Shearer Jessica (Orcid ID: 0000-0001-5521-5322)
Parker Richard (Orcid ID: 0000-0003-4888-8670)
Tapper Elliot B (Orcid ID: 0000-0002-0839-1515)
Rowe Ian A (Orcid ID: 0000-0003-1288-0749)

TITLE PAGE

Systematic Review: Development of a consensus code set to identify cirrhosis in electronic health records

Running Head: Systematic Review: Code set for cirrhosis in EHR

Authors: Jessica Shearer[1,2], Juan J. Gonzalez[3], Thazin Min[1], Richard Parker[1], Rebecca Jones[1], Grace L. Su[3], Elliot B Tapper[3], Ian A. C. Rowe[1, 2]

[1] Leeds Liver Unit, Leeds Teaching Hospitals NHS Trust, Leeds, UK

[2] Leeds Institute for Data Analytics, University of Leeds, Leeds, UK

[3] Division of Gastroenterology and Hepatology, University of Michigan, Ann Arbor, MI 4809, United States

Correspondence: Dr Ian A. C. Rowe MBChB, MRCP (UK), PhD

Leeds Institute for Data Analytics

Room 6.1

Level 6

Clinical Sciences Building

St. James's University Hospital

Beckett Street

Leeds

United Kingdom

LS9 7TF

Email: i.a.c.rowe@leeds.ac.uk

Telephone: +44 (0)113 206 5667

Abstract

Background: Electronic health records (EHR) collate longitudinal data that can be used to facilitate large scale research in patients with cirrhosis. However, there is no consensus code set to define the presence of cirrhosis in EHR. This systematic review aims to evaluate the validity of diagnostic coding in cirrhosis and to synthesize a comprehensive set of ICD-10 codes for future EHR research.

Method: MEDLINE and EMBASE databases were searched for studies that used EHR to identify cirrhosis and cirrhosis-related complications. Validated code sets were summarised, and the performance characteristics were extracted. Citation analysis was done to inform development of a consensus code set. This was then validated in a cohort of patients.

Results: 1626 records were screened, and eighteen studies identified. The positive predictive value (PPV) was the most frequently reported statistical estimate and was ≥80% in 17/18 studies. Citation analyses showed continued variation in those used in contemporary research practice. Nine codes were identified as those most frequently used in the literature and these formed the consensus code set. This was validated in diverse patient populations from Europe and North America and showed high PPV (83-89%) and greater sensitivity for the identification of cirrhosis than the most often used code set in the recent literature.

Conclusion: There is variation in code sets used to identify cirrhosis in contemporary research practice. A consensus set has been developed and validated, showing improved performance, and is proposed to align EHR study designs in cirrhosis to facilitate international collaboration and comparisons.

Introduction

Cirrhosis is recognised as a growing public health burden, accounting for 1.3 million deaths worldwide each year [1]. The economic impact of cirrhosis is considerable with higher rates of unemployment, years of life lost and reduced quality of life [2].

The ability to identify large cohorts of patients with chronic liver disease can improve understanding of the natural history of cirrhosis and liver-related complications. Electronic health records (EHR) and administrative databases collate longitudinal data generated throughout the course of routine clinical care, often abstracted using diagnostic and procedure coding systems such as ICD-9 and ICD-10 [3]. This data is easily accessible and can provide comprehensive information regarding 'real-world' care patterns, costs, and outcomes [4-6].

The meaning and value of these data are directly related to both their validity and applicability to the population with cirrhosis. Several studies have evaluated the validity of diagnostic codes in identifying patients with cirrhosis. As there are many codes relating to liver disease and its complications there is variation amongst studies in terms of the codes used to define the presence of cirrhosis. The increasing importance of EHR based research and the role of real-world evidence in clinical decision making demands a critical appraisal of the tools used to identify cirrhosis.

The aim of this systematic review was therefore to evaluate the current evidence assessing the validity of diagnostic coding to identify cirrhosis using electronic health record databases. The review aims to synthesize and validate a comprehensive code set which can be used for future studies using EHR to study patients with cirrhosis by comparing definitions of cirrhosis based upon sets of existing diagnostic and procedural codes across studies and countries.

Methods

Data sources and search strategy

A search was completed using the OVID platforms of MEDLINE and EMBASE electronic bibliographic databases from inception (1946 and 1947 respectively) to March 2020 including 'In-Process' citations of all peer reviewed literature and conference abstracts. The full search strategy is included in the **Supplementary Table 1**. The search was limited to articles published in English and human studies, and the studies were de-duplicated prior to evaluation. To identify additional studies, bibliography lists were hand-searched. Once the search was completed, abstracts were screened for relevance and the identified studies were reviewed in full text and assessed for eligibility against the inclusion and exclusion criteria.

The systematic review protocol was prospectively registered with PROSPERO (International prospective register of systematic reviews) registration ID: CRD 42019118848. It was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines and checklist **(Supplementary Table 2).**

Study selection

Studies were evaluated for inclusion in two stages. In the first stage all identified titles and abstracts were screened. In the second stage relevant studies were retrieved and a full text review was done on all studies which met the predefined inclusion and exclusion criteria. We included all observational cohort and cross-sectional validation studies, which assessed the validity of diagnostic and procedural codes (ICD-9 and ICD-10) used to identify cirrhosis. Studies had to report the code set or algorithm employed to search the electronic database.

Inclusion and exclusion criteria

A study was included in the systematic review if it met the following predefined criteria: age >18 years, information regarding hospital admissions stored in electronic records as part of routine care, ICD-9 or ICD-10 codes explicitly defined and validated in medical record review. Studies using laboratory data to identify and define those patients with cirrhosis were excluded, as this data is not routinely available through EHR data alone. Where conference abstracts and full manuscripts of the same study are identified, data were extracted from the full manuscript.

Data extraction and quality assessment

The full text of each article was reviewed. Data were extracted, tabulated and summarised onto a standardised template. The information gathered included study author, year of publication and site, start date and duration of data collection, electronic data source, sample size, ICD codes or algorithm employed.

If statistical estimates were not reported in the original study, estimates were calculated from the available data. This included sensitivity, specificity, positive predictive value, negative predictive value, and kappa value (a measure of agreement beyond that expected by chance). As there is no validated quality assessment tool for non-comparator retrospective studies, we used an adaptation of the QUADAS-2 tool (Quality Assessment of Diagnostic Accuracy Studies) to evaluate the quality of the included studies [7].

