BMJ Open

Validation of an algorithm using inpatient electronic health records to determine liver disease severity in patients with hepatocellular carcinoma

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-028571
Article Type:	Research
Date Submitted by the Author:	18-Dec-2018
Complete List of Authors:	Driver, Robert; University of Leeds Leeds Institute of Biomedical Clinical Sciences; Leeds Teaching Hospitals NHS Trust Balachandrakumar, Vinay; Royal Liverpool and Broadgreen University Hospitals NHS Trust Burton, Anya; National Cancer Registration and Analysis Service Shearer, Jessica; Leeds Teaching Hospitals NHS Trust Downing, A; University of Leeds, Leeds Institute for Data Analytics Cross, Tim; Royal Liverpool and Broadgreen University Hospitals NHS Trust Morris, Eva; University of Leeds, Leeds Institute for Data Analytics Rowe, Ian; University of Leeds; Leeds Teaching Hospitals NHS Trust
Keywords:	Hepatology < INTERNAL MEDICINE, Hepatobiliary tumours < ONCOLOGY, Epidemiology < ONCOLOGY

Validation of an algorithm using inpatient electronic health records to determine liver disease severity in patients with hepatocellular carcinoma

R J Driver[1,2], V K Balachandrakumar[3], A Burton[4], J Shearer[1,2], A Downing[5], T Cross[3], E Morris[5], I A C Rowe[1,2]

1. Leeds Teaching Hospitals NHS Trust, Leeds, UK
2. Leeds Institute of Biomedical and Clinical Sciences, University of Leeds, UK
3. Royal Liverpool and Broadgreen University Hospitals NHS Trust, Liverpool, UK
4. National Cancer Registration and Analysis Service, Public Health England, Bristol, UK
5. Leeds Institute for Data Analytics, University of Leeds, UK

Corresponding author's email address: r.j.driver@leeds.ac.uk

ABSTRACT

Objectives Outcomes in hepatocellular carcinoma are determined by both cancer characteristics and liver disease severity. This study aims to validate the use of inpatient electronic health records to determine liver disease severity from treatment and procedure codes.

Design Retrospective observational study.

Setting Two National Health Service (NHS) cancer centres in England.

Participants 339 patients with a new diagnosis of hepatocellular carcinoma between 2007 and 2016.

Main Outcome Using inpatient electronic health records, we have developed an optimised algorithm to identify cirrhosis and determine liver disease severity in a population with hepatocellular carcinoma. The diagnostic accuracy of the algorithm was optimised using clinical records from one NHS Trust and it was externally validated using anonymised data from another centre.

Results The optimised algorithm has a positive predictive value (PPV) of 99% for identifying cirrhosis in the derivation cohort, with a sensitivity of 86% (95% confidence interval, CI: 82%-

90%) and a specificity of 98% (95% CI: 96%-100%). The sensitivity for detecting advanced stage cirrhosis is 83% (95% CI: 78%-89%) and specificity is 96% (95% CI: 93%-99%), with a PPV of 89%.

Conclusions Our optimised algorithm, based on inpatient electronic health records, reliably identifies and stages cirrhosis in patients with HCC. This highlights the potential of routine health data in population studies to stratify patients with hepatocellular carcinoma according to liver disease severity.

ARTICLE SUMMARY: STRENGTH AND LIMITATIONS OF THIS STUDY

- First study to use inpatient electronic health records to identify and stage cirrhosis severity in a population with hepatocellular carcinoma.
- The presence of cirrhosis predicted by inpatient electronic health records is accurate and advanced stage disease identified by the algorithm is associated with increased disease severity scores in validation.
- A potential limitation is a variation in coding practices between centres and over time.
- This algorithm may be used in population studies to understand outcomes in hepatocellular carcinoma, which require an assessment of liver disease severity.

INTRODUCTION

Primary liver cancer accounts for 2% of all cancers diagnosed in the UK, with approximately 5700 new cases each year [1] and these are most commonly hepatocellular carcinoma (HCC). It is estimated that 70-90% of HCC occurs in the background of cirrhosis [2, 3] and global outcomes are poor despite a number of treatment options [4]. Curative treatments may be limited by poor liver function due to underlying cirrhosis, or late presentation of advanced cancer in patients not known to have cirrhosis. Therefore to understand outcomes of patients with HCC, it is essential to consider the presence and severity of cirrhosis in all analyses.

Population-based cancer registry data are used to describe trends in cancer incidence and mortality in a number of cancer sites, as well as regional variation in clinical outcomes [5]. In HCC research, registry data have been used to describe geographical variation in incidence, survival and treatment allocation in France [6]. In England, the National Cancer Registration and Analysis Service (NCRAS) dataset contains patient-level information about individuals

with HCC, but information on the presence of cirrhosis is not currently included. Also, blood test results are not collected, so cirrhosis severity using tools such as the Child Pugh score and the Model for End-Stage Liver Disease (MELD) score cannot be calculated.

Previous international studies have outlined methods to use electronic health records (EHR) to identify cirrhosis [7-10]. In the UK, Ratib and colleagues used a combination of inpatient and outpatient records to identify cirrhosis and its complications, including oesophageal varices and ascites [11]. These complications relate to advanced stage or "decompensated" cirrhosis and they often result in admitted patient care. In England these records are captured by the Hospital Episode Statistics (HES) database, which is linked to the cancer registry data within NCRAS.

We present a clinical validation study using EHRs from two regional cancer centres in England to assess the performance of an algorithm to determine liver disease severity using the local inpatient HES records, which are subsequently transmitted to the national HES dataset. This study aims to demonstrate the use of routinely-collected diagnosis and treatment codes to identify the presence or absence of cirrhosis, and cirrhosis severity in individuals diagnosed with HCC, for use in registry applications to improve outcomes for this patient group.

METHODS

All patients diagnosed with HCC between 1st January 2007 and 31st December 2016 and resident in the secondary care catchment area of Leeds Teaching Hospitals NHS Trust (LTHT) were identified from clinical audit data. Patients seen as tertiary referrals were excluded to avoid selection bias and because the LTHT HES record would not contain the complete cirrhosis follow-up. The local HES records were searched to identify inpatient episodes containing codes related to cirrhosis within the ICD10 (International Classification of Diseases, tenth revision) and OPCS4 (Office of Population, Censuses and Surveys' Classification, fourth revision), together with the corresponding time interval from the HCC diagnosis date. An algorithm was developed to characterise patients from these codes, and comparison made with the clinical records. External validation of the algorithm was undertaken using the same search within the local HES records for patients diagnosed with HCC between 1st January 2013 and 31st December 2014 and local to Royal Liverpool and Broadgreen University Hospitals NHS Trust (RLBUHT).

This retrospective study comprises an assessment of the accuracy of clinical coding of inpatient episodes for service evaluation and as such does not require formal ethical approval. All patient data were anonymised and permission was granted from the Caldicott Guardian for sharing of routinely-collected anonymised data.

Patient and Public Involvement

Patients and public were not involved in this study.

Identification of Cirrhosis

To determine the presence of cirrhosis at HCC diagnosis, episodes containing cirrhosis-related codes which occurred up to five years before the HCC diagnosis date were initially included. However, to improve the sensitivity of the algorithm by maximising the number of available inpatient codes, additional episodes occurring after HCC diagnosis were subsequently included. This approach assumes that if an inpatient cirrhosis code occurs after the HCC diagnosis, the patient is likely to have had cirrhosis at the time of HCC diagnosis. The timeframe post-HCC diagnosis of included episodes was increased incrementally and the performance of the algorithm tested to validate this assumption.

Different definitions of cirrhosis within ICD10 have been used in population studies [12]. Some investigators [7, 8] used cirrhosis diagnosis codes only, whereas others [9] also included varices codes. Ratib and colleagues [13] additionally included OPCS4 procedure codes for treatment of varices and version 1 of our algorithm is based on this approach. Patients are classified as cirrhotic if they had inpatient episodes containing the diagnosis and treatment codes for cirrhosis or varices outlined in Table 1. In version 2, a broader definition of cirrhosis proposed by Leon and colleagues [14] was used, including codes for “alcoholic liver disease” (ALD, K70.9) and “alcoholic hepatic failure” (AHF, K70.4). To assess the accuracy of including ascites as a cirrhosis-defining condition in HCC, codes for ascites and paracentesis were included in version 3 of the algorithm. Previously, some investigators [9, 13] excluded ascites in their definitions because this may be due to malignancy in the absence of cirrhosis in a general population. In version 4, only ascites codes occurring before the HCC diagnosis date were included.

The clinical records were reviewed by three clinical investigators, each with greater than two years specialist experience, to identify patients with clinical, radiological or histological evidence of cirrhosis at the time of HCC diagnosis. This was used as the gold standard for testing different versions of the algorithm to classify cirrhosis status. For comparison, published algorithms [7-9, 13] were also tested in the LTHT cohort of patients with HCC.

