THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

The Accuracy and Utility of Using Administrative Healthcare Databases to Identify People with Epilepsy: A Protocol for a Systematic Review and Meta-Analysis

OPEN ACCESS

BMJ Open

Accuracy and utility of using administrative healthcare databases to identify people with epilepsy: a protocol for a systematic review and meta-analysis

Gashirai K Mbizvo,[1,2] Kyle Bennett,[1] Colin R Simpson,[3,4] Susan E Duncan,[1,2] Richard F M Chin[1,5]

[1]Muir Maxwell Epilepsy Centre, University of Edinburgh Centre for Clinical Brain Sciences, Edinburgh, UK
[2]Department of Clinical Neurosciences, Western General Hospital, Edinburgh, UK
[3]Faculty of Health, Victoria University of Wellington, Wellington, New Zealand
[4]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh School of Molecular Genetic and Population Health Sciences, Edinburgh, UK
[5]Department of Neurosciences, Royal Hospital for Sick Children, Edinburgh, UK

Correspondence to
Dr Gashirai K Mbizvo;
gashirai.mbizvo@ed.ac.uk

ABSTRACT

Introduction In an increasingly digital age for healthcare around the world, administrative data have become rich and accessible tools for potentially identifying and monitoring population trends in diseases including epilepsy. However, it remains unclear (1) how accurate administrative data are at identifying epilepsy within a population and (2) the optimal algorithms needed for administrative data to correctly identify people with epilepsy within a population. To address this knowledge gap, we will conduct a novel systematic review of all identified studies validating administrative healthcare data in epilepsy identification. We provide here a protocol that will outline the methods and analyses planned for the systematic review.

Methods and analysis The systematic review described in this protocol will be conducted to follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. MEDLINE and Embase will be searched for studies validating administrative data in epilepsy published from 1975 to current (01 June 2018). Included studies will validate the International Classification of Disease (ICD), Ninth Revision (ICD-9) onwards (ie, ICD-9 code 345 and ICD-10 codes G40–G41) as well as other non-ICD disease classification systems used, such as Read Codes in the UK. The primary outcome will be providing pooled estimates of accuracy for identifying epilepsy within the administrative databases validated using sensitivity, specificity, positive and negative predictive values, and area under the receiver operating characteristic curves. Heterogeneity will be assessed using the I^2 statistic and descriptive analyses used where this is present. The secondary outcome will be the optimal administrative data algorithms for correctly identifying epilepsy. These will be identified using multivariable logistic regression models. 95% confidence intervals will be quoted throughout. We will make an assessment of risk of bias, quality of evidence, and completeness of reporting for included studies.

Ethics and dissemination Ethical approval is not required as primary data will not be collected. Results will be disseminated in peer-reviewed journals, conference presentations and in press releases.

PROSPERO registration CRD42017081212.

Strengths and limitations of this study

► The protocol describes what will be the first systematic review to conduct a worldwide assessment of the accuracy of administrative data in identifying epilepsy and the optimal disease-identification algorithms.

► This protocol also describes what will be the first systematic review to make an assessment of risk of bias, quality of evidence, and completeness of reporting for studies validating administrative healthcare data in epilepsy identification.

► The review described in this protocol will be limited to assessing the use of administrative data in diagnosing epilepsy within observational studies, which are more prone to bias than randomised controlled trials.

► A systematic review of the diagnostic accuracy of administrative data within randomised controlled trials in epilepsy remains to be completed and is out of the scope of the current review.

INTRODUCTION

Administrative healthcare databases are electronic data sources that consist of demographic, diagnostic and clinical information routinely collected about patients when they use a healthcare service.[1] They are often national and mandatory, and therefore they have the potential to provide a relatively cheap, widely available and less intrusive resource for medical research.[1] However, the accuracy of the information held in an administrative database needs to be validated before such use can be made. This is because the administrative data were not originally collected for research, but for other purposes such as assisting in health insurance claims. The clinical information held may therefore lack the rigour in accuracy that might be expected in scientifically collected data.

Furthermore, the data may be limited by inaccurate or incomplete hospital discharge letters or clinical coding transcription errors.[2]

The validation of administrative data involves comparing the diagnostic codes held within the administrative database against a reference standard (such as medical records) in order to quantify the number of instances in which the administrative diagnosis made matches the diagnosis in the reference standard (deemed to be the true diagnosis). In this way, the administrative database can be handled like a diagnostic test and measures of disease-identification accuracy calculated. These measures usually include the sensitivity, specificity and the positive or negative predictive value (PPV or NPV, respectively). Optimal disease-identification algorithms can also be determined by making relative comparisons of predictive values after adding in data from other variables recorded in an administrative database, such as drug combinations, investigations, and procedures.

