Estimating the Prevalence of Treated Epilepsy Using Administrative Health Data and Its Validity: ESSENCE Study

Seo-Young Lee\textsuperscript{a}, Soo-Eun Chung\textsuperscript{b}, Dong Wook Kim\textsuperscript{c}, So-Hee Eun\textsuperscript{d}, Hoon Chul Kang\textsuperscript{a}, Yong Won Cho\textsuperscript{e}, Sang Do Yi\textsuperscript{f}, Heung Dong Kim\textsuperscript{g}, Ki-Young Jung\textsuperscript{h}, Hae-Kwan Cheong\textsuperscript{b} on behalf of the Committee on Epidemiology of Korean Epilepsy Society

\textsuperscript{a}Department of Neurology, Kangwon National University College of Medicine, Chuncheon, Korea
\textsuperscript{b}Department of Social and Preventive Medicine, Sungkyunkwan University School of Medicine, Suwon, Korea
\textsuperscript{c}Department of Neurology, Konkuk University School of Medicine, Seoul, Korea
\textsuperscript{d}Department of Pediatrics, Kangwon National University College of Medicine, Ansan, Korea
\textsuperscript{e}Department of Pediatrics, Yonsei University College of Medicine, Seoul, Korea
\textsuperscript{f}Department of Neurology, Keimyung University School of Medicine, Daeegu, Korea
\textsuperscript{g}YiSangDo Neurology Clinic, Daeegu, Korea
\textsuperscript{h}Department of Neurology, Seoul National University College of Medicine, Seoul, Korea

\textbf{Background and Purpose} Few of the epidemiologic studies of epilepsy have utilized well-validated nationwide databases. We estimated the nationwide prevalence of treated epilepsy based on a comprehensive medical payment database along with diagnostic validation.

\textbf{Methods} We collected data on patients prescribed of antiepileptic drugs (AEDs) from the Health Insurance Review and Assessment service, which covers the entire population of Korea. To assess the diagnostic validity, a medical records survey was conducted involving 6,774 patients prescribed AEDs from 43 institutions based on regional clusters and referral levels across the country. The prevalence of treated epilepsy was estimated by projecting the diagnostic validity on the number of patients prescribed AEDs.

\textbf{Results} The mean positive predictive value (PPV) for epilepsy was 0.810 for those prescribed AEDs with diagnostic codes that indicate epilepsy or seizure (Diagnosis-E), while it was 0.066 for those without Diagnosis-E. The PPV tended to decrease with age in both groups, with lower values seen in females. The prevalence was 3.84 per 1,000, and it was higher among males, children, and the elderly.

\textbf{Conclusions} The prevalence of epilepsy in Korea was comparable to that in other East Asian countries. The diagnostic validity of administrative health data varies depending on the method of case ascertainment, age, and sex. The prescriptions of AEDs even without relevant diagnostic codes should be considered as a tracer for epilepsy.

\textbf{Key Words} epilepsy, seizure, prevalence, health data, epidemiology, validation.

\section*{INTRODUCTION}

Epilepsy is a common neurologic disorder, with a reported prevalence ranging from 2.2 to 17.6 per 1,000.\textsuperscript{1} Until recently, epidemiologic reports on the prevalence of epilepsy in Korea were lacking. Previously the authors reported the prevalence of treated epilepsy based on the diagnostic codes for claims and the antiepileptic drugs (AEDs) prescribed, using the database from the National Health Insurance (NHI), public third party payment system.\textsuperscript{2} That study found that the overall prevalence was 2.41 per 1,000, which is very near to the bottom of the range of the rates found in other countries. However, because the diagnostic codes of the NHI database were generated for claims, it is crucial to validate the accuracy of these codes. Although the use of administrative health data for epidemiologic studies has been increasing, few studies have estimated the prevalence adjusted by diagnostic validity. The diagnostic validity may depend on demographic, institutional, or clinical variables, and applying validity that is specific for subgroups will allow more accurate estimation of prevalence.

\textsuperscript{c}This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
In this study we aimed to estimate the age- and sex-specific prevalence of epilepsy with less bias, by using a nationwide database supplemented by extensive validation performed using a medical records survey of the representative sample.