Classification of Cirrhosis Severity

Cirrhosis severity was classified using the Baveno IV consensus [15]. Compensated cirrhosis is defined by Baveno stage 1 (no ascites or varices) and stage 2 (non-bleeding varices). Decompensated cirrhosis is defined by Baveno stage 3 (ascites, with or without varices) and stage 4 (bleeding varices, with or without ascites). In this model of the natural history of cirrhosis [3], patients progress to a higher Baveno stage over time, but do not return to a lower stage. For each hospital episode, the Baveno stage and compensation status were calculated using the diagnosis and treatment codes for ascites and varices in Table 1. Three definitions of bleeding varices were tested; version A (based on Goldberg and colleagues [10]) contains ICD10 codes for variceal bleeding, version B (based on Ratib and colleagues [11]) also includes OPCS4 codes for treatment of varices, and version C limits the inclusion of these treatment codes to those occurring in a hospital episode with a concurrent ICD10 code for gastrointestinal haemorrhage (K92.0, K92.1, and K92.2). This is to distinguish between bleeding varices and the prophylactic treatment of non-bleeding varices.

Cirrhosis severity at the time of HCC diagnosis was determined by the highest Baveno stage recorded in hospital episodes occurring in the five years before HCC diagnosis. In order to increase the accuracy of this assessment, additional episodes occurring after the HCC diagnosis date were also included. The timeframe post-HCC diagnosis of included episodes was increased incrementally up to four months. The clinical records were reviewed to determine the true Baveno stage at the time of HCC diagnosis, along with routine blood tests for calculation of Child Pugh and MELD scores.

Statistical Analysis

Data management and statistical analysis was performed using Stata version 15.1 (StataCorp, College Station, TX). The diagnostic accuracy of the algorithm to identify cirrhosis status and

decompensation status involved comparison of sensitivity and specificity derived from 2 x 2 contingency tables [16]. For Baveno stage, agreement between the algorithm and the clinical records were assessed using the kappa statistic. This is used to assess observer agreement for categorical variables and allows for agreement occurring by chance [17, 18].

RESULTS

Study Population

During the study period, 289 patients (median age 69, 79% male) with a new diagnosis of HCC were included. Review of the clinical record identified 191 (66%) of these as cirrhotic at HCC diagnosis, 48 (25%) of whom had evidence of previous decompensation. In the external validation cohort at RLBUHT, 50 patients meeting the inclusion criteria were assessed (median age 71, 82% male), 31 (62%) of whom were cirrhotic and 11 (35%) with previous decompensation.

Identification of Cirrhosis

Limiting the inclusion of episodes to those occurring before the HCC diagnosis results in a sensitivity of less than 50% for cirrhosis detection (Table 2). When additional episodes are included up until three years after the HCC diagnosis, the sensitivity increases to greater than 80% for all versions of the algorithm, without significant loss of specificity.

The sensitivity of algorithm 1 is increased by including ALD and AHF (version 2), and further increased by including ascites (version 3). However, the inclusion of ascites also reduces the specificity. This is overcome by limiting the inclusion of ascites to episodes that occurred before the HCC diagnosis (version 4). Using this optimised algorithm and including records up to three years post HCC diagnosis, the sensitivity is 86% (95% confidence interval, CI: 82%-90%) and the specificity is 98% (95% CI: 96%-100%), with a positive predictive value (PPV) of 99% and negative predictive value (NPV) of 79% (95% CI: 74%-83%) (Supplementary Table 1). For external validation, when version 4 of the algorithm was applied to the RLBUHT cohort with three years of follow-up, the sensitivity was 79% and specificity was 100%. Additionally, version 4 of the algorithm outperformed published algorithms for cirrhosis detection when they were applied to the LTHT cohort of HCC patients (Table 3).

Classification of Cirrhosis Severity

Table 4 shows the performance of the three versions of the algorithm for determining cirrhosis severity according to Baveno stage. Compared to version A, there is slightly less agreement between the calculated Baveno stage and the clinical record in version B, where Baveno stage 4 is defined by procedure codes for varices. Similarly, the sensitivity for detecting decompensation (defined by Baveno stages 3 and 4) is increased in version B, but with reduced sensitivity (Supplementary Table 2). Agreement between the algorithm and the clinical record is optimised in version C, when bleeding varices are defined by a concurrent gastrointestinal haemorrhage code. Agreement was further improved when episodes occurring within 60 days of the registered HCC diagnosis were included.

Using version C with a 60 day interval in the LTHT cohort, agreement between the clinical record and calculated Baveno stage was 84%, with a kappa coefficient of 0.74 (95% CI: 71%-77%). The sensitivity for detecting prior decompensation is 83% (95% CI: 78%-89%) and specificity is 96% (95% CI: 93%-99%), with a PPV of 89% (95% CI: 84% - 94%) and NPV of 93% (95% CI: 90%-97%). When this version was applied to the RLBUHT cohort for external validation, the agreement of Baveno stage with the clinical record was 81% (kappa 0.70). The sensitivity for detecting decompensation was 73% and specificity was 90%.

Finally, among the 167 LTHT patients identified as cirrhotic by the algorithm, 45 (27%) were coded with prior decompensation. At the time of HCC diagnosis, Child Pugh class and MELD scores were each higher in those individuals identified with decompensation (figure 1).

DISCUSSION

Main findings

This study demonstrates the reliability of an algorithm using inpatient HES records to identify and stage cirrhosis in patients with HCC. This is the first such algorithm validated in a UK population that uses only inpatient codes. Using inpatient codes from the whole follow-up period improves the sensitivity of the algorithm in cirrhosis identification, without loss of specificity. This validates the assumption that if a patient had an inpatient cirrhosis code during follow-up, they had cirrhosis at the time of HCC diagnosis. Using a broad definition of cirrhosis (versions 2-4) improves sensitivity and accounts for variations in coding practice in

which ALD and AHF are coded synonymously with cirrhosis. Excluding ascites after HCC diagnosis (version 4) improves the specificity; ascites in liver disease without HCC is most likely to be due to cirrhosis, whereas it may be malignant ascites in the context of HCC. Algorithm 4 is an improvement over published algorithms for cirrhosis detection when they are applied to our cohort of HCC patients. However, these algorithms also have a higher PPV in our cohort than in the corresponding validation studies in general populations due to the high prevalence of cirrhosis among patients with HCC. Algorithm C (for assessing cirrhosis severity) also outperformed published versions in this population. Inclusion of a concurrent gastrointestinal haemorrhage code alongside variceal procedures distinguishes between treatment of bleeding varices and treatment of non-bleeding varices for primary prevention.

Strengths and limitations

The strengths of this study are the systematic development of an algorithm which uses routinely available inpatient episode codes, and its applicability to large population studies in HCC. These patients often require hospital admission to manage complications in advanced cirrhosis to receive HCC therapies. Limiting to inpatient episodes has therefore not reduced the performance of the algorithm compared to the method employed by Ratib and colleagues [11], which used a combination of inpatient and outpatient records. This study benefits from robust case note evaluation, using both a development and external validation cohort. In the UK, previous validation of inpatient coding was achieved using free text analysis of primary care and death certification data [11], and the original case note validation of the cirrhosis algorithm included only 36 patients [19].

The algorithm benefits from exploiting the 'anchor point' of the HCC diagnosis date, so that inpatient codes can be associated with a time interval. This has led to optimised cirrhosis detection and severity classification. The algorithm for cirrhosis detection was optimised using three years of follow-up after HCC diagnosis, but the high sensitivity and specificity using one year of follow-up may be sufficient in some settings.

The limitations include its location in specialist cancer centres, which may not reflect coding practices throughout the UK and these may change over time. However, portal hypertensive complications are common and often result in inpatient care and, since these are high cost procedures, we anticipate them to be reliably coded. The analysis was limited to patients

local to the two centres, in order to capture cirrhosis-related episodes. Additional episodes may have been missed if patients were admitted elsewhere, but these would be captured by the algorithm when extended to a national dataset.

Implications

This algorithm can be applied to population cancer registries, enabling the identification and staging of cirrhosis in patients with HCC. This is essential for assessing clinical outcomes in population-based studies of individuals with HCC both in the UK and elsewhere. It is hoped that this will lead to a better understanding of outcomes in HCC, including progression of underlying liver disease severity as well as overall survival. The algorithm may also be used in other population-based applications which require the identification of cirrhosis and an assessment of severity.

In this study, we demonstrated the use of inpatient HES records to determine the cirrhosis severity at the time of HCC diagnosis. The algorithm may be adapted to classify the Baveno stage at different time intervals following HCC diagnosis or date of treatment, so that subsequent cirrhosis decompensation events can be identified over time.

Conclusion

This study demonstrates the reliability of an algorithm based on inpatient EHRs to stratify patients with HCC according to liver disease severity. It may be used in routine health data in order to assess outcomes in HCC in population studies.