There are many administrative databases worldwide in which the process of validation has been successfully performed for many diagnostic codes.[3–9] There are also examples of where the results of these have been pooled successfully into systematic review to increase confidence in the estimates made and scrutinise the quality of evidence, and this has led to changes in practice.[4] There has been limited systematic review of the validation of administrative databases in capturing epilepsy as a diagnosis. The only systematic review[10] on this subject included only studies from the USA or Canada and therefore excluded 121 studies because the data sources were not from these two countries. Furthermore, the 11 studies included were published between 2000 and 2010, making the conclusions nearly a decade old now. With health informatics now at the forefront of epidemiological disease surveillance, it is important to have an update on performance of the administrative disease-identification codes. Only one of the included studies evaluated the performance of the International Classification of Disease (ICD), 10th Revision (ICD-10) system in capturing epilepsy within administrative datasets[11]; the remainder evaluated the older ICD, 9th Revision (ICD-9) system.[10] The review also made no assessment of risk of bias, quality of evidence and completeness of reporting for included studies. This limits the confidence with which conclusions can be interpreted. There is now need for a more contemporary systematic review of the validation of administrative databases in capturing epilepsy. This should include evaluating performance of the ICD-10 system, as well as other non-ICD disease classification systems used, such as Read Codes in the UK.[12 13] The review should include studies from anywhere in the world in order to give clinicians and researchers a representative picture of the performance of administrative data in capturing epilepsy as a diagnosis and in order to allow more generalisable diagnostic algorithms to be suggested. Furthermore, the review should make an assessment of risk of bias, quality of evidence and completeness of reporting for included studies. These are

the aims of the proposed systematic review described in this protocol. This will help researchers and clinicians better understand the accuracy of global estimates for incidence, prevalence and population characteristics in epilepsy, which have largely been made using administrative data.

Aims and objectives

The study hypotheses are:

1. Administrative data can correctly identify people with epilepsy within a population with a high degree of accuracy. We predict the PPV to be above 80%.
2. The optimal disease-identification algorithms for epilepsy within administrative datasets take into account diagnoses, investigations and drug combinations.[11 14–20]

The aim of the systematic review is to quantify the disease-identification accuracy and algorithm performance of administrative healthcare data in epilepsy. To this end, the research questions are:

1. How accurately do administrative data identify epilepsy within a population as measured by sensitivity, specificity, PPV, NPV, or area under the receiver operating characteristic curve (AUC) analysis (which are the approved accuracy measures described in the Standards for Reporting Diagnostic Accuracy (STARD) statement)?[21] This will be the primary outcome.
2. What are the optimal administrative data algorithms for correctly identifying epilepsy within a population? This will be the secondary outcome.

A preliminary feasibility search of the MEDLINE database via PubMed identifies at least nine studies validating diagnostic epilepsy codes held within administrative databases around the world that could be used to answer these research questions.[11 13–20]

METHODS AND ANALYSIS

This protocol follows the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) checklist.[22 23] The systematic review will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.[24]

Eligibility criteria

We will include studies according to the following criteria:

Language: there will be no language restriction on full-length articles, although abstracts will need to be in English to allow the authors to screen them. We will seek translations for full-length articles not written in English that appear eligible in abstract. These will remain in the section for 'studies still awaiting classification' and will feature in subsequent updates to the review if not translated by the time of initial publication. *Setting:* there will be no restrictions by study location worldwide. Where possible, we will show pooled accuracy and best algorithm data for administrative datasets from individual countries in addition to pooled global estimates of these measures.

Databases: the data sources will be routine administrative healthcare databases. This means that the data should have been routinely and passively collected without an a priori research question.[4] We will include databases containing diagnostic codes for epilepsy classified on the ICD system, where we will restrict this to studies using the ICD-9 onwards (active from 1975 to 1994).[25] This is because although the ICD system is currently in its 10th revision (ICD-10; active from 1994 to present) and the primary coding system is used by many countries around the world, a significant proportion of countries, particularly developing ones, still use the ICD-9 system.[11] ICD-9 code 345 and ICD-10 codes G40 and or G41 will be used to identify epilepsy diagnoses. We will provide summary measures of accuracy and best algorithms for any other disease classification systems used in studies separately, for example, the primary care Read Code system used in the UK.[12 13]

Study design: prospective or retrospective observational studies including cohort or case–control designs that are community-based/population-based or primary/secondary/tertiary care-based and have used administrative databases.