**METHODS**

**Study design**
We conducted two surveys in parallel: 1) a Health Insurance Review and Assessment service (HIRA) database survey and 2) a medical records survey. Data on potential epilepsy patients were extracted from the main HIRA database based on our working criteria described below. We also performed a nationwide survey of medical records of patients sampled from representative health institutes. Using the results of the medical records survey, we assessed positive predictive value (PPV), the probability of the HIRA database patients having epilepsy, according to diagnostic codes, age group, and sex. Finally, we estimated the prevalence of treated epilepsy by projecting the PPV derived from the medical records survey involving the number of initially presumed epilepsy patients from the HIRA database after stratifying for covariates (Fig. 1).

This study was approved by the Institutional Review Board (IRB) of the lead institutes (approval number: AN10221-001) and all participating institutes with an IRB.

**Data sources**
The national health system of Republic of Korea is based on the NHI system, registration in which is mandatory for the entire population and for all medical facilities, and Medicaid (MA), which is provided through a social welfare fund for registrants who are unable to pay the NHI premiums. The population coverage has been greater than 98% since 2005. This is basically a fee-for-service system, and all medical expenditures including for medications, medical services, and revenue are submitted to HIRA as separate claims. HIRA is a governmental third-party agency that has assessed all claims based on diagnostic codes and medical records from NHI.

---

**Fig. 1.** Study design. AEDs: antiepileptic drugs, HIRA: Health Insurance Review and Assessment service.
Prevalence of Treated Epilepsy

and MA since 2000, and from the Veterans Administration since 2008. Costs for virtually all diagnostic and therapeutic practices associated with epilepsy are covered by the system, and all related records are stored in the HIRA database. We excluded patients under 1 year of age because we found that it is often difficult to determine from reviews of medical records whether they have epilepsy or acute symptomatic seizures at this age.

Potential patient dataset using the HIRA database

Our working criterion for extracting potential patients with epilepsy from the HIRA database was a prescription for at least one AED according to the list of claims in 2009. The list of AEDs, based on those available in 2009 in Korea, included carbamazepine, clobazam, ethosuximide, gabapentin, lamotrigine, levetiracetam, oxcarbazepine, phenobarbital, phenytoin, pregabalin, primidone, topiramate, vigabatrin, valproate, and zonisamide. Clonazepam was excluded because it is rarely used as a monotherapy for epilepsy and is used more frequently in other conditions. Other AEDs, including primidone, felbamate, eslicarbazepine, lacosamide, and tiagabine, were not available in 2009.

Potential epilepsy patients were classified into two categories: those prescribed AEDs with and without diagnostic codes indicating epilepsy or seizure (Diagnosis-E), based on either the principal or additional diagnostic code for the claim. Diagnosis-E included G40* (epilepsy), G41* (status epilepticus), F803 (Landau-Kleffner syndrome), and R56 (convulsion) according to the 10th version of the International Classification of Diseases and Related Health Problems (ICD-10).

In total, 7,282,236 claims from patients prescribed AEDs during 2009 were retrieved. The claims were sorted by individual based on anonymous personal identifiers and hospital identifiers, in order to determine the actual number of patients for 2009 (Fig. 1).

Medical records survey and validation of diagnostic codes for epilepsy

We collected clinical information on epilepsy patients by surveying the medical records for the validation of diagnostic codes. The survey consisted of two steps: 1) data abstraction by certified health record administrators and 2) subsequent verification by epileptologists. The health record administrators were trained to review medical records and extract data related to epilepsy. All cases were scrutinized and assessed by a board of epileptologists based on detailed clinical information abstracted by the health record administrators. Diagnoses were categorized as follows based on the 1993 International League Against Epilepsy guidelines: epilepsy, single seizure, uncertain whether epilepsy or seizure, nonepileptic, and undetermined. This survey method demonstrated a high validity in a preliminary study, with a sensitivity of 90% and a specificity of 97%, and kappa statistics for interrater and test-retest reliabilities of 0.907 and 0.975, respectively.

Multistage cluster sampling was performed in three geographic regions (the capital, midland, and southeast areas) categorized based on approximately even distributions of institutes and population. Based on a pilot survey of three hospitals, the sample size was estimated at 2,000 for patients prescribed AEDs with Diagnosis-E, and 4,000 for those prescribed without Diagnosis-E. Finally, we surveyed 6,774 patients from 43 institutes distributed throughout the country, comprising 4 tertiary referral hospitals, 29 general hospitals, 8 hospitals, and 2 private clinics.

