Seizure 52 (2017) 195–198

Contents lists available at ScienceDirect

Seizure

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

Validating epilepsy diagnoses in routinely collected data

Beata Fonferko-Shadrach[a,b,c], Arron S. Lacey[a,c,d], Catharine P. White[a,b], H.W. Rob Powell[a,b], Inder M.S. Sawhney[a,b], Ronan A. Lyons[d], Phil E.M. Smith[a,e], Mike P. Kerr[a,f], Mark I. Rees[a,c], W. Owen Pickrell[a,c,e,*]

[a] Wales Epilepsy Research Network, Institute of Life Science, Swansea University, Swansea, UK
[b] Morriston Hospital, Abertawe Bro Morgannwg University Health Board, Swansea, UK
[c] Neurology and Molecular Neuroscience Research Group, Institute of Life Science, Swansea University Medical School, Swansea University, Swansea, UK
[d] Farr Institute, Swansea University Medical School, Swansea University, Swansea, UK
[e] University Hospital of Wales, Cardiff and Vale University Health Board, Cardiff, UK
[f] Institute of Psychological Medicine and Clinical Neuroscience, Cardiff University, Cardiff, UK

ARTICLE INFO

Article history:
Received 11 August 2017
Received in revised form 20 September 2017
Accepted 12 October 2017
Available online xxx

Keywords:
Diagnosis
Validation
Routinely collected data
Epilepsy

ABSTRACT

Purpose: Anonymised, routinely-collected healthcare data is increasingly being used for epilepsy research. We validated algorithms using general practitioner (GP) primary healthcare records to identify people with epilepsy from anonymised healthcare data within the Secure Anonymised Information Linkage (SAIL) databank in Wales, UK.

Method: A reference population of 150 people with definite epilepsy and 150 people without epilepsy was ascertained from hospital records and linked to records contained within SAIL (containing GP records for 2.4 million people). We used three different algorithms, using combinations of GP epilepsy diagnosis and anti-epileptic drug (AED) prescription codes, to identify the reference population.

Results: Combining diagnosis and AED prescription codes had a sensitivity of 84% (95% ci 77–90) and specificity of 98% (95–100) in identifying people with epilepsy; diagnosis codes alone had a sensitivity of 86% (80–91) and a specificity of 97% (92–99); and AED prescription codes alone achieved a sensitivity of 92% (70–83) and a specificity of 73% (65–80). Using AED codes only was more accurate in children achieving a sensitivity of 88% (75–95) and specificity of 98% (88–100).

Conclusion: GP epilepsy diagnosis and AED prescription codes can be confidently used to identify people with epilepsy using anonymised healthcare records in Wales, UK.

1. Introduction

Vast amounts of electronic, routinely-collected, medical and related administrative data are generated in modern healthcare systems. These data can be anonymised, linked and used for healthcare research [1,2]. Large numbers of individuals can be studied without having to specifically recruit individuals for projects, which can be expensive, time-consuming and introduce selection bias. Records can also be linked from a wide variety of different sources, enabling a wide breadth of data to be analysed. Routinely-collected data are increasingly being used for high quality epilepsy studies [3–5].

Every individual in the United Kingdom (UK) is entitled to register with a primary care General Practitioner (GP) and there is evidence that almost everyone in the UK does register with a GP [6]. GPs have a central role in providing primary care for people with epilepsy through assessment, diagnosis, appropriate referral to secondary and tertiary services, managing and prescribing medications (including the vast majority of anti-epileptic drugs) and creating and maintaining a centralised health care record. GPs are the patient's primary contact point for access to specialist services. GP health records contain details of encounters with GPs and other healthcare providers using Read codes.

Read codes are the current clinical terminology coding system used in UK primary care systems to record symptoms, diagnosis and prescriptions [7]. Read codes are hierarchical (with increasing level of detail with increasing digits) e.g: *F25* is used to record epilepsy, *F25A.* is used for juvenile myoclonic epilepsy and *F2540* for temporal lobe epilepsy.

* Corresponding author at: Neurology and Molecular Neuroscience, Swansea University Medical School, Institute of Life Science, Swansea University, Swansea, Wales SA2 8PP, UK.
E-mail address: w.o.pickrell@swansea.ac.uk (W. O. Pickrell).

https://doi.org/10.1016/j.seizure.2017.10.008
1059-1311/

GP records have been used as the basis for epilepsy studies within data repositories such as the clinical practice research datalink (CPRD) and the Secure Anonymised Information Linkage (SAIL) databank [4,8,9]. One of the limitations of using routinely-collected data for epilepsy studies is the possibility of including incorrectly recorded epilepsy diagnoses. In particular, it's not known how accurately epilepsy diagnoses made by hospital specialists are recorded in GP records. Guidelines advise that algorithms used for case ascertainment in routinely-collected data studies are validated in each population studied [10]. The accuracy of UK GP diagnosis codes has been validated for many diseases but, to our knowledge, has only been partially validated for epilepsy diagnosis [8,9,11]. In this study we specifically aimed to validate the accuracy of algorithms using GP records to identify people with epilepsy from anonymised, linked, routinely collected Welsh healthcare data.