Participants: people with epilepsy of all ages. Where available, we will additionally report data for adults and children separately.

Observations and outcomes: studies will need to employ a validation process for diagnostic epilepsy codes, that is, estimate the disease-identification accuracy of the epilepsy codes held within the database using sensitivity, specificity, PPV, NPV or AUC analysis.[21] In this, true positives (TPs) and false negatives (FNs) will be considered as the patient has the disease and the administrative diagnosis is positive or negative, respectively. True negatives (TNs) and false positives (FPs) will be considered as the patient does not have the disease and the administrative diagnosis is negative or positive, respectively. Sensitivity will be considered as the ability of the administrative database to correctly identify those patients with the disease (TP/(TP+FN)). Specificity will be considered as the ability of the administrative database to correctly identify those patients without the disease (TN/(TN+FP)). PPV will be considered as how likely it is that a patient has the disease if the administrative diagnosis is positive (TP/(TP+FP)). NPV will be considered as how likely it is that this patient does not have the disease if the administrative diagnosis is negative (TN/(TN+FN)). AUC analysis will be considered as where TP and FP are plotted against each other in a perfect correlation as reference to show poor test accuracy, then accuracy measured as the area under the curve created by plotting the actual values against each other.[26]

Studies may also use diagnostic as well as other variables (eg, admissions, drugs or investigations) to calculate optimal disease-identification algorithms for epilepsy within the database. Studies will need to provide a clinical reference standard. An appropriate clinical reference standard will be medical records, clinical assessment, or a validated disease registry.[21]

Timeframe: studies conducted from 1 January 1975 to 01 June 2018. The year 1975 represents the advent of ICD-9.[25]

We will exclude studies according to the following criteria:

Data reported in systematic reviews unless we can identify the primary data, for example, by contacting authors of the original source.

Conference proceedings abstracts or studies not written in English where we are unable to obtain the meta-data from authors or full-length manuscript translations remain awaited, respectively.

Information sources

Studies will be identified from the following sources

▶ *Electronic databases:* we will search 01 January 1975–01 June 2018 for studies meeting the inclusion criteria within the MEDLINE (Ovid interface) and Embase (Ovid interface) databases. The search strategies are outlined in table 1.

▶ *Conference proceedings:* for conference abstracts that appear to meet the inclusion criteria but do not have a full-length article published, we will contact authors directly to request metadata.

▶ *Reference lists:* we will also identify studies meeting the inclusion criteria from the reference lists of included studies and relevant reviews identified through the electronic database searches.

Study records

Data management

Literature search results will be uploaded onto Review Manager 5, an internet-based software program that facilitates collaboration among reviewers. Citation titles and abstracts will be uploaded. GKM will then screen titles and abstracts to identify and exclude duplicate publications. Duplicate publications will be identified by comparing author names, study titles, sample sizes, outcomes used, and any other information held in the abstracts. All reviewer authors will have access to the systematic review process via the internet-based review software program, and this will create an audit trail of studies included/excluded, data analysis steps, and subsequent manuscript revisions. All data will be held within the management software and password protected.

Selection process

Once duplicates have been excluded, two review authors (GKM and KB) will independently screen the titles and abstracts yielded from the databases searches against the inclusion criteria. Where titles and abstracts indicate that a study may meet the inclusion criteria or where there is uncertainty about this, the full-length manuscripts will be

