Title page:

Validity of health insurance data to identify people with epilepsy

Merel Wassenaar, PhD[1] , Johannes A. Carpay, MD,PhD[2,3], Josemir W. Sander, MD, PhD, FRCP [1,4] , Roland D. Thijs, MD, PhD [1,3,4]

[1] Stichting Epilepsie Instellingen Nederland (SEIN), Heemstede, The Netherlands

[2] Department of Neurology, Tergooi Hospitals, Hilversum, The Netherlands

[3] Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands

[4] NIHR University College London Hospitals, Biomedical Research Centre, UCL Institute of Neurology, London WC1N 3BG and Chalfont Centre for Epilepsy, Chalfont St Peter SL9 0RJ, UK

Correspondence to:

Roland D. Thijs, Stichting Epilepsie Instellingen Nederland (SEIN)

P.O. Box 540, 2130 AM Hoofddorp, The Netherlands

Tel: + 31-235588948

Fax: + 31-235588159

rthijs@sein.nl

Number of figures: 1; Number of tables: 3;

Word count text: 2390 (excluding, title page, abstract, references, figures and tables)

Word count abstract: 258. Number of references: 34

Abstract

Objective: Large administrative databases may prove useful to assess epilepsy-related comorbidity and mortality. Despite their increased use, their validity as data source in epilepsy is yet under-ascertained.

Methods: Achmea is a large Dutch health insurance company covering about 25% of the population. We performed a retrospective cohort study using data from the Achmea Health Insurance Database (AHID) over the period 2006-2009. To assess the validity of epilepsy codes in the AHID, we randomly invited 1000 individuals (age 18-75 years insured by Achmea), attending an epilepsy centre or a district hospital during 2006-2009, to participate. Informed consent was provided by 293 eligible for inclusion. We compared the diagnostic codes for epilepsy in AHID with the diagnosis in their case-notes (reference standard). As additional measure of validity, we compared prevalence of epilepsy codes in AHID (based on anonymized data of all 26.297 subjects with this code in AHID) with epilepsy prevalence rates in the general Dutch population to estimate an age-specific standardized prevalence ratio.

Results: We identified 293 participants with an epilepsy code in AHID. The majority (278) of them had a definite or possible diagnosis of epilepsy in the case-notes; i.e. a positive predictive value of 0.95 (95% CI 0.92-0.97). The overall prevalence of epilepsy codes in the AHID was slightly higher than the putative prevalence in the general Dutch population (7.4/1.000 vs. 6.8/1.000) with a Standardized Prevalence Ratio of 1.08 (95% CI: 1.08-1.09).

Conclusions: Our findings establish the validity of AHID data for a diagnosis of epilepsy and confirm previous work on using administrative data for epilepsy research.

Keywords: epilepsy, administrative health data, accuracy, prevalence

Introduction

Epilepsy is a neurological condition characterized by recurring seizures usually requiring antiepileptic drugs (AED). (Loscher and Schmidt, 2011; Sander, 2003) People with epilepsy often have comorbid conditions. (Gaitatzis et al., 2012; Hermann et al., 2008; Keezer et al., 2016) The risk of premature mortality is 2-3 times higher than in the general population. (Neligan et al., 2011) Premature mortality may be explained by several factors such as epilepsy-related causes including accidents, Sudden Unexpected Death in Epilepsy (SUDEP), or may be related to comorbid conditions such as a cerebrovascular disease. (Novy et al., 2013; Surges et al., 2009) Precise figures on epilepsy mortality and comorbidity are, however, still needed. Case-control studies in high risk populations (from specialized epilepsy centres) may suffer from selection bias, small sample sizes and a disregard of comorbid conditions. Population-based cohorts are also often limited by relative small samples of the epilepsy population and have a too long study run time for surveillance of the actual burden and risk profiles of epilepsy-related comorbidities and mortality.

