician notes and records, conservatively excluding patients with documentation of a definitively acknowledged epilepsy diagnosis or a potential seizure before 2012.

Claims Diagnostic Codes

The goal of model development was to predict the probability of having epilepsy using the longitudinal administrative Medicare data, including the Medicare claims. To this end, we first used expert clinical knowledge to create lists of codes suggestive of epilepsy care (provided in the Table, Supplemental Digital Content 3, http://links.lww.com/MLR/B718).

TABLE 1. Characteristics of Reconstructed Sample

Characteristics	Sample A: At Least 1 Epilepsy Claim (n = 534)	Sample B: No Epilepsy Claims and Not Seen by Neurologist (n = 16,065)	Sample C: No Epilepsy Claims But Seen by Neurologist (n = 4346)
Demographic*			
Age [mean (SD)]	77.3 (7.2)	75.6 (39.4)	73.7 (19.1)
Female [n (%)]	322 (60.3)	9446 (58.8)	2582 (59.4)
Treatment [n (%)]†			
Epilepsy-specific AED	183 (34.3)	161 (1.0)	235 (5.4)
AED	271 (50.7)	4016 (25.0)	1660 (38.2)
Suspected diagnosis [n (%)]*			
No epilepsy	295 (55.2)	15,712 (97.8)	4215 (97.0)
Epilepsy	234 (43.8)	97 (0.6)	104 (2.0)
Missing data	5 (1.0)	257 (1.6)	26 (1.0)

Characteristics of reconstructed sample (n = 20,945) following EHR review. Missing data (n = 288) were reported when there was incomplete patient EHR information to ascertain an epilepsy diagnosis.

*Information abstracted from EHRs.

†Information derived from Medicare claims.

AED indicates antiepileptic drug; EHR, electronic health record.

TABLE 2. Candidate Variables Constructed From Claims

Variables	Description
Suspected epilepsy*	
Primary epilepsy diagnosis	Epilepsy as a primary diagnosis
Epilepsy diagnosis†	Weighted; epilepsy as primary, secondary, or tertiary diagnosis, and gives an additional weight to the occurrence of ≥ 2 primary codes 60 d apart
Algorithmic epilepsy diagnosis[18]‡	Any epilepsy diagnosis
Algorithmic epilepsy diagnosis[2]‡	Convulsions at least 30 d apart coinciding with inpatient stays, outpatient visits, or physician visits within 1 y
Potential false-positive*	
Withdrawal seizure	Diagnosis of seizure due to alcohol or drug withdrawal
Drug reaction seizure	Diagnosis of seizure due to a drug reaction
Syncope and collapse event	Diagnosis of "syncope or collapse"
Alteration of consciousness event	Diagnosis of alteration of consciousness event
Sleep disturbance event	Diagnosis of sleep disturbance event
Dementia	Diagnosis of dementia
Migraine event	Diagnosis of migraine event
Psychogenic event	Diagnosis of psychogenic event
Cerebrovascular event	Diagnosis of cerebrovascular event or cognitive deficit
Nonepileptic event diagnosis	Potential false-positive diagnosis, excluding withdrawal and drug reaction seizures
Nonepileptic event or seizure diagnosis	Any potential false-positive diagnosis
Specialty involvement§	
Neurology	Epilepsy-related claim indicative of neurology specialty involvement
Site of care§	
Urgent care and ER	Epilepsy-related claim indicative of service provided at an urgent care facility or hospital ER
Procedure‖	
Brain imaging	Epilepsy-related claim indicative of a brain imaging procedure
EEG	Epilepsy-related claim indicative of an EEG procedure
Treatment	
Epilepsy-specific AED	Specific AED prescription (levetiracetam, lamotrigine, valproic acid, phenytoin, or carbamazepine) with a concurrent epilepsy diagnosis code
AED†	All AED prescriptions

List of variables constructed from ICD-9, CPT, and service codes for creation of models. Codes used to create each variable are listed in Supplemental Digital Content 2 (Table; http://links.lww.com/MLR/B717). Medicare claims were pulled between the dates January 2012 and December 2014.

*ICD-9 codes.

†Variables used to recreate Holden et al[18] algorithm.

‡Variables created based on Faught et al[2] algorithm.

§Service codes.

‖CPT codes.