Table 1 Search strategies for MEDLINE and Embase

Ovid MEDLINE Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE Daily and Ovid MEDLINE 1946 to 01 June 2018	Embase 1974 to 01 June 2018
1. Epilepsy, Complex Partial/or Epilepsy, Reflex/or Epilepsy, Absence/or Drug Resistant Epilepsy/or Epilepsy/or Epilepsy, Rolandic/or Epilepsy, Partial, Motor/or Epilepsy, Benign Neonatal/or Epilepsy, Tonic-Clonic/or Epilepsy, Post-Traumatic/or Epilepsy, Partial, Sensory/or Epilepsy. mp. or Epilepsy, Temporal Lobe/or Epilepsy, Frontal Lobe/or Myoclonic Epilepsy, Juvenile/or Epilepsy, Generalized/	1. reflex epilepsy/or photosensitive epilepsy/or grand mal epilepsy/or epilepsy.mp. or drug resistant epilepsy/or experimental epilepsy/or severe myoclonic epilepsy in infancy/or childhood absence epilepsy/or benign childhood epilepsy/or catamenial epilepsy/or symptomatic epilepsy/or startle epilepsy/or generalized epilepsy/or epilepsy/or mesial temporal lobe epilepsy/or rolandic epilepsy/or traumatic epilepsy/or myoclonic astatic epilepsy/or temporal lobe epilepsy/or intractable epilepsy/or focal epilepsy/or 'seizure, epilepsy and convulsion'/or myoclonus epilepsy/or lateral temporal lobe epilepsy/or frontal lobe epilepsy/
2. Databases, Factual/	2. factual database/or data base/
3. 'Reproducibility of Results'/	3. reproducibility/
4. Algorithms/or algorithm*.mp.	4. algorithm/or algorithm*.mp.
5. 3 or 4	5. 3 or 4
6. 2 and 5	6. 2 and 5
7. 1 and 6	7. 1 and 6
8. Administrative Claims, Healthcare/or administrativ*.mp. or insurance data*.mp. or claims data*.mp. or Veterans Health Administration.mp.	8. 'administrative claims (health care)"/or administrative*.mp. or insurance data*.mp. or claims data*.mp. or Veterans Health Administration.mp.
9. administrat* data*.mp.	9. administrat* data*.mp.
10. routin* data*.mp.	10. routin* data*.mp.
11. big data.mp.	11. big data.mp.
12. 8 or 9 or 10 or 11	12. 8 or 9 or 10 or 11
13. 1 and 12	13. 1 and 12
14. Algorithms/or algorithm*.mp.	14. algorithm/or algorithm*.mp.
15. 'Sensitivity and Specificity'/or 'Predictive Value of Tests'/or predictive value.mp.	15. (specificity or sensitivity).mp. or 'sensitivity and specificity'/or positive* predict* value*.mp. or negative* predict* value*. mp.
16. positive* predict* value*.mp.	16. area under the curve/or area* under* curve*.mp.
17. negative* predict* value*.mp.	17. roc curve/or receiver operating characteristic/or ROC curve*.mp.
18. sensitivity.mp.	18. 14 or 15 or 16 or 17
19. specificity.mp.	19. code*.mp. or 'Read code'/
20. area* under* curve*.mp. or Area Under Curve/	20. ICD-10.mp. or 'International Classification of Diseases'/or ICD-10/or disease classification/or ICD-9.mp. or ICD-9/
21. ROC Curve/or ROC curve*.mp.	21. 19 or 20
22. 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21	22. 18 and 21
23. code*.mp.	23. 1 and 22
24. (ICD-9 or ICD-10).mp. or 'International Classification of Diseases'/or Clinical Coding/or read code*.mp.	24. validation study/or validation Process/or validat*.mp. or validity/or predictive validity/or validity.mp.
25. 23 or 24	25. electronic medical record/or medical record/or medical record*.mp. or medical case note*.mp.
26. 22 and 25	26. Registries.mp. or register/
27. 1 and 26	27. 25 or 26
28. (validat* or validity).mp. or Validation Studies/	28. 24 and 27
29. Medical Records/or medical record*.mp. or medical case note*.mp.	29. 1 and 28
30. electronic health records.mp. or Medical Records Systems, Computerized/or Electronic Health Records/	30. 7 or 13 or 23 or 29
31. Registries.mp. or Registries/	31. limit 30 to yr='1975 -Current'
32. 29 or 30 or 31	32. animal/not human/
33. 28 and 32	33. 31 not 32
34. 1 and 33	
35. 7 or 13 or 27 or 34	
36. limit 35 to yr='1975 -Current'	
37. Animals/not Humans/	
38. 36 not 37	

downloaded and used to help decide. Where details in the manuscript are still insufficient for a decision to be made about eligibility, we will seek additional information from the study authors and automatically exclude studies where there is no response from authors after three weeks. We will record the reasons for excluding all excluded studies. The two review authors will compare their list of included and excluded trials and any disagreements

will be resolved by mutual discussion and, where necessary, adjunction by a third reviewer (RFMC/SED/CRS). Review authors will not be blind to the journal titles, study authors, or institutions.

Data collection process and items

Two review authors (GKM and KB) will independently abstract data about the primary and secondary outcomes from included studies using the data collection tool shown below. The additional information extracted maps onto the items contained within the STARD guidelines modified for epilepsy studies reporting diagnostic accuracy of administrative databases.[21] This will allow us to extract sufficient information to make a quality assessment of the completeness of included studies against the STARD checklist.[21] Any disagreements in the contents of data abstraction will be resolved by mutual discussion and, where necessary, adjunction by a third reviewer (RFMC/SED/CRS).