Large population based studies using administrative health insurance data provide opportunities to increase power of data and obtain comprehensive surveillance data. (England et al., 2012; Hesdorffer et al., 2013; Smeets et al., 2011) This "big data" approach has been successfully used to address epilepsy-related questions, e.g. concerning psychiatric comorbidity in premature mortality (Fazel et al., 2013) or epilepsy-related psychiatric and somatic comorbidities e.g. migraine, autism, stroke, diabetes. (Chen et al., 2011; Chou et al., 2016; Oh et al., 2017; Selassie et al., 2014; Sundelin et al., 2016; Wannamaker et al., 2015) Previous work validated the diagnostic codes for epilepsy from administrative databases. (Christensen et al., 2007; Ertl et al., 2016; Jette et al., 2010; Parko and Thurman, 2009; Reid et al., 2012) Yet the validity of epilepsy diagnostic codes using health insurance databases is still under-ascertained and needs confirmation in every database given the variety in health care and coding systems. (Thurman et al., 2011)

Dutch health insurance databases are a potentially useful source for epilepsy research. The Netherlands has a health care system with a mandatory and ubiquitous health insurance coverage. Insurers register all reimbursed health care visits based on diagnostic codes provided by the treating physician. (Braithwaite et al., 2017; Schoen et al., 2007) We assessed the validity of diagnostic codes for epilepsy derived from Achmea, one of the major health insurance companies in the Netherlands. The Achmea Health Insurance Database (AHID) was previously found representative of the health care utilization of the total Dutch population with respect to age, gender and socioeconomic status. (Smeets et al., 2011) We assessed accuracy by comparing diagnostic codes for epilepsy in AHID with the diagnosis in case-notes (reference standard). As additional measure of validity, we compared the prevalence of epilepsy codes in AHID with epilepsy prevalence rates in the general Dutch population.

Methods

Population and setting

Retrospective cohort data, with respect to demographics and health care utilization over the period 2006-2009, was retrieved from AHID, which covers around 4.1 million policyholders (25% of the total Dutch population). All health care reimbursement claims are collected and continuously monitored to ensure an accurate and valid database. (Smeets et al., 2011) During the period 2006-2009 the yearly average number of people ≥ 18 years in AHID was 3.247.887.

To assess the validity of the diagnostic epilepsy codes, we set to enrol around 1% of epilepsy cases in AHID based on an estimated prevalence of 0.7% (Gommer AM, 2010). From two hospital registries: 1) a tertiary epilepsy referral centre and 2) a district hospital, we selected all individuals with at least one epilepsy-related visit to a neurologist between 2006 and 2009. This selection was made using the hospital based claims database. All subjects aged 18-75 years, insured by Achmea for at least one year during the study period, were eligible. Those with incomplete contact details and those in residential care were excluded. Of 2916 potentially eligible subjects, we randomly invited by post, 1000 (70% from tertiary care and remaining from the district hospital) to participate. This included an information letter and a request to provide informed consent to access their case-notes and AHID data for comparison. A total of 293 respondents were eligible for inclusion.

We also had access to anonymized AHID data of all 26.297 adults with an epilepsy diagnostic code in the same period. This allowed the delineation of the AHID epilepsy-cohort and to estimate epilepsy prevalence amongst them. The study was fully compliant with Dutch regulations and approved by the Medical Ethics Committee of Leiden University Medical Center and by Achmea's Scientific and Privacy Committee.

Data collection

AHID data included demographics (sex, year of birth and death (if applicable) and duration of insurance by Achmea) and health care utilization over the period of interest. This included information on all primary, secondary and tertiary care visits and drug prescriptions (ATC codes and number of daily defined dose (DDD)), with corresponding dates. Hospital visits are coded by a Diagnostic Treatment Protocol (DTP, in Dutch: 'Diagnose Behandel Combinatie'): an administrative code combining hospital registration of diagnoses (International Classification of Disease 9th revision (ICD-9) with therapeutic interventions i.e. classification in generalized and focal epilepsy based on the

ICD 345.xx coding for epilepsy. For all participants in the validation subset, we requested a personal identifiable code to link the AHID data to information retrieved from the case-notes.