Cirrhosis Diagnoses (ICD10):	**Codes**
Cirrhosis	K70.3, K71.7, K72.1, K74.4, K74.5, K74.6, K76.6, K72.1, K72.9
Alcoholic hepatic failure	K70.4
Alcoholic liver disease	K70.9
Ascites	R18.X
Varices	I85.9, I86.4, I98.2
Bleeding varices	I85.0, I98.3
Cirrhosis Treatments (OPCS4):	
Treatment of ascites	T46.1, T46.2, J06.1, J06.2
Treatment of varices	G10.4, G10.8, G10.9, G14.4, G17.4, G43.4, G43.7, J06.1, J06.2
Gastrointestinal Haemorrhage (ICD10):	
Gastrointestinal haemorrhage	K92.0, K92.1, K92.2

Table 1. Treatment and procedure codes included in the algorithm to determine cirrhosis status and cirrhosis severity.

	Algorithm 1 No Ascites - ALD - AHF		**Algorithm 2 No Ascites + ALD + AHF**		**Algorithm 3 Ascites + ALD + AHF**		**Algorithm 4 Pre-HCC Ascites + ALD + AHF**	
Time after HCC Diagnosis/ days	sens	spec	sens	spec	sens	spec	sens	spec
0	0.45	1.00	0.47	1.00	0.49	1.00	0.49	1.00
30	0.52	1.00	0.55	1.00	0.57	0.99	0.57	1.00
60	0.60	1.00	0.64	1.00	0.67	0.98	0.66	1.00
90	0.66	1.00	0.70	1.00	0.73	0.97	0.72	1.00
120	0.68	1.00	0.72	1.00	0.74	0.97	0.74	1.00
150	0.71	1.00	0.75	1.00	0.78	0.96	0.77	1.00
180	0.72	0.99	0.76	0.99	0.79	0.95	0.78	0.99
365	0.76	0.99	0.80	0.99	0.82	0.94	0.82	0.99
730	0.80	0.98	0.84	0.98	0.87	0.92	0.85	0.98
1095	0.81	0.98	0.85	0.98	0.88	0.92	0.86	0.98
Total Follow-up	0.81	0.98	0.85	0.98	0.88	0.92	0.86	0.98

Table 2. Performance of different versions of the cirrhosis status algorithm. Sens = sensitivity, spec = specificity, ALD = Alcoholic Liver Disease, AHF = Alcoholic Hepatic Failure

Algorithm	Sensitivity (%)	Specificity (%)	PPV (%)
Kramer *et al.*[7]	72	100	100
Jepsen *et al.*[8]	71	100	100
Nehra *et al.*[9]	80	98	99
Ratib *et al.*[13]	80	98	99

Table 3. Performance of different published algorithms for cirrhosis detection in the LTHT cohort of patients with HCC. PPV = positive predictive value.

	Algorithm A Variceal bleeding codes		**Algorithm B Variceal bleeding codes or treatment codes**		**Algorithm C Variceal bleeding codes or treatment codes + UGIB**	
Time after HCC Diagnosis/ days	Correct Baveno Stage (%)	K-statistic	Correct Baveno Stage (%)	K-statistic	Correct Baveno Stage (%)	K-statistic
0	80	0.67	80	0.67	81	0.70
30	82	0.70	81	0.70	83	0.73
60	83	0.71	82	0.71	84	0.74
90	81	0.69	80	0.69	82	0.71
120	81	0.69	80	0.69	82	0.71

Table 4. Performance of different versions of the Baveno stage algorithm. UGIB = Upper gastrointestinal bleeding, K = kappa statistic.

Author Contributions

IR and RD had the original idea for the study and all authors contributed to its design and planning. RD, JS and VK performed the case note reviews. RD was responsible for data management, statistical analyses and wrote the first draft of the paper. IR reviewed the paper critically and all authors approved the final version.

Funding

The research received no specific grants from any funding agency. RD and JS were funded by clinical research fellowships at Leeds Teaching Hospitals NHS Trust. AB is an analyst within Public Health England and receives funding from the British Association for the Study of the Liver (BASL) as part of the HCC-UK Partnership. No industrial sponsor had input into study planning, discussion of results or manuscript preparation

Competing Interests

VK has received an unrestricted educational grant from Bayer, Bristol-Myers Squibb and Sirtex.

Patient Consent

Not applicable

Ethical approval

This retrospective study comprises an assessment of the accuracy of clinical coding for service evaluation and as such does not require formal ethical approval. All patient data were anonymised and permission was granted from the Caldicott Guardian for sharing of routinely-collected anonymised data.

Data Statement

Statistical code is available from the corresponding author RD

Acknowledgements

This work uses data provided by patients and collected by the NHS as part of their care and support. The planning of this project was discussed with members of the HCC-UK Partnership, which is co-ordinated by AB.

Word Count: 2665

REFERENCES

1. *Cancer Research UK. Liver Cancer Statistics*. 2015 [cited 2018 05/07/2018]; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/liver-cancer#heading-Zero.
2. El-Serag, H.B. and K.L. Rudolph, *Hepatocellular carcinoma: epidemiology and molecular carcinogenesis.* Gastroenterology, 2007. **132**(7): p. 2557-76.
3. D'Amico, G., G. Garcia-Tsao, and L. Pagliaro, *Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies.* Journal of Hepatology, 2006. **44**(1): p. 217-31.
4. El-Serag, H.B., *Hepatocellular carcinoma.* New England Journal of Medicine, 2011. **365**(12): p. 1118-27.
5. Coleman, M.P., et al., *Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data.* The Lancet, 2011. **377**(9760): p. 127-138.
6. Goutté, N., et al., *Geographical variations in incidence, management and survival of hepatocellular carcinoma in a Western country.* Journal of Hepatology, 2017. **66**(3): p. 537-544.
7. Kramer, J.R., et al., *The validity of viral hepatitis and chronic liver disease diagnoses in Veterans Affairs administrative databases.* Aliment Pharmacol Ther, 2008. **27**(3): p. 274-82.
8. Jepsen, P., et al., *Clinical course of alcoholic liver cirrhosis: a Danish population-based cohort study.* Hepatology, 2010. **51**(5): p. 1675-82.
9. Nehra, M.S., et al., *Use of administrative claims data for identifying patients with cirrhosis.* J Clin Gastroenterol, 2013. **47**(5): p. e50-4.
10. Goldberg, D., et al., *Validation of three coding algorithms to identify patients with end-stage liver disease in an administrative database.* Pharmacoepidemiol Drug Saf, 2012. **21**(7): p. 765-769.
11. Ratib, S., et al., *1 and 5 year survival estimates for people with cirrhosis of the liver in England, 1998-2009: a large population study.* Journal of Hepatology, 2014. **60**(2): p. 282-9.
12. Ratib, S., J. West, and K.M. Fleming, *Liver cirrhosis in England—an observational study: are we measuring its burden occurrence correctly?* BMJ Open, 2017. **7**(7).
13. Ratib, S., et al., *Diagnosis of Liver Cirrhosis in England, a Cohort Study, 1998–2009: A Comparison With Cancer.* The American Journal Of Gastroenterology, 2014. **109**: p. 190.
14. Leon, D.A. and J. McCambridge, *Liver cirrhosis mortality rates in Britain from 1950 to 2002: an analysis of routine data.* The Lancet, 2006. **367**(9504): p. 52-56.
15. de Franchis, R., *Evolving consensus in portal hypertension. Report of the Baveno IV consensus workshop on methodology of diagnosis and therapy in portal hypertension.* Journal of Hepatology, 2005. **43**(1): p. 167-76.
16. Alberg, A.J., et al., *The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests.* J Gen Intern Med, 2004. **19**(5 Pt 1): p. 460-5.
17. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* Biometrics, 1977. **33**(1): p. 159-74.

18. Cohen, J., *A coefficient of agreement for nominal scales.* Educational and psychological measurement, 1960. **20**(1): p. 37-46.
19. Fleming, K.M., et al., *Incidence and prevalence of cirrhosis in the United Kingdom, 1992–2001: A general population-based study.* Journal of Hepatology, 2008. **49**(5): p. 732-738.

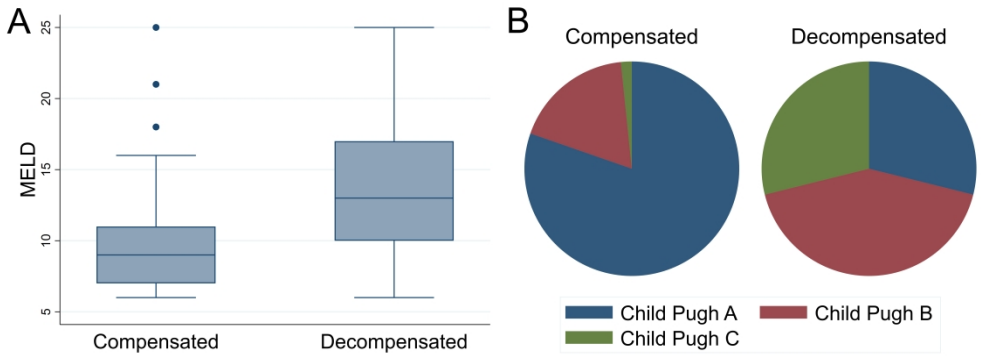

Figure 1. Box-and-whisker plots showing the distribution of MELD scores (A) and pie graphs showing the distribution of Child Pugh class (B) within compensated and decompensated cirrhosis groups determined by the algorithm.