Data collection tool

► What is the study title?
► Who are the study authors?
► What is the year of study publication?
► What is the journal of publication?
► What country(s) was the study conducted in?
► Does the study explicitly identify as utilising 'administrative data' (yes/no)?
    – If not, how is this identified by the reviewer? For example, from descriptions given of the databases utilised and background knowledge about them or from correspondence with authors.
► Does the abstract provide a structured summary of study design, methods, results and conclusions (yes/no/unclear)?
► Does the introduction give a scientific and clinical background including intended use and clinical role of administrative data (yes/no/unclear)?
► Are the study objectives and hypotheses described (yes/no/unclear)?
    – If so, what are they?
► What was the intended study sample size and how was it determined?
► What is the study design?
► Was a study/participant flow diagram used?
► What are the eligibility criteria for participation in the validation cohort (ie, the cohort of patients to which the reference standard will be applied)?
► Where, when and how were potentially eligible validation cohort participants identified? Include within this:
    – What is the name of the administrative database(s) on which the validation cohort was identified?
    – What are the setting and location of the administrative database(s) from which the validation cohort was identified? For example, is it primary care, secondary care, tertiary care, outpatient care and emergency care?

    – What are the names of any hospitals/organisations affiliated with or using the administrative database routinely?
    – What is the size of the administrative database(s) on which the validation cohort was identified? That is, how many people/records does it hold in total?
    – What were the epilepsy ICD codes (or other disease classification system codes) used to identify the validation cohort within the administrative database(s)? That is, what are the diagnostic epilepsy codes that will be validated by the study?
    – What was the size of the validation cohort identified by the epilepsy codes? That is, give the number of participants identified by these diagnostic epilepsy codes.
    – Did the validation cohort include identifying a sample of people (1) without epilepsy and (2) with epilepsy 'mimicker codes'?
        – If so, give details of the codes used and the number of participants for each of these groups.
        – Samples of people without epilepsy are often used to help calculate the specificity and NPV of an administrative database. 'Mimicker codes' are often interrogated as the conditions may resemble epilepsy. These may include, for example, classical migraine (ICD-9 code 346.x and ICD-10 code G43.1), transient cerebral ischaemia (ICD-9 code 435 and ICD-10 code G45), syncope (ICD-9 code 780.2 and ICD-10 code R55) or convulsion (ICD-9 code 780.3 and ICD-10 code R56.0 or R56.8), which are intended to be used for organic convulsions but not for epilepsy.[11]
► What other information was obtained about an individual to help identify epilepsy on the administrative database? For example, describe if they linked an individual's ICD epilepsy codes with the investigations they underwent (such as an electroencephalogram (EEG)) or the antiepileptic drug (AED) they were taking.
► What were the demographic and clinical characteristics of the validation cohort? That is, age, gender, type of epilepsy, comorbidities and AEDs.
► Describe the reference standard. Include the following:
    – Name of the reference standard used.
    – What type of reference standard it was, for example, clinical assessment, medical records, or validated disease registry.
    – Any rational given for choosing this reference standard (if alternatives exist).
    – What were the number, training and expertise of persons reading the reference standard?
    – If more than one person read the reference standard, what were the measures of consistency given? For example, kappa statistic.
► Describe any methods used to blind persons reading the reference standard to how the validation cohort

were coded diagnostically on the administrative database; that is, how a person reviewing the medical records diagnosis of an individual was made unaware of their administrative ICD diagnosis.

► Describe any methods used to blind persons reading the diagnostic administrative data codes to results of the reference standard diagnoses.

► What method was used to estimate the disease-identification accuracy of the administrative database (ie, sensitivity, specificity, PPV, NPV or AUC)?
  – What were the results of this?
    – That is, provide figures for these estimates and, where possible, also the individual TP, FP, TN and FN figures.
    – Include the results of any cross-tabulation of the administrative data diagnoses results against the results of the reference standard diagnoses.
    – Describe the methods used and results of the measures used to estimate variability or precision of the diagnostic accuracy results (eg, 95% CI).

► What method was used to compare the diagnostic accuracy of variables within the administrative database? That is, describe the method used to determine an optimal diagnostic algorithm.
  – What were the results of this?
  – Describe the methods used and results of the measures used to estimate variability or precision of the algorithm estimates (eg, 95% CI).

