**Performance of models to predict hepatocellular carcinoma risk among UK patients with cirrhosis and cured hepatitis C infection.**

AUTHORS: Hamish Innes1,2,3; Peter Jepsen3,4,5; Scott McDonald1,2; John Dillon6; Victoria Hamill1,2; Alan Yeung1,2; Jennifer Benselin7; April Went2; Andrew Fraser8,9; Andrew Bathgate10; M. Azim Ansari11; Stephen Barclay12; David Goldberg1,2; Peter C Hayes10; Philip Johnson13; Eleanor Barnes11; William Irving7; Sharon Hutchinson1,2\*; Indra Neil Guha7\*.

\*joint senior authors

AFFILIATIONS:

1. School of Health and Life Sciences; Glasgow Caledonian University. Glasgow UK.
2. Public Health Scotland, Glasgow, UK.
3. Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK.
4. Department of Hepatology and Gastroenterology, Aarhus University Hospital, Aarhus, Denmark
5. Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark
6. Division of Molecular and Clinical Medicine; School of Medicine, University of Dundee, UK.
7. NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, UK.
8. Aberdeen Royal Infirmary, Aberdeen, UK.
9. Queen Elizabeth University Hospital, Glasgow, UK
10. Royal Infirmary of Edinburgh, Edinburgh, UK
11. Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine and the Oxford NIHR Biomedical Research Centre, Oxford University, UK.
12. Glasgow Royal Infirmary, Glasgow, UK.
13. Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, UK.

CONFLICTS OF INTEREST: There are no conflicts of interest to disclose.

FINANCIAL SUPPORT: This study has three main sources of financial support: a) Medical Research Foundation (Grant ID:C0825); b) Public Health Scotland; and c) Medical Research Council (Grant ID: MR/K01532X/1). For more detailed information, please see our “Funding and acknowledgements” paragraph.

AUTHOR CONTRIBUTIONS: a) Study concept: all authors; b) study design: HI, JP, SH, ING, WI, EB, SM; c) acquisition of data: HI, WI, SH, ING, EB; d) resources: All authors; e) Statistical analysis: HI, JP, SM; f) drafting manuscript: HI, JP; g) critical revision of manuscript: all authors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

WORD COUNT: 5,850 words (including abstract; introduction, methods, results, discussion; references and figure/table legends).

CORRESPONDING AUTHOR INFORMATION:

Dr Hamish Innes; Glasgow Caledonian University, George Moore Building, Room M403A; Cowcaddens Road; G4 0BA. Tel:0141-5332950; Email:Hamish.innes@gcu.ac.uk

ABBREVIATIONS (order of appearance):

aMAP, age-Male gender-ALBI-platelet count score; HCV, hepatitis C Virus; THRI, Toronto HCC Risk Index; VHA, Veteran Health Affairs; HCC, hepatocellular carcinoma; PNPLA3, Patatin-like phospholipase domain-containing protein 3; STOPHCV, STratified medicine to Optimise Treatment of Hepatitis C Virus; SMR06, Scottish cancer register; SMR01, Scottish inpatient hospital admission database; ICD, international classification of diseases; SVR, sustained virological response; C-index, Concordance index.

DATA AVAILABILITY STATEMENT:

The Scottish data used in this study are not publicly available, but can be acquired through successful application to the Public Benefit and Privacy Panel for Health and Social Care (<https://www.informationgovernance.scot.nhs.uk/pbpphsc/home/for-applicants/> ). The STOPHCV consortium welcomes collaboration with interested parties. Anonymised samples and clinical data held on the study database are accessible through application to the HCVRUK Tissue Data Access Committee. Contact Professor Indra Neil Guha (neil.guha@nottingham.ac.uk) or Professor William Irving (will.irving@nottingham.ac.uk) for more information. Please note that we cannot share linked NHS digital data with other research groups, or any data variables derived from linked NHS digital data.

**ABSTRACT:**

BACKGROUND:

Hepatocellular carcinoma (HCC) prediction models can inform clinical decisions about HCC screening provided their predictions are robust. We conducted an external validation of six HCC prediction models for UK patients with cirrhosis and a hepatitis C virus (HCV) virological cure.

METHODS:

Cirrhosis patients with cured HCV were identified from the Scotland HCV clinical database (N=2139) and the STOPHCV study (N=606). We calculated patient values for four competing non-genetic HCC prediction models, plus two genetic models (for STOPHCV cohort only). Follow-up began at the date of SVR achievement. HCC diagnoses were identified through linkage to nation-wide cancer, hospitalisation and mortality registries. We compared discrimination and calibration measures between prediction models.