Supplementary Tables

		True Status		
		Non-cirrhotic	Cirrhotic	Total
Cirrhosis Algorithm	Negative for Cirrhosis	96	26	122
	Positive for Cirrhosis	2	165	167
	Total	98	191	289

Supplementary Table 1. 2 x2 Contingency table for cirrhosis identification by optimised algorithm version 4 with three years of follow-up.

	Algorithm A Variceal bleeding codes		**Algorithm B Variceal bleeding codes or treatment codes**		**Algorithm C Variceal bleeding codes or treatment codes + UGIB**	
Time after HCC Diagnosis/ days	sens	spec	sens	spec	sens	spec
0	0.77	0.97	0.81	0.92	0.79	0.96
30	0.79	0.97	0.83	0.92	0.81	0.96
60	0.81	0.97	0.85	0.92	0.83	0.96
90	0.81	0.96	0.85	0.91	0.83	0.94
120	0.81	0.96	0.85	0.91	0.83	0.92

Supplementary Table 2. Performance of different versions of the Baveno stage algorithm for predicting decompensation. Sens = sensitivity, spec = specificity, UGIB = Upper gastrointestinal bleeding.

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	1
Objectives	3	State specific objectives, including any pre-specified hypotheses	2
Methods			
Study design	4	Present key elements of study design early in the paper	2
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	2-3
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	2
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	2-4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*2-4*
Bias	9	Describe any efforts to address potential sources of bias	2
Study size	10	Explain how the study size was arrived at	4-5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	4-5
		(*b*) Describe any methods used to examine subgroups and interactions	
		(*c*) Explain how missing data were addressed	
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	5
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	4-6
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	5-6
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	6
Discussion			
Key results	18	Summarise key results with reference to study objectives	6
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	7
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	7-8
Generalisability	21	Discuss the generalisability (external validity) of the study results	8
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	12

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-028571.R1
Article Type:	Research
Date Submitted by the Author:	26-Mar-2019
Complete List of Authors:	Driver, Robert; University of Leeds Leeds Institute of Biomedical Clinical Sciences; Leeds Teaching Hospitals NHS Trust Balachandrakumar, Vinay; Royal Liverpool and Broadgreen University Hospitals NHS Trust Burton, Anya; National Cancer Registration and Analysis Service Shearer, Jessica; Leeds Teaching Hospitals NHS Trust Downing, A; University of Leeds, Leeds Institute for Data Analytics Cross, Tim; Royal Liverpool and Broadgreen University Hospitals NHS Trust Morris, Eva; University of Leeds, Leeds Institute for Data Analytics Rowe, Ian; University of Leeds; Leeds Teaching Hospitals NHS Trust
Primary Subject Heading:	Gastroenterology and hepatology
Secondary Subject Heading:	Epidemiology, Oncology
Keywords:	Hepatology < INTERNAL MEDICINE, Hepatobiliary tumours < ONCOLOGY, Epidemiology < ONCOLOGY

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

R J Driver[1,2], V K Balachandrakumar[3], A Burton[4], J Shearer[1,2], A Downing[5], T Cross[3], E Morris[5], I A Rowe[1,2]

1. Leeds Teaching Hospitals NHS Trust, Leeds, UK
2. Leeds Institute of Biomedical and Clinical Sciences, University of Leeds, UK
3. Royal Liverpool and Broadgreen University Hospitals NHS Trust, Liverpool, UK
4. National Cancer Registration and Analysis Service, Public Health England, Bristol, UK
5. Leeds Institute for Data Analytics, University of Leeds, UK

Corresponding author's email address: r.j.driver@leeds.ac.uk

ABSTRACT

Objectives Outcomes in hepatocellular carcinoma are determined by both cancer characteristics and liver disease severity. This study aims to validate the use of inpatient electronic health records to determine liver disease severity from treatment and procedure codes.

Design Retrospective observational study.

Setting Two National Health Service (NHS) cancer centres in England.

Participants 339 patients with a new diagnosis of hepatocellular carcinoma between 2007 and 2016.

Main Outcome Using inpatient electronic health records, we have developed an optimised algorithm to identify cirrhosis and determine liver disease severity in a population with hepatocellular carcinoma. The diagnostic accuracy of the algorithm was optimised using clinical records from one NHS Trust and it was externally validated using anonymised data from another centre.

Results The optimised algorithm has a positive predictive value (PPV) of 99% for identifying cirrhosis in the derivation cohort, with a sensitivity of 86% (95% confidence interval, CI: 82%-90%) and a specificity of 98% (95% CI: 96%-100%). The sensitivity for detecting advanced stage cirrhosis is 80% (95% CI: 75%-87%) and specificity is 98% (95% CI: 96%-100%), with a PPV of 89%.

Conclusions Our optimised algorithm, based on inpatient electronic health records, reliably identifies and stages cirrhosis in patients with HCC. This highlights the potential of routine health data in population studies to stratify patients with hepatocellular carcinoma according to liver disease severity.

ARTICLE SUMMARY: STRENGTH AND LIMITATIONS OF THIS STUDY

- First study to use inpatient electronic health records to identify and stage cirrhosis severity in a population with hepatocellular carcinoma.
- The presence of cirrhosis predicted by inpatient electronic health records is accurate and advanced stage disease identified by the algorithm is associated with increased disease severity scores in validation.
- A potential limitation is a variation in coding practices between centres and over time.
- This algorithm may be used in population studies to understand outcomes in hepatocellular carcinoma, which require an assessment of liver disease severity.

INTRODUCTION

Primary liver cancer accounts for 2% of all cancers diagnosed in the UK, with approximately 5700 new cases each year [1] and these are most commonly hepatocellular carcinoma (HCC). It is estimated that 70-90% of HCC occurs in the background of cirrhosis [2, 3] and global outcomes are poor despite a number of treatment options [4]. Curative treatments may be limited by poor liver function due to underlying cirrhosis, or late presentation of advanced cancer in patients not known to have cirrhosis. Therefore to understand outcomes of patients with HCC, it is essential to consider the presence and severity of cirrhosis in all analyses.

Population-based cancer registry data are used to describe trends in cancer incidence and mortality in a number of cancer sites, as well as regional variation in clinical outcomes [5]. In HCC research, registry data have been used to describe geographical variation in incidence,

survival and treatment allocation in France [6]. In England, the National Cancer Registration and Analysis Service (NCRAS) dataset contains patient-level information about individuals with HCC, but information on the presence of cirrhosis is not currently included. Also, blood test results are not collected, so cirrhosis severity using tools such as the Child Pugh score and the Model for End-Stage Liver Disease (MELD) score cannot be calculated.

Previous international studies have outlined methods to use electronic health records (EHR) to identify cirrhosis [7-11]. In the UK, Ratib and colleagues used a combination of inpatient and outpatient records to identify cirrhosis and its complications, including oesophageal varices and ascites [12]. These complications relate to advanced stage or "decompensated" cirrhosis and they often result in admitted patient care. In England all inpatient records are captured by the Hospital Episode Statistics (HES) database, which is linked to the cancer registry data within NCRAS.

We present a clinical validation study using EHRs from two regional cancer centres in England to assess the performance of an algorithm to determine the presence and severity of cirrhosis using the local inpatient HES records, which are subsequently transmitted to the national HES dataset. This study aims to demonstrate that the use of routinely-collected diagnosis and treatment codes from inpatient records alone is sufficient to identify cirrhosis and grade its severity in patients with HCC. This will facilitate future studies of outcomes for patients with HCC by considering the severity of any underlying cirrhosis.

METHODS

All patients diagnosed with HCC between 1st January 2007 and 31st December 2016 and resident in the secondary care catchment area of Leeds Teaching Hospitals NHS Trust (LTHT) were identified. The diagnosis of HCC was confirmed for all patients in a weekly Hepatobiliary Cancer Multidisciplinary Team (MDT) meeting and the reporting of all cases to the national cancer registry is mandatory. HCC was usually diagnosed by radiology, using the European Association for the Study of the Liver (EASL) non-invasive criteria [13], and if indicated a targeted biopsy was performed. Live minutes are taken at these meetings and details collected into the clinical records along with a confirmed date of diagnosis. The cohort was identified from the data submitted to the central registry. We only had access to the inpatient codes from hospital episodes which occurred at LTHT. Therefore, only those patients

registered with a Clinical Commissioning Group local to LTHT were included, where we would expect them to have received their inpatient cirrhosis care. The local HES records were searched to identify inpatient episodes containing codes related to cirrhosis within the ICD10 (International Classification of Diseases, tenth revision) and OPCS4 (Office of Population, Censuses and Surveys' Classification, fourth revision), together with the corresponding time interval from the HCC diagnosis date. These codes are used routinely for reimbursement and are submitted to the national HES dataset. An algorithm was developed to characterise patients from these codes, and comparison made with the clinical records. External validation of the algorithm was undertaken using the same search within the local HES records for patients diagnosed with HCC between 1st January 2013 and 31st December 2014 and local to Royal Liverpool and Broadgreen University Hospitals NHS Trust (RLBUHT).

