Published in final edited form as:

Epilepsy Res. 2020 October ; 166: 106414. doi:10.1016/j.epilepsyres.2020.106414.

Accuracy of ICD-10-CM claims-based definitions for epilepsy and seizure type

Jason R. Smith, BA^a, Felipe J.S. Jones, MD^a, Brandy E. Fureman, PhD^b, Jeffrey R. Buchhalter, MD^c, Susan T. Herman, MD^d, Neishay Ayub, MD^a, Christopher McGraw, MD, PhD^{a,e}, Sydney S. Cash, MD, PhD^{a,e}, Daniel B. Hoch, MD, PhD^{a,e}, Lidia M.V.R. Moura, MD, MPH^{a,e,f}

^aDepartment of Neurology, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA

^bResearch and New Therapies, Epilepsy Foundation, 8301 Professional Place West, Suite 230, Landover, MD 20785, USA

^cDepartment of Pediatrics, Cumming School of Medicine, University of Calgary, 3330 Hospital Drive NW, Calgary, AB T2N 4N1, CA

^dDepartment of Neurology, Barrow Neurological Institute, 350 W Thomas Road, Phoenix, AZ 85013, USA

^eDepartment of Neurology, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA

^fDepartment of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA

Abstract

Objective: To evaluate the accuracy of ICD-10-CM claims-based definitions for epilepsy and classifying seizure types in the outpatient setting.

Methods: We reviewed electronic health records (EHR) for a cohort of adults aged 18+ years seen by six neurologists who had an outpatient visit at a level 4 epilepsy center between 01/2019–09/2019. The neurologists used a standardized documentation template to capture the diagnosis of epilepsy (yes/no/unsure), seizure type (focal/generalized/unknown), and seizure frequency in the EHR. Using linked ICD-10-CM codes assigned by the provider, we assessed the accuracy of claims-based definitions for epilepsy, focal seizure type, and generalized seizure type against the reference-standard EHR documentation by estimating sensitivity (Sn), specificity (Sp), positive predictive value (PPV), and negative predictive value (NPV).

Corresponding author: Jason R. Smith, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, Phone: (617) 726-9218, Fax: (617) 726-9250, jason.smith@mgh.harvard.edu.

Data statement:

The authors are not authorized to make EHR data publicly accessible.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Results: There were 673 eligible outpatient encounters. After review of EHRs for standardized documentation, an analytic sample consisted of 520 encounters representing 402 unique patients. In the EHR documentation, 93.5% (n=486/520) of encounters were with patients with a diagnosis of epilepsy. Of those, 66.0% (n=321/486) had 1 focal seizure, 41.6% (n=202/486) had 1 generalized seizure, and 7% (n=34/486) had 1 unknown seizure. An ICD-10-CM definition for epilepsy (i.e., ICD-10 G40.X) achieved Sn=84.4% (95% CI 80.8–87.5%), Sp=79.4% (95% CI 62.1–91.3%), PPV=98.3% (95% CI 96.6–99.3%), and NPV=26.2% (95% CI 18.0–35.8%). The classification of focal vs generalized/unknown seizures achieved Sn=69.8% (95% CI 64.4–74.8%), Sp=79.4% (95% CI 72.4–85.3%), PPV=86.8% (95% CI 82.1–90.7%), and NPV=57.5% (95% CI 50.8–64.0%).

Conclusions: Claims-based definitions using groups of ICD-10-CM codes assigned by neurologists in routine outpatient clinic visits at a level 4 epilepsy center performed well in discriminating between patients with and without a diagnosis of epilepsy and between seizure types.

Keywords

Administrative Claims; Healthcare [D000067575]; Clinical Coding [D059019]; International Classification of Diseases [D038801]; Population Surveillance [D011159]; Validation Study [D023361]

1. Introduction

Epilepsy is a chronic, heterogeneous neurological disorder that affects approximately 3.4 million people in the United States (Zack and Kobau, 2017). Epilepsy can vary markedly in the age of onset and severity, and can substantially decrease the quality of life of those affected (e.g., lower income potential, unemployment) (IOM, 2012). Yet, despite the high prevalence and heterogeneity of epilepsy, there is limited national-level data on the quality and value of epilepsy care.

To appropriately measure and surveil care quality for people with epilepsy at the population level, researchers and other stakeholders require accurate data. Currently, the most powerful tool for population-based research is the use of large administrative claims datasets, which can include diagnosis codes recorded after a clinical encounter for internal administrative purposes or codes billed for health insurance (Jette et al., 2010a). These data are increasingly used to obtain population-level estimates of incidence and prevalence (Faught et al., 2012; Helmers et al., 2015; Jette et al., 2010a), measure financial burden (Begley and Durgin, 2015; Pisu et al., 2019), and measure the value of health care provided to people with epilepsy (AHRQ, 2019; Burneo et al., 2016; Hill et al., 2019). Claims data overcome many methodological challenges by enabling systematic collection at both the regional and national levels (Thurman et al., 2011).

Claims-based definitions for epilepsy using International Classification of Diseases-10th Revision (ICD-10) have been validated in care settings in Canada and elsewhere (Jette et al., 2010b; Reid et al., 2012; Tan et al., 2015; Tu et al., 2014). The accuracy of claims-based algorithms varies considerably by region and health care system (Moura et al., 2019; Moura

et al., 2017), however, which could hinder the ability of studies that use claims data to generate valid estimates of epilepsy prevalence/incidence or measure the quality of care in epilepsy (Faught et al., 2012; Hill et al., 2019). Thus, the validity of ICD-10 claims-based definitions in the U.S. for epilepsy, seizure type (Scheffer et al., 2017), and epilepsy severity (e.g., uncontrolled seizures) remains uncertain.

To bridge this gap, we conducted a cross-sectional study to evaluate the accuracy of ICD-10-CM claims-based definitions of the diagnosis of epilepsy and seizure type using groups of codes assigned by neurologists during routine outpatient clinical care.