AED indicates antiepileptic drug; CPT, current procedural terminology; EEG, electroencephalography; ER, emergency room; ICD-9, International Classification of Diseases, Ninth Revision.

Next, we combined varying criteria (such as occurrence, position, or counts) to create several ordinal candidate variables (Table 2). Some ordinal variables could increase the probability of epilepsy, for example, 2 codes for epilepsy as a primary, secondary, or tertiary diagnosis 60 days apart. Conversely, others might decrease the probability, for example, diagnosis for recurrent syncope, as they may trigger work-up for loss of awareness that may lead to miscoding for epilepsy. We constructed these variables to reflect the potential diversity in epilepsy care, as patients can receive care and treatment at varying levels (such as primary care physician offices, general neurology clinics, or epilepsy subspecialty clinics). We optimized variable performance by accounting for varying levels of granularity in these discrete encounters.

Finally, we used inverse probability sampling weights to recreate the population (ie, we used the EHR review sample to efficiently reconstruct the full sample but with EHR-based epilepsy diagnoses). We accounted for age and sex and analyzed the distribution of clinical characteristics with confirmed epilepsy and nonconfirmed epilepsy.

Model Development

To develop and validate models in a split-sample approach we randomly divided the reconstructed sample into 2 subsets: (1) model development dataset (n = 10,518; 50.9%), which included the indicator variable for EHR-based diagnosis for each case; and (2) validation dataset (n = 10,139; 49.1%), in which the indicator variable for EHR-based diagnosis was removed. We checked covariate balance to assess the adequacy of the randomization.

We first developed several multivariable logistic models to predict the probability of reference-standard prevalent epilepsy. We used clinical judgement to choose 1 variable indicative of an epilepsy diagnosis ("epilepsy diagnosis," Table 2). "Epilepsy diagnosis" is an ordinal variable with "weights" ranging from 0 to 4, that adds one weight for the presence of ICD-9 claims coded as primary, secondary, or tertiary diagnosis of epilepsy, convulsions, syncope, or

collapse (ie, 3 weights total if at least 1 claim in the primary, secondary, and tertiary diagnosis) and gives an extra weight for the presence of >1 primary epilepsy claim at least 60 days apart (ie, maximum of 4 "weights").

We also included variables that could represent potential false-positives (ie, alternative conditions on the list of differential diagnoses). Several of these variables were tested in varying combinations to evaluate initial model performance [measured by area under the receiver operating characteristic curve (AUROC)]. We also included variables that reflected procedures performed, specialty of the physician making the diagnosis, and/or site of care variables. To conclude, we added treatment variables: (1) "epilepsy-specific antiepileptic drug (AED)" (Table 2), which functioned by assigning a weight of 1 for the presence of at least 1 epilepsy-specific drug (ie, lamotrigine, levetiracetam, valproic acid, phenytoin, and carbamazepine), and a weight of 0 if otherwise; or (2) "AED" (Table 2), which functioned exactly as epilepsy-specific AED, but included all AEDs (ie, AEDs that, after expert clinical review, are more frequently prescribed for indications aside from epilepsy, including gabapentin, topiramate, benzodiazepines). Different models used different epilepsy-related covariates to reflect varying potential uses (eg, some researchers might have access to diagnosis and drug data, others might only have access to diagnosis data).

Statistical Analysis and Model Validation

Using the validation dataset, we assessed the performance of the different prediction models. Our performance measures included sensitivity (ie, the percentage of total epilepsy cases correctly identified as epilepsy cases by the model), specificity (ie, the percentage of total nonepilepsy cases correctly identified as nonepilepsy by the model), and the AUROC (ie, a measure of classifying a true epilepsy case higher than a false epilepsy case). We also calculated the positive predictive values (PPV; ie, model-identified "true" epilepsy cases) and negative predictive values (NPV; ie, mode-identified "true" nonepilepsy cases), which by definition are sample-specific measures that vary with the underlying prevalence. While we report all of the performance measures and they tended to track together, we primarily use the AUROC as the most relevant measure.[24–26] We include additional details on statistical code in Supplemental Digital Content 1 (http://links.lww.com/MLR/B716).