Data synthesis and citation analysis

Data was synthesised qualitatively, with the authors reviewing the data extraction table and then re-reviewing the relevant articles. Citation analysis was conducted using the web resource Scopus to assess the impact, geographical reach, and applicability of the studies. This analysis was conducted in September 2020. Abstracts were excluded and only those studies in which the primary objective was validation of codes within liver disease were included, as it was felt that this would be a more accurate reflection of the impact and use of these validated code sets. Only those studies published at least five years ago were included, and citations were analysed per publication year.

Validation of the consensus code set

ICD-9 codes were converted to the closest possible ICD-10 equivalent and the most common codes and definitions used across all studies were identified and considered for inclusion in the consensus code set. The consensus code set was validated using four independent cohorts. To determine sensitivity, a cohort of 300 patients (UK cohort [sensitivity]) from a secondary / tertiary care centre at the Leeds Liver Unit, United Kingdom with advanced chronic liver disease and median liver stiffness as measured by transient elastography of ≥15kPa between 2012 to 2017. Only patients with codes occurring after transient elastography were included in the primary analysis and out-patient codes were not used. In a sensitivity analysis, patients with decompensation before transient elastography (n = 33) were also included to describe the sensitivity of the consensus code set. Second, we evaluated a cohort of 113 patients seen at the University of Michigan Hepatology Clinic (US cohort [sensitivity]) who were enrolled prospectively in a chronic disease monitoring system between 2010-2015 and followed for at least 3 years. As described elsewhere [8] all patients had a CT scan within 365 days of enrolment and received their diagnosis of cirrhosis based on imaging, laboratory and/or histological

parameters from a board-certified transplant hepatologist and were followed clinically thereafter. All diagnosis codes were entered in or mapped to ICD-10 in the electronic medical record. In each case the full medical record was reviewed. Basic demographic information was extracted and all events following the identification of fibrosis were recorded. This included out-patient visits in the hepatology clinic and admissions to hospital with decompensation (variceal bleeding, ascites, and hepatic encephalopathy). Data held within the EHR was extracted and coded information relating to hospital admissions, investigations, and procedures was collected.

Following this we evaluated the positive predictive value (PPV) of the consensus code set. We evalulated PPV in a separate cohort because the above described cohort did not include patients without cirrhosis, making it impossible to assess PPV. Firstly, a cohort of 335 patients admitted to Leeds Teaching Hospital NHS Trust (UK cohort [PPV]), United Kingdom in 2019 with one or more codes from the consensus code set. Two experienced clinicians (JS and TM) independently reviewed the medical record to confirm if the diagnosis of cirrhosis was correct. A positive diagnosis of cirrhosis was made following review on one or more of the following criteria: histological confirmation of cirrhosis, portal hypertension on imaging (varices/ascites), documentation in medical record by a Specialist Gastroenterologist or Hepatologist of an episode of decompensation (ascites, variceal bleeding, hepatic encephalopathy), or synthetic dysfunction consistent with cirrhosis (Albumin ≤30, Bilirubin ≥20, INR ≥1.2). Additionally, we evaluated PPV in 241 patients identified by any one or more of the codes in the consensus code set with an outpatient encounter in May or June 2021 at the University of Michigan (US cohort [PPV]). The full medical record was reviewed by an experienced clinician (EBT) to determine if the patient had a confirmed diagnosis of cirrhosis, based on the criteria outlined above.

Results

Study characteristics

A total of 1975 abstracts were identified. After de-duplication 1626 abstracts remained. 138 studies were reviewed in full text. A further twenty-nine studies were identified and reviewed through hand-searching of reference lists. Of the discounted records, sixty-six were conference abstracts, which did not contain sufficient information for analysis. Overall, eighteen studies met the inclusion criteria and were included in the final qualitative analysis. No additional suitable studies were identified through hand searching of bibliographies. A flowchart showing the number of studies screened and included is shown in **Figure 1**. The studies and a description of their characteristics are shown in **Table 1**.

The sample size ranged between 84 to 6714 people, with a total of 18,704 patients included. Twelve studies were conducted in the United States [9-20], two in Denmark [21,22], two in Canada [23,24], and two in the United Kingdom [25,26]. Of those studies from the United States, five used cohorts from the Veterans Administration (VA) population [9,12,14,18,27]. In two studies the evaluation was carried out in a single hospital setting [15,19].

Seventeen of the studies used medical record review to validate the diagnosis of cirrhosis [9-19,21-24,26,27]. In these studies, the full medical record was retrieved and compared with the diagnostic codes of interest. Amongst the seventeen studies, thirteen outlined an explicit definition of their primary outcome measure [9-11,13-15,17-19,21,23,25,26]. All of these included histological and/or radiological evidence of liver disease and five also included specific laboratory parameters [9,13,17,18,21]. One study searched primary and secondary care records and death registry data for codes or free-text terms relating to cirrhosis as their validation standard [25].

Ten studies evaluated codes using electronic health records [10,13-16,18,19,25-27] and seven used administrative databases [9,11,12,17,21,23,24], the majority of which reported on in-patient and out-patient data. One study used a national registry database [22]. Validation was the primary outcome measure in fourteen studies [9-16,18,19,22-24,26]. Two of these studies focussed on validation of the comorbidity variables which constitute the Charlson index, of which liver disease was extracted separately [22,24]. Seven of the validation studies analysed disease severity i.e. codes representing decompensation events in addition to cirrhosis codes [11,13-15,18,23,26]. One study validated an algorithm using ICD codes with and without the addition of a natural language processing algorithm [10]. A description of the validation standards is included in **Table 1**.

Study quality

Study quality was assessed using an adapted QUADAS tool [7]. A detailed copy of the tool and a breakdown of individual scores for each study are included in **Supplementary Table 3 and 4**. The QUADAS scores ranged from 7 to 11 with a maximum of 14 (median 10). Three studies used a selected population of patients; patients enrolled in the chronic hepatitis cohort study [13,16] and patients with an ICD-10 code for hepatocellular carcinoma [26]. Two studies did not adequately describe their selection criteria in detail [22,24]. Three studies used a random selection from their total sample to verify as a gold standard comparison [10,17,20]. Seven studies stated that the individual abstracting data from the medical record was blinded to the database coding [9,12,14,19,20,23,24], whilst the rest did not specify. Seven studies used a single clinician to conduct chart review [11,12,14,15,17,19,24] the remaining ten studies used more than one clinician often in addition to an arbitrator.

Quality of coding sets

Details of the type and number of codes used are shown in **Table 2**. Ten studies used ICD-9 codes [9-12,15,17-19,24,27], three used ICD-10 codes [14,21,22] and the remaining five used a combination of ICD-9 and/or ICD-10 codes combined with procedural codes [13,16,23,25,26]. Aside from one study, which specified that it used only the primary diagnostic code [22], none of the remaining studies commented upon whether the code was the designated primary diagnosis code or one of the subsequent twenty diagnosis codes which can be associated with an in-patient or out-patient encounter. It is therefore assumed that the codes of interest occurred at any position.