2. Methods

This study is part of an ongoing quality improvement initiative at a level 4 epilepsy center (Massachusetts General Hospital) (Labiner et al., 2010). Beginning in 01/2019, six epilepsy providers volunteered to use a standardized documentation template that was built into the EHR to capture the diagnosis of epilepsy, seizure frequency, and seizure type as part of routine epilepsy care for all patients they treated (henceforth, reference-standard documentation). We retrospectively analyzed these data to examine the accuracy of claims-based definitions for the diagnosis of epilepsy (i.e., yes vs no) and seizure type (i.e., focal vs generalized/unknown; generalized vs focal/unknown).

This study received approval from the Partners Healthcare institutional review board and need for informed consent was waived.

2.1. Data sources

Using one source of longitudinal data (Partners EHRs) (<https://www.partners.org/>), we created two datasets: (1) reference-standard documentation of provider-reported data for the diagnosis of epilepsy, seizure type, and seizure frequency generated during the encounter; (2) ICD-10-CM diagnosis codes provided by the treating neurologist generated at the end of the encounter. For outpatient visits scheduled with an epilepsy fellow, the fellows provide the ICD-10-CM diagnosis codes and the attendings subsequently review.

The Partners EHR contains all patient medical records, including inpatient and outpatient encounter information, procedures, prescriptions, laboratory values, and diagnosis codes for encounters across all patients seen within Partners Healthcare. We obtained the reference-standard documentation of provider-reported data through manual review of the EHR in the specific outpatient encounters between 01/2019–09/2019 (Supplemental Appendix 1).

We queried ICD-10-CM diagnosis codes for each encounter through the Partners Research Patient Data Registry (RPDR) (Supplemental Appendix 1) (<https://rpdrssl.partners.org/Information>). The Partners RPDR is a centralized data repository that captures demographic and clinical encounter data (e.g., primary diagnosis codes for encounters) from the EHRs of all patients seen within the Partners Healthcare system for inpatient and outpatient visits. For our purposes, we obtained all ICD-10-CM diagnosis codes assigned by the treating provider (i.e., visit diagnosis codes) at the end of the outpatient encounter at the level 4 epilepsy

center. We linked provider-reported reference-standard documentation and ICD-10-CM codes using encounter-level identifiers.

2.2. Patient population

Figure 1 details our patient population. We established our study period from 01/20/2019–09/27/2019 and defined eligibility as all outpatient encounters (including both new and follow-up encounters) with adult patients aged 18+ years seen by the providers (six in total) at the level 4 epilepsy center that adopted the reference-standard documentation (n=673). After excluding encounters with missing reference-standard documentation in the EHR (n=153; 22.7%), there were 520 encounters comprising 402 unique adult patients with ICD-10-CM data linked with the documentation in the analytic sample.

2.3. EHR Review for Reference-standard Documentation of Epilepsy Diagnosis and Seizure Type

Providers documented patient's clinical information pertaining to epilepsy with an interactive, standardized documentation template (i.e., the reference-standard documentation) built into EHR notes for routine clinical care. The reference-standard documentation template with response skip-logic is provided in Supplemental Appendix 2. All participating providers were neurologists, and all but one (LM) were blinded to study aims during the generation of these data throughout the study period (Supplemental Appendix 1). The providers used all relevant clinical, EEG, and imaging data to assess patients regarding the diagnosis of epilepsy (Fisher et al., 2014), seizure classification (Fisher et al., 2017; Scheffer et al., 2017) and diagnostic certainty (Loring et al., 2011), and seizure frequency. All treating providers, moreover, had similar access to levels of available data.

Two reviewers abstracted these reference-standard documentation data from the patient's outpatient encounter note in the EHR (Supplemental Appendix 1) and defined incomplete reference-standard documentation for the encounter as either 1) absence of reference-standard documentation in the encounter note in the EHR, or 2) an incomplete answer for at least one required question within the skip logic in the reference-standard documentation. For example, if a provider recorded "Yes" for the diagnosis of epilepsy, recorded one seizure type (e.g., tonic-clonic), but did not provide an answer for seizure frequency, the reference-standard documentation would be considered incomplete.

The reviewers classified reference-standard epilepsy as a provider-reported response of "Does the patient have epilepsy?: Yes" in the documentation. Among patients with a diagnosis of epilepsy, the reviewers classified reference-standard generalized seizures as patients who had provider-reported generalized-onset seizures (e.g., tonic-clonic seizures), and reference-standard focal seizures as those patients with provider-reported focal-onset seizures (e.g., focal aware without impairment of consciousness, with observable motor components). And the reviewers classified reference-standard unknown seizures as those patients with provider-reported unknown-onset seizure type.

2.4. ICD-10-CM Claims-based Definitions

All providers at the level 4 epilepsy center select up to five diagnosis ICD-10-CM codes at the end of outpatient encounters (Supplemental Appendix 1). All but one of the providers (LM) were blinded to the study aims during data generation throughout the study period and we used these provider-selected diagnosis codes to generate our claims-based definitions.

Supplemental Table 1 provides a detailed description of the ICD-10-CM definitions of the diagnosis of epilepsy and seizure types; each definition utilized up to five diagnosis code fields. We created two claims-based definitions of epilepsy: (1) ICD-10-CM codes G40.X only (epilepsy and recurrent seizures) in any position; (2) ICD-10-CM codes G40.X and/or R56.X (convulsions, not elsewhere classified) in any position. The latter definition therefore catches those coded with only G40.X, only R56.X, or both G40.X and R56.X. Although R56.X is intended for coding of non-epileptic convulsions, the inclusion of these codes is recommended for identifying epilepsy under specific conditions, for example when the certainty level of the diagnosis of epilepsy is not probable (Jette et al., 2010b; Thurman et al., 2011). We defined generalized seizure type as codes G40.3X, G40.4X, G40.AX, G40.BX, G40.81X, or G40.82X in any position and codes G40.0X, G40.1X, G40.2X in any position for focal seizure type.

