Epilepsy & Behavior 51 (2015) 65–72

Contents lists available at ScienceDirect

Epilepsy & Behavior

journal homepage: www.elsevier.com/locate/yebeh

Development and validation of an epidemiologic case definition of epilepsy for use with routinely collected Australian health data

Michael Tan [a,*], Ian Wilson [a], Vanessa Braganza [b], Sophia Ignatiadis [a], Ray Boston [a,c], Vijaya Sundararajan [a], Mark J. Cook [a], Wendyl J. D'Souza [a]

[a] Department of Medicine, St Vincent's Hospital, The University of Melbourne, PO Box 2900, Fitzroy, 3065 Victoria, Australia
[b] The University of Virginia, Charlottesville, VA, USA
[c] Department of Clinical Studies, New Bolton Centre, School of Veterinary Medicine, University of Pennsylvania, 382 West Street Rd., Kennett Square, PA 19348, USA

ARTICLE INFO

Article history:
Received 30 April 2015
Revised 15 June 2015
Accepted 16 June 2015
Available online 8 August 2015

Keywords:
Validation
ICD-10AM
Epilepsy
Classification
Data linkage

ABSTRACT

Objectives: We report the diagnostic validity of a selection algorithm for identifying epilepsy cases.
Study design and setting: Retrospective validation study of International Classification of Diseases 10th Revision Australian Modification (ICD-10AM)-coded hospital records and pharmaceutical data sampled from 300 consecutive potential epilepsy-coded cases and 300 randomly chosen cases without epilepsy from 3/7/2012 to 10/7/2013. Two epilepsy specialists independently validated the diagnosis of epilepsy. A multivariable logistic regression model was fitted to identify the optimum coding algorithm for epilepsy and was internally validated.
Results: One hundred fifty-eight out of three hundred (52.6%) epilepsy-coded records and 0/300 (0%) nonepilepsy records were confirmed to have epilepsy. The kappa for interrater agreement was 0.89 (95% CI = 0.81–0.97). The model utilizing epilepsy (G40), status epilepticus (G41) and ≥1 antiepileptic drug (AED) conferred the highest positive predictive value of 81.4% (95% CI = 73.1–87.9) and a specificity of 99.9% (95% CI = 99.9–100.0). The area under the receiver operating curve was 0.90 (95% CI = 0.88–0.93).
Conclusion: When combined with pharmaceutical data, the precision of case identification for epilepsy data linkage design was considerably improved and could provide considerable potential for efficient and reasonably accurate case ascertainment in epidemiological studies.

1. Introduction

Data linkage is an emerging powerful tool, particularly for ascertaining low incidence events, enabling medical diseases and health outcomes to be connected using routinely collected centralized databases. For the study of epilepsy, it has the advantage of efficiently identifying large samples of patients with epilepsy [1–3]; however, misclassification of cases with epilepsy in administrative databases is a major issue, limiting its utility as an epidemiologic instrument for disease surveillance and research.

International Classification of Diseases (ICD)-coded data are utilized internationally as an epidemiological tool for collection of national health statistics [4]; however, their use requires validation. Previous studies that utilize this classification system have used ICD-9 and ICD-10 versions [5–10]. As ICD-10 revisions are country-specific and coding practices may vary between regions, it has been recommended that specificity and predictive values be evaluated for each population studied [11]. Current epidemiological guidelines suggest that a probable diagnosis of epilepsy can be made if 1 of the following 3 conditions is met: one medical encounter with a 3-digit code of G40.x (epilepsy); ≥2 medical encounters on separate days coded with G41 (status epilepticus) or with a 4-digit code R56.8 (other and unspecified convulsions); and a single medical encounter coded as other and unspecified convulsions (R56.8) and an antiepileptic drug prescription for three or more months [11]. A suspected diagnosis of epilepsy can be made with single episodes coded with R56.8 or G41 [11].

The coding of Australian Modification (AM) version of ICD-10 (ICD-10AM) clinical data in Australia exists for diagnoses and procedures of acute admitted patient episodes only [12]. At present, epilepsy coding for the ICD-10AM has not been validated. We sought to estimate the diagnostic accuracy of different ICD-10AM coding algorithms to identify patients with epilepsy in an Australian hospital setting and to develop an algorithm combining other routinely nationally collected data to maximize its precision for potential epidemiologic surveillance and research.

2. Methodology

Following a discharge from a public or a private hospital in Australia, all principal diagnoses and additional diagnoses (up to 99) in medical records are coded with ICD-10AM codes and submitted to the National Hospital Morbidity Database for statistical reporting [13,14]. We

* Corresponding author. Tel.: +61 3 9288 2211.
E-mail address: michael.tan@svha.org.au (M. Tan).

http://dx.doi.org/10.1016/j.yebeh.2015.06.031
1525-5050/

retrospectively recruited a validation cohort by identifying cases with epilepsy and their common differential diagnoses in a metropolitan adult hospital setting with a large neurology unit including epilepsy subspecialists from 3 July 2012 to 10 July 2013. All hospital records were identified by the Health Information Services (HIS) Department at St Vincent's Hospital, Melbourne for episodes coded with the ICD codes of interest (see Table 1). Potential cases with epilepsy were defined as patients coded with epilepsy (ICD-10AM G40.xx), status epilepticus (G41.xx), other and unspecified convulsions (R56.8x), and acquired aphasia with epilepsy (F80.3x). Controls ("mimickers") were identified with the most common alternative diagnoses (Table 1): syncope and collapse (R55.xx) and conversion disorders (F44.xx), among others [7]. Ethics approval was obtained from St Vincent's Hospital, Melbourne HREC.