Clinical data for the period participants were insured by Achmea were extracted from the hospital case-notes. This included information on epilepsy diagnosis (history, seizure description, MRI, EEG etc.), antiepileptic drug (AED) use, comorbid conditions and co-medication. Neurological diagnosis were classified into three categories: (1) *definite or probable epilepsy* in case diagnostic work-up of the treating neurologist supported a diagnosis and no alternative diagnosis suggested (Thurman et al., 2011); (2) *suspect epilepsy* in case epilepsy was thought most likely but other alternatives were considered (Thurman et al., 2011) (3) *no epilepsy* in case other conditions like syncope or psychogenic non-epileptic seizures were considered to most likely. For each category, we recorded whether AED were prescribed (i.e. all drugs within the AED subgroup of the ATC classification system except for benzodiazepines), thus resulting in a total of six categories expressing the likelihood of the epilepsy diagnosis including its treatment status. From the AHID we included all those with a DTP for (generalized and focal) epilepsy, corresponding to the ICD 345.xx codes for epilepsy and treatment status. Data extraction and classification was performed by one author (MW). An audit was performed on a subset by RT and JC to minimize potential misclassification. In case of uncertainty RT & JC reviewed and discussed to reach consensus.

Statistical analysis

We assessed baseline demographics, epilepsy and AED prescription data. The validation subset was compared to the total number of epilepsy cases, aged 18-75 years, in AHID to assess the representativeness of our sample. The prevalence of epilepsy in AHID was compared to epilepsy prevalence data of Statistics Netherlands based on DTP codes for epilepsy in the general population. (CBSStatline) An age-specific standardized prevalence ratio and 95% CI was calculated.

We assessed the percentage of correct DTPs for epilepsy in AHID validation subset by comparing them to the reference standard (case-notes). Its positive predictive value (PPV and 95% CI) is presented as measure of diagnostic accuracy (number of correctly classified cases). The PPV refers to the proportion of epilepsy cases (category 1-3 from the case-notes) from all those identified with a DTP for epilepsy in the AHID. As a sensitivity analysis we assessed the effect of reclassifying suspect cases (category 3 or 4) as epilepsy or as non-epilepsy cases. All statistical analyses were performed with IBM SPSS Statistics for Windows, version 22.0 (Armonk, NY: IBM Corp).

Results

A total of 293 respondents had a diagnostic code for epilepsy in AHID and were eligible for inclusion i.e. for review of their case-notes over 2006-2009 (Figure 1). Baseline characteristics of 293 participants are listed in Table 1. These were compared to 23.493 individuals, aged 18-75 years, with an epilepsy code in AHID. Participants in the validation study were slightly younger than the total cohort of people with an epilepsy code (median age 46 vs. 47 years, p-value 0.003 and more likely to have an AED prescription (92% vs. 76%, p-value <0.001).

Of the average 3.247.887 people of 18 years and older in AHID during 2006-2009, 24.188 persons per year had at least one DTP for epilepsy. These numbers were used to estimate prevalence. This does not equal the total number of people (26.297, of whom 23.493 between 18 and 75 years) found with an epilepsy code due to variation in the duration of the insurance policy. During the four year period 19% of all insured switched to an alternative insurance provider, emigrated or died. The overall epilepsy prevalence using DTP codes in AHID was slightly higher than the prevalence rate for the same codes in the general population (Statistics Netherlands): 7.4/ 1.000 vs. 6.8/1.00 with a Standardized Prevalence Ratio of 1.08 (95% CI: 1.08-1.09). The prevalence increased with advancing age in both groups (Table 2).

Of the total 293 cases with an epilepsy code in AHID, 278 were considered to have epilepsy according to the case-notes (category 1-3). An epilepsy code in AHID accurately predicted a diagnosis of epilepsy in 95% (PPV) (95% CI: 0.92-0.97), (Table 3). The PPV was 1.0 and 0.85 respectively for a tertiary epilepsy centre and district hospital). As a sensitivity analysis, we reclassified suspect cases i.e. we analysed only all definite or probable epilepsy cases (i.e. excluding category 3) and also including all suspect cases as epilepsy (i.e. including category 4). This resulted in a PPV of 0.93 (95% CI: 0.89-0.95) and 0.99 (95% CI:0.97-1.00) respectively i.e. only a minor difference when compared to the originally calculated PPV (0.95).