2.5. Statistical analysis

We conducted a complete-case analysis of these data following matching of provider-reported reference-standard documentation and ICD-10-CM coded diagnosis data. First, we summarized analytic sample demographic characteristics among unique patients (n=402) and clinical characteristics (i.e., using reference-standard documentation data) at the encounter level (n=520). We reported parametric data with mean (SD), nonparametric variables with median (IQR), and categorical data with absolute frequency (n, %). To assess study representativeness of the analytic sample, we compared demographic characteristics between included and excluded patients (i.e. missing reference-standard documentation) using χ^2 test of independence, Fisher's exact test, and Mann-Whitney U test with a two-sided alpha level of 0.05.

In the main analysis, we assessed the performance of claims-based definitions of the diagnosis of epilepsy (i.e., yes vs no) and seizure types (i.e., generalized vs focal/unknown; focal vs generalized/unknown) using encounters as the level of analysis. It's reasonable to expect that patients with a diagnosis of epilepsy constitute a larger share of the patient population at a level 4 epilepsy center, and would thus influence predictive values. Within this context, we estimated prevalence, sensitivity (Sn), specificity (Sp), positive predictive value (PPV), and negative predictive value (NPV) and 95% confidence intervals (CI) for each.

We then conducted several sensitivity analyses. First, because previous studies demonstrated that the accuracy of epilepsy claims-based definitions can markedly vary when applied externally to older adult (i.e., 65 years) populations (Moura et al., 2017; Moura et al., 2019), we analyzed definition performance between older adult and younger adult (i.e., 18 to 64 years) patients at the encounter level. Second, because one provider was un-blinded to the

study aims during data generation, we examined definition performance after excluding encounters with the un-blinded provider.

Third, because some of the reference-standard characteristics of the excluded encounters are unknown (e.g., reference-standard diagnosis of epilepsy), we examined the boundaries of this potential missing data bias in two ways: a) assuming all excluded encounters had a reference-standard diagnosis of epilepsy with no code suggestive of epilepsy; and b) assuming all excluded encounters had no reference-standard diagnosis of epilepsy with a code suggestive of epilepsy.

Fourth, because the number of seizures a patient has might impact diagnosis coding accuracy, we analyzed definition performance for seizure type between those with only one seizure type and those with more than one seizure type. For example, a patient's history might include multiple seizure types, yet only be coded for one seizure type at the end of the encounter.

And fifth, because diagnosis coding can be dynamic (i.e., one patient might have well-controlled epilepsy in January yet return in March with the report of a seizure, which should be reflected in the ICD-10-CM code for the specific encounter), we analyzed definition performance using individual patients as the unit of analysis (i.e., among patients with multiple encounters, use only final encounter).

We also performed two secondary analyses. First, we summarized descriptive statistics of the demographic and clinical characteristics of patients with claims-based definitions for “convulsions, not elsewhere classified” (i.e., R56.X) with respect to reference-standard diagnosis of epilepsy and seizure type.

Second, we explored the performance of an ICD-10-CM claims-based definition for epilepsy severity by estimating the accuracy of one, albeit partial, definition of uncontrolled seizures (Sapkota et al., 2018; Zack and Kobau, 2017). We defined the reference-standard of uncontrolled seizures as having a diagnosis of epilepsy plus a seizure frequency of at least one within the past year. Due to the lack of a one-to-one match of uncontrolled seizures in ICD-10-CM codes (i.e., ICD-10-CM codes do not specify seizure frequency of relative date of occurrence), we selected ICD-10-CM codes that included “intractable epilepsy” indicators (more commonly referred to as drug-resistant epilepsy)(Kwan et al., 2010) (G40.01X, G40.11X, G40.21X, G40.31X, G40.A1X, G40.B1X, G40.41X, G40.803, G40.804, G40.813, G40.814, G40.823, G40.824, G40.91X) for a claims-based definition of uncontrolled seizures. Although there is potential to conflate uncontrolled seizures with “intractable”, or drug-resistant, epilepsy when treatment and outcomes vary (Hao et al., 2013), this signifies a novel and important first step towards systematic identification of epilepsy severity, and a potentially more feasible target for intervention, at the population level using claims data (Patel et al., 2018).

We used Stata 16 (StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC) to merge datasets and to perform statistical analysis. We provide our code in Supplemental Appendix 3.

3. Results

3.1. Demographic and clinical characteristics of the analytic sample

The analytic sample consisted of 520 new and follow-up outpatient encounters (402 adult patients) with reference-standard documentation between 01/2019–09/2019. Most patients in the analytic sample were English-speaking ($n=375/402$; 93.3%), young and middle-aged adults (median=43.8 years, IQR=29.8–57.5 years; $n=219$ female, 54.5%) (Table 1). There was no significant difference in age, sex, or primary language between the analytic sample and excluded patients ($n=126$ unique patients; $n=153$ total encounters; Supplemental Table 2). The proportion of encounters by each of the six providers in the analytic sample ranged from 2% ($n=12/520$) to 28% ($n=145/520$).

Among all encounters in the analytic sample, 93.5% ($n=486/520$) involved patients with a diagnosis of epilepsy (i.e., provider-reported reference-standard documentation), 4.4% ($n=23/520$) with no diagnosis of epilepsy, and 2.1% ($n=11$) with an unsure diagnosis. Among those with a diagnosis of epilepsy, 66.0% ($n=321$) had at least one focal seizure, 41.6% ($n=202$) had at least one generalized seizure, and 7.0% ($n=34$) had at least one unknown seizure type per reference-standard documentation (Table 2). 38.1% ($n=185$) of those encounters involved patients with more than one seizure type and 10.3% ($n=50$) involved patients with both focal and generalized seizures. Over half of those with a diagnosis of epilepsy had 1 seizure within the past year ($n=264$; 54.3%).