RESULTS

Mean follow-up was 3.4-3.9 years, with 118 (Scotland) and 40 (STOPHCV) incident HCCs observed. The aMAP (Age-Male gender-Albi score-Platelet count) model showed the best discrimination; e.g. C-index in Scottish cohort was 0.77 (95%CI:0.73-0.81). However, for all models, discrimination varied by cohort (better for Scottish cohort) and by age (better for younger patients). Also, genetic models performed better in HCV genotype 3 patients.

The observed 3-year HCC risk was 3.3% (95%CI: 2.6-4.2) and 5.1% (3.5-7.0%) in the Scottish and STOPHCV cohort respectively. These were most closely matched by aMAP, where the mean predicted 3-year risk was 3.6% and 5.0% in the Scottish and STOPHCV cohorts, respectively.

CONCLUSIONS:

aMAP was the best performing model in terms of both discrimination and calibration and should be used as a benchmark for rival models to surpass. This study underlines the opportunity for “real world” risk stratification in cirrhosis patients with cured HCV. However, auxiliary research is needed to help translate a HCC risk prediction into a HCC screening decision.

LAY SUMMARY: Patients with cirrhosis and cured hepatitis C are at high risk of liver cancer- but the risk varies substantially from one patient to the next. Risk calculator tools can alert clinicians to high risk patients and thereby influence decision-making. In this study we tested the performance of 6 risk calculators in more than 2500 patients with cirrhosis and cured hepatitis C. We show that some risk calculators are considerably better than others. Overall, we found that the “aMAP" calculator worked the best, but more work is needed to convert predictions into clinical decisions.

KEYWORDS:

Prognosis; Risk prediction; Primary liver cancer; External validation; Genetic risk scores.

**INTRODUCTION**

Patients with hepatitis C virus (HCV) related cirrhosis remain at high risk of hepatocellular carcinoma (HCC) after a virological cure4-6, which does not appear to diminish over time.7 HCC has amongst the worst five-year survival probabilities of any cancer.8 However, if detected at an early stage - i.e. when curative treatments can be administered - five-year survival can exceed 70%.8 The current standard-of-care-for early HCC detection is biannual abdominal ultrasound surveillance with or without alpha fetoprotein. 9,10 Although existing clinical guidelines recommend this intervention for all cirrhosis patients after HCV eradication, there is growing recognition that a more targeted approach is needed – i.e. where clinicians focus their finite resources on patients who stand to benefit the most from surveillance.11,12

It is against this backdrop that HCC prediction models are now emerging that can estimate a patient’s risk of developing HCC from routine data. Currently such scores include the Albi-male-age-platelet count (aMAP)13, the Toronto HCC risk Index (THRI)14, models derived from the US Veteran Health Affairs (VHA) cohort15, and models from the French prospective ANRS-CO12 Cir Vir cohort.16 In addition, two genetic prediction models for HCC have recently been published,17-19 drawing on common genetic polymorphisms such as the rs738409 variant in *Patatin-like phospholipase domain-containing protein 3 (PNPLA3)*.20

To enhance clinical decision making, it is crucial that HCC prediction models are able to accurately predict HCC risk in a given patient. Inaccurate predictions have the potential to do harm. For example, underestimating HCC risk could lead to higher-risk patients being denied biannual ultrasound screening, and vice versa, overestimating HCC risk could lead to unnecessary screening in lower risk patients.

At present, there are uncertainties regarding the performance of existing HCC prediction models. First, the acid-test of a prediction model’s accuracy is validation in a cohort that is independent from the one used to “train” the model (known as external validation).21 Studies show that model performance is systematically better when measured on the same dataset used to train the model, versus when measured on an “unseen” dataset,22 Existing HCC prediction models have not been rigorously validated in external cohorts (with the exception of the aMAP13) and thus their performance could be overly optimistic. Second, studies have not adopted a competing risk perspective when evaluating model performance. This may be important because cirrhosis patients are at high risk of dying from causes unrelated to HCC, such as liver failure23 and non-HCC cancer24; failing to take this into account could lead to biased estimates of prognosis25. Third, the question of whether model performance is the same for all patients, or if it varies according to clinical characteristics, has not been explored. Fourth, genetic prediction models have practical advantages over non-genetic models (e.g. risk score is constant over time, and hence only needs to be measured once). However, it is not clear how they compare performance-wise to their non-genetic counterparts. With these issues in mind, the aim of this study was to investigate the performance of selected HCC prediction models for patients with cured HCV cirrhosis in two separate UK cohorts.

**METHODS:**

HCC PREDICTION MODELS:

This study focuses on six HCC prediction models that are suitable or potentially suitable for patients with cirrhosis after HCV virological cure.