Discussion

Our findings suggest good validity of an AHID epilepsy code. Previous work also indicated good validity of administrative diagnostic codes for epilepsy but applied different criteria for case ascertainment, and used different coding methods and different data sources. (Christensen et al., 2007; Jette et al., 2010; Lee et al., 2016; Parko and Thurman, 2009; Reid et al., 2012; Tu et al., 2014) Although mainly all studies used ICD 9 or 10 coding, they applied differences in for example number and type of visits for their case definition, and codes may have been provided by different health providers or administrators, which is frequently not mentioned. Coding practice not only varied per

study, but also per country as a result of differences in billing practices, weakening comparisons. (Thurman et al., 2011) The most accurate algorithm to identify epilepsy cases was found to be 2 physician claims or 1 hospitalization in 2 years coded. (Reid et al., 2012) In our study, we assessed both in- as well as outpatient diagnostic codes over a period of 4 years, in which >90% had at least 2 diagnostic codes for epilepsy. Ideally, a diagnosis of epilepsy is based on prospective assessment and includes expert interviews and (video-) EEG-recordings of events. We aimed to account for this inevitable limitation in diagnostic uncertainty by comparing data from two different sources: a tertiary epilepsy centre (in which we presumed diagnostic certainty to be above average) and a district hospital. We found only small differences in validity between both centres (a PPV of 1.00 and 0.85, respectively) suggesting limited bias by diagnostic uncertainty.

Studies also varied in terms of individual selection. We selected those with at least one hospital code for an epilepsy-related visit as a proxy for the diagnostic codes in the AHID. We therefore could not determine the negative predictive value of the epilepsy codes in AHID.

In only 5% of our sample the epilepsy codes in AHID did not match the information in case-notes. Mismatches in coding may be due to diagnostic errors but may also be due to failures in adapting an initial code, given after the first clinical visit, when, over time, a more appropriate diagnosis emerges. Assessment of reasons behind the chosen DTP codes was outside the study scope.

The prevalence of (active) epilepsy in The Netherlands as derived from AHID was 7.4/1000 persons. This was slightly higher than the prevalence of the same codes for the Dutch population. (CBSStatline) Health care insurance is mandatory in the Netherlands. All insurance companies offer a core insurance package including epilepsy care at a fixed price, with optional packages for e.g. physiotherapy, alternative medicine or psychological care without indication or referral by a physician. Bias based on health status leading to the choice for a certain health insurance company is thus not likely. The prevalence was also a little higher in comparison to the prevalence of 7.0/ 1000 persons estimated from a sample of GP registries in 2007. (Gommer AM, 2010). Prevalence rates for active as well as lifetime epilepsy are known to vary, even within studies of similar age groups or socio-economic level. (Ngugi et al., 2010) There is, thus, no reason to assume that these small differences in prevalence rates hampered our results with respect to its validity.

Our study may be prone to selection bias due to low participation rate. Yet we consider our sample size (> 1% of the epilepsy cases in the AHID) combined with the quality of the data suited to assess the validity of AHID data. Our study population was slightly younger than the average epilepsy population in the AHID. This might be a proxy for epilepsy severity and willingness to participate in research. Indeed, there was some bias towards more severe epilepsy cases, as the study sample was

more likely to use AEDs compared to the average epilepsy population in AHID. This may have resulted in the categorization of more people in group 1-3, but it is unlikely to have influenced our results with respect to the accuracy of the AHID data, as this is independent of the categorization process itself.

One of the strengths of our study is that it was performed in a country with equal and accessible healthcare system with abundant available data. (Schoen et al., 2007; Schut and van de Ven, 2011) The Netherlands has a unique system in which diagnosis and subsequent treatment are combined in one DTP code, the diagnostic information contained is ICD-based in line with coding applied in most countries. Another strength includes the fact that codes were inserted by the treating neurologist, thus contrasting studies where GPs or administrators classified the diagnosis. The Dutch 'DTP coding' system has two unique features that avoid false epilepsy codes: 1) special codes are available if no definite diagnosis can be made (e.g. code for 'other paroxysmal events'), 2) people can see their own diagnosis when accessing their health insurance reimbursements. In the Netherlands there is also relatively little epilepsy-related stigma (Baker et al., 2000; Brigo et al., 2015), which is likely to result in a more representative sample.