Of those encounters involving patients with reference-standard “unsure” diagnosis of epilepsy, 36.4% ($n=4/11$) were assigned an ICD-10-CM diagnosis code suggestive of epilepsy (G40.X), 27.3% ($n=3/11$) were assigned the “convulsion, not elsewhere classified” (R56.X) code, 27.3% ($n=3/11$) were assigned non-specific epilepsy codes (G40.802, other epilepsy, not intractable, without status epilepticus; G40.804, other epilepsy, intractable, without status epilepticus) that were not included in the claims-based definitions for focal or generalized seizures; and 9.0% ($n=1/11$) were assigned codes that were not potentially suggestive of epilepsy.

3.2. Performance of claims-based definitions

Table 3 displays the performance of ICD-10-CM claims-based definitions of epilepsy and seizure type. In a cohort of adult patients who had an outpatient encounter with a neurologist at a level 4 epilepsy center between 01/2019–09/2019, a claims-based definition of having at least one provider-assigned ICD-10-CM code suggestive of epilepsy (i.e., G40.X) in any position achieved a PPV of 98.3% (95% CI 96.6–99.3%), NPV of 26.2% (95% CI 18.0–35.8%), Sn of 84.4% (95% CI 80.8–87.5%) and Sp of 79.4% (95% CI 62.1–91.3%). Expanding the claims-based definition of epilepsy to include “convulsions, not elsewhere classified” (R56.X) yielded a PPV of 96.2% (95% CI 94.2–97.7%), NPV of 57.7% (95% CI 36.9–76.6%), Sn of 97.7% (95% CI 91.0–95.4%), and Sp of 44.1% (95% CI 27.2–62.1%) (Table 3).

The claims-based definitions for focal seizure type (versus generalized and unknown type) achieved a PPV of 86.8% (95% CI 82.1–90.7%), and the claims-based definition for generalized seizures attained a PPV of 84.9% (95% CI 75.5–91.7%) (Table 3).

3.3. Sensitivity analysis

Repeated analyses of performance metrics are reported in Supplemental Tables 3 through 7 and the overall performance of claims-based definitions remained largely unchanged.

3.4. Secondary analyses

Among all outpatient encounters, 14.6% (n=76) contained a provider-assigned primary diagnosis code of R56.X (“convulsions, not elsewhere classified”) (Table 4 and Supplemental Table 8). Of those, 85.5% (n=65) had a reference-standard diagnosis of epilepsy, 50.8% (n=33) had reference-standard focal seizures, and 44.6% (n=29) had reference-standard generalized seizures.

There were 127 encounters with a provider-reported code suggestive of uncontrolled seizures. The exploratory claims-based definition of uncontrolled seizures achieved a PPV of 78.0% (95% CI 69.7–84.8%) for patients with reference-standard uncontrolled seizures, achieved Sn=37.5% (95% CI 31.6–43.6%), Sp=87.4% (95% CI 82.3–91.5%), and NPV=54.0% (95% CI 48.7–59.3%).

4. Discussion

There is limited data on the accuracy of ICD-10-CM claims-based definitions for epilepsy, as well as seizure type and epilepsy severity. In this study, we used a standardized documentation template to capture the diagnosis of epilepsy, seizure type, and seizure frequency among a sample of adult patients that received routine outpatient epilepsy care at a level 4 epilepsy center and linked this data with diagnosis codes indicated by the provider for the encounter. Our findings indicate ICD-10-CM claims-based definitions performed well in identifying patients with epilepsy and those with focal seizure type in this outpatient setting.

To our knowledge, this is the first study examining the accuracy of ICD-10-CM definitions in the U.S., so it remains to be seen how favorably our results compare among different inpatient (e.g., emergency department, epilepsy monitoring unit) or outpatient (e.g., general neurology, primary care) settings within the country. Our findings should be interpreted with respect to the large share of our outpatient population with a diagnosis of epilepsy (94% of all encounters with an epilepsy provider; n=486). The accuracy of these definitions in capturing true negatives in other settings, for example, is limited even with a high sensitivity. Only 4% (n=23/520) of the patients seen by the six participating neurologists did not have a diagnosis of epilepsy, meaning the NPV and specificity estimates for claims-based definitions of the diagnosis of epilepsy were based on a limited number of negative cases. Therefore these estimates should be interpreted cautiously.

Country-specific code differences and different care settings might further pose challenges for direct comparisons to outside findings (St Germaine-Smith et al., 2012; Thurman et al., 2011). In terms of ICD-10 coding, two studies in Canada, however, attained similar accuracy of ICD-10 coded definitions (G40+G41, excluding convulsion codes) using different case ascertainment methods: among seizure monitoring unit admissions, Sn=98.8% (95% CI 93.3–99.8%) and Sp of 69.6% (95% CI 55.2–80.9%) (Jette et al., 2010b); among patients

seen by primary care physicians, Sn=90.5% (95% CI 84.6–96.4%), Sp=98.0 (95% CI 97.7–98.3%) (Tu et al., 2014). And in Australia, ICD-10-AM coded definitions (G40+G41, excluding convulsion codes) among inpatient admissions achieved Sn=67.7% (95% CI 59.8–74.9%) and Sp=99.9% (95% CI 99.9–99.9%) with excellent discriminatory accuracy (AUROC=0.84; 95% CI 0.80–0.87) (Tan et al., 2015).

We observed that 15.6% of patients with the diagnosis of epilepsy were incorrectly coded with R56.X (i.e., convulsions, not elsewhere classified) and most of these patients presented with at least one focal seizure, although we did not calculate significance levels for this comparison. Moreover, the PPV (96.2%) did not significantly decrease when including R56.X codes in the claims-based definition for epilepsy (as shown in Table 2), likely stemming from the high prevalence of epilepsy in our patient population (i.e., patients seen for routine outpatient epilepsy care, of whom 93.0% had a true provider diagnosis of epilepsy).