This retrospective study comprises an assessment of the accuracy of clinical coding of inpatient episodes for service evaluation and as such does not require formal ethical approval. All patient data were anonymised and permission was granted from the Caldicott Guardian for sharing of routinely-collected anonymised data.

Patient and Public Involvement

Patients and public were not involved in this study.

Identification of Cirrhosis

To determine the presence of cirrhosis at HCC diagnosis, episodes containing cirrhosis-related codes which occurred up to five years before the HCC diagnosis date were initially included. However, to improve the sensitivity of the algorithm by maximising the number of available inpatient codes, additional episodes occurring after HCC diagnosis were subsequently included. This approach assumes that if an inpatient cirrhosis code occurs after the HCC diagnosis, the patient is likely to have had cirrhosis at the time of HCC diagnosis. The timeframe post-HCC diagnosis of included episodes was increased incrementally and the performance of the algorithm tested to validate this assumption.

Different definitions of cirrhosis within ICD10 have been used in population studies [14]. Some investigators [7, 8] used cirrhosis diagnosis codes only, whereas others [9, 11] also included varices codes. Ratib and colleagues [15] additionally included OPCS4 procedure codes for

treatment of varices and version 1 of our algorithm is based on this approach. Patients are classified as cirrhotic if they had inpatient episodes containing the diagnosis and treatment codes for cirrhosis or varices outlined in Table 1. In version 2, a broader definition of cirrhosis proposed by Leon and colleagues [16] was used, including codes for "alcoholic liver disease" (ALD, K70.9) and "alcoholic hepatic failure" (AHF, K70.4). To assess the accuracy of including ascites as a cirrhosis-defining condition in HCC, codes for ascites and paracentesis were included in version 3 of the algorithm. Previously, some investigators [9, 15] excluded ascites in their definitions because this may be due to malignancy in the absence of cirrhosis in a general population. In version 4, only ascites codes occurring before the HCC diagnosis date were included.

The clinical records were reviewed between April and August 2018 and data abstracted by three clinical investigators (RJD, VKB, JS), each experienced hepatology fellows working in this field for at least two years. A standard abstraction form was used and discrepancies were resolved by consensus view. Cirrhosis at the time of HCC diagnosis was identified based on explicit mention of cirrhosis in the clinical record or MDT minutes, evidence of portal hypertension on radiological imaging or endoscopy reports, explicit mention of cirrhosis on liver biopsy or a consistent result on transient elastography. This was used as the gold standard for testing different versions of the algorithm to classify cirrhosis status. For comparison, published algorithms [7-9, 15] were also tested in the LTHT cohort of patients with HCC.

Classification of Cirrhosis Severity

Cirrhosis severity was classified using the Baveno IV consensus [17]. Compensated cirrhosis is defined by Baveno stage 1 (no ascites or varices) and stage 2 (non-bleeding varices). Decompensated cirrhosis is defined by Baveno stage 3 (ascites, with or without varices) and stage 4 (bleeding varices, with or without ascites). In this model of the natural history of cirrhosis [3], patients progress to a higher Baveno stage over time, but do not return to a lower stage. For each hospital episode, the Baveno stage and compensation status were calculated using the diagnosis and treatment codes for ascites and varices in Table 1. Three definitions of bleeding varices were tested; version A (based on Goldberg and colleagues [10]) contains ICD10 codes for variceal bleeding, version B (based on Ratib and colleagues [12]) also

includes OPCS4 codes for treatment of varices, and version C limits the inclusion of these treatment codes to those occurring in a hospital episode with a concurrent ICD10 code for gastrointestinal haemorrhage (K92.0, K92.1, and K92.2). This is to distinguish between bleeding varices and the prophylactic treatment of non-bleeding varices.

Cirrhosis severity at the time of HCC diagnosis was determined by the highest Baveno stage recorded in hospital episodes occurring in the five years before HCC diagnosis. In order to increase the accuracy of this assessment, additional episodes occurring after the HCC diagnosis date were also included. The timeframe post-HCC diagnosis of included episodes was increased incrementally up to four months. The clinical records were reviewed to determine the true Baveno stage at the time of HCC diagnosis, along with routine blood tests for calculation of Child Pugh and MELD scores. Baveno stage 2 was identified by non-bleeding varices explicitly mentioned in the clinical records or endoscopy reports, but excluded a report of portal hypertensive gastropathy. Baveno stage 3 was identified by explicit mention of ascites in the clinical record, requiring diuretic therapy or paracentesis, but a small volume of ascites only visible on cross-sectional imaging was excluded. Baveno stage 4 was identified by explicit mention of variceal haemorrhage in the clinical record or endoscopy reports. Clinical evidence of decompensation was identified by the presence of bleeding varices or ascites, as per the Baveno IV classification.

Statistical Analysis

Data management and statistical analysis was performed using Stata version 15.1 (StataCorp, College Station, TX). The diagnostic accuracy of the algorithm to identify cirrhosis status and decompensation status involved comparison of sensitivity and specificity derived from 2 x 2 contingency tables [18]. For Baveno stage, agreement between the algorithm and the clinical records were assessed using the kappa statistic. This is used to assess observer agreement for categorical variables and allows for agreement occurring by chance [19, 20].

RESULTS

Study Population

During the study period, 289 patients (median age 69, 79% male) with a new diagnosis of HCC were included (Table 2) and 249 (86.2%) of these had an inpatient record. Review of the

clinical record identified 191 (66%) of these as cirrhotic at HCC diagnosis, 50 (26%) of whom had evidence of previous decompensation. The median age of the cirrhotic group was 67 compared with 73 in the non-cirrhotic group (P < 0.001). An additional 15 patients had histological evidence of advanced fibrosis but cirrhosis was not mentioned explicitly in the clinical records. Among the patients who did not have an inpatient record, 12 had cirrhosis according to outpatient case note review. In the external validation cohort at RLBUHT, 50 patients meeting the inclusion criteria were assessed (median age 71, 82% male), 31 (62%) of whom were cirrhotic and 11 (35%) with previous decompensation.

Identification of Cirrhosis

Limiting the inclusion of episodes to those occurring before the HCC diagnosis results in a sensitivity of less than 50% for cirrhosis detection (Table 3). When additional episodes are included up until three years after the HCC diagnosis, the sensitivity increases to greater than 80% for all versions of the algorithm, without significant loss of specificity.

The sensitivity of algorithm 1 is increased by including ALD and AHF (version 2), and further increased by including ascites (version 3). However, the inclusion of ascites also reduces the specificity. This is overcome by limiting the inclusion of ascites to episodes that occurred before the HCC diagnosis (version 4). Using this optimised algorithm and including records up to three years post HCC diagnosis, the sensitivity is 86% (95% confidence interval, CI: 82%-90%) and the specificity is 98% (95% CI: 96%-100%), with a positive predictive value (PPV) of 99% and negative predictive value (NPV) of 79% (95% CI: 74%-83%) (Supplementary Table 1). For external validation, when version 4 of the algorithm was applied to the RLBUHT cohort with three years of follow-up, the sensitivity was 79% and specificity was 100%. Additionally, version 4 of the algorithm outperformed published algorithms for cirrhosis detection when they were applied to the LTHT cohort of HCC patients (Table 4).

Classification of Cirrhosis Severity

Table 5 shows the performance of the three versions of the algorithm for determining cirrhosis severity according to Baveno stage. Compared to version A, there is slightly less agreement between the calculated Baveno stage and the clinical record in version B, where Baveno stage 4 is defined by procedure codes for varices. Similarly, the sensitivity for

detecting decompensation (defined by Baveno stages 3 and 4) is increased in version B, but with reduced specificity (Supplementary Table 2). Agreement between the algorithm and the clinical record is optimised in version C, when bleeding varices are defined by a concurrent gastrointestinal haemorrhage code. Agreement was further improved when episodes occurring within 60 days of the registered HCC diagnosis were included. The performance characteristics of the component codes are summarised in Supplementary Tables 3 and 4; the sensitivity for detecting bleeding varices is increased in algorithm B, but the PPV and overall agreement with the Baveno stage is reduced due to the misclassification of non-bleeding varices. The sensitivity for detecting ascites is increased when both diagnosis and paracentesis procedure codes are included.

Using version C with a 60 day interval in the LTHT cohort, agreement between the clinical record and calculated Baveno stage was 84%, with a kappa coefficient of 0.74 (95% CI: 71%-77%). The sensitivity for detecting prior decompensation is 80% (95% CI: 75%-85%) and specificity is 98% (95% CI: 96%-100%), with a PPV of 89% (95% CI: 85% - 93%) and NPV of 96% (95% CI: 94%-98%). When this version was applied to the RLBUHT cohort for external validation, the agreement of Baveno stage with the clinical record was 81% (kappa 0.70). The sensitivity for detecting decompensation was 73% and specificity was 90%.