These six models were: aMAP (2020);13 THRI (2018);14 VHA cirrhosis SVR score (2018);15 ANRS CO12 CirVir score (2017);16 Dongiovanni et al Genetic risk score (2020);17,18 and Gellert-Kristensen et al Genetic risk score (2020).19

For each prediction model, we extracted information from published articles relating to the following aspects of model derivation: sample size, average duration of follow-up, number of HCCs observed, proportion with HCV aetiology; specific prognostic factors selected and the discrimination performance reported see Table S1-S2 and Appendix A).

DATA SOURCES FOR EXTERNAL VALIDATION:

Model performance was assessed on cirrhosis SVR patients from two UK cohorts, both followed from the date of SVR achievement.

STOPHCV COHORT

The STratified medicine to Optimise Treatment of Hepatitis C Virus (STOP-HCV) cirrhosis cohort is a prospective cohort study of 1255 patients with liver cirrhosis and a history of chronic HCV infection. Participants were recruited from 31 liver clinics across the UK (except Northern Ireland) in Jan 2015-July 2016- i.e. coinciding with the introduction of direct acting antivirals. Cirrhosis was defined on the basis of: a) Histological assessment (Ishak 5/6 or Metavir 4); OR b) imaging results consistent with cirrhosis, including Fibroscan greater than 15kPa; OR c) validated serum biomarker consistent with cirrhosis (including APRI>2 and ELF test>10.48). Detailed clinical and laboratory information were collected on participants at the time of study enrolment and during subsequent annual study visits. Participants also provided a blood sample at enrolment which was used to generate genotyping information using the Affymetrix UK Biobank array, which directly characterises individuals with respect to more than 800,000 genetic variants. Participants have also been linked to health registries covering England and/or Wales, including the Hospital Episodes Statistics Admitted patients care dataset; cancer registrations collected by Public Health England; and death registrations. The study was approved by the West Midlands Research Ethics Committee (application reference: 14/WM/1128); Informed consent was obtained from all participants.

SCOTTISH HCV CLINICAL DATABASE

The Scottish HCV clinical database has been described extensively elsewhere.2,26,27 It is a retrospective cohort study of c.25,000 patients in Scotland who have attended a specialist liver clinic appointment for care/management of chronic HCV infection. The database records information collected during routine clinical care including antiviral treatment episodes, diagnosis of cirrhosis and the results of laboratory tests. It is also linked routinely to national health registries in Scotland, including the hospital, mortality and cancer registers. Approval to link these registries and perform data analysis was granted by the Privacy Public Benefit Panel for Health and Social Care in NHS Scotland (application number: 1516-0457).

Liver cirrhosis was defined as compensated or decompensated cirrhosis diagnosed during routine clinical investigation. Diagnoses were typically made following liver biopsy; transient elastography; abdominal ultrasound; clinical examination; and routine liver function tests, according to clinical guidelines at that time. Of note, no information on genetic risk factors/polymorphisms were available in the Scottish cohort.

INCLUSION CRITERIA:

For both cohorts, we included all patients with cirrhosis prior to initiating antiviral therapy and who subsequently achieved SVR. All SVRs were included irrespective of antiviral treatment regimen. If a patient had more than one treatment episode resulting in SVR, then the first episode was selected.

EXCLUSION CRITERIA:

Of those satisfying our aforementioned inclusion criteria, we then excluded patients as follows.

For the STOPHCV cohort, we first excluded participants recruited from Scottish and Welsh clinics where linkage data to health registries were unavailable or not complete. This exclusion also ensured there was no patient overlap between the STOPHCV and Scottish cohorts. Second, we excluded participants with a diagnosis of HCC prior to completing antiviral therapy. Third, we excluded participants who had already achieved an SVR at the time of STOPHCV study enrolment. This exclusion was applied to prevent immortal time bias28 between SVR achievement and STOPHCV study enrolment.

For the Scottish cohort, we excluded individuals with a history of HCC prior to achieving SVR. No other exclusions were made.

HCC RISK PREDICTIONS:

Many of the prognostic factors included in the non-genetic HCC risk prediction models are dynamic insofar as they change over time. For prognostic factors based on laboratory tests (i.e. albumin, platelet count etc), we selected the most recent test on or prior to the start of antiviral treatment. Tests conducted more than twelve months before initiating treatment were excluded. We considered using laboratory tests conducted up to twelve months prior to SVR achievement to align with the date of follow-up commencement (see later paragraph on “definition of risk sets”), however we decided against this because antiviral treatment can cause acute and temporary changes in liver blood test values that may not necessarily reflect long-term risk profile. Age however, was based on age at the time of SVR achievement (i.e. time zero). Information on Gamma Glutamyl transferase was not available in the STOPHCV cohort, precluding calculating values for the ANRS C012 CirVir model.16

PRIMARY OUTCOME EVENT

The primary outcome event was diagnosis of hepatocellular carcinoma (HCC), identified through linkage to relevant administrative health databases. Specifically, for the STOP-HCV study, we used data from the England Admitted Patient Care database, the national cancer registry and mortality register to identify incident cases of HCC.