Conclusion

Our findings demonstrates the validity of the Achmea health Insurance database for a diagnosis of epilepsy, based on a comparison with specialists' diagnostic information and confirms previous work on using administrative data for epilepsy research. Dutch health insurance databases, as AHID, may accurately identify people with epilepsy, based on diagnosis and treatment data, for large scale epidemiological studies to deepen our understanding in epilepsy-related comorbidity and mortality.

Highlights:

- Achmea is a large Dutch healt insurance company covering about 25 of the population.
- The Achmea Health Insurance data (AHID) can be reliably used to identify people with epilepsy
- Epilepsy codes in Achmea health data proofed quite accurate (PPV 0.95)
- This confirms previous work on using administrative health data for epilepsy research.

Acknowledgements:

We are grateful to Anne Hollinga and Hugo Smeets for providing the AHID data and for their support in preparation of the analysis files and to Dr. Gail S. Bell and Arnold van der Lee for their critical review of the manuscript. JWS is based at UCLH/UCL Comprehensive Biomedical Research Centre, which receives a proportion of funding from the UK Department of Health's National Institute for Health Research Biomedical Research Centres funding scheme. He receives support from the Dr Marvin Weil Epilepsy Research Fund.

Study funding: No funding was obtained for this study

Conflict of interests:

No author has any conflict of interest in respect to this work. MW and JAC have no disclosures to make. RDT receives research support from the Dutch National Epilepsy Fund, ZonMW, NUTS Ohra Fund, Medtronic and AC Thomson Foundation and has received fees for lectures from Medtronic, UCB and GSK.JWS has received research grants and honoraria from UCB, Eisai, Teva, Lundbeck and GSK which are involved in the manufacturing of AEDs.

Ethical statement:

We confirm that we have read the journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines. The study has been approved by the local ethics committee and was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

Contributorship:

All authors were involved in the conceptualization and design of the study. Data was collected and analyzed by MW. All authors were involved in data interpretation. The manuscript was drafted by MW. The manuscript was critically revised for intellectual content by RDT, JAC and JWS. All authors had full access to the data and take responsibility for the integrity of the data, the accuracy of the data analysis and the conduct of the research and approved the submitted version of the manuscript.

References

Baker, G.A., Brooks, J., Buck, D., Jacoby, A., 2000. The stigma of epilepsy: a European perspective. Epilepsia 41, 98-104.

Braithwaite, J., Hibbert, P., Blakely, B., Plumb, J., Hannaford, N., Long, J.C., Marks, D., 2017. Health system frameworks and performance indicators in eight countries: A comparative international analysis. SAGE Open Med 5, 2050312116686516.

Brigo, F., Igwe, S.C., Ausserer, H., Tezzon, F., Nardone, R., Otte, W.M., 2015. Epilepsy-related stigma in European people with epilepsy: correlations with health system performance and overall quality of life. Epilepsy & behavior : E&B 42, 18-21.

CBSStatline, Medisch Specialistische Zorg; DBC's naar diagnose (detail)

Chen, Y.J., Wu, C.Y., Lin, M.W., Chen, T.J., Liao, K.K., Chen, Y.C., Hwang, C.Y., Chu, S.Y., Chen, C.C., Lee, D.D., Chang, Y.T., Wang, W.J., Liu, H.N., 2011. Comorbidity profiles among patients with bullous pemphigoid: a nationwide population-based study. The British journal of dermatology 165, 593-599.

Chou, I.C., Wang, C.H., Lin, W.D., Tsai, F.J., Lin, C.C., Kao, C.H., 2016. Risk of epilepsy in type 1 diabetes mellitus: a population-based cohort study. Diabetologia 59, 1196-1203.

Christensen, J., Vestergaard, M., Olsen, J., Sidenius, P., 2007. Validation of epilepsy diagnoses in the Danish National Hospital Register. Epilepsy research 75, 162-170.

England, M.J., Liverman, C.T., Schultz, A.M., Strawbridge, L.M., 2012. Epilepsy across the spectrum: promoting health and understanding. A summary of the Institute of Medicine report. Epilepsy & behavior : E&B 25, 266-276.

Ertl, J., Hapfelmeier, J., Peckmann, T., Forth, B., Strzelczyk, A., 2016. Guideline conform initial monotherapy increases in patients with focal epilepsy: A population-based study on German health insurance data. Seizure 41, 9-15.