► How were indeterminate or missing administrative database diagnoses or reference standard results handled?

► What were the time interval and any clinical interventions given between reference standard diagnosis and administrative database diagnosis?

► Describe any adverse events found from using the administrative database or reference standard.

► Summarise the study limitations described by the authors including any sources of potential bias, statistical uncertainty, generalisability limitation, and what they described as implications for practice.

► What were the study's sources of funding?

Systematic review and meta-analysis outcomes
The primary outcome will be providing pooled disease-identification accuracy estimates of the included administrative databases using sensitivity, specificity, PPV, NPV and AUC as the measures of accuracy. This will answer research question 1: how accurately do administrative data identify epilepsy within a population? We will provide an overall estimate of the accuracy of the ICD-9 coding system and that of the ICD-10 coding system in correctly identifying epilepsy cases. This will be done by pooling together individual accuracy estimates from all included studies in which ICD-9 or ICD-10 were used provided there is no significant heterogeneity between studies (as measured using the I^2 statistic). The preferred estimators will be means with standard errors

(SEs) or medians with interquartile ranges (IQRs), dependent on distribution. We will also quote the 95% CIs for estimates. Where there is significant heterogeneity (I^2 statistic >50%), we will provide a descriptive analysis of the results and include ranges. Just as there may be heterogeneity introduced by making comparisons across different trial designs, we might expect there to be heterogeneity introduced by making comparisons across different healthcare systems. This is because there are likely to be differences in the accuracy of administrative healthcare data owing to differences in coding practice and/or the quality of reference standards between different healthcare systems. Therefore, we will also conduct subgroup analysis in which diagnostic accuracy results are pooled together for studies that have used the same or similar healthcare systems, for example, the National Health Service in the UK, Veterans Health Administration in the USA, healthcare systems with geographical overlap, and private-funded versus state-funded healthcare systems. We will also create subgroups in which results are pooled together within the following study design groups: prospective cohorts, retrospective cohorts, case–control studies, primary care studies, secondary care studies, tertiary care studies, paediatric studies (age <18 years), adult studies, and studies from the same country. Differences in the results for each subgroup may provide an important guide for future studies in the field, and they may also help to explain any statistical heterogeneity seen.

The secondary outcome will be the optimal administrative data algorithms for correctly identifying epilepsy within a population. For this, we will assign a dummy variable with a binary 0 = 'no' or 1 = 'yes' category to participants having the following:

A. A reference standard diagnosis of epilepsy (yes/no).
B. An administrative diagnosis code for epilepsy (yes/no).
C. Multiple administrative epilepsy diagnoses codes over time (yes/no).
D. Having previously had an EEG (yes/no).
E. Having previously had a computed tomography (CT) or magnetic resonance image (MRI) of the brain (yes/no).
F. Having previously had epilepsy surgery (yes/no);
G. Being on an individual AED (yes/no).
H. Being on two or more AEDs (yes/no).

Multivariable logistic regression models with A as the outcome variable and B–H individually and in combinations as the independent variables will be used in order to demonstrate the algorithm(s) best fitting the data across the included studies and to assess the significance of each variable's contribution to the model. The results of the logistic models will be displayed as sensitivity, specificity, PPV, NPV and AUC where possible, with 95% CIs and measures of interstudy heterogeneity provided using the I^2 statistic.

Table 2 Risk of bias and applicability judgements in QUADAS-2

Domain	Patient selection	Administrative database	Reference standard	Flow and timing
Description	Describe methods of patient selection: what is the review question?	Describe the administrative database and how it was used and interpreted: where available, include comment on how coding was done, by whom and whether there was reimbursement for coding.	Describe the reference standard and how it was conducted and interpreted: where available, include comment on quality of the reference standard, including the level of experience of clinicians making the diagnosis, access to diagnostic tests such as electroencephalography and telemetry and the thresholds/criteria used to make a diagnosis of epilepsy.	Describe any patients in the validation cohort who were not found within the reference standard or who were excluded from cross-tabulation of the administrative data diagnoses results against the results of the reference standard diagnoses: describe the time interval and any interventions between administrative database diagnosis and reference standard diagnosis.
Signalling questions (yes/no/unclear)	Was a consecutive or random sample of patients enrolled?	Were the administrative database diagnosis results interpreted without knowledge of the results of the reference standard diagnosis?	Is the reference standard likely to correctly classify the epilepsy?	Was there an appropriate interval between administrative database diagnosis and reference standard diagnosis?
	Was a case–control design avoided?	If a diagnostic threshold was used, was it prespecified?	Were the reference standard results interpreted without knowledge of the results of the administrative database diagnosis?	Did all patients receive a reference standard?
	Did the study avoid inappropriate exclusions?			Did all patients receive the same reference standard?
				Were all patients included in the analysis?
Risk of bias: high/low/unclear	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the administrative database have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability: high/low/unclear	Are there concerns that the included patients do not match the review question?	Are there concerns that the administrative database, its conduct or interpretation differ from the review question?	Are there concerns that epilepsy, as defined by the reference standard, does not match the review question?	

QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies 2.

Risk of bias analysis

We will use the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2)[27] tool to assess risk of bias within and across studies, modified for studies validating administrative data. This is summarised in the table 2 and will be completed independently by two review authors (GKM and KB) for each study, with disagreements resolved by mutual discussion and, where necessary, adjunction by a third reviewer (RFMC/SED/CRS). The tool consists of four key domains (see row 1) covering: (1) patient selection, (2) the administrative database, (3) the reference standard and (4) flow of patients through the study and timing of the administrative database and reference standard. Each domain is assessed in terms of the risk of bias (graded as high, low or unclear; see row 4) and the first three domains are also assessed in terms of concerns regarding applicability (see row 5). The description (see row 2) contains information used to support the risk of bias judgement. To help reach a judgement on the risk of bias, signalling questions are included (see row 3). These flag aspects of study design related to the potential for bias and aim to help reviewers make risk of bias judgements. If all signalling questions for a domain are answered 'yes', then risk of bias is judged 'low'. If any signalling question is answered 'no', then risk of bias is judged 'high'. If any signalling question is answered 'unclear', then risk of bias is judged 'unclear'. Applicability sections are structured in a similar way to the bias sections but do not include signalling questions. Review authors are asked

to record the information on which the judgement of applicability is made and then to rate their concern that the study does not match the review question.[27 28]

On completing the QUADAS-2 table, we will provide a risk of bias and applicability concerns graph demonstrating the review authors' judgements about each domain, presented as percentages across included studies. We will also provide a risk of bias and applicability concerns summary demonstrating review authors' judgements about each domain for each included study. We will use the Deek's test[29] to interrogate for publication bias. This test is specifically designed for detecting funnel plot asymmetry in reviews of diagnostic studies.[28]

Confidence in cumulative evidence

We will use the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach to assess strength of the body of evidence.[30] The GRADE system classifies the quality of evidence into one of four grades:

1. *High*: further research is very unlikely to change our confidence in the estimate of effect.
2. *Moderate*: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
3. *Low*: further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
4. *Very low*: any estimate of effect is very uncertain.[31]

A judgement is made on the individual studies used to provide the pooled effect estimates, and the quality of evidence is then downgraded by the cumulative presence of: (1) bias (see risk of bias analysis), (2) inconsistency (ie, heterogeneity present on I^2 statistic), (3) indirectness (ie, high concerns regarding applicability; see table 2), (4) imprecision (small sample sizes, wide CIs, and inadequately powered studies) and (5) publication bias (see Deek's test[29] comments).[32] GRADE classifications will be independently conducted by two review authors (GKM and KB) with any disagreements resolved by mutual discussion and, where necessary, adjunction by a third reviewer (RFMC/SED/CRS).

We will rate the completeness of reporting for each study out of 30 using the STARD 2015 checklist.[21] A score of 0–10, 11–20 and 21–30 will indicate a low, moderate and high quality of completeness of reporting, respectively.

PATIENT AND PUBLIC INVOLVEMENT

Patients and the public were not involved in development of the research question and outcome measures, nor the study design. The study does not involve patient recruitment, and patients were not involved in conduct of the study. We plan to liaise closely with patients, special interest groups, and charities in the dissemination of our results in printed and electronic media. Meta-data and information about the study will also be made available through our website (www.muirmaxwell-centre.com).

Ethics and dissemination

Ethical approval is not required as primary data will not be collected. Results will be disseminated in peer-reviewed journals, conference presentations, and in press releases. Meta-data and information about the study will also be made available through our website (www.muir-maxwellcentre.com) and via social media.