We conducted several further analyses on the best-performing prediction model. We first altered the diagnosis threshold (ie, we increased, then decreased, the threshold for an individual to be classified as having epilepsy by the model) to assess the impact on sensitivity and specificity. Next, to imitate conditions in which less data might be available, we examined predictive power after restricting claims-based variables to a 1-year window (January 2012–December 2012). And third, we examined performance after conservatively excluding reference-standard nonepilepsy cases due to "incomplete chart but no evidence to indicate epilepsy" (n = 181).

Finally, among those patients with epilepsy identified by the best-performing model (ie, prevalent epilepsy between January 2012 and December 2014), we analyzed the 3-year incidence rate using 2 methods. In the first method, we applied washout criteria (no claim indicative of epilepsy and no epilepsy-specific AED prescription) for periods of 6, 12, 18, 24, 36, and 48 months before the index event.[27] In the second method, we applied washout criteria exclusively to January 2009–December 2011. The best-performing method was chosen by its level of agreement with diagnosis dates extracted from review of EHRs, measured by the Cohen κ coefficient.

Sensitivity Analyses

We conducted 3 types of sensitivity analyses. First, because more severely ill patients, or patients seen by a neurologist, may yield richer (or more accurate) epilepsy care documentation, we performed stratified analyses on the best-performing model. We separately analyzed the patients seen by a neurologist (n = 6785) and those without neurologist involvement (n = 16,160). Second, we examined the accuracy of 2 previously published algorithms[2,18] when applied to our reconstructed sample of Medicare beneficiaries to evaluate accuracy and generalizability: (1) one claim for epilepsy as primary diagnosis (345.xx) or at least 2 claims for convulsions (789.3×) at least 30 days apart[2]; (2) diagnosis codes potentially indicative of epilepsy and AED prescriptions.[18] The variables we constructed to recreate these algorithms are detailed in Table 2. And third, because epilepsy is not always diagnosed concurrently with a first seizure diagnosis, we analyzed the performance of our methods to measure 3-year incidence rates by altering the index event of the date of first diagnosis to "date of first seizure" and reporting agreement with EHR review (measured by κ coefficient).

RESULTS

Patient Characteristics

In the reconstructed sample, 2.1% or 21/1000 patients had a reference-standard epilepsy diagnosis over the 3-year period (Fig. 1 and Table 1). Of those patients with prevalent epilepsy [mean age = 76.4 y (SD = 8.5 y)], 230 (52.9%) were female and 205 (47.1%) were male. We provide descriptive information about each EHR review stratum in the Supplemental Digital Content 4 (http://links.lww.com/MLR/B726) (Text) and Supplemental Digital Content 5–7 (http://links.lww.com/MLR/B719) (Tables).

Claims Diagnostic Codes

We report the distribution of all claims-based variables stratified by epilepsy diagnosis (reference-standard) in the table of Supplemental Digital Content 8 (http://links.lww.com/MLR/B720). For instance, all suspected claims-based epilepsy variables (eg, epilepsy diagnosis, primary epilepsy diagnosis) occurred more often among patients with epilepsy compared with no epilepsy (eg, primary epilepsy diagnosis: 59.8% and 13.5%, respectively; $P < 0.0001$).

Models

Epilepsy diagnosis was evenly and randomly distributed across the model development and model validation subset: 10,336 patients with no epilepsy (51.1%) were distributed in the development subset and 9886 (48.9%) were

TABLE 3. Characteristics of Highest Performing Claims-based Models

Characteristics	Model 1*	Model 2†	Model 3‡
Suspected epilepsy			
Epilepsy diagnosis	Yes	Yes	Yes
Potential false-positive			
Nonepileptic event or seizure diagnosis	No	Yes	Yes
Site of care			
Urgent care and ER	No	Yes	Yes
Procedure			
EEG	No	Yes	Yes
Brain imaging	No	Yes	Yes
Specialty involvement			
Neurology	No	Yes	Yes
Treatment			
Epilepsy-specific AED	Yes	Yes	No

Claims-based variables used to construct best-performing models.

*Model 1: logit (epilepsy) = $\beta_0+\beta_1\times$(epilepsy diagnosis)+$\beta_2\times$(epilepsy-specific AED); where $\beta_0=-3.1$, $\beta_1=0.9$, and $\beta_2=1.9$; number of observations read = 1518; number of observations used = 1516; *c*-statistic: 0.927.