Estimation of prevalence

The PPV for epilepsy was calculated as the number of patients with true epilepsy divided by the total number of patients prescribed AEDs, for age, sex, and diagnostic categories.

The total number of the cases was estimated as

\[ N = N_{Di} \times P_{Di} + N_{Mi} \times P_{Mi}, \]

where

\[ P_{Di} = \frac{E_{Di}}{D_i}, \]

\[ P_{Mi} = \frac{E_{Mi}}{M_i}, \]

\( N_{Di} \) is the number of patients prescribed AEDs with Diagnosis-E according to the HIRA claims data, \( N_{Mi} \) is the number of patients prescribed AEDs without Diagnosis-E according to the HIRA claims data, \( P_{Di} \) is the PPV for patients prescribed AEDs with Diagnosis-E, \( P_{Mi} \) is the PPV for patients prescribed AEDs without Diagnosis-E, \( D_i \) is the number of patients prescribed AEDs with Diagnosis-E, \( M_i \) is the number of patients prescribed AEDs without Diagnosis-E, \( E_{Di} \) is the number of cases deemed to be true epilepsy among \( D_i \), and \( E_{Mi} \) is the number of cases deemed to be true epilepsy among \( M_i \), based on the medical records survey.

We applied the PPV derived from the medical records survey to groups of subjects extracted from the HIRA database according to specific strata. We used the age- and sex-specific national population estimated for 2009 from the Korean national statistical office as the denominator when calculating the prevalence.

RESULTS

Validity of diagnostic codes

Among the total of 50,290,771 beneficiaries, 1,119,360 patients (2.2%) were prescribed AEDs during 2009, of which 155,307 (13.9%) had Diagnosis-E and 964,053 (86.1%) did not have Diagnosis-E. The PPV for epilepsy was 0.460–0.978 (overall 0.810) for the group with Diagnosis-E and 0.021–0.324 (overall 0.066) for the group without Diagnosis-E. The
ranges indicate the variations with age and sex: the PPV tended to decrease with age in both groups, with consistently lower values in females over 30 years of age based on the medical records survey (Table 1).

The number of patients prescribed AEDs with Diagnosis-E but deemed not to have epilepsy was estimated at 28,712 nationwide. The false-positive cases were determined to be acute symptomatic seizures, single seizure, or uncertain whether epilepsy or seizure (41.9%); therapeutic trial for possible seizures (7.07%); prophylactic use (18.2%); misassignment of the diagnostic codes for pain (23.7%); other purpose (4.29%); or due to an unknown cause (4.80%).

The number of patients prescribed AEDs without Diagnosis-E but determined as having epilepsy was estimated at 60,902 nationwide. The false-negative cases had diagnostic codes suggesting central nervous system pathology potentially related to epilepsy (53.6%); psychiatric illness (5%); paroxysmal disorders requiring differentiation from epilepsy, such as syncope, dystonia, cramps, and headache (5%); and those unrelated to epilepsy (36.4%).

Prevalence
The estimated number of epilepsy patients treated with AEDs was 187,497 [95% confidence interval (CI)=185,512–188,697]. The overall prevalence was 3.84 per 1,000 (95% CI=3.81–3.87). The prevalence was higher in males (4.20 per 1,000; 95% CI=4.17–4.23) than in females (3.47 per 1,000; 95% CI=3.44–3.51) for all age groups except patients aged 1–9 years (males: 4.57 per 1,000; females: 5.78 per 1,000). The prevalence was relatively constant across age groups, with a tendency to be higher in patients younger than 15 years and in the elderly (Fig. 2) (Supplementary Table 1 in the online-only Data Supplement).

The prevalence of treated epilepsy with Diagnosis-E was 2.59 per 1,000 (males: 2.94 per 1,000; females: 2.32 per 1,000), and the prevalence of those without Diagnosis-E was 1.25 per 1,000 (males: 1.26 per 1,000; females: 1.23 per 1,000). Age- and sex-specific prevalence patterns differed between groups with and without Diagnosis-E. A plot of the prevalence against age had an inverted U-shape peaking at adolescence for patients with Diagnosis-E, and an upright U-shape for the group without Diagnosis-E (Fig. 3).