Fifteen studies reported specific ICD codes used to define liver disease in their cohort [9-15,17,18,21-23,25-27]. The remaining three studies [16,19,24] did not specify the codes however it was possible to obtain the information from other related studies [28-30]. Seven studies adopted ICD code sets which had previously been used and validated by other authors [12,13,16,22,24-26], whilst eleven studies developed their own selection of codes [9-11,14,15,17-19,21,23,27]. Quan *et al* used a coding algorithm developed previously by Deyo *et al* [28], which included fourteen ICD-9 codes in total. The 'mild' liver disease category included 3 codes for cirrhosis, and this was therefore combined with the codes for 'moderate or severe' liver disease. Thygesen *et al* used a larger number of codes to define 'mild' liver disease which included codes we considered to be less specific for cirrhosis (K71; K74; K76.0) [22]. For this reason, we included only the coding algorithm which was employed for 'moderate or severe liver disease'.

There was significant variation in the number and type of codes used. Overall, there were a total of sixty-three ICD-9 codes and fifty-four ICD-10 codes as well as seventy-seven procedural codes used to identify cirrhosis in the included studies (**Supplementary Tables 5 - 8**). Of those studies using the ICD-10 classification, this included codes from five disease

manifestation categories (B15.0-94.2; C22; E80-E84.5; I81-I98.3; K22-K92.2) and two symptom-related and external causation categories (R16-R18.8; T86). Three ICD-9 and four ICD-10 codes appeared as clustered codes denoting that all the sub-codes were used. Five studies incorporated procedural codes into their code sets. In one study the specific procedural codes were unavailable [16]. In the remaining four studies the number of procedural codes used ranged between 7 and 60 [13,23,25,26]. Whilst there were similarities between some of the code sets used, none of the studies used the same codes from the same ICD dictionary.

Assessment of validation in the literature

The validation statistics are shown in **Table 3**. The positive predictive value (PPV) was available in all but one study [23] and was >90% in ten studies with a range of 71-100% [9-11,14,18,19,22,25-27]. Negative predictive value (NPV) was reported in seven studies [9,10,15,16,18,24,26] with a range of 72-99%. Nine studies reported sensitivity and/or specificity values [10,13,15-18,23,24,26], the range for which were 20-98% and 43-99% respectively. Kappa values were reported in only four studies and the values ranged from 0.48-0.71 [9,15,18,24]. Of the ten studies which reported a PPV of >90%, six of these included codes taken from both the in-patient and out-patient setting (**Table 3**).

The median number of codes used was thirteen. There was no improvement in the statistical estimates in those studies that used more codes within their definition (≤13 codes PPV range 71-100%; >13 codes PPV range 71-91%). However, four studies which validated diagnostic codes found that combinations of codes improved sensitivity in comparison to a single code [11,15,18,23]. There was no difference in the range of PPV between studies using ICD-9 codes (71-95%) and those using ICD-10 codes (71-100%). There was also no discernible difference in PPV depending upon the type of database from which coded information was extracted

(administrative database 71-94%; electronic health record 71-99%). The study which used the Danish national registry reported PPV of 100%, although only 50 patient records were reviewed. We observed an increase in the minimal value of the PPV range in the five studies conducted in the VA population (89-93%) in comparison to the remaining studies (71-100%). The eighteen studies included were published over a 17-year period (2002-2019). The range of time for data collection varied widely from 1 to 14 years with a median length of 4 years and four of the studies collected data from over ten years [11,12,21,25]. None of the studies commented upon any longitudinal changes in statistical estimates during the study collection period. It was noted that there was no difference in the trend in PPV in later years compared to earlier years; in the six earliest studies published between 2002-2012 [9,11,12,17,18,20-22,24,30], the PPV ranged between 71-100% whilst in the most recent studies published between 2013-2018 the PPV ranged between 71-99% [10,13-16,23,25,26].

Citation Analysis

We conducted citation analysis focussing on those manuscripts cited most frequently over the last 3 years. The total number of citations per study, mean number of citations per year over that period and the field-weighted citation impact (FWCI), which compare how a frequently a document is cited in comparison to similar documents (values greater than 1.00 indicate that a publication is cited more than expected according to the average) [31], are shown in **Table 4**. Over that period, the code set most frequently cited was from Kramer *et al*, but those from Nehra *et al*, and Goldberg *et al* were also often reported [9,11,15]. This use of different code sets between studies highlights the need for a consensus approach to EHR research in the identification of patients with cirrhosis.

Consensus code set synthesis

The most common codes used across all studies (**Table 5**) were considered for inclusion in the consensus code set (**Table 6**). The most frequently used codes were (when mapped to ICD-10) K70.3 – alcoholic cirrhosis, and K74.6 – other / unspecified cirrhosis. Other commonly used codes related to complications of cirrhosis and portal hypertension, including the presence of oesophageal varices and ascites. Since ascites can occur in conditions unrelated to liver disease (e.g., cardiac or renal failure, or intra-abdominal malignancy) we considered this code to be of low specificity and it was excluded from the proposed consensus code set to evaluate for future use. This is supported in previous studies [15,32], which have found that using the code for ascites alone rather than in combination with other codes for chronic liver disease yields a PPV between 43-63%.

Validation of code set

We used two independent samples to validate the sensitivity of the consensus code set. In the UK and US cohorts (sensitivity), a result was positive if the EHR contained one or more of these codes, either as an in-patient or out-patient where available. This was compared to the code set used most frequently from Kramer and colleagues [9]. In the UK cohort (sensitivity) 300 patients were included. 63% were male, the mean age at time of diagnosis was 55 years, and the majority had either non-alcoholic fatty liver disease or alcohol associated liver disease. In the US cohort (sensitivity), 113 patients were included. The mean age was 64 years, 59% were male, and the commonest liver disease aetiology was hepatitis C virus infection. Further details are included in **Supplementary Table 9.**

The sensitivity for individual codes within the consensus code set was low (**Table 5**). There were three codes (K74.4, K74.5 and K72.1) which did not appear within either the UK or US

cohorts (sensitivity). Given the additional benefit gained from including these codes was likely to be negligible, these were subsequently excluded from the proposed consensus code set (**Table 6).**

The final consensus code set improved the sensitivity in the UK cohort from 44% using the Kramer *et al* code set to 61% using the consensus code set (p < 0.0001, McNemar's test). The consensus code set was further evaluated in the subset of the UK validation cohort using different liver stiffness measurements (LSM) to define cirrhosis. When using a threshold of >20kPa rather than >15kPa, the sensitivity for the detection of cirrhosis was improved from 61% to 68% in 227 patients. If the threshold was raised to >25kPa LSM the sensitivity improved to 74% in 156 individuals. In comparison to the Kramer *et al* codes the sensitivity was 51% and 58% for patients with a liver stiffness measurement of >20kPa and >25kPa respectively. Sensitivity in the US cohort was also improved from 89% to 100% (p = 0.0015, McNemar's test) highlighting the utility of the consensus code set in diverse patient populations.