In terms of seizure types, classification between types in claims data is important for epidemiological surveillance and anticonvulsant comparative effectiveness research, yet there has been no clear path forward for researchers due to the limited validity of seizure type (Thurman et al., 2011). The prior International League Against Epilepsy (ILAE) classification (2014) posed several problems for classifying epilepsy and seizure type (Jette et al., 2015) in previous studies that examined the accuracy of ICD-10 codes provided by general practitioners or neurologists against the prior ILAE classification and terminology (Jette et al., 2010b; Reid et al., 2012; Tan et al., 2015; Tu et al., 2014). Petit mal seizure terminology of ICD-10-CM, for example, had no ILAE equivalent (Tan et al., 2015).

We found that our ICD-10-CM claims-based definitions performed well in identifying patients with true focal seizures (PPV=86.8%; Sp=79.4%) and generalized seizure type (PPV=84.9%, Sp=95.4%); using these ICD-10-CM code groups for research in the outpatient epilepsy clinic setting might capture a high percentage of those with true seizure types. This is one modest step forward towards ascertainment of seizure types at the population level, which has historically suffered from misclassification (WHO, 2019). Accurate population-level estimates of the prevalence of specific seizure types, for example, are highly useful in estimating risk factors and determinates of disease. However, the definition for generalized seizures yielded poor sensitivity (36.1%) despite the fact that the diagnosis codes were provided by neurologists at a level 4 epilepsy center (i.e., where we might expect higher diagnostic accuracy and coding). These findings in part underscore the rift between the clinical guidelines and classifications used by practicing physicians (i.e., ILAE classification) and the ICD-10 coding scheme; epilepsy type, for example, is how ICD-10 codes are grouped and do not definitively distinguish seizure types. The upcoming ICD-11 coding scheduled to release in 2022 in the U.S. though aims to bridge this gap through increased harmonization with the 2017 ILAE classification (Bergen et al., 2012; ICD-11 Coding Tool; Jette et al., 2015).

Our study has a number of strengths. We used a highly accurate reference-standard to classify the diagnosis of epilepsy, seizure type, and seizure frequency. Neurologists with sub-specialization in epilepsy or in epilepsy fellowship training also made the diagnoses for

all patients and used a reference-standard documentation template aligned with updated 2017 ILAE classification for epilepsy and seizure type (including diagnostic certainty). And in contrast to prior studies in which professional coders provided the ICD-10 coding (Jette et al., 2010b; Tan et al., 2015; Tu et al., 2014), the neurologist provided the diagnosis codes for the outpatient encounter. Taken together these support the internal validity of our findings, as measurement error and exposure misclassification bias are common in epidemiological studies of epilepsy (Faught et al., 2012; Helmers et al., 2015; Jette et al., 2010b; Moura et al., 2017; Tan et al., 2015; Tu et al., 2014; WHO, 2019). Future learnings on the accuracy of diagnostic codes for the diagnosis of epilepsy and seizure type will help facilitate the use of large datasets to answer key questions in epilepsy epidemiology or even health services research at the population level.

4.1. Limitations

We recognize several limitations in our study. Whether or not our findings generalize to other settings is possibly limited by a few elements. First, we used a sample for primary analysis that was not randomly selected, reflecting only those patients seen by volunteering providers using the reference-standard documentation, and further limited to those that provided complete documentation. However the demographic and clinical characteristics of our sample (Tables 1, 2) compare favorably to a typical comprehensive epilepsy center population (Labiner et al., 2010), where we would assume that a greater number of individuals with more severe or complex epilepsies (e.g., active or drug-resistant) might be referred, which supports the validity of our findings in this setting. Second, it is possible that providers that are more likely to adopt a standardized documentation template during routine outpatient visits are also more likely to provide accurate diagnosis coding; the process of completing the documentation itself might also lead to more accurate coding, though with no reference-standard available prior to this new implementation at our epilepsy center we did not estimate a possible effect. And third, neurologists provided the diagnosis codes; these providers could yield more accurate epilepsy coding than a provider that does not specialize in epilepsy. These latter two points might not accurately represent coding practices across settings and our findings should be interpreted cautiously.

Our data, moreover, could still indicate variance in coding patterns among epilepsy providers and fellows even though it is reasonable to assume that they provide more precise epilepsy coding. For example, we cannot exclude the possibility that they might be more likely to document and/or code only one seizure type that is the more severe or in which they have greater diagnostic certainty (e.g., documenting/coding only generalized seizure in patients with both generalized and focal seizures). However, we found no differences in the performance of our claims-based definitions when comparing patients with only one seizure to those presenting with two or more seizures.

Future studies should examine increasingly complex ICD-10 coded definitions to identify epilepsy type, and could incorporate anticonvulsants, provider level (e.g., epileptologist versus epilepsy fellow), conditions that can mimic epilepsy symptomology (e.g., psychogenic non-epileptic seizures, syncope), or temporality of primary diagnosis codes between encounters. Because the diagnosis of epilepsy and classification of seizure type

might not be refined during an initial encounter with a neurologist, the stability of ICD-10-CM diagnosis codes over time by the same treating provider should also be investigated. And to bolster the representativeness of findings reliant upon a standardized documentation tool, studies could also examine coding accuracy before the implementation of documentation.

One final area for future investigation should include further development of more nuanced definitions of uncontrolled seizures, which could plausibly address one of the increasing challenges of measuring gaps in epilepsy care at the population level. (Hill et al., 2019; US DHHS, 2014) Although our novel exploratory analysis investigated just a fractional definition of uncontrolled seizures (Sapkota et al., 2018; Thurman et al., 2011; Zack and Kobau, 2017), and did not precisely align with the claims-based definition, this is a critical first step towards operationally defining aspects of quality epilepsy care and analyzing the management of people with epilepsy (Jones and Patel, 2018; Patel et al., 2018, US DHHS, 2014). This is particularly relevant as other data sources relevant to disease severity (e.g., patient-reported outcome data) are not always widely available for systematic collection at the population level. Further refinement of definitions for uncontrolled seizures should include successful anticonvulsant trials, drug dosing during trials, adherence to therapy, and potentially five years of patient-level data to ascertain the most recent seizure (Kwan et al., 2010; Thurman et al., 2011).