Initial power calculations indicated that 242 subjects per group would provide 90% power (two-sided $\propto = 0.05$) to detect a 95% positive predictive value when compared with the gold standard of an epilepsy specialist's diagnosis. We chose to oversample and included 300 consecutive potential epilepsy cases and randomly sampled 300 differential diagnostic cases from either the ED or an inpatient unit.

During the approximate 12-month time period from 3 July 2012 to 10 July 2013, 42,760 patients were seen through St Vincent's Hospital, Melbourne. Data extraction yielded 3272 potential epilepsy episodes for 3014 individuals (1966 emergency department (ED) and 1306 inpatient) from 3/7/2012 to 10/7/2013. Emergency department episodes were selected if they were solely managed in this setting for the selected episode and not subsequently admitted into an inpatient unit.

Information pertaining to gender, postcode, age, indigenous status (Aboriginals and Torres Strait Islanders), as well as previous admissions for up to 5 years for each patient was included in the request from the hospital's HIS. As an indicator of socioeconomic status, each patient's postcode of residence was scored into one of five ordered categories according to the Socioeconomic Index for Area (SEIFA) constructed by the Australian Bureau of Statistics (using the subcategory for relative advantage/disadvantage) [15]. Socioeconomic Index for Area provides a range of measures to rank geographic areas based on their relative social and economic well-being.

A computer database was developed to facilitate standardized data collection and validate entries on completion. A pilot study of 30 subjects was performed, with MT and IW initially reviewing 30 records (medical charts, ambulance notes, investigations, and correspondence) and entering standardized data (e.g., diagnosis and seizure details as listed in Box 1) into the database. Any disagreements in any data element entered into the database required the record to be reviewed by a third reviewer (WD). Disagreements between MT and IW were automatically detected by the database, and notification was sent to the third reviewer to conduct a review. Following this, MT, IW, and WD discussed the case. Feedback from MT, IW, and WD was sought to optimize the standardized data collection method, with the final standardized questions listed in Box 1. The pilot data were discarded from further analysis.

All assessors were blinded to the designated ICD-10AM codes. The primary outcome measured was true disease status determined independently by two epilepsy specialists with formal general neurology and epilepsy subspecialty training (MT and IW) using epilepsy case definition based on published guidelines [11]. Independent agreement between assessors for each data item was considered to confirm the final diagnosis. Assessors were instructed to disregard information that was not available to the treating clinician at the time of the presentation. Any discordance between assessors in any of the data items listed in Box 1 was reviewed by a third epilepsy specialist (WD) before a final consensus was reached.

Statistical analysis was performed using Stata 12. The diagnostic characteristics of a test (i.e., sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV)) were defined according to published epilepsy guidelines [11]. Kappa scores for interrater agreement were calculated to assess agreement between assessors. The kappa statistic represents the level of agreement above that due to chance alone where almost perfect > 0.81, substantial = 0.61–0.8, moderate = 0.41–0.60, fair = 0.21–0.40, slight = 0.00–0.20, and

Table 1
Frequency of cases and common differential diagnoses by primary ICD-10AM diagnosis.

ICD-10AM	Description of code	n	Epilepsy specialist diagnosis on medical record review							
			Established epilepsy	Incident epilepsy	First unprovoked seizure	Acute symptomatic seizure	Syncope	Psychogenic nonepileptic seizure	Other determined diagnoses	Nature of episode — uncertain
Cases										
G40	Epilepsy	139	94	3	4	12	4	9	3	10
G41	Status epilepticus	10	9	0	0	0	0	1	0	0
R56.8	Other and unspecified convulsions	151	50	2	16	46	5	13	9	10
F80.3	Acquired aphasia with epilepsy (Landau–Kleffner)	0	0	0	0	0	0	0	0	0
	Total	300	153	5	20	58	9	23	12	20
Mimickers										
F41	Other anxiety disorders	56	0	0	0	0	0	0	55	1
F44	Dissociative (conversion) disorders	9	0	0	0	0	0	6	3	0
F51	Nonorganic sleep disorders	0	0	0	0	0	0	0	0	0
G43	Migraine	21	0	0	0	0	0	0	21	0
G45	Transient cerebral ischemic attacks and related syndromes	14	0	0	0	0	1	1	12	0
G47	Sleep disorders	79	0	0	0	0	0	0	79	0
H81	Disorders of vestibular function	14	0	0	0	0	0	0	14	0
R55	Syncope and collapse	107	0	0	0	1	79	4	14	9
	Total	300	0	0	0	1	80	11	198	10
Additional variables										
	AEDs ≥ 1		139	1	5	23	5	21	27	13
	EEG		41	1	10	18	4	13	0	11
	MRI brain		14	1	4	11	2	0	8	4
	CT brain		31	0	15	34	17	5	28	16
	Previous admissions ≥ 1		17	0	0	2	1	2	9	1
	AED level		38	1	1	5	1	4	2	7