To understand whether relevant information was lost by excluding the term for ascites, we repeated the analyses including this code. In these analyses the sensitivity was not significantly changed; in the UK cohort the sensitivity was 60%, whilst in the US dataset sensitivity was maintained at 100%. To determine if the inclusion of patients with evidence of prior decompensation altered the performance characteristics, we reviewed the medical record of an additional 33 patients with decompensation events prior to index transient elastography. Twenty-three of these patients would have been subsequently identified by the consensus code set as being cirrhotic. When combined with the UK cohort the overall sensitivity was unchanged at 61% (204/333 patients correctly identified).

We used two further independent samples to validate the positive predictive value of the code set. In the UK cohort (PPV), 335 patients were included. In the US cohort (PPV), 241 patients were included, and in both cohorts alcohol-related liver disease was the most common underling aetiology. Additional clinical information is included in **Supplementary Table 9.** Of the 335 patients in the UK cohort, 278 patients had cirrhosis confirmed in the medical records, giving a PPV of 83%. In the US cohort 214 out of 241 patients had a confirmed diagnosis of cirrhosis, equating to a PPV of 89%.

Discussion

Accurate assessments of the population burden and the impact of cirrhosis in EHR research depend on the performance and validity of the coding algorithms used to identify cases. The aim of this study was to synthesise and validate an approach that can be used to facilitate future research to improve the applicability of EHR research findings internationally. We found that there was substantial variation in the codes used to define cirrhosis. We extracted the most frequently used and relevant codes and combined them into a consensus code set (**Table 6**), with a positive result indicated by the presence of one or more of the included codes in in-patient or out-patient records. This code set was validated in two diverse patient populations from Europe and North America. In contrast to the most frequently used code set for cirrhosis, we found that our consensus code set improved sensitivity for the identification of cirrhosis with maintained high PPV. It is intended that this code set is used in future EHR research, where cirrhosis is defined by the presence of one or more of the codes in the set in the inpatient or outpatient setting. The code set will enable researchers to collaborate internationally and compare diverse populations of patients with cirrhosis using EHR data.

The purpose and context of diagnostic coding

The increasing importance of EHR based research and the role of real-world evidence in clinical decision making demands a critical appraisal of the tools used to identify cirrhosis in such studies. When reviewing the literature to determine the validity of diagnostic coding one must consider the study purpose, location, and the data source from which the codes were extracted. The provision of healthcare and the databases in use vary considerably worldwide, and in developed countries the most important factor to consider is the role of medical billing. In the UK and most Scandinavian countries healthcare is financed through tax payments. European countries such as Germany and France use insurance systems and Canada employs

a government led publicly funded model, with the option of privately paid insurance as a supplement. In the United States there are numerous systems in place, the majority of which rely upon medical billing and coding. Administrative and physicians claims databases were developed primarily for the purpose of billing and financial re-payment. Whilst the accuracy of these databases in identifying diseases has been widely reported upon [33-35] how accurately these findings translate to those countries where databases and healthcare systems differ, and medical billing does not exist remains unclear.

The need for a consensus code set

We identified important differences in the sensitivity between our validation cohorts. This highlights the challenges in translating coding approaches derived from one dataset to another and the importance of reporting validation from different settings when these approaches are being developed and used. The lack of OP codes in the UK validation cohort likely impacted on the comparatively low sensitivity. Whilst diagnosis and procedural codes are included in the Hospital Episodes Statistics (HES) OP dictionary they are not frequently included alongside OP attendances[36], and this has been highlighted as an important area of improvement for studies using HES-derived datasets[37]. Where available, both IP and OP codes should be used.

The most widely used coding algorithm within the literature to date is adopted from Kramer *et al* [9]. The VA system differs from the rest of healthcare provided in the US, both in terms of structure, funding and demographically. The vast majority of VA patients with cirrhosis are middle-aged males with a higher prevalence of hepatitis C and comorbidities than the general population [38,39]. Despite this, more than half of studies citing the Kramer code set were from outside the VA system suggesting wide adoption of these codes for EHR research particularly in the US. However, to facilitate international collaboration and comparison a consensus code

set that is better able to identify cirrhosis has several advantages and indeed these have gained traction in several other disease areas [40-42].

Assessing code set performance

There was variation in the measures of performance of the various code sets reported. Most frequently the positive predictive value was reported, and this was often related to the study design, in which the medical records reviewed were already selected to enrich for the presence of cirrhosis. Several factors can improve the sensitivity of code sets, recognising that there is a balance to be found between sensitivity and PPV. Increasing numbers of codes used, codes from both the inpatient and outpatient setting, and codes that encompass the whole range of cirrhosis complications all yield improvements in the sensitivity of the described code sets. This increase in sensitivity however must be considered in the light of any reductions in the PPV. For example, Nehra and colleagues reported that the inclusion of multiple codes relating to liver decompensation, except for ascites, maximised the sensitivity for the detection of cirrhosis with an acceptable PPV (78.0%). Additionally, they found that almost 5% of patients with cirrhosis had a code for a complication of cirrhosis without a specific cirrhosis code, supporting their inclusion within a code set [15]. The consensus code set incorporates each of these aspects in response to the observations made during the review.

Limitations

There are several limitations to this systematic review. First, the studies reported were often of small validation sets from single institutions without external validation with inherent bias in the assessment of the presence of cirrhosis in the medical chart review. Second, the weight of importance of the individual codes analysed in the primary reports was seldom reported meaning that a quantitative analysis was not possible to define the codes carrying the most

information in the EHR and how this varied between studies. Third, developing a consensus code set that can be used across all healthcare systems is a challenge and we recognise that no two systems are the same. Fourth, validation using chart review has inherent limitations with the potential for misclassification though the extraction was done blind to the code set evaluation. The approaches taken in the qualitative synthesis recognise these limitations and validation in four diverse patient populations addresses, to some extent, issues regarding the validity of the consensus code set across healthcare systems. Fifth, it is recognised that the sensitivity in the UK cohort was comparatively low at 60%. This was in part owing to the population, which comprised of patients who had undergone transient elastography in the out-patient setting, and due to the lack of out-patient coded data meaning a proportion of patients did not have any coded information that could be used. Sixth, as the patients in the assessment of PPV were identified using the consensus code set we were unable to assess its specificity or negative predictive value since no code set negative cases were identified to enter the cohort. This is also a limitation to the description of existing code sets where these measures are infrequently reported. The potential impact of the uncertainty regarding the specificity of the consensus code set should be considered in the design of EHR-based studies. Finally, as the validation was conducted in two tertiary care systems, further evaluation of the performance of the consensus code set in other healthcare systems would be appropriate.