Table 1. Positive predictive values for epilepsy in patients prescribed antiepileptic drugs according to diagnostic code category, age, and sex, based on a medical records survey (n=6,774)

<table>
<thead>
<tr>
<th>Age, years</th>
<th>With diagnostic code indicating epilepsy or seizure (n=2,374)</th>
<th>Without diagnostic code indicating epilepsy or seizure (n=4,400)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
</tr>
<tr>
<td>1–4</td>
<td>0.857</td>
<td>0.833</td>
</tr>
<tr>
<td>5–9</td>
<td>0.936</td>
<td>0.923</td>
</tr>
<tr>
<td>10–14</td>
<td>0.940</td>
<td>0.925</td>
</tr>
<tr>
<td>15–19</td>
<td>0.978</td>
<td>0.897</td>
</tr>
<tr>
<td>20–29</td>
<td>0.897</td>
<td>0.943</td>
</tr>
<tr>
<td>30–39</td>
<td>0.897</td>
<td>0.852</td>
</tr>
<tr>
<td>40–49</td>
<td>0.877</td>
<td>0.840</td>
</tr>
<tr>
<td>50–59</td>
<td>0.784</td>
<td>0.692</td>
</tr>
<tr>
<td>60–69</td>
<td>0.633</td>
<td>0.556</td>
</tr>
<tr>
<td>≥70</td>
<td>0.536</td>
<td>0.460</td>
</tr>
<tr>
<td>≥65</td>
<td>0.562</td>
<td>0.505</td>
</tr>
<tr>
<td>Total</td>
<td>0.825</td>
<td>0.790</td>
</tr>
</tbody>
</table>

*p<0.001 for analysis of variance between males and females without diagnostic codes.
Prevalence of Treated Epilepsy

**DISCUSSION**

This study determined the nationwide prevalence of treated epilepsy across all ages except infants utilizing a nationwide database and validated by a medical records survey. Only a few studies have estimated the prevalence of epilepsy for the entire population of a country, for which various data sources can be used. A door-to-door survey is a traditional gold standard, but it is not practical in a large population. Although a population-level self-report survey has less selection bias and is less demanding than a door-to-door survey, its diagnostic accuracy is uncertain. A records-based survey is advantageous over a direct population survey in that it avoids recall bias, although selection bias may be an issue. In countries with universal health care systems, the utilization of administrative health databases is useful for nationwide epidemiologic studies with a relatively low selection bias.

The main concerns when using administrative health data for epilepsy epidemiology studies are the correct identification of epilepsy cases and the determination of their validity. Diagnostic codes are commonly used as tracers for epilepsy, and only diagnostic codes for the primary underlying condition are entered into the system. Physicians may defer coding epilepsy or seizure in children due to the social stigma, whereas the prevalence was U-shaped in patients without Diagnosis-E, which suggests that Diagnosis-E was assigned more frequently to elderly and pediatric patients with epilepsy. This is perhaps because symptomatic epilepsy is more common in these populations, particularly in elderly individuals, and only diagnostic codes for the primary underlying condition, such as stroke, head trauma, or neurodegenerative disease, were entered into the system. Physicians may defer coding epilepsy or seizure in children due to the social stigmatization and negative attitudes toward epilepsy.

Our study was subject to several limitations. Because we used the HIRA database, and thus included only patients actively seeking medical attention, we were able to estimate
the prevalence of treated epilepsy but not that of all epilepsy cases. This may have resulted in underestimation of the actual prevalence of epilepsy in Korea, although the treatment gap is anticipated to be minimal because of the high accessibility to health institutions and the 70–100% coverage of costs for AEDs in Korea. Another overlooked group was patients who used AEDs that were not recorded in the HIRA database; for example, patients who covered the costs themselves to avoid stigma and those who were reimbursed by means other than NHI, such as automobile insurance or industrial accident compensation. Our medical records survey was performed only at institutes that agreed to join the study, which could result in a selection bias in the estimation of diagnostic validity. In spite of these limitations, our study had strengths in that it was performed nationwide and covered the entire population, and the prevalence was estimated by diagnostic validation according to strata based on an extensive survey of medical records.

Conclusions
We have estimated the nationwide prevalence of epilepsy using claims data from administrative health database complemented by a medical records survey. We were able to calculate reliable age- and sex-specific prevalence, by applying stratified diagnostic validity from a medical records survey of a large number of samples. The overall prevalence was comparable to those found in other East Asian countries. The age-specific prevalence appeared as a U-shaped graph, being higher among the young and the elderly. Administrative health data may reveal variability in diagnostic validity according to the method of case ascertainment, age, and sex, and a significant number of treated patients with epilepsy may not show diagnostic codes for epilepsy or seizure. Our study may provide a basis for further epidemiologic studies using administrative health databases.