5. Conclusions

In a population of patients seen for routine outpatient epilepsy care at a level 4 epilepsy center, ICD-10-CM claims-based definitions performed well in identifying patients with a diagnosis of epilepsy as well as capturing seizure type. To systematically identify other aspects of epilepsy severity at the population level, further improvements in the current claims data system or inclusion of other data are likely necessary.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

The authors would like to thank Dr. Sahar Zafar, a member of our multidisciplinary team that contributed to data collection.

Disclosure of funding:

This study was funded by the NIH (1K08AG053380-01A1) to design and conduct all aspects of the study (i.e., data collection and management, data analysis and interpretation, manuscript preparation and review). This project is also part of a local quality improvement project, which has been endorsed by the Epilepsy Learning Healthcare System and Epilepsy Foundation and the American Academy of Neurology (Health Service Research Subcommittee).

Declaration of interest:

J.R.S. reports no disclosures. F.J.S.J. reports no disclosures. B.E.F. receives salary support from the Epilepsy Foundation, and funding from UCB Biopharma (Human Epilepsy Project 2) and PCORI (PPRN-1306-04577 (Rare Epilepsy Network) and the sponsors of the Epilepsy Learning Healthcare System (detailed below), and reports no other disclosures. J.R.B. receives funding from Epilepsy Foundation (Epilepsy Learning Healthcare System),

Epilepsy Study Consortium, Pediatric Learning Healthcare System, Epilog, UCB, and reports no other disclosures. S.T.H. receives funding from UCB Biopharma, Epilepsy Therapy Development Project, NIH (R01NS047029), and Epilepsy Foundation (Epilepsy Learning Healthcare System), and reports no other disclosures. N.A. reports no disclosures. C.M. reports no disclosures. S.S.C receives funding from NIH and NINDS (R01NS062092, K24 NS088568) and reports no other disclosures. D.B.H reports no disclosures. L.M.V.R.M receives funding from NIH (1K08AG053380–01A1 and 1R01AG062282–01), Epilepsy Foundation, and reports no other disclosures.

The Epilepsy Learning Healthcare System is supported by grants from PCORI and the James M. Anderson Center for Health Systems Excellence (RI-PCC-2017 (sub: 303699), the CDC (1NU58DP006256–02–00), the National Association of Epilepsy Centers, Greenwich Biosciences, Eisai, and Sunovion. Although funded in part by the CDC, the contents of this publication are solely the responsibility of the Epilepsy Foundation and do not necessarily represent the views of the CDC.

REFERENCES

- ICD-11 Coding Tool. World Health Organization Mortality and Morbidity Statistics (MMS). Available at: https://icd.who.int/ct11_2018/icd11_mms/en/release#/.
- Institute of Medicine. Epilepsy across the spectrum: promoting health and understanding. Washington, DC: The National Academy Press; 2012 <https://www.nap.edu/read/13379/chapter/3#25>.
- Partners Health Care Research Patient Data Registry. Information: RPDR Overview. <https://rpdrrsl.partners.org/Information>.
- Partners Health Care. <https://www.partners.org/>.
- US Department of Health and Human Services. Healthy People 2020. Barriers to health care: DH-6. Office of Disease Prevention and Health Promotion; 2014 Available at: <https://www.healthypeople.gov/2020/topics-objectives/topic/disability-andhealth/objectives>.
- World Health Organization. 2019 Epilepsy: A Public Health Imperative. Global Burden of Epilepsy; Pages 10–18. Available at: https://www.who.int/mental_health/neurology/epilepsy/report_2019/en/.
- 2019 Agency for Healthcare Research and Quality. 2018 National healthcare quality and disparities report. Available at: <https://www.ahrq.gov/research/findings/nhqrdr/nhqrdr18/index.html>.
- Begley CE, Durgin TL, 2015 The direct cost of epilepsy in the United States: A systematic review of estimates. *Epilepsia* 56, 1376–1387. [PubMed: 26216617]
- Bergen DC, Beghi E, Medina MT, 2012 Revising the ICD-10 codes for epilepsy and seizures. *Epilepsia* 53 Suppl 2, 3–5.
- Burneo JG, Shariff SZ, Liu K, Leonard S, Saposnik G, Garg AX, 2016 Disparities in surgery among patients with intractable epilepsy in a universal health system. *Neurology* 86, 72–78. [PubMed: 26643546]
- Faught E, Richman J, Martin R, Funkhouser E, Foushee R, Kratt P, Kim Y, Clements K, Cohen N, Adoboe D, Knowlton R, Pisu M, 2012 Incidence and prevalence of epilepsy among older US Medicare beneficiaries. *Neurology* 78, 448. [PubMed: 22262750]
- Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, Engel J Jr., Forsgren L, French JA, Glynn M, Hesdorffer DC, Lee BI, Mathern GW, Moshe SL, Perucca E, Scheffer IE, Tomson T, Watanabe M, Wiebe S, 2014 ILAE official report: a practical clinical definition of epilepsy. *Epilepsia* 55, 475–482. [PubMed: 24730690]
- Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE, Lagae L, Moshe SL, Peltola J, Roulet Perez E, Scheffer IE, Zuberi SM, 2017 Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. *Epilepsia* 58, 522–530. [PubMed: 28276060]
- Hao X, Goldberg D, Kelly K, Stephen L, Kwan P, Brodie MJ, 2013 Uncontrolled epilepsy is not necessarily the same as drug-resistant epilepsy: differences between populations with newly diagnosed epilepsy and chronic epilepsy. *Epilepsy Behav* 29, 4–6. [PubMed: 23911352]
- Helmert SL, Thurman DJ, Durgin TL, Pai AK, Faught E, 2015 Descriptive epidemiology of epilepsy in the U.S. population: A different approach. *Epilepsia* 56, 942948.
- Hill CE, Lin CC, Burke JF, Kerber KA, Skolarus LE, Esper GJ, Magliocco B, Callaghan BC, 2019 Claims data analyses unable to properly characterize the value of neurologists in epilepsy care. *Neurology* 92, e973–e987. [PubMed: 30674587]