2. Method

In Wales, anonymised GP primary care electronic health records are collated and linked with other data within the Secure Anonymised Information Linkage (SAIL) databank [1,12]. We searched the SAIL databank on 13th April 2016, at this time GP records were available up to 31st December 2015 and there were records for 73% of GP practices across Wales (approx. 2.4 million people). GP records can be tracked over time, so that individual patient's records can be analysed longitudinally through multiple GP practices. We used combinations of epilepsy diagnosis and anti-epileptic drug (AED) prescription codes to create three epilepsy case ascertainment algorithms.

We anonymously uploaded and linked a list of 150 individuals with epilepsy and 150 individuals without epilepsy (reference population) to existing SAIL records, using an established and validated split-file approach [1,12]. We then compared the performance of the three different epilepsy case ascertainment algorithms within SAIL in identifying the reference population.

2.1. The reference population

The Swansea Epilepsy Database currently holds detailed clinical information (including diagnosis, medications, imaging and EEG results) for 960 patients seen by a clinician with a specialist interest in epilepsy (neurologist or paediatric neurologist) treated at Morriston Hospital, Swansea. 283 (29%) of these patients have generalised epilepsy, 510 (53%) have focal epilepsy, 125 (13%) have unclassifiable epilepsy and 42 (4%) have an uncertain diagnosis.

Between January and March 2015, we examined the database and used a random number generator to select a sample of 100 adults (50 men and 50 women, who were over 16 at their last consultation date) and 50 children (25 boys and 25 girls, who were 16 and under at their last consultation date) with a clinically definite diagnosis of epilepsy from the database. The clinical record and investigation results for each of these 150 individuals were reviewed to confirm a clinically definite diagnosis as per the International League Against Epilepsy's (ILAE) practical clinical definition of epilepsy. These 150 individuals formed the reference population of people *with epilepsy*.

To ascertain a control cohort, 300 patients were reviewed from consecutive general neurology clinics run by neurologists and paediatric neurologists. Their diagnosis was checked using clinic letters stored in an electronic format on the hospital system. Patients with a diagnosis of epilepsy were excluded. Using a random number generator, we randomly selected a sample of 100 adults (50 men and 50 women, who were over 16 at their last consultation date) and 50 children (25 boys and 25 girls, who were 16 and under at their last consultation date) from these 300 patients. These 150 individuals formed the reference population of people *without* epilepsy.

We have previously estimated the sensitivity of an epilepsy case ascertainment algorithm at 90% using GP diagnosis and AED prescription [9]. Based on this, a sample size of 150 provides a 95% confidence interval of 10% for sensitivities (proportions) of 90%.

Table 1
Proportion of epilepsy cases (n = 145) and cases without epilepsy (n = 143) identified within SAIL using three different algorithms: A – Individuals with a primary care epilepsy diagnosis code and at least two consecutive codes for prescription of an anti-epileptic drugs (AED); B – Individuals with an epilepsy diagnosis code only; C – Individuals with at least two consecutive codes for prescription of an AED. See method section for definitions of positive predictive value, sensitivity, false positive rate, specificity and Youden's Index.*We included 145 (97 adults, 48 children) people with a hospital diagnosis of epilepsy and 143 (98 adults and 45 children) people without a hospital diagnosis of epilepsy.

Patients within SAIL identified as having epilepsy			Hospital neurology service diagnosis of epilepsy*		Positive predictive value (95% CI)	Sensitivity (95% CI)	False positive rate (95% CI)	Specificity (95% CI)	Youden's Index (J)
Algorithm Used			Yes	No					
A – Epilepsy diagnosis & AED	All patients	Yes	122	2	98% (94–100)	84% (77–90)	1% (0–5)	99% (95–100)	0.83
		No	23	141					
	Adults	Yes	84	2	98% (92–100)	87% (78–93)	2% (0–7)	98% (93–100)	0.85
		No	13	96					
	Children	Yes	38	0	100% (91–100)	79% (65–90)	0% (0–8)	100% (92–100)	0.79
		No	10	45					
B – Epilepsy diagnosis only	All patients	Yes	125	5	96% (91–99)	86% (80–91)	3% (1–8)	97% (92–99)	0.83
		No	20	138					
	Adults	Yes	85	2	98% (92–100)	88% (80–93)	2% (0–7)	98% (93–100)	0.86
		No	12	96					
	Children	Yes	40	3	93% (81–99)	83% (70–93)	7% (1–18)	93% (82–99)	0.76
		No	8	42					
C – AED only	All patients	Yes	133	39	77% (70–83)	92% (86–96)	27% (20–35)	73% (65–80)	0.65
		No	12	104					
	Adults	Yes	91	38	71% (63–78)	94% (87–98)	39% (30–49)	61% (51–71)	0.55
		No	6	60					
	Children	Yes	42	1	98% (94–100)	88% (75–95)	2% (0–12)	98% (88–100)	0.86
		No	6	44					