†Model 2: logit (epilepsy) = $\beta_0+\beta_1\times$(epilepsy diagnosis)+$\beta_2\times$(epilepsy-specific AED) $\beta_3\times$(alteration of consciousness)+$\beta_4\times$(sleep disturbance)+$\beta_5\times$(dementia)+$\beta_6\times$(nonepileptic event or seizure diagnosis)+$\beta_7\times$(urgent care and emergency room)+$\beta_8\times$(EEG)+$\beta_9\times$(brain imaging)+$\beta_{10}\times$(neurology); detailed statistical output values omitted from this footnote.

‡Model 3: logit (epilepsy) = $\beta_0+\beta_1\times$(epilepsy diagnosis)+$\beta_2\times$(alteration of consciousness)+$\beta_3\times$(sleep disturbance)+$\beta_4\times$(dementia)+$\beta_5\times$(nonepileptic event or seizure diagnosis) +$\beta_6\times$(urgent care and emergency room)+$\beta_7\times$(EEG)+$\beta_8\times$(brain imaging)+$\beta_9\times$(neurology).

AED indicates antiepileptic drug; EEG, electroencephalography; ER, emergency room.

distributed into the validation subset; 182 patients with epilepsy (41.8%) were randomly distributed into the development subset and 253 (58.2%) into the validation subset.

Three models were chosen based on their ability to predict the neurologist's EHR-based diagnosis of epilepsy (Tables 3, 4). The best-performing model (model 1), defined by the highest AUROC, included: an epilepsy diagnosis code variable and epilepsy-specific AED variable [sensitivity = 87.8%, 95% confidence interval (CI), 80.4%–93.2%; specificity = 98.4%, 95% CI, 98.2%–98.5%; AUROC = 0.93, 95% CI, 0.90–0.96; PPV = 23.1%, 95% CI, 19.2%–27.4%]. Increasing the diagnosis threshold of model 1 from 0.80 to 0.90 maximized sensitivity and AUROC over specificity (sensitivity = 100.0%, 95% CI, 93.5%–100.0%; specificity = 98.2%, 95% CI, 98.0%–98.3%; AUROC = 0.99, 95% CI, 0.99–0.99; PPV = 12.6%, 95% CI, 9.6%–16.1%; NPV = 100.0, 95% CI, 100.0%–100.0%), meanwhile decreasing the diagnosis threshold to 0.05 maximized specificity and PPV over sensitivity (sensitivity = 18.3%, 95% CI, 16.5%–20.3%; specificity = 99.3%, 95% CI, 99.2%–99.4%; AUROC = 0.59, 95% CI, 0.58–0.60; PPV = 69.6%, 95% CI, 65.0%–73.8%; NPV = 93.3%, 95% CI, 93.0%–93.7%). This indicates that when constructing prediction models, different diagnosis thresholds should be examined to achieve balanced sensitivity and specificity.

Of those patients with epilepsy identified by model 1 (n = 114), the most accurate method to identify incident epilepsy was to apply washout criteria exclusively to January 2009–December 2011 ($k = 0.79$, 95% CI, 0.66–0.92). The other tested approach (ie, washout periods of no claims indicative of epilepsy and no epilepsy-specific AED claims directly before the identified index event) for substantiating an incident epilepsy case resulted in a severe drop to the measure of agreement. A 48-month washout period resulted in a *k* coefficient of 0.02 (95% CI, −0.00 to 0.05); 6 months: $k = -0.30$ (95% CI, −0.38 to −0.21); 12 months: $k = -0.38$ (95% CI, −0.52 to −0.24); 18 months: $k = -0.47$ (−0.63 to −0.31); 24 months: $k = -0.49$ (95% CI, −0.66 to −0.32); and 36 months: $k = -0.52$ (95% CI, −0.71 to −0.34).

Supplemental Digital Content 4 (http://links.lww.com/MLR/B726) (Text) and Supplemental Digital Content 9 (http://links.lww.com/MLR/B721) (Table) provide further information on 2 additional high-performing models, and further analyses on model 1.