Fazel, S., Wolf, A., Langstrom, N., Newton, C.R., Lichtenstein, P., 2013. Premature mortality in epilepsy and the role of psychiatric comorbidity: a total population study. Lancet 382, 1646-1654.

Gaitatzis, A., Sisodiya, S.M., Sander, J.W., 2012. The somatic comorbidity of epilepsy: a weighty but often unrecognized burden. Epilepsia 53, 1282-1293.

Gommer AM, P.M., Carpay JA, 2010. Hoe vaak komt epilepsie voor en hoeveel mensen sterven eraan? Volksgezondheid Toekomst Verkenning, Nationoaal Kompas Volksgezondheid. RIVM, Bilthoven dec 2010.

Hermann, B., Seidenberg, M., Jones, J., 2008. The neurobehavioural comorbidities of epilepsy: can a natural history be developed? The Lancet. Neurology 7, 151-160.

Hesdorffer, D.C., Beck, V., Begley, C.E., Bishop, M.L., Cushner-Weinstein, S., Holmes, G.L., Shafer, P.O., Sirven, J.I., Austin, J.K., 2013. Research implications of the Institute of Medicine Report, Epilepsy Across the Spectrum: Promoting Health and Understanding. Epilepsia 54, 207-216.

Jette, N., Reid, A.Y., Quan, H., Hill, M.D., Wiebe, S., 2010. How accurate is ICD coding for epilepsy? Epilepsia 51, 62-69.

Keezer, M.R., Sisodiya, S.M., Sander, J.W., 2016. Comorbidities of epilepsy: current concepts and future perspectives. The Lancet. Neurology 15, 106-115.

Lee, S.Y., Chung, S.E., Kim, D.W., Eun, S.H., Kang, H.C., Cho, Y.W., Yi, S.D., Kim, H.D., Jung, K.Y., Cheong, H.K., Committee on Epidemiology of Korean Epilepsy, S., 2016. Estimating the Prevalence of Treated Epilepsy Using Administrative Health Data and Its Validity: ESSENCE Study. J Clin Neurol 12, 434-440.

Loscher, W., Schmidt, D., 2011. Modern antiepileptic drug development has failed to deliver: ways out of the current dilemma. Epilepsia 52, 657-678.

Neligan, A., Bell, G.S., Johnson, A.L., Goodridge, D.M., Shorvon, S.D., Sander, J.W., 2011. The long-term risk of premature mortality in people with epilepsy. Brain 134, 388-395.

Ngugi, A.K., Bottomley, C., Kleinschmidt, I., Sander, J.W., Newton, C.R., 2010. Estimation of the burden of active and life-time epilepsy: a meta-analytic approach. Epilepsia 51, 883-890.