- Jette N, Beghi E, Hesdorffer D, Moshe SL, Zuberi SM, Medina MT, Bergen D, 2015 ICD coding for epilepsy: past, present, and future--a report by the International League Against Epilepsy Task Force on ICD codes in epilepsy. *Epilepsia* 56, 348–355. [PubMed: 25684068]
- Jette N, Quan H, Hemmelgarn B, Drosler S, Maass C, Moskal L, Pao W, Sundararajan V, Gao S, Jakob R, Ustun B, Ghali WA, 2010a The development, evolution, and modifications of ICD-10: challenges to the international comparability of morbidity data. *Medical care* 48, 1105–1110. [PubMed: 20978452]
- Jette N, Reid AY, Quan H, Hill MD, Wiebe S, 2010b How accurate is ICD coding for epilepsy? *Epilepsia* 51, 62–69. [PubMed: 19682027]
- Jones LK Jr., Patel AD, 2018 The road to quality: Paved with more than good intentions. *Neurology* 91, 817–819. [PubMed: 30282767]
- Kwan P, Arzimanoglou A, Berg AT, Brodie MJ, Allen Hauser W, Mathern G, Moshe SL, Perucca E, Wiebe S, French J, 2010 Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies. *Epilepsia* 51, 1069–1077. [PubMed: 19889013]
- Labiner DM, Bagic AI, Herman ST, Fountain NB, Walczak TS, Gumnit RJ, 2010 Essential services, personnel, and facilities in specialized epilepsy centers--revised 2010 guidelines. *Epilepsia* 51, 2322–2333. [PubMed: 20561026]
- Loring DW, Lowenstein DH, Barbaro NM, Fureman BE, Odenkirchen J, Jacobs MP, Austin JK, Dlugos DJ, French JA, Gaillard WD, Hermann BP, Hesdorffer DC, Roper SN, Van Cott AC, Grinnon S, Stout A, 2011 Common data elements in epilepsy research: development and implementation of the NINDS epilepsy CDE project. *Epilepsia* 52, 1186–1191. [PubMed: 21426327]
- Moura L, Smith JR, Blacker D, Vogeli C, Schwamm LH, Cole AJ, HernandezDiaz S, Hsu J, 2019 Epilepsy Among Elderly Medicare Beneficiaries: A Validated Approach to Identify Prevalent and Incident Epilepsy. *Medical care* 57, 318–324. [PubMed: 30762723]
- Moura LM, Price M, Cole AJ, Hoch DB, Hsu J, 2017 Accuracy of claims-based algorithms for epilepsy research: Revealing the unseen performance of claims-based studies. *Epilepsia* 58, 683–691. [PubMed: 28199007]
- Patel AD, Baca C, Franklin G, Herman ST, Hughes I, Meunier L, Moura L, Munger Clary H, Parker-McFadden B, Pugh MJ, Schultz RJ, Spanaki MV, Bennett A, Josephson SA, 2018 Quality improvement in neurology: Epilepsy Quality Measurement Set 2017 update. *Neurology* 91, 829–836. [PubMed: 30282773]
- Pisu M, Richman J, Szaflarski JP, Funkhouser E, Dai C, Juarez L, Faught E, Martin RC, 2019 High health care costs in minority groups of older US Medicare beneficiaries with epilepsy. *Epilepsia* 60, 1462–1471. [PubMed: 31169918]
- Reid AY, St Germaine-Smith C, Liu M, Sadiq S, Quan H, Wiebe S, Faris P, Dean S, Jette N, 2012 Development and validation of a case definition for epilepsy for use with administrative health data. *Epilepsy Res* 102, 173–179. [PubMed: 22727659]
- Sapkota S, Kobau R, Pastula DM, Zack MM, 2018 Close to 1 million US adults aged 55 years or older have active epilepsy-National Health Interview Survey, 2010, 2013, and 2015. *Epilepsy Behav* 87, 233–234. [PubMed: 30078516]
- Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, Hirsch E, Jain S, Mathern GW, Moshe SL, Nordli DR, Perucca E, Tomson T, Wiebe S, Zhang YH, Zuberi SM, 2017 ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. *Epilepsia* 58, 512–521. [PubMed: 28276062]
- St Germaine-Smith C, Metcalfe A, Pringsheim T, Roberts JI, Beck CA, Hemmelgarn BR, McChesney J, Quan H, Jette N, 2012 Recommendations for optimal ICD codes to study neurologic conditions: a systematic review. *Neurology* 79, 1049–1055. [PubMed: 22914826]
- Tan M, Wilson I, Braganza V, Ignatiadis S, Boston R, Sundararajan V, Cook MJ, D'Souza WJ, 2015 Development and validation of an epidemiologic case definition of epilepsy for use with routinely collected Australian health data. *Epilepsy Behav* 51, 65–72. [PubMed: 26262935]
- Thurman DJ, Beghi E, Begley CE, Berg AT, Buchhalter JR, Ding D, Hesdorffer DC, Hauser WA, Kazis L, Kobau R, Kroner B, Labiner D, Liow K, Logroscino G, Medina MT, Newton CR, Parko

K, Paschal A, Preux PM, Sander JW, Selassie A, Theodore W, Tomson T, Wiebe S, Epidemiology I.C.o., 2011 Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia* 52 Suppl 7, 2–26. [PubMed: 21899536]

Tu K, Wang M, Jaakkimainen RL, Butt D, Ivers NM, Young J, Green D, Jette N, 2014 Assessing the validity of using administrative data to identify patients with epilepsy. *Epilepsia* 55, 335–343. [PubMed: 24417710]

Zack MM, Kobau R, 2017 National and State Estimates of the Numbers of Adults and Children with Active Epilepsy - United States, 2015. *MMWR. Morbidity and mortality weekly report* 66, 821–825. [PubMed: 28796763]

Highlights

- Claims-based definitions using groups of ICD-10-CM codes performed well in identifying epilepsy in an outpatient setting.
- Groups of ICD-10-CM codes accurately distinguished between focal and generalized seizure types.
- Further research is needed to measure aspects related to epilepsy severity (e.g., uncontrolled seizures) at the population level.