Conclusions

A large number of diagnostic codes and combinations of these codes have been proposed to define cirrhosis in EHR research. In this systematic review we have defined a consensus code list of nine codes that increase sensitivity for the identification of cirrhosis in patients from both Europe and the US with maintained high positive predictive values. This consensus code set is proposed to align EHR study designs in cirrhosis to facilitate international collaboration and comparisons.

Author (year)	Country	Study years	Source population	Type of database	Sample size	Records validated	Definition of validation	Validator
Quan et al [24] (2002)	Canada	1996-1997	Patients admitted to one of three hospitals within the Calgary Regional Health Authority	AD	1200	1200	Details not given in study	1 clinician
Hachem et al [12] (2008)	US	1995-2005	Veterans registered at VA medical clinics in Houston, Texas	AD	84	84	Pathology +/- radiology +/- evidence in medical records	1 clinician
Kramer et al [9] (2008)	US	1998-2004	Veterans registered at VA medical clinics in Houston, Texas	AD	331	331	Stage 4 cirrhosis on liver biopsy or ≥2 of cirrhosis, ascites/peritonitis, varices, HCC, HRS, HE on imaging (CT/MRI/USS) or in notes or ≥2 albumin <30g/L, bilirubin>2.0mg/dL, INR >1.2 (or 1 of laboratory parameters with one of above)	1 clinician, 20% by 2nd clinician, 10% by 3rd clinician
Re et al [18] (2011)	US	2005	Patients enrolled in the Veterans Aging Cohort Study	EHR	137	137	Radiological evidence of ascites (CT/MRI/USS) or evidence of peritoneal fluid analysis +/- polymorphonuclear leukocyte count ≥250 cells/mL or bacterascites or bleeding varices on endoscopy report or documentation of mental confusion in absence of non-hepatic causes or diagnosis of HCC on biopsy or radiology (CT/MRI)	1 non-clinician, results reviewed by 2 clinicians
Thygesen et al [22] (2011)	Denmark	1998-2007	Patients registered in the Danish National Registry in the North Jutland Region, Denmark	NR	950	50	Discharge summary/medical record describing exact diagnosis	1 clinician, 1 arbitrator
Singal et al [19] (2011)	US	2008-2009	Patients admitted to one hospital in Dallas County	EHR	1589	1589	Consistent histology +/- cirrhotic-appearing liver on imaging with evidence portal hypertension (ascites, HE, varices or splenomegaly with thrombocytopenia) †	1 clinician
Goldberg et al [11] (2012)	US	1997-2011	Patients receiving IP or OP care at two tertiary care hospitals in Pennsylvania	AD	266	244	Liver biopsy demonstrating cirrhosis or radiological evidence of cirrhosis (CT/MRI/USS), or documentation of cirrhosis based on biopsy/radiology	1 clinician
Kanwal et al [27] (2012)	US	2000-2007	Patients receiving IP or OP care at 3 VA medical centres and 15 clinics in the Midwest	EHR	774	300	Documentation, laboratory or radiological evidence of ascites, HE, in-patient GI bleeding, paracentesis or SBP	1 clinician, 10% by 2nd clinician
Rakoski et al [17] (2012)	US	2008	Patients enrolled in the national Health and Retirement Study & receiving care at University of Michigan	AD	317	100	Liver biopsy demonstrating cirrhosis or radiological evidence of cirrhotic liver with splenomegaly + platelet count of <120,000mm/^{3} or evidence of decompensated cirrhosis with HE, HRS, ascites or variceal bleeding	1 clinician

Fialla et al [21] (2012)	Denmark	1996-2006	Patients enrolled in the Funen Patient Administrative System registry in Denmark	AD	1369	1369	Consistent histology cirrhosis or evidence of portal hypertension with hepatic wedge pressure of >8mmHg or INR >1.5 or cirrhotic liver on USS or perioperatively or evidence of complications such as varices, ascites +/- HE	N/A
Rabin et al [16] (2013)	US	2013	Patients enrolled in the Chronic Hepatitis Cohort Study in Detroit, Michigan*	EHR	283	283	Radiology, laboratory parameters, biopsy and clinical events	2 clinicians, 1 arbitrator
Nehra et al [15] (2013)	US	2008-2011	Patients receiving IP or OP care at one hospital in Dallas County	EHR	2893	2893	Stage 4 cirrhosis on liver biopsy or radiological evidence of cirrhosis + evidence of portal hypertension on imaging or clinical evidence of portal hypertension/complications (ascites, varices, HE, HCC)	1 clinician
Ratib et al [25] (2014)	England	1998-2009	Patients enrolled in primary and secondary registries in England	EHR	5118	2282	Search of primary and secondary care records and ONS death registry data for codes related to liver disease + examination of FTD for any of the following terms: "cirrhosis", "ascites", "varices", "liver", "portal hypertension", "hepatic", "jaundice" or "paracentesis"	N/A
Chang et al [10] (2016)	US	2013-2015	Patients receiving IP or OP care at 4 hospitals in Los Angeles	EHR	5343	168	Stage 4 cirrhosis on liver biopsy, radiological evidence of cirrhosis (CT/MRI/USS) or documented clinical diagnosis	1 clinician, 1 non-clinician
Lu et al [13](2017)	US	2015-2016	Patients enrolled in the Chronic Hepatitis Cohort Study in Detroit, Michigan	EHR	296	296	Documented evidence of HE or GI bleeding due to portal hypertension or jaundice with bilirubin >2.5mg/dL or ascites/hydrothorax due to portal hypertension, or HCC	2 clinicians, 1 arbitrator
Mapakshi et al [14] (2018)	US	2015-2016	Patients with data stored within the VA Corporate Data Warehouse	EHR	325	325	Stage 4 cirrhosis on liver biopsy or documentation of cirrhosis or complications in medical record, radiological or endoscopic evidence of cirrhosis	1 clinician
Lapointe-Shaw et al [23] (2018)	Canada	2006-2013	Patients receiving IP or OP care at two tertiary care hospitals in Ontario, Canada	AD	6714	6714	Stage 4 cirrhosis on liver biopsy or cirrhotic appearance on USS, non-invasive test result consistent with F4 fibrosis or evidence in clinical record of ascites, bleeding varices, encephalopathy, use of spironolactone or nadolol without alternative indication or explicit mention of cirrhosis/decompensation/non-bleeding varices	2 clinicians, 1 arbitrator, 5% by 2nd clinician
Driver et al [26](2019)	UK	2007-2016	Patients diagnosed with hepatocellular carcinoma in two NHS cancer centres in England	EHR	339	339	Documentation of cirrhosis in MR or MDT minutes, radiological/endoscopic evidence of portal hypertension, cirrhosis on liver biopsy, consistent TE result	3 clinicians