Sensitivity Analysis

Models derived from previously published algorithms[2,18] performed poorly in our reconstructed sample. One approach[2] maximized specificity at the cost of diminished sensitivity (sensitivity = 2.7%, 95% CI, 2.5%–3.0%; specificity = 100.0%, 95% CI, 99.9%–100.0%; PPV = 100.0%, 95% CI, 99.2%–100.0%; NPV = 22.6%, 95% CI 22.1%–23.2%; AUROC = 0.51, 95% CI 0.51–0.51). Another approach,[18] with which we set a conservative diagnosis threshold of $P > 0.10$, likewise maximized specificity and PPV over sensitivity and AUROC (sensitivity = 17.3%, 95% CI, 15.1%–19.8%; specificity = 98.7%, 95% CI, 98.5%–98.8%; PPV = 40.2%, 95% CI, 35.6%–44.9%; NPV = 95.9%, 95% CI, 95.6%–96.1%; AUROC = 0.58, 95% CI, 0.57–0.59).

Analyzing incident epilepsy cases using the date of first seizure diagnosis yielded results consistent with the date of first epilepsy diagnosis. Using the method of no claims between January 2009–December 2011, $k = 0.76$ (95% CI, 0.62–0.90); 48-month washout period, $k = 0.02$ (95% CI, −0.00 to 0.04); 6 months, $k = -0.29$ (95% CI, −0.37 to −0.21); 12 months, $k = -0.36$ (95% CI, −0.50 to −0.22); 18 months, $k = -0.44$ (95% CI, −0.60 to −0.29); 24 months, $k = 0.47$ (95% CI, −0.63 to −0.30); and 36 months, $k = -0.50$ (95% CI, −0.68 to −0.32). Supplemental Digital Content 4 (http://links.lww.com/MLR/B726) (Text) supplies final

TABLE 4. Performance Metrics of Highest Performing Claims-based Models

Model	AUROC (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Diagnosis Threshold
Model 1	0.93 (0.90–0.96)	87.8 (80.4–93.2)	98.4 (98.2–98.5)	23.1 (19.2–27.4)	99.9 (99.9–100.0)	> 0.80
Model 2	0.91 (0.88–0.94)	83.9 (76.0–90.0)	98.4 (98.2–98.5)	22.6 (18.8–26.8)	99.9 (99.9–99.9)	> 0.80
Model 3	0.80 (0.72–0.88)	80.0 (59.3–93.2)	79.5 (77.1–81.7)	7.4 (4.6–11.2)	99.5 (98.8–99.8)	> 0.80

Performance metrics of best-performing models are determined by the AUROC.

AUROC indicates area under the receiver operating characteristic curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

sensitivity analysis on model 1, where we stratify our sample by neurology specialty involvement.

DISCUSSION

In the first rigorous validation of models for epilepsy identification in the elderly using large datasets, we found that longitudinal Medicare administrative data can accurately detect prevalent and incident epilepsy using a multivariate prediction modeling approach. The most accurate model required a claims-based diagnosis code for epilepsy and evidence of epilepsy-specific drug use. Incorporating other relevant clinical data such as differential diagnosis and procedure data yielded similarly high levels of accuracy but were arguably more complex. In short, the general prediction approach and range of potential methods illustrate multiple viable options that others could match with their needs and data contexts (eg, surveillance studies with limited access to certain data).

The most recent estimates of epilepsy incidence and prevalence in the elderly population using Medicare administrative data in the United States revealed an annual incidence of 6.1/1000 individuals and prevalence of 15.2/1000 individuals.[14] This study utilized an unvalidated algorithm that, when applied to a nationwide sample of Medicare beneficiaries, yielded an annual prevalence of 10.8/1000.[2] Our study showed this approach[2] strongly favored specificity over sensitivity (specificity = 100%, 95% CI, 99.9%–100%; sensitivity = 2.7%, 95% CI, 2.5%–3.0%) when applied to our dataset, which suggests this algorithmic approach is not versatile to identify individuals with epilepsy in outside datasets and might not be suited for tracking patient care nationwide.

Previous algorithms have reported moderate performance albeit often requiring difficult tradeoffs. One recent study used data from the health information services department at a hospital in Melbourne, Australia to validate an algorithm utilizing ICD-10 codes for epilepsy and ≥ 1 AED, favoring high specificity at the cost of sensitivity (sensitivity = 60.8%, specificity = 99.9%, PPV = 81.4%).[17] Conversely, others reported high sensitivity (85%–95%) and specificity (87%–99%), but their models performed poorly when applied to a second dataset, suggesting overfitting.[11,12,15,16] Our predictive models were developed and validated in a split-sample approach, and attained high accuracy for the function of identifying epilepsy.