Figure 1. Identification of the analytic sample at a level 4 epilepsy center

Abbreviations: EHR, electronic health record.

Identification of analytic sample with reference-standard documentation to validate claims-based definitions. Demographics are presented for unique patients in the analytic sample (Table 1) and clinical characteristics are presented for each encounter (Table 2). Demographics of the excluded, unique patients are presented in Supplemental Table 2.

Table 1.

Demographics of the analytic sample

Demographics ^a	Analytic sample (n=402) ^b
Age, median (IQR)	43.8 (29.8–57.5)
Sex, n (%)	
Male	183 (45.5)
Female	219 (54.5)
Primary language, n (%)	
English	375 (93.3)
Other than English	27 (6.7)

Abbreviations: IQR, interquartile range.

^aData obtained from the research patient data registry;

^bunique patients.

Table 2.

Clinical characteristics of the analytic sample with a diagnosis of epilepsy

Clinical characteristics ^a	Diagnosis of epilepsy (n=486) ^b
Seizure type, n (%)	
At least one focal	321/486 (66.0)
Definite ^c	196/321 (61.0)
Probable ^d	106/321 (33.0)
Possible ^e	16/321 (5.0)
Unknown ^f	3/321 (1.0)
At least one generalized	202/486 (41.6)
Definite ^c	151/202 (74.7)
Probable ^d	45/202 (22.3)
Possible ^e	5/202 (2.5)
Unknown ^f	1/202 (0.5)
At least one unknown	34/486 (7.0)
Definite ^c	15/34 (44.1)
Probable ^d	13/34 (37.2)
Possible ^e	5/34 (14.7)
Unknown ^f	1/34 (2.9)
Number of discrete seizures in past year, n (%)	
One	301/486 (61.9)
Two or more	185/486 (38.1)
Both focal and generalized seizures	50/486 (10.3)
1 seizure in past year, n (%)	
Yes	264/486 (54.3)
No	222/486 (45.7)

^aDerived from reference-standard documentation in the EHR;^boutpatient encounters with patients with a reference-standard diagnosis of epilepsy;^cdefinite diagnostic certainty = the summary of evidence suggests 100% confidence level;^dprobable diagnostic certainty = the summary of evidence suggests greater than 50% confidence level;^epossible diagnostic certainty = the summary of evidence suggests less than 50% confidence level;^funknown diagnostic certainty = the summary of evidence is not sufficient to support a finding.

Clinical characteristics of outpatient encounters with reference-standard epilepsy at first captured encounter with reference-standard documentation. 6.5% (n=34/520) of the analytic sample had no or unsure reference-standard epilepsy. Definitions for diagnostic certainty were derived from the National Institute for Neurological Disorders and Stroke Epilepsy Common Data Elements (Loring et al., 2011).

Table 3.

Accuracy of the ICD-10-CM claims-based definitions

Claims-based definition	Prevalence % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
Epilepsy (G40 only) ^a	93.0 (91.0–95.4)	84.4 (80.8–87.5)	79.4 (62.1–91.3)	98.3 (96.6–99.3)	26.2 (18.0–35.8)
Epilepsy (G40 and/or R56) ^b	93.0 (91.0–95.4)	97.7 (91.0–95.4)	44.1 (27.2–62.1)	96.2 (94.2–97.7)	57.7 (36.9–76.6)
Focal seizures ^{c,d}	66.0 (62.0–70.3)	69.8 (64.4–74.8)	79.4 (72.4–85.3)	86.8 (82.1–90.7)	57.5 (50.8–64.0)
Generalized seizures ^{e,f}	36.1 (29.5–43.2)	36.1 (29.5–43.2)	95.4 (92.3–97.5)	84.9 (75.5–91.7)	67.8 (62.9–72.3)

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; Sn, sensitivity; Sp, specificity.

^a Sn=410/486, Sp=27/34;

^b Sn=475/486, Sp=15/34;

^c encounters with ICD-10-CM codes suggestive of focal versus generalized or unknown seizures;

^d Sn=224/321, Sp=131/165

^e encounters with ICD-10-CM codes suggestive of generalized versus focal or unknown seizures;

^f Sn=73/202, Sp=271/284.

Performance metrics of ICD-10-CM claims-based definitions in classifying diagnosis of epilepsy and seizure types against reference-standard documentation in patient EHRs.

Table 4.

Clinical characteristics of the analytic sample between encounters with or without a provider-assigned “convulsions, not elsewhere classified” ICD-10-CM code

Clinical characteristics ^a	Convulsions, not elsewhere classified (R56.X)	
	Yes (n=76)	No (n=444)
Epilepsy, n (%)		
Yes	65 (85.5)	421 (94.8)
No/unsure	11 (14.5)	23 (5.2)
Seizure type, n (%); n= 486		
Focal	33 (50.8)	288 (68.4)
Generalized	29 (44.6)	123 (29.2)
1 seizure in past year, n (%); n= 486		
Yes	29 (38.2)	235 (52.9)
No	47 (61.8)	209 (47.1)

Abbreviations: EHR, electronic health record.

^aDerived from reference-standard documentation in the EHR.

Clinical characteristics of patients with or without a provider-assigned ICD-10-CM code for “convulsions, not elsewhere classified” (R56.X).