Author (year)	Codes used	Case definition	No. of codes
Quan et al [24] (2002)	ICD-9	≥ 1 code (IP only)	14
Hachem et al [12] (2008)	ICD-9	≥ 1 code (IP or OP)	2†
Kramer et al [9] (2008)	ICD-9	≥ 1 code (IP or OP)	3
Re et al [18] (2011)	ICD-9	1 IP+2 OP codes	22†
Thygesen et al [22] (2011)	ICD-10	1st listed code (IP or OP)	11
Singal et al [19] (2011)	ICD-9	≥ 3 codes	11‡
Goldberg et al [11] (2012)	ICD-9	≥ 2 codes (IP or OP)	58†
Kanwal et al [27] (2012)	ICD-9	≥ 2 codes (IP or OP)	12
Rakoski et al [17] (2012)	ICD-9	≥ 1 code (IP or OP)	12†
Fialla et al [21] (2012)	ICD-10	≥ 1 code (IP or OP)	4
Rabin et al [16] (2013)	ICD-9 +CPT	≥ 1 code	41
Nehra et al [15] (2013)	ICD-9	≥ 1 code (IP or OP)	11
Ratib et al [25] (2014)	ICD-10 + OPCS4	≥ 1 code	21
Chang et al [10] (2016)	ICD-9	≥ 1 code (IP or OP)	16
Lu et al [13] (2017)	ICD-9/10 + CPT	≥ 1 code (IP or OP)	43
Mapakshi et al [14] (2018)	ICD-10	≥ 1 code (IP or OP)	7
Lapointe-Shaw et al [23] (2018)	ICD-9/10+ CCP	≥ 1 code (IP or OP)	40
Driver et al [26] (2019)	ICD-10- +OPCS4	≥ 1 code (IP only)	33

Table 2: Details of code dictionary and number of codes used in each study.
ICD; international classification of diseases, CPT; current procedural terminology, ONS; office for national statistics, CCP; Canadian classification of diagnostic, therapeutic and surgical procedures
† Information not in original abstract-deduced from subsequent paper (30)
‡ Paper uses ICD-9-CM (clinical modification) classification

Author (year)	Se (%)	Sp (%)	PPV (%)	NPV (%)	Kappa (κ)
Quan et al [24] (2002)	72	99	80	99†	0.75
Hachem et al [12] (2008)	-	-	89	-	-
Kramer et al [9] (2008)	-	-	90	87†	0.70
Re et al [18] (2011)	20‡	99‡	91	99‡	0.48‡
Thygesen et al [22] (2011)	-	-	100	-	-
Singal et al [19] (2011)	-	-	95	-	-
Goldberg et al [11] (2012)	-	-	94	-	-
Kanwal et al [27] (2012)	-	-	91	-	-
Rakoski et al [17] (2012)	67	-	88	-	-
Fialla et al [21] (2012)	-	-	71	-	-
Rabin et al [16] (2013)	91	72	71	91†	-
Nehra et al [15] (2013) ‡	98	43§	78	91§	0.71
Ratib et al [25] (2014)	-	-	90	-	-
Chang et al [10] (2016)	47	97	92	72†	-
Lu et al [13] (2017) ¶	83	89	85	-	-
Mapakshi et al [14] (2018)	-	-	93	-	-
Lapointe-Shaw et al [23] (2018) ††	67-82	77-90	-	-	-
Driver et al [26] (2019)	86	98	99	79†	-

Table 3: Performance characteristics of each study
Se; Sensitivity, Sp; Specificity, PPV; positive predictive value, NPV; negative predictive value
† NPV defined as probability that cirrhosis was absent amongst those patients without a code
‡Estimated performance statistics using random sample of 100 patients without codes/hepatic decompensation
§ Authors validated sensitivity using cohort of 285 patients prospectively determined to have cirrhosis. NPV validated using 116 patients with liver disease but no codes for cirrhosis
¶ Paper uses a specific combination of codes to achieve these performance characteristics
 Range given as results separated into 3 separate cohorts

Author (year)	Total number of citations	Number of citations within last 3 years (2018, 2019, 2020)	Field-weighted citation impact	Mean number of citations per year
Kramer et al [9] (2008)	166	56 (18, 21, 17)	2.67	12.8
Re et al [18] (2011)	76	29 (10, 7 12)	1.87	8.4
Goldberg et al [11] (2012)	77	46 (8, 15 23)	1.45	9.6
Nehra et al [15] (2013)	86	46 (8, 20 18)	2.97	10.3

Table 4. Details of citation analysis
Total number of citations since publication is shown alongside the number of citations within the most recent three years.

Author Manuscript

ICD-9 code	ICD-10 code	Description (ICD-10 version)	Number of authors using code	Sensitivity of individual codes in validation group (total 413 patients), sensitivity (n)
571.5	K74.6	Other and unspecified cirrhosis of the liver	16	43% (177)
571.2	K70.3	Alcoholic cirrhosis of the liver	16	18 % (74)
456 -456.0 -456.1 -456.2 -456.21 -456.20	I85 -I85.0 -I85.9 I98 -I98.2 -I98.3	Oesophageal varices -with bleeding -without bleeding Oesophageal varices in diseases classified elsewhere -without bleeding -with bleeding	14	24% (99)
572.3	K76.6	Portal hypertension	13	37% (153)
572.2	K72.9	Hepatic failure, unspecified	12	7% (29)
572.4	K76.7	Hepatorenal syndrome	9	1% (4)
571.6	K74.4 K74.5	Secondary biliary cirrhosis Biliary cirrhosis, unspecified	9	0
572.8	K72.1	Chronic hepatic failure	8	0
789.5	R18.0	Ascites	8	14% (58)

Table 5: Most common codes used to identify cirrhosis with sensitivity for the prediction of cirrhosis in combined UK and US cohorts (sensitivity).
Approximate conversions from ICD-9 to ICD-10 dictionary have been used to determine the most appropriate code(s). The number of authors using the code includes those papers which used the code in either ICD-9 or ICD-10 format. In the sensitivity calculation an individual patient can have multiple codes contributing to the identification of cirrhosis.