REFERENCES

1. Garratt E, Barnes H, Dibben C. *Health administrative data: exploring the potential for academic research*. St Andrews: Administrative Data Liaison Service, 2010.
2. Loke YK. Use of databases for clinical research. *Arch Dis Child* 2014;99:587–9.
3. St Germaine-Smith C, Metcalfe A, Pringsheim T, et al. Recommendations for optimal ICD codes to study neurologic conditions: a systematic review. *Neurology* 2012;79:1049–55.
4. Abraha I, Montedori A, Eusebi P, et al. The current state of validation of administrative healthcare databases in Italy: a systematic review. *Pharmacoepidemiology and Drug Safety* 2012;21:400–00.
5. Traversa G, Bianchi C, Da Cas R, et al. Cohort study of hepatotoxicity associated with nimesulide and other non-steroidal anti-inflammatory drugs. *BMJ* 2003;327:18–22.
6. Hottes TS, Skowronski DM, Hiebert B, et al. Influenza vaccine effectiveness in the elderly based on administrative databases: change in immunization habit as a marker for bias. *PLoS One* 2011;6:e22618.
7. Lalmohamed A, Vestergaard P, Cooper C, et al. Timing of stroke in patients undergoing total hip replacement and matched controls: a nationwide cohort study. *Stroke* 2012;43:3225–9.
8. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323–37.
9. Krarup LH, Boysen G, Janjua H, et al. Validity of stroke diagnoses in a national register of patients. *Neuroepidemiology* 2007;28:150–4.

10. Kee VR, Gilchrist B, Granner MA, *et al*. A systematic review of validated methods for identifying seizures, convulsions, or epilepsy using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:183–93.
11. Jetté N, Reid AY, Quan H, *et al*. How accurate is ICD coding for epilepsy? *Epilepsia* 2010;51:62–9.
12. Benson T. The history of the read codes: the inaugural James Read memorial lecture 2011. *Inform Prim Care* 2011;19:173–82.
13. Fonferko-Shadrach B, Lacey AS, White CP, *et al*. Validating epilepsy diagnoses in routinely collected data. *Seizure* 2017;52:195–8.
14. Tu K, Wang M, Jaakkimainen RL, *et al*. Assessing the validity of using administrative data to identify patients with epilepsy. *Epilepsia* 2014;55:335–43.
15. Franchi C, Giussani G, Messina P, *et al*. Validation of healthcare administrative data for the diagnosis of epilepsy. *J Epidemiol Community Health* 2013;67:1019–24.
16. Reid AY, St Germaine-Smith C, Liu M, *et al*. Development and validation of a case definition for epilepsy for use with administrative health data. *Epilepsy Res* 2012;102:173–9.
17. Parko K, Thurman DJ. Prevalence of epilepsy and seizures in the Navajo Nation 1998-2002. *Epilepsia* 2009;50:2180–5.
18. Pugh MJ, Van Cott AC, Cramer JA, *et al*. Trends in antiepileptic drug prescribing for older patients with new-onset epilepsy: 2000-2004. *Neurology* 2008;70(22 Pt 2):2171–8.
19. Christensen J, Vestergaard M, Olsen J, *et al*. Validation of epilepsy diagnoses in the Danish National Hospital Register. *Epilepsy Res* 2007;75(2-3):162–70.
20. Tan M, Wilson I, Braganza V, *et al*. Development and validation of an epidemiologic case definition of epilepsy for use with routinely collected Australian health data. *Epilepsy Behav* 2015;51:65–72.
21. Bossuyt PM, Reitsma JB, Bruns DE, *et al*. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Clin Chem* 2015;61:1446–52.
22. Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
23. Shamseer L, Moher D, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647.
24. Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
25. World Health Oraganisation. Classification of Diseases (ICD). 2016. www.who.int/classifications/icd/en/
26. Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin Biochem Rev* 2008;29(Suppl 1):S83–7.
27. Whiting PF, Rutjes AW, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
28. Leeflang MM, Deeks JJ, Takwoingi Y, *et al*. Cochrane diagnostic test accuracy reviews. *Syst Rev* 2013;2:82.
29. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;58:882–93.
30. Schünemann H, Broek J, Oxman A. *GRADE handbook for grading quality of evidence and strength of recommendation*. 3.2, 2009.
31. Mbizvo GK, Dixon P, Hutton JL, *et al*. Levetiracetam add-on for drug-resistant focal epilepsy: an updated Cochrane Review. *Cochrane Database Syst Rev* 2012;9:CD001901.
32. Singh S, Chang SM, Matchar DB, *et al*. Chapter 7: grading a body of evidence on diagnostic tests. *J Gen Intern Med* 2012;27(Suppl 1):47–55.