Yet, there are a host of additional difficulties that belie simple epilepsy and nonepileptic event identification in the elderly. Of note, preexisting dementia, for example, can obscure epilepsy symptoms (eg, preexisting confusional states, delirium). At the same time, events common in the elderly population (eg, syncope, transient ischemic events) often mimic symptoms of epilepsy and widen the differential diagnoses. Compared with prior validation studies that reviewed and assessed epilepsy diagnosis via nonspecialist physicians or researchers,[12,28] our reference-standard defined by a neurologist decreased the chance for erroneous assessment.

Further compounding the problem of diagnosis and treatment is difficulty in differentiating between nonepileptic and epileptic seizures based on a single claim. Our review of EHRs, which yielded an accurate portrayal of the elderly population with epilepsy, elucidated that nonepileptic events (ie, events such as syncope or altered mental status) are common among the elderly population [eg, 987 patients (4.9%) were diagnosed with a nonepileptic event], and these, in turn, represented most of the claims-based false-positives. The data are consistent with clinical practice and previous studies showing that the differentiation between epileptic and nonepileptic events remains a challenge, particularly in the elderly.[29]

Limitations

We designed our methods of sampling and data collection with prevalence in mind. Prediction of disease onset and development of future disease merit attention in future work. Similarly, our strategy of using claims data over a 3-year period for disease status prediction also relies on the diagnostic and treatment patterns within the US Medicare population aged 65 years or older. Thus, our findings have unclear generalizability to populations in other countries or in younger populations, which might have different care patterns.

In addition, sampling patients enrolled in the ACO may limit generalizability. These patients could have a higher (or lower) level of education and/or more severe disease (ie, a continuous need for Medicare enrollment); thus, patients with specific levels of education and/or less severe disease might be underrepresented. Although we did not have access to data on patient race or socioeconomic status through the ACO, we did document a broad range of both epilepsy syndromes/etiologies and differential diagnoses through EHR review (Tables, Supplemental Digital Content 5 (http://links.lww.com/MLR/B719) and Supplemental Digital Content 6 (http://links.lww.com/MLR/B719), which provide a breakdown of epilepsy syndrome/etiologies and differential diagnoses of patients in the reconstructed sample) (but we acknowledge that we are unable to assess the distribution of these factors).

Moreover, our study had a finite number of additional discriminatory analyses that we could feasibly perform. Although we had access to a large sample of patients spanning several hospitals and health centers, we did not interrogate the consistency of our model's performance across each discrete center. Further, we did not assess how accurately ICD-9 claims data captured the specific epilepsy syndrome, etiology, seizure type, or differential diagnosis (ie, subdecimal coding). The very limited studies that do classify epilepsy patients by etiologies based on claims still require validation.[27,30] Future validation is critical to provide accurate reports of epilepsy prevalence by etiology and syndromes,[31] and there is the potential to increase the accuracy of identifying nonepileptic events in the elderly to promote more nuanced treatment.

Finally, we used only one neurologist reviewer from a leading neurology department in this study. Because of potential interreviewer variation in determining epilepsy status, reproducibility of our findings might be limited.

CONCLUSIONS

In this validation study, prediction models using longitudinal data from large administrative datasets were able to identify epilepsy status accurately. These types of large dataset electronic disease signatures represent the first step in

monitoring epilepsy nationally and gathering real-world evidence on care for patients with epilepsy.

ACKNOWLEDGMENTS

The authors thank biostatistician Hang Lee, PhD, of Massachusetts General Hospital, who revised statistical analysis through the Harvard Catalyst statistical consulting program.