ICD-10 code	Description
K74.6	Other and unspecified cirrhosis of the liver
K70.3	Alcoholic cirrhosis of the liver
I85 *-I85.0* *-I85.9* I98 *-I98.2* *-I98.3*	Oesophageal varices *-with bleeding* *-without bleeding* Oesophageal varices in diseases classified elsewhere *-without bleeding* *-with bleeding*
K76.6	Portal hypertension
K72.9	Hepatic failure, unspecified
K76.7	Hepatorenal syndrome

Table 6: Consensus code set
Final code set used to define cirrhosis in electronic health records

References

1. Mortality GBD, Causes of Death C. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet.* 2016;388(10053):1459-1544.

2. Stepanova M, De Avila L, Afendy M, et al. Direct and Indirect Economic Burden of Chronic Liver Disease in the United States. *Clin Gastroenterol Hepatol.* 2017;15(5):759-766 e755.

3. World Health Organisation. http://apps.who.int/classifications/icd10/browse/2010/en (Accessed 24th September 2018). Accessed.

4. Ratib S, West J, Fleming KM. Liver cirrhosis in England - An observational study: Are we measuring its burden occurrence correctly? *BMJ Open.* 2017;7 (7) (no pagination)(e013752).

5. Jepsen P, Vilstrup H, Sorensen HT. Alcoholic cirrhosis in Denmark - population-based incidence, prevalence, and hospitalization rates between 1988 and 2005: a descriptive cohort study. *BMC Gastroenterol.* 2008;8:3.

6. Leon DA, McCambridge J. Liver cirrhosis mortality rates in Britain from 1950 to 2002: an analysis of routine data. *Lancet.* 2006;367(9504):52-56.

7. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003;3:25.

8. Tapper EB, Zhang P, Garg R, et al. Body composition predicts mortality and decompensation in compensated cirrhosis patients: A prospective cohort study. *JHEP Rep.* 2020;2(1):100061.

9.   Kramer JR, Davila JA, Miller ED, Richardson P, Giordano TP, El-Serag HB. The validity of viral hepatitis and chronic liver disease diagnoses in Veterans Affairs administrative databases. *Alimentary Pharmacology and Therapeutics*. 2008;27(3):274-282.

10.   Chang EK, Yu CY, Clarke R, et al. Defining a Patient Population With Cirrhosis: An Automated Algorithm With Natural Language Processing. *J Clin Gastroenterol*. 2016;50(10):889-894.

11.   Goldberg D, Lewis JD, Halpern SD, Weiner M, Lo Re IV. Validation of three coding algorithms to identify patients with end-stage liver disease in an administrative database. *Pharmacoepidemiology and Drug Safety*. 2012;21(7):765-769.

12.   Hachem CY, Kramer JR, Kanwal F, El-Serag HB. Hepatitis vaccination in patients with hepatitis C: Practice and validation of codes at a large veterans administration medical centre. *Alimentary Pharmacology and Therapeutics*. 2008;28(9):1078-1087.

13.   Lu M, Chacra W, Rabin D, et al. Validity of an automated algorithm using diagnosis and procedure codes to identify decompensated cirrhosis using electronic health records. *Clinical Epidemiology*. 2017;9:369-376.

14.   Mapakshi S, Kramer JR, Richardson P, El-Serag HB, Kanwal F. Positive Predictive Value of International Classification of Diseases, 10th Revision, Codes for Cirrhosis and Its Related Complications. *Clinical Gastroenterology and Hepatology*. 2018.

15.   Nehra MS, Ma Y, Clark C, Amarasingham R, Rockey DC, Singal AG. Use of administrative claims data for identifying patients with cirrhosis. *Journal of Clinical Gastroenterology*. 2013;47(5):E50-E54.

16.   Rabin D, Chacra W, Gordon S, Yang J, Rupp L. Use of billing and procedure codes to predict stage of liver disease in patients with chronic viral hepatitis. *American Journal of Gastroenterology*. 2013;1):S147.

17. Rakoski MO, McCammon RJ, Piette JD, et al. Burden of cirrhosis on older Americans and their families: analysis of the health and retirement study. *Hepatology.* 2012;55(1):184-191.

18. Re VL, Lim JK, Goetz MB, et al. Validity of diagnostic codes and liver-related laboratory abnormalities to identify hepatic decompensation events in the Veterans Aging Cohort Study. *Pharmacoepidemiology and Drug Safety.* 2011;20(7):689-699.

19. Singal AG, Ma Y, Nehra M, Gopal P, Amarasingham R. The validity of administrative codes for identifying patients with cirrhosis. *Hepatology.* 2011;1):586A.

20. Kanwal F, Kramer JR, Buchanan P, et al. The quality of care provided to patients with cirrhosis and ascites in the Department of Veterans Affairs. *Gastroenterology.* 2012;143(1):70-77.

21. Fialla AD, De Muckadell OBS, Touborg Lassen A. Incidence, etiology and mortality of cirrhosis: A population-based cohort study. *Scandinavian Journal of Gastroenterology.* 2012;47(6):702-709.

22. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sorensen HT. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. *BMC Med Res Methodol.* 2011;11:83.

23. Lapointe-Shaw L, Georgie F, Carlone D, et al. Identifying cirrhosis, decompensated cirrhosis and hepatocellular carcinoma in health administrative data: A validation study. *PLoS One.* 2018;13(8):e0201120.

24. Quan H, Parsons GA, Ghali WA. Validity of information on comorbidity derived from ICD-9-CCM administrative data. *Medical care.* 2002;40(8):675-685.

25. Ratib S, West J, Crooks CJ, Fleming KM. Diagnosis of liver cirrhosis in England, a cohort study, 1998-2009: a comparison with cancer. *Am J Gastroenterol.* 2014;109(2):190-198.

26. Driver RJ, Balachandrakumar V, Burton A, et al. Validation of an algorithm using inpatient electronic health records to determine the presence and severity of cirrhosis in patients with hepatocellular carcinoma in England: an observational study. *BMJ Open.* 2019;9(7):e028571.

27. Kanwal F, Kramer JR, Buchanan P, et al. The quality of care provided to patients with cirrhosis and ascites. *Hepatology.* 2011;1):408A.

28. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol.* 1992;45(6):613-619.

29. Gordon SC, Pockros PJ, Terrault NA, et al. Impact of disease severity on healthcare costs in patients with chronic hepatitis C (CHC) virus infection. *Hepatology.* 2012;56(5):1651-1660.

30. Singal AG, Rahimi RS, Clark C, et al. An automated model using electronic medical record data identifies patients with cirrhosis at high risk for readmission. *Clinical Gastroenterology and Hepatology.* 2013;11(10):1335-1341.