Novy, J., Belluzzo, M., Caboclo, L.O., Catarino, C.B., Yogarajah, M., Martinian, L., Peacock, J.L., Bell, G.S., Koepp, M.J., Thom, M., Sander, J.W., Sisodiya, S.M., 2013. The lifelong course of chronic epilepsy: the Chalfont experience. Brain 136, 3187-3199.
Oh, A., Thurman, D.J., Kim, H., 2017. Comorbidities and risk factors associated with newly diagnosed epilepsy in the U.S. pediatric population. Epilepsy & behavior : E&B 75, 230-236.
Parko, K., Thurman, D.J., 2009. Prevalence of epilepsy and seizures in the Navajo Nation 1998-2002. Epilepsia 50, 2180-2185.
Reid, A.Y., 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 research 102, 173-179.
Sander, J.W., 2003. The epidemiology of epilepsy revisited. Curr Opin Neurol 16, 165-170.
Schoen, C., Osborn, R., Doty, M.M., Bishop, M., Peugh, J., Murukutla, N., 2007. Toward higher-performance health systems: adults' health care experiences in seven countries, 2007. Health Aff (Millwood) 26, w717-734.
Schut, F., van de Ven, W., 2011. Health care reform in the Netherlands: the fairest of all? J Health Serv Res Policy 16, 3-4.
Selassie, A.W., Wilson, D.A., Martz, G.U., Smith, G.G., Wagner, J.L., Wannamaker, B.B., 2014. Epilepsy beyond seizure: a population-based study of comorbidities. Epilepsy research 108, 305-315.
Smeets, H.M., de Wit, N.J., Hoes, A.W., 2011. Routine health insurance data for scientific research: potential and limitations of the Agis Health Database. J Clin Epidemiol 64, 424-430.
Sundelin, H.E., Larsson, H., Lichtenstein, P., Almqvist, C., Hultman, C.M., Tomson, T., Ludvigsson, J.F., 2016. Autism and epilepsy: A population-based nationwide cohort study. Neurology 87, 192-197.
Surges, R., Thijs, R.D., Tan, H.L., Sander, J.W., 2009. Sudden unexpected death in epilepsy: risk factors and potential pathomechanisms. Nature reviews. Neurology 5, 492-504.
Thurman, D.J., Beghi, E., Begley, C.E., Berg, A.T., Buchhalter, J.R., Ding, D., Hesdorffer, D.C., Hauser, W.A., Kazis, L., Kobau, R., Kroner, B., Labiner, D., Liow, K., Logroscino, G., Medina, M.T., Newton, C.R., Parko, K., Paschal, A., Preux, P.M., Sander, J.W., 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.
Tu, K., Wang, M., Jaakkimainen, R.L., Butt, D., Ivers, N.M., Young, J., Green, D., Jette, N., 2014. Assessing the validity of using administrative data to identify patients with epilepsy. Epilepsia 55, 335-343.
Wannamaker, B.B., Wilson, D.A., Malek, A.M., Selassie, A.W., 2015. Stroke after adult-onset epilepsy: a population-based retrospective cohort study. Epilepsy & behavior : E&B 43, 93-99.

TABLE 1 Baseline characteristics of the participants in the validation study compared to the people with a diagnostic code for epilepsy in the Achmea Health Insurance Database (AHID)			
	Participants validation study (n=293)	AHID epilepsy population (18-75 years) (n=23493)	P-value*
Sex (% males)	143 (49%)	12783 (54%)	
Age in 2006 (median, IQR)	46 (34-55)	47 (34-59)	0.003
At least 2 DTP codes for epilepsy (2 or more epilepsy related visits) in the period 2006-2009	266 (91%)	18143 (77%)	<0.001
AED use in the period 2006-2009	269 (92%)	17869 (76%)	<0.001
Number of different AED types (average per year) (median, IQR) in the period 2006-2009	1.3 (1-2)	0.8 (0.3-1)	<0.001

TABLE 2: Age-specific prevalence rates for the different age groups AHID vs general Dutch population			
Age groups	Prevalence in AHID	Prevalence general population*	Standardized prevalence ratio
18 to 29 years	0.59	0.60**	0.99 (0.97-1.00)
30 to 39 years	0.59	0.55	1.07 (1.05-1.09)
40 to 49 years	0.72	0.63	1.15 (1.13-1.17)
50 to 59 years	0.84	0.75	1.12 (1.10-1.13)
60 to 69 year	0.86	0.81	1.05 (1.04-1.07)
70 to 79 years	0.99	0.94	1.06 (1.04-1.08)
80 years and older	0.87	0.71	1.23 (1.20-1.26)
Total	0.74	0.68	1.08 (1.08-1.09)
* Prevalence calculation based on number of DTPs for epilepsy in the Dutch population; data available for the period 2008 and 2009 ** age group 15-29 years as data are only provided per 5-year age groups			

TABLE 3 Diagnostic accuracy table

		Diagnosis from case-notes						
Diagnostic codes from the Achmea Health Insurance Database (AHID)		1 definite or probable epilepsy & AED	2 definite or probable epilepsy & no AED	3 suspect epilepsy & AED	4 suspect epilepsy & no AED	5 no epilepsy & AED	6 no epilepsy & no AED	Total
	1 DTP epilepsy & AED filled	258	3	5	2	1	0	269
	2 DTP epilepsy & no AED filled	6*	5	1*	11	0	1	24
	TOTAL	264	8	6	13	1	1	293

PPV = 0.95

*Non-compliance i.e. AEDs were prescribed by neurologist but never filled