Finally, among the 167 LTHT patients identified as cirrhotic by the algorithm, 45 (27%) were coded with prior decompensation. At the time of HCC diagnosis, Child Pugh class and MELD scores were each higher in those individuals identified with decompensation (figure 1).

DISCUSSION

Main findings

This study demonstrates the reliability of an algorithm using inpatient HES records to identify and stage cirrhosis in patients with HCC. This is the first such algorithm validated in a UK population that uses only inpatient codes. Using inpatient codes from the whole follow-up period improves the sensitivity of the algorithm in cirrhosis identification, without loss of specificity. This validates the assumption that if a patient had an inpatient cirrhosis code during follow-up, they had cirrhosis at the time of HCC diagnosis. Using a broad definition of cirrhosis (versions 2-4) improves sensitivity and accounts for variations in coding practice in

which ALD and AHF are coded synonymously with cirrhosis. Excluding ascites after HCC diagnosis (version 4) improves the specificity; ascites in liver disease without HCC is most likely to be due to cirrhosis, whereas it may be malignant ascites in the context of HCC. Algorithm 4 is an improvement over published algorithms for cirrhosis detection when they are applied to our cohort of HCC patients. Algorithm C (for assessing cirrhosis severity) also outperformed published versions in this population. Inclusion of a concurrent gastrointestinal haemorrhage code alongside variceal procedures distinguishes between treatment of bleeding varices and treatment of non-bleeding varices for primary prevention.

Strengths and limitations

The strengths of this study are the systematic development of an algorithm which uses routinely available inpatient episode codes, and its applicability to large population studies in HCC. These patients often require hospital admission to manage complications in advanced cirrhosis to receive HCC therapies, or day case procedures such as paracentesis and endoscopy which are also coded in the HES dataset. The high performance characteristics (particularly the PPVs) derived from inpatient codes here are in part a consequence of the high pre-test probability of cirrhosis in patients with HCC. This observation is supported by the improved PPVs seen in existing algorithms in our cohort. In summary, this suggests that inpatient episodes are sufficient for high quality analyses of the impact of cirrhosis and its severity on the outcomes of patients with HCC.

This study benefits from robust case note evaluation, using both a development and external validation cohort. In the UK, previous validation of inpatient coding was achieved using free text analysis of primary care and death certification data [12], and the original case note validation of the cirrhosis algorithm included only 36 patients [21]. The algorithm benefits from exploiting the 'anchor point' of the HCC diagnosis date, so that inpatient codes can be associated with a time interval. This has led to optimised cirrhosis detection and severity classification. The algorithm for cirrhosis detection was optimised using three years of follow-up after HCC diagnosis, but the high sensitivity and specificity using one year of follow-up may be sufficient in some settings.

The limitations include its location in specialist cancer centres, which may not reflect coding practices throughout the UK and these may change over time. However, portal hypertensive

complications are common and often result in inpatient care and, since these are high cost procedures, we anticipate them to be reliably coded. The analysis was limited to patients local to the two centres, in order to capture cirrhosis-related episodes. Additional episodes may have been missed if patients were admitted elsewhere, but these would be captured by the algorithm when extended to a national dataset. The majority of patients had an inpatient record, suggesting high rates of hospital admission in patients with cirrhosis and those undergoing HCC treatment. The limitations of using inpatient codes alone are common to other studies which have utilised the linked inpatient HES dataset to produce impactful analyses [22, 23].

The proportion of patients with cirrhosis identified from their clinical records was 66% and this is lower than previous reports [2, 3]. By limiting to inpatient codes, the algorithm missed 12/191 (6.3%) patients with cirrhosis and those with histological evidence of advanced fibrosis were classified as non-cirrhotic. Many patients were diagnosed with HCC in the absence of known liver disease; 68.4% of those without cirrhosis had no known underlying liver disease aetiology (Table 2). If patients had advanced cancer at presentation their clinical record may not have explicitly stated the presence of cirrhosis. Additionally, they may have not been investigated further to establish a diagnosis of cirrhosis if not clinically appropriate. It is also notable that there was a high proportion of patients aged over 80 years who were not identified to have cirrhosis. Finally, the definition of decompensation using the Baveno IV classification is limited because it does not capture hepatic encephalopathy (HE), which may occur without variceal bleeding or ascites. HE can be coded in ICD10 code as “hepatic coma”, but we found that this was used uncommonly in our cohort and so we did not broaden our definition of decompensation beyond that used by Ratib and colleagues [12].

Implications

This algorithm can be applied to population cancer registries in the UK, enabling the identification and staging of cirrhosis in patients with HCC. This is essential for assessing clinical outcomes in population-based studies of individuals with HCC both in the UK and elsewhere. It is anticipated that this will lead to a better understanding of outcomes in HCC, including progression of underlying liver disease severity as well as overall survival. The

algorithm may also be used in other population-based applications which require the identification of cirrhosis and an assessment of severity.

In this study, we demonstrated the use of inpatient HES records to determine the cirrhosis severity at the time of HCC diagnosis. The algorithm may be adapted to classify the Baveno stage at different time intervals following HCC diagnosis or date of treatment, so that subsequent cirrhosis decompensation events can be identified over time. This approach is likely to have value in other health systems and we anticipate that the algorithm described will be evaluated by other investigators in outcomes oriented research in cirrhosis and HCC.

Conclusion

This study demonstrates the reliability of an algorithm based on inpatient EHRs to stratify patients with HCC according to the presence and severity of cirrhosis. It may be used in routine health data in order to assess outcomes in HCC in population studies.

Cirrhosis Diagnoses (ICD10):	**Codes**
Cirrhosis	K70.3, K71.7, K72.1, K74.4, K74.5, K74.6, K76.6, K72.1, K72.9
Alcoholic hepatic failure	K70.4
Alcoholic liver disease	K70.9
Ascites	R18.X
Varices	I85.9, I86.4, I98.2
Bleeding varices	I85.0, I98.3
Cirrhosis Treatments (OPCS4):	
Treatment of ascites	T46.1, T46.2, J06.1, J06.2
Treatment of varices	G10.4, G10.8, G10.9, G14.4, G17.4, G43.4, G43.7, J06.1, J06.2
Gastrointestinal Haemorrhage (ICD10):	
Gastrointestinal haemorrhage	K92.0, K92.1, K92.2

Table 1. Treatment and procedure codes included in the algorithm to determine cirrhosis status and cirrhosis severity.

		Total N (%)	No Cirrhosis N (%)	Cirrhosis N (%)	P-value
Characteristic:		289	98 (33.9%)	191 (66.1%)	
Age Group	<50	22 (7.6)	10 (10.2)	12 (6.3)	0.26
	50-59	49 (17.0)	10 (10.2)	39 (20.4)	0.04
	60-69	81 (28.0)	18 (18.4)	63 (33.0)	0.03
	70-79	92 (31.8)	31 (31.6)	61 (31.9)	0.95
	80+	45 (15.6)	29 (29.6)	16 (8.4)	<0.001
Sex	Male	228 (78.0)	76 (77.6)	152 (79.6)	0.83
	Female	61 (21.1)	22 (22.4)	39 (20.4)	0.73
Ethnicity	White	252 (87.1)	87 (88.8)	165 (86.4)	0.86
	Black	12 (4.2)	5 (5.1)	7 (3.7)	0.58
	South Asian	12 (4.2)	2 (2.0)	10 (5.2)	0.21
	Chinese	4 (1.4)	0	4 (2.1)	0.15
	Other Ethnic Group	4 (1.4)	1 (1.0)	3 (1.6)	0.70
	Not Stated	5 (1.7)	3 (3.1)	2 (1.0)	0.22
Aetiology	HCV	44 (15.2)	4 (4.1)	40 (20.9)	<0.001
	HBV	17 (5.9)	5 (5.1)	12 (6.3)	0.69
	PBC	7 (2.4)	0	7 (3.7)	0.06
	AIH	3 (1.0)	0	3 (1.6)	0.21
	Haemochromatosis	19 (6.6)	5 (5.1)	14 (7.3)	0.48
	Alcohol	68 (23.5)	4 (4.1)	64 (33.5)	<0.001
	NAFLD	43 (14.9)	13 (13.3)	30 (15.7)	0.60
	Other/ unknown	88 (30.4)	67 (68.4)	21 (11.0)	<0.001
MELD	<10			90 (47.1)	
	10-14			73 (38.2)	
	15-19			21 (11.0)	
	20+			7 (3.7)	
Child Pugh	A			131 (68.6)	
	B			44 (23.0)	
	C			16 (8.4)	
Previous Decompensation	Ascites			37 (19.3)	
	Variceal bleed			13 (6.8)	

Table 2. Baseline characteristics of the LTHT cohort. HCV = hepatitis C, HBV = hepatitis B, PBC = primary biliary cirrhosis, AIH = autoimmune hepatitis, NAFLD = non-alcoholic fatty liver disease.