31. Elsevier Scopus Website. https://www.elsevier.com/solutions/scopus?dgcid=RN_AGCM_Sourced_300005030. Accessed 20/09/2020.

32. Bengtsson B, Askling J, Ludvigsson JF, Hagstrom H. Validity of administrative codes associated with cirrhosis in Sweden. *Scand J Gastroenterol.* 2020;55(10):1205-1210.

33. Bernatsky S, Lix L, O'Donnell S, Lacaille D, Network C. Consensus statements for the use of administrative health data in rheumatic disease research and surveillance. *The Journal of rheumatology.* 2013;40(1):66-73.

34. Hanly JG, Thompson K, Skedgel C. The use of administrative health care databases to identify patients with rheumatoid arthritis. *Open access rheumatology : research and reviews.* 2015;7:69-75.

35. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of myocardial infarction diagnoses in administrative databases: a systematic review. *PLoS One.* 2014;9(3):e92286.

36. Thorn JC, Turner E, Hounsome L, et al. Validation of the Hospital Episode Statistics Outpatient Dataset in England. *Pharmacoeconomics.* 2016;34(2):161-168.

37. Spencer SA, Davies MP. Hospital episode statistics: improving the quality and value of hospital data: a national internet e-survey of hospital consultants. *BMJ Open.* 2012;2(6).

38. Dominitz JA, Boyko EJ, Koepsell TD, et al. Elevated prevalence of hepatitis C infection in users of United States veterans medical centers. *Hepatology.* 2005;41(1):88-96.

39. Das SR, Kinsinger LS, Yancy WS, Jr., et al. Obesity prevalence among veterans at Veterans Affairs medical facilities. *Am J Prev Med.* 2005;28(3):291-294.

40. Rimland JM, Abraha I, Luchetta ML, et al. Validation of chronic obstructive pulmonary disease (COPD) diagnoses in healthcare databases: a systematic review protocol. *BMJ Open.* 2016;6(6):e011777.

41. Khokhar B, Jette N, Metcalfe A, et al. Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations. *BMJ Open.* 2016;6(8):e009952.

42. McCormick N, Bhole V, Lacaille D, Avina-Zubieta JA. Validity of Diagnostic Codes for Acute Stroke in Administrative Databases: A Systematic Review. *PLoS One.* 2015;10(8):e0135834.

Systematic Review: Development of a consensus code set to identify cirrhosis in electronic health records

Systematic review
Diagnostic codes used to identify cirrhosis synthesised into a **consensus code set**

International validation
Use of the **consensus code set** increases sensitivity above currently used code sets with high positive predictive values

It is intended that this code set is used to define cirrhosis in EHR research where cirrhosis is defined by the presence of one or more of the codes in the set from either inpatient or outpatient settings

AP&T
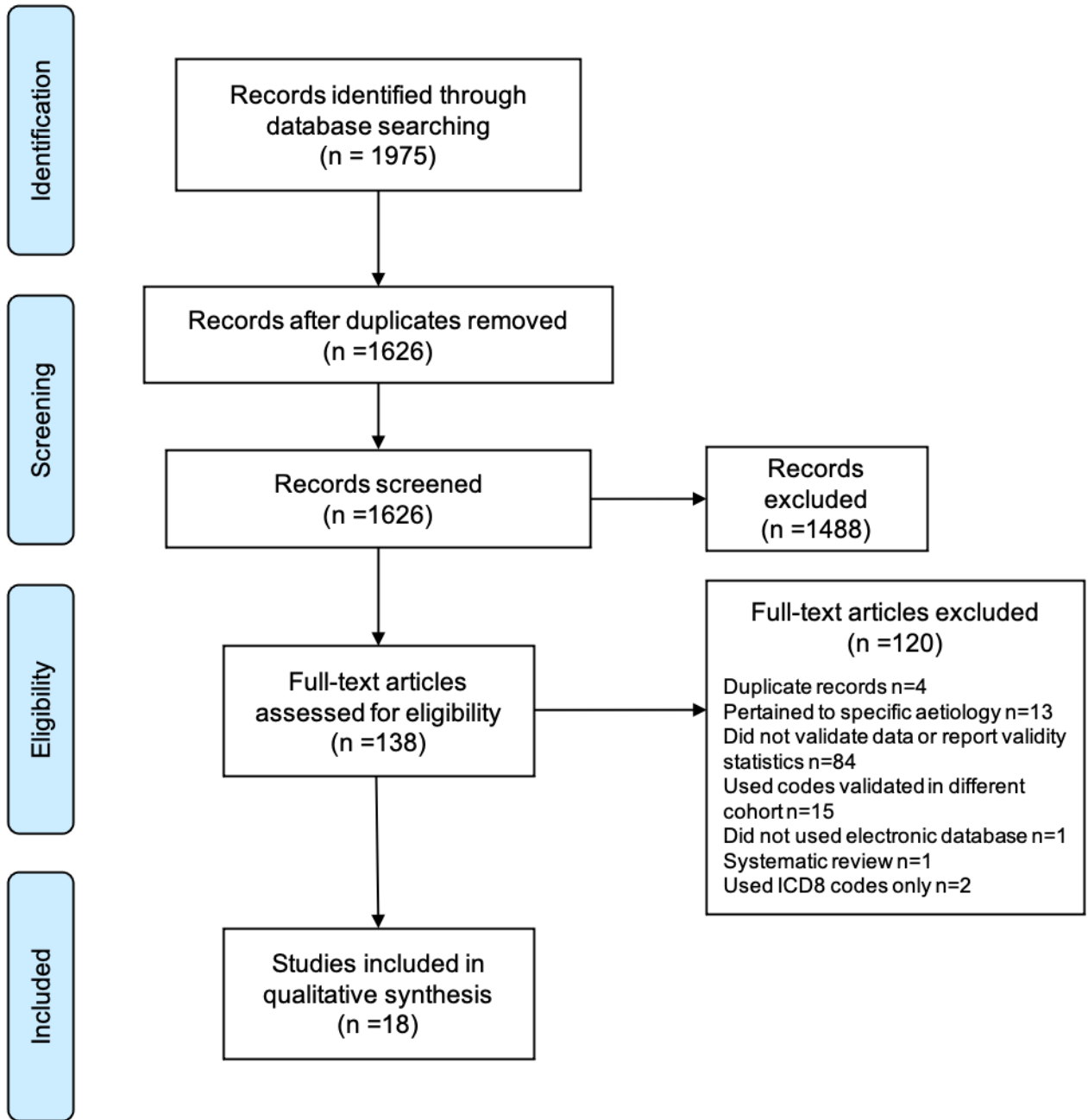
Figure 1: Study flow chart. ICD; international classification of diseases

APT_16806_Figure 1.tiff