Collaborators
Medical records review board: Hoon-Chul Kang, Department of Pediatrics, Yonsei University School of Medicine; So-Hee Eun, Department of Pediatrics, Korea University College of Medicine; Byung Chan Lim, Department of Pediatrics, Seoul National University College of Medicine; Joong-Hyun Bin, Department of Pediatrics, The Catholic University of Korea College of Medicine; Sang Hyeon Jang, Department of Neurology, Eulji University College of Medicine; Yunsook Jhang, Myongji Hospital; Seo-Young Lee, Department of Neurology, Kangwon National University College of Medicine; Bong Su Kang, Department of Neurology, Korea University College of Medicine; Dongwook Kim, Department of Neurology, Konkuk University School of Medicine; Ki-Young Jung, Department of Neurology, Seoul National University College of Medicine.

Participant investigators and institutes involved in the medical records review (in alphabetical order):
Bong Su Kang, Seogwipo Medical Center;
Boo Jung, Jeju Hanmaeum Hospital;
Chung Eun Ji, Daegu Medical Center;
Do Hyoun Kwon, Han-II Hospital;
Dong Wook Kim, Konkuk University Hospital;
Donghoon Oh, Incheon Medical Center;
Gyeong Sik Shin, Jeil Neurology Clinic;
Heon Han, Sokcho Medical Center;
Hong Seok Shon, Yeong Gwang Sin-Ha Christian Hospital;
Hoon-Chul Kang, Yonsei University Severance Hospital;
Hyang Woon Lee, Ewha Womans University Mokdong Hospital;
Hyun Kyung Kim, Eunpyeong Hospital;
Hyunwoo Nam, Boramae Medical Center;
Il Keun Lee, Seoul Brain Neurology Clinic;
Jae Kook Yoo, Korea University Anam Hospital;
Jang Joon Lee, Daegu Fatima Hospital;
Jee Hyun Kim, Dankook University Hospital;
Jeong Seok Lee, Jeju National University Hospital;
Jeongyeon Kim, Inje University Sanggye Paik Hospital;
Ji Eun Kim, Daegu Catholic University Hospital;
Ji Hyun Kim, Korea University Guro Hospital;
Jung-Ju Lee, Seoul Eulji General Hospital;
Jun-Young Lee, Chonbuk National University Hospital;
Ki-Hwan Ji, Inje University Busan Paik Hospital;
Kwang Ki Kim, Dongguk University Ilsan Hospital;
Kyung-Il Park, Inje University Seoul Paik Hospital;
Myeong-Kyu Kim, Chonnam National University Hospital;
Oh-Young Kwon, Gyeongsang National University Hospital;
Sang Ho Kim, Dong-A University Hospital;
Sang Hyun Jang, Eulji University Hospital;
Sang Kun Lee, Seoul National University Hospital;
Sang-Ahm Lee & Seok Ho Hong, Asan Medical Center;
Seon Woo Nam, Sun General Medical Center;
Seo-Young Lee, Kangwon National University Hospital;
Seung Bong Hong, Sungkyunkwan University School of Medicine;
So-Hee Eun, Korea University Ansan Hospital;
Sung-Hyun Lee & Shin-Hye Baek, Chungbuk National University Hospital;
Wonchul Shin, Kyung Hee University Hospital at Gangdong;
Yong Won Cho, Keimyung University Dongsan Medical Center;
Young Joo No, Gangneung Asan Hospital;
Young Mi Oh, Hankook General Hospital;
Young-Min Shon, The Catholic University of Korea, Yeou-
do St. Mary’s Hospital; Yunsook Jhang, SahmYook Hospital

Contributors
Reviewing and collecting the information from medical records (in alphabetical order):
All of these people are certified health record administrators who are experienced in reviewing medical records for epidemiologic studies and health registries.
Retrieval of claims from the HIRA database:
Sang Min Jeong, Department of Research and Statistics, HIRA (Health Insurance Review and Assessment service).

Supplementary Materials
The online-only Data Supplement is available with this article at http://dx.doi.org/10.3988/jcn.2016.12.4.434.

Conflicts of Interest
The authors have no financial conflicts of interest.

Acknowledgements
This work was partially supported by Korean Epilepsy Society and Kang-won National University.

REFERENCES