	Algorithm 1 No Ascites - ALD - AHF				Algorithm 2 No Ascites + ALD + AHF				Algorithm 3 Ascites + ALD + AHF				Algorithm 4 Pre-HCC Ascites + ALD + AHF			
Time post HCC Diagnosis / days	Sens	95% CI	Spec	95% CI	Sens	95% CI	Spec	95% CI	Sens	95% CI	Spec	95% CI	Sens	95% CI	Spec	95% CI
0	0.45	0.39-0.51	1.00	1.00-1.00	0.47	0.41-0.52	1.00	1.00-1.00	0.49	0.43-0.54	1.00	1.00-1.00	0.49	0.43-0.54	1.00	1.00-1.00
30	0.52	0.47-0.58	1.00	1.00-1.00	0.54	0.49-0.60	1.00	1.00-1.00	0.57	0.51-0.63	0.99	0.98-1.00	0.57	0.51-0.63	1.00	1.00-1.00
60	0.60	0.55-0.66	1.00	1.00-1.00	0.64	0.58-0.69	1.00	1.00-1.00	0.66	0.61-0.72	0.98	0.96-1.00	0.66	0.61-0.71	1.00	1.00-1.00
90	0.66	0.61-0.72	1.00	1.00-1.00	0.70	0.65-0.75	1.00	1.00-1.00	0.73	0.68-0.78	0.97	0.95-0.99	0.72	0.67-0.77	1.00	1.00-1.00
120	0.69	0.64-0.74	1.00	1.00-1.00	0.73	0.68-0.78	1.00	1.00-1.00	0.75	0.70-0.80	0.97	0.95-0.99	0.75	0.70-0.80	1.00	1.00-1.00
150	0.72	0.67-0.77	1.00	1.00-1.00	0.76	0.72-0.81	1.00	1.00-1.00	0.79	0.74-0.83	0.96	0.94-0.98	0.78	0.73-0.83	1.00	1.00-1.00
180	0.73	0.68-0.78	0.99	0.98-1.00	0.77	0.73-0.82	0.99	0.98-1.00	0.80	0.75-0.84	0.95	0.92-0.97	0.79	0.74-0.84	0.99	0.98-1.00
365	0.76	0.71-0.81	0.99	0.98-1.00	0.81	0.76-0.85	0.99	0.98-1.00	0.83	0.78-0.87	0.94	0.91-0.97	0.82	0.78-0.87	0.99	0.98-1.00
730	0.80	0.76-0.85	0.98	0.96-1.00	0.84	0.80-0.88	0.98	0.96-1.00	0.87	0.83-0.91	0.95	0.93-0.98	0.85	0.81-0.89	0.98	0.96-1.00
1095	0.81	0.77-0.86	0.98	0.96-1.00	0.85	0.81-0.89	0.98	0.96-1.00	0.88	0.84-0.92	0.92	0.89-0.95	0.86	0.82-0.90	0.98	0.96-1.00
Total Follow-up	0.81	0.77-0.86	0.98	0.96-1.00	0.85	0.81-0.89	0.98	0.96-1.00	0.88	0.84-0.92	0.92	0.89-0.95	0.86	0.82-0.90	0.98	0.96-1.00

Table 3. Performance of different versions of the cirrhosis status algorithm. Sens = sensitivity, Spec = specificity, ALD = Alcoholic Liver Disease, AHF = Alcoholic Hepatic Failure. CI = confidence interval.

Algorithm	Sensitivity (%)	95% CI Lower	95% CI Upper	Specificity (%)	95% CI Lower	95% CI Upper	PPV (%)	95% CI Lower	95% CI Upper
Kramer *et al.*[7]	72	67	77	100	100	100	100	100	100
Jepsen *et al.*[8]	71	66	76	100	100	100	100	100	100
Nehra *et al.*[9]	80	76	85	98	96	100	99	97	100
Ratib *et al.*[15]	80	76	85	98	96	100	99	97	100
Algorithm 4	86	82	90	98	96	100	99	97	100

Table 4. Performance of different published algorithms for cirrhosis detection in the LTHT cohort of patients with HCC. PPV = positive predictive value. CI = confidence interval.

	Algorithm A Variceal bleeding codes		**Algorithm B Variceal bleeding codes or treatment codes**		**Algorithm C Variceal bleeding codes or treatment codes + UGIB**	
Time after HCC Diagnosis/ days	Correct Baveno Stage (%)	K-statistic	Correct Baveno Stage (%)	K-statistic	Correct Baveno Stage (%)	K-statistic
0	80	0.67	80	0.67	81	0.70
30	82	0.70	81	0.70	83	0.73
60	83	0.71	82	0.71	84	0.74
90	81	0.69	80	0.69	82	0.71
120	81	0.69	80	0.69	82	0.71

Table 5. Performance of different versions of the Baveno stage algorithm. UGIB = Upper gastrointestinal bleeding, K = kappa statistic.

Author Contributions
IR and RD had the original idea for the study and AB, AD, TC and EM contributed to its design and planning. RD, JS and VK performed the case note reviews. RD was responsible for data management, statistical analyses and wrote the first draft of the paper. IR reviewed the paper critically and VK, AB, JS, AD, TC and EM all approved the final version.

Funding
The research received no specific grants from any funding agency. RD and JS were funded by clinical research fellowships at Leeds Teaching Hospitals NHS Trust. AB is an analyst within Public Health England and receives funding from the British Association for the Study of the Liver (BASL) as part of the HCC-UK Partnership. No industrial sponsor had input into study planning, discussion of results or manuscript preparation.

Competing Interests
VK has received an unrestricted educational grant from Bayer, Bristol-Myers Squibb and Sirtex.

Patient Consent
Not applicable

Ethical approval
This retrospective study comprises an assessment of the accuracy of clinical coding for service evaluation and as such does not require formal ethical approval. All patient data were anonymised and permission was granted from the Caldicott Guardian for sharing of routinely-collected anonymised data.

Data Statement
Statistical code is available from the corresponding author RD

Acknowledgements
This work uses data provided by patients and collected by the NHS as part of their care and support. The planning of this project was discussed with members of the HCC-UK Partnership, which is co-ordinated by AB.

Word Count: 3303

REFERENCES

1. *Cancer Research UK. Liver Cancer Statistics*. 2015 [cited 2018 05/07/2018]; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/liver-cancer#heading-Zero.
2. El-Serag, H.B. and K.L. Rudolph, *Hepatocellular carcinoma: epidemiology and molecular carcinogenesis.* Gastroenterology, 2007. **132**(7): p. 2557-76.
3. D'Amico, G., G. Garcia-Tsao, and L. Pagliaro, *Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies.* Journal of Hepatology, 2006. **44**(1): p. 217-31.
4. El-Serag, H.B., *Hepatocellular carcinoma.* New England Journal of Medicine, 2011. **365**(12): p. 1118-27.
5. Coleman, M.P., et al., *Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data.* The Lancet, 2011. **377**(9760): p. 127-138.
6. Goutté, N., et al., *Geographical variations in incidence, management and survival of hepatocellular carcinoma in a Western country.* Journal of Hepatology, 2017. **66**(3): p. 537-544.
7. Kramer, J.R., et al., *The validity of viral hepatitis and chronic liver disease diagnoses in Veterans Affairs administrative databases.* Aliment Pharmacol Ther, 2008. **27**(3): p. 274-82.
8. Jepsen, P., et al., *Clinical course of alcoholic liver cirrhosis: a Danish population-based cohort study.* Hepatology, 2010. **51**(5): p. 1675-82.
9. Nehra, M.S., et al., *Use of administrative claims data for identifying patients with cirrhosis.* J Clin Gastroenterol, 2013. **47**(5): p. e50-4.
10. Goldberg, D., et al., *Validation of three coding algorithms to identify patients with end-stage liver disease in an administrative database.* Pharmacoepidemiol Drug Saf, 2012. **21**(7): p. 765-769.
11. Lapointe-Shaw, L., et al., *Identifying cirrhosis, decompensated cirrhosis and hepatocellular carcinoma in health administrative data: A validation study.* PLoS One, 2018. **13**(8): p. e0201120.
12. Ratib, S., et al., *1 and 5 year survival estimates for people with cirrhosis of the liver in England, 1998-2009: a large population study.* Journal of Hepatology, 2014. **60**(2): p. 282-9.
13. *EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma.* J Hepatol, 2012. **56**(4): p. 908-43.
14. Ratib, S., J. West, and K.M. Fleming, *Liver cirrhosis in England—an observational study: are we measuring its burden occurrence correctly?* BMJ Open, 2017. **7**(7).
15. Ratib, S., et al., *Diagnosis of Liver Cirrhosis in England, a Cohort Study, 1998–2009: A Comparison With Cancer.* The American Journal Of Gastroenterology, 2014. **109**: p. 190.
16. Leon, D.A. and J. McCambridge, *Liver cirrhosis mortality rates in Britain from 1950 to 2002: an analysis of routine data.* The Lancet, 2006. **367**(9504): p. 52-56.
17. de Franchis, R., *Evolving consensus in portal hypertension. Report of the Baveno IV consensus workshop on methodology of diagnosis and therapy in portal hypertension.* Journal of Hepatology, 2005. **43**(1): p. 167-76.