REFERENCES

1. Stovner LJ, Hoff JM, Svalheim S, et al. Neurological disorders in the Global Burden of Disease 2010 study. *Acta Neurol Scand Suppl*. 2014;198:1–6.
2. Faught E, Richman J, Martin R, et al. Incidence and prevalence of epilepsy among older US Medicare beneficiaries. *Neurology*. 2012;78:448–453.
3. Rowan AJ, Ramsay RE, Collins JF, et al. New onset geriatric epilepsy: a randomized study of gabapentin, lamotrigine, and carbamazepine. *Neurology*. 2005;64:1868–1873.
4. Leppik IE. Epilepsy in the elderly: scope of the problem. *Int Rev Neurobiol*. 2007:1–14.
5. Silveira DC, Jehi L, Chapin J, et al. Seizure semiology and aging. *Epilepsy Behav*. 2011;20:375–377.
6. Morillo LE. Temporal lobe epilepsy in the elderly. *Epilepsy Res Treat*. 2012;2012:641323.
7. Ramsay RE, Rowan AJ, Pryor FM. Special considerations in treating the elderly patient with epilepsy. *Neurology*. 2004;62:S24–S29.
8. Berg AT. Risk of recurrence after a first unprovoked seizure. *Epilepsia*. 2008;49(suppl 1):13–18.
9. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. *N Engl J Med*. 1988;26:1701–1707.
10. Tsiropoulos I, Gichangi A, Andersen M, et al. Trends in utilization of antiepileptic drugs in Denmark. *Acta Neurol Scand*. 2006;113:405–411.
11. Reid AY, St Germaine-Smith C, Liu M, et al. Development and validation of a case definition for epilepsy for use with administrative health data. *Epilepsy Res*. 2012;102:173–179.
12. Franchi C, Giussani G, Messina P, et al. Validation of healthcare administrative data for the diagnosis of epilepsy. *J Epidemiol Community Health*. 2013;67:1019–1024.
13. Betjemann JP, Josephson SA, Lowenstein DH, et al. Trends in status epilepticus-related hospitalizations and mortality: redefined in US practice over time. *JAMA Neurol*. 2015;72:650–655.
14. Ip Q, Malone DC, Chong J, et al. An update on the prevalence and incidence of epilepsy among older adults. *Epilepsy Res*. 2018;139:107–112.
15. Moura LM, Price M, Cole AJ, et al. Accuracy of claims-based algorithms for epilepsy research: revealing the unseen performance of claims-based studies. *Epilepsia*. 2017;58:683–691.
16. Jette N, Reid AY, Quan H, et al. How accurate is ICD coding for epilepsy? *Epilepsia*. 2010;51:62–69.
17. Tan M, Wilson I, Braganza V, et al. Development and validation of an epidemiologic case definition of epilepsy for use with routinely collected Australian health data. *Epilepsy Behav*. 2015;51:65–72.
18. Holden EW, Thanh Nguyen H, Grossman E, et al. Estimating prevalence, incidence, and disease-related mortality for patients with epilepsy in managed care organizations. *Epilepsia*. 2005;46:311–319.
19. Wassenaar M, Carpay JA, Sander JW, et al. Validity of health insurance data to identify people with epilepsy. *Epilepsy Res*. 2018;139:102–106.
20. Choi H, Pack A, Elkind MSV, et al. Predictors of incident epilepsy in older adults: the Cardiovascular Health Study. *Neurology*. 2017;88:870–877.
21. Yun H, Kilgore ML, Curtis JR, et al. Identifying types of nursing facility stays using medicare claims data: an algorithm and validation. *Health Serv Outcomes Res Methodol*. 2010;10:100–110.
22. Fisher RS, Cross JH, French JA, et al. Operational classification of seizure types by the International League Against Epilepsy: position paper of the ILAE Commission for Classification and Terminology. *Epilepsia*. 2017;58:522–530.
23. Scheffer IE, Berkovic S, Capovilla G, et al. ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. *Epilepsia*. 2017;58:512–521.
24. Toll DB, Janssen KJ, Vergouwe Y, et al. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol*. 2008;61:1085–1094.
25. Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ*. 2016;353:i3140.
26. Debray TP, Damen JA, Snell KI, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ*. 2017;356:i6460.
27. Thurman DJ, Beghi E, Begley CE, et al. Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia*. 2011;52(suppl 7):2–26.
28. Fonferko-Shadrach B, Lacey AS, White CP, et al. Validating epilepsy diagnoses in routinely collected data. *Seizure*. 2017;52:195–198.
29. Leppik IE. Epilepsy in the elderly: scope of the problem. *Int Rev Neurobiol*. 2007;81:1–14.
30. Kurth T, Lewis BE, Walker AM. Health care resource utilization in patients with active epilepsy. *Epilepsia*. 2010;51:874–882.
31. Fiest KM, Sauro KM, Wiebe S, et al. Prevalence and incidence of epilepsy. *Neurology*. 2017;88:296–303.