18. Alberg, A.J., et al., *The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests.* J Gen Intern Med, 2004. **19**(5 Pt 1): p. 460-5.
19. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* Biometrics, 1977. **33**(1): p. 159-74.
20. Cohen, J., *A coefficient of agreement for nominal scales.* Educational and psychological measurement, 1960. **20**(1): p. 37-46.
21. Fleming, K.M., et al., *Incidence and prevalence of cirrhosis in the United Kingdom, 1992–2001: A general population-based study.* Journal of Hepatology, 2008. **49**(5): p. 732-738.
22. Downing, A., et al., *Early mortality from colorectal cancer in England: a retrospective observational study of the factors associated with death in the first year after diagnosis.* Br J Cancer, 2013. **108**(3): p. 681-5.
23. Morris, E.J., et al., *A population-based comparison of the survival of patients with colorectal cancer in England, Norway and Sweden between 1996 and 2004.* Gut, 2011. **60**(8): p. 1087-93.

FIGURE LEGENDS

Table 6. Treatment and procedure codes included in the algorithm to determine cirrhosis status and cirrhosis severity.

Table 7. Baseline characteristics of the LTHT cohort. HCV = hepatitis C, HBV = hepatitis B, PBC = primary biliary cirrhosis, AIH = autoimmune hepatitis, NAFLD = non-alcoholic fatty liver disease.

Table 8. Performance of different versions of the cirrhosis status algorithm. Sens = sensitivity, Spec = specificity, ALD = Alcoholic Liver Disease, AHF = Alcoholic Hepatic Failure. CI = confidence interval.

Table 9. Performance of different published algorithms for cirrhosis detection in the LTHT cohort of patients with HCC. PPV = positive predictive value. CI = confidence interval.

Table 10. Performance of different versions of the Baveno stage algorithm. UGIB = Upper gastrointestinal bleeding, K = kappa statistic.

Figure 1. Box-and-whisker plots showing the distribution of MELD scores (A) and pie graphs showing the distribution of Child Pugh class (B) within compensated and decompensated cirrhosis groups determined by the algorithm.

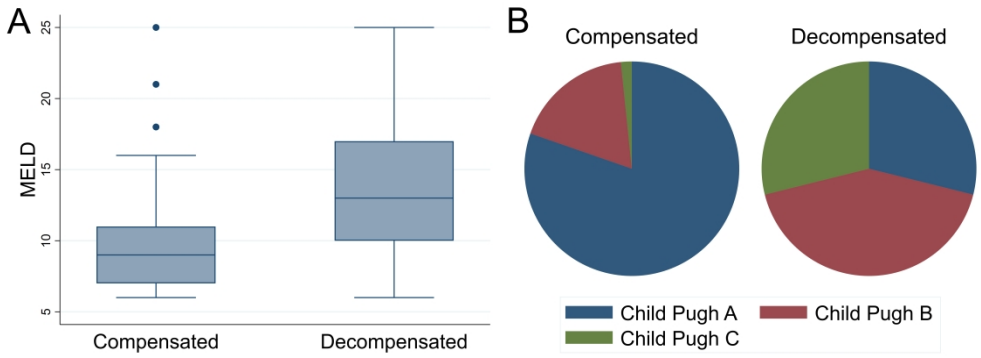

Figure 1. Box-and-whisker plots showing the distribution of MELD scores (A) and pie graphs showing the distribution of Child Pugh class (B) within compensated and decompensated cirrhosis groups determined by the algorithm.

Supplementary Tables

		True Status		
		Non-cirrhotic	Cirrhotic	Total
Cirrhosis Algorithm	Negative for Cirrhosis	96	26	122
	Positive for Cirrhosis	2	165	167
	Total	98	191	289

Supplementary Table 1. 2 x2 Contingency table for cirrhosis identification by optimised algorithm version 4 with three years of follow-up.

	Algorithm A Variceal bleeding codes				**Algorithm B Variceal bleeding codes or treatment codes**				**Algorithm C Variceal bleeding codes or treatment codes + UGIB**			
Time after HCC Diagnosis / days	Sens	95% CI	Spec	95% CI	Sens	95% CI	Spec	95% CI	Sens	95% CI	Spec	95% CI
0	0.74	0.69-0.79	0.98	0.97-1.00	0.78	0.73-0.83	0.96	0.94-0.98	0.76	0.71-0.81	0.98	0.96-1.00
30	0.76	0.71-0.81	0.99	0.97-1.00	0.80	0.75-0.85	0.82	0.78-0.86	0.78	0.73-0.83	0.98	0.96-1.00
60	0.78	0.73-0.83	0.99	0.97-1.00	0.82	0.76-0.86	0.96	0.94-0.98	0.80	0.75-0.85	0.98	0.96-1.00
90	0.78	0.73-0.83	0.98	0.96-1.00	0.82	0.76-0.86	0.95	0.93-0.98	0.80	0.75-0.85	0.97	0.95-0.99
120	0.78	0.73-0.83	0.98	0.96-1.00	0.82	0.76-0.86	0.95	0.93-0.98	0.80	0.75-0.85	0.97	0.95-0.99

Supplementary Table 2. Performance of different versions of the Baveno stage algorithm for predicting decompensation. Sens = sensitivity, spec = specificity, UGIB = Upper gastrointestinal bleeding.

	Algorithm A Variceal bleeding codes					Algorithm B Variceal bleeding codes or treatment codes					Algorithm C Variceal bleeding codes or treatment codes + UGIB				
Clinical Condition	Sens	95% CI	Spec	95% CI	PPV	Sens	95% CI	Spec	95% CI	PPV (%)	Sens	95% CI	Spec	95% CI	PPV
Varices	0.76	0.71-0.81	1.00	0.99-1.00	92	0.62	0.56-0.67	1.00	0.99-1.00	90	0.76	0.71-0.81	1.00	0.99-1.00	96
Bleeding Varices	0.31	0.25-0.36	1.00	1.00-1.00	80	0.92	0.89-0.95	0.96	0.94-0.99	54	0.62	0.56-0.67	0.99	0.97-1.00	67

Supplementary Table 3. Performance of different versions of the Baveno stage algorithm at 60 days post-HCC diagnosis for detecting varices and bleeding varices. Sens = sensitivity, spec = specificity, CI = confidence interval, PPV = positive predictive value.

	Ascites detection using Algorithm C (ICD10 codes and OPCS4 codes)					Ascites detection using ICD10 code R18.X only				
Clinical Condition	Sens	95% CI	Spec	95% CI	PPV (%)	Sens	95% CI	Spec	95% CI	PPV (%)
Ascites	0.73	0.68-0.78	0.98	0.96-0.99	73	0.57	0.51-0.62	0.98	0.97-1.00	84

Supplementary Table 4. Performance of ICD10 code R18.X for detection of ascites compared with the optimised algorithm C which includes additional OPCS4 codes for paracentesis (T46.1 and T46.2). Sens = sensitivity, spec = specificity, CI = confidence interval, PPV = positive predictive value.

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	1
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	2-3
	4	Study objectives and hypotheses	3
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	3-4
Participants	**6**	Eligibility criteria	3
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	3
	8	Where and when potentially eligible participants were identified (setting, location and dates)	3
	9	Whether participants formed a consecutive, random or convenience series	3
Test methods	**10a**	Index test, in sufficient detail to allow replication	4-6
	10b	Reference standard, in sufficient detail to allow replication	4-6
	11	Rationale for choosing the reference standard (if alternatives exist)	5-6
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	5-6
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	5-6
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	5-6
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	5-6
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	6
	15	How indeterminate index test or reference standard results were handled	5
	16	How missing data on the index test and reference standard were handled	5-6
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	6-7, suppl tables
	18	Intended sample size and how it was determined	5
RESULTS			
Participants	**19**	Flow of participants, using a diagram	-
	20	Baseline demographic and clinical characteristics of participants	7, Table 2
	21a	Distribution of severity of disease in those with the target condition	7, Table 2
	21b	Distribution of alternative diagnoses in those without the target condition	7, Table 2
	22	Time interval and any clinical interventions between index test and reference standard	-
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	8,9, Tables 2,4
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	8,9, Tables 2,4
	25	Any adverse events from performing the index test or the reference standard	-
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	9,10
	27	Implications for practice, including the intended use and clinical role of the index test	10
OTHER INFORMATION			
	28	Registration number and name of registry	-
	29	Where the full study protocol can be accessed	-
	30	Sources of funding and other support; role of funders	14

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.