For the Scottish HCV clinical database, we used the equivalent National inpatient hospital admission database (SMR01), cancer registry (SMR6) and mortality register to identify HCC cases. For all registries/administrative databases, we used the standard ICD10 “C22.0” or ICD9:155.0 code in the primary diagnostic/cause of death position to define HCC.

STATISTICAL ANALYSIS:

DEFINITION OF RISK SETS

All statistical analyses were underpinned by survival analysis methods. Follow-up time began at the date of SVR achievement. This was defined as six months after the treatment completion date for episodes initiated before the year 2014 (i.e. SVR24); and three months after the treatment completion date for episodes initiated from 2014 onwards (i.e. SVR12). This aligns with how SVR was defined by clinicians during the time period of this study. Follow-up ended at the date of incident HCC (if at all), mortality (if at all) or the date of study completion. For both cohorts, the study completion date was 1st Jan 2020, corresponding to the date the hospital admission registers were complete to. Unless indicated otherwise, non-HCC mortality was treated as a competing risk in all analyses.

MULTIPLE IMPUTATION

Multiple imputation was used to replace missing data for the individual components of each HCC prediction model with plausible imputed estimates.29 We generated 20 imputations for each missing prediction using predictive mean matching. The following variables were used to predict these imputed values: a) the Nelson-Aalen estimate of the baseline cumulative hazard, b) the outcome variable (i.e. HCC status), c) gender, d) decompensated cirrhosis, e) age, f) alcohol use and g) type of antiviral treatment (i.e. IFN-based or not). Rubin’s rules were used to combine statistics of interest across imputation datasets.30 Similarly, cumulative incidence curves by risk tertile are based on the average estimate across the 20 imputation datasets created. Risk “tertiles” refer to three groups; a) those whose prediction is in the 33rd percentile or lower; b) those in the 33-67th percentile, and c) those in the 68-100th percentile.

PREDICTION MODEL PERFORMANCE

Each prediction model was assessed in terms of two main aspects of prognostic model performance21:

* Discriminative ability (i.e. ability to differentiate between patients who develop HCC and those who do not).
* Calibration (agreement between the 3-year risk of HCC predicted by the model versus the 3-year HCC risk actually observed).

DISCRIMINATION

The discriminative ability of each HCC prediction model was investigated in three ways.

First, we assessed the discrimination of each model visually, by plotting the cumulative incidence of HCC for individuals with low, moderate and high scores. Categorisation into low, medium and high groups was based on risk tertiles, as described earlier. Cumulative incidence was computed non-parametrically using the “stcompet” command within Stata v16.31 Non-HCC mortality was treated as the competing risk event.

Secondly, we determined each prediction model’s discriminative ability quantitatively using the Concordance-index (C-index), which provides an overall summary of a risk score’s discriminative ability. Specifically, the C-index measures the proportion of all possible “participant-pairs” that are “concordant”. A “participant pair” refers to a random selection of two individuals from the dataset, and this pair is said to be “concordant” if the individual with the higher risk score develops the outcome event of interest sooner than the individual with the lower risk score.32 In our base-case analysis, we used a version of the C-index adapted for a competing-risk scenario, as previously described by Wolbers et al.33 The key difference between the standard C-index and competing-risk adjusted C-index is that in the latter, individuals with a competing risk event are assumed to have an infinite survival time. In addition, we also calculated the standard Harrell C-index which does not account for competing risks. For all versions of the C-index, higher values indicate better discrimination; a value of 0.5 indicates zero discrimination (i.e. no better than chance), whilst a value of 1.0 indicates perfect discrimination.

Thirdly, we assessed if the C-index of each prediction model varied according to selected patient characteristics. These characteristics were as follows: age<60years; gender; history of heavy alcohol use (defined as consumption of >50 units/week for a sustained period of >6 months before SVR); genotype 3; and SVR through IFN-free therapy.

CALIBRATION:

Calibration measures how closely the predicted risk of HCC matches the observed risk of HCC.21,34 For this analysis, we calculated the 3-year predicted and 3-year observed risk of HCC for individuals with low, moderate and high predictions (again defined according to risk tertiles).

The predicted 3-year probability of HCC was calculated using standard Cox regression, as prescribed by the authors of each risk score. Namely:1-S0(t)exp(linear predictor)

Where t=3 years, and S0(t) refers to the estimated three-year HCC-free survival for individuals with zero for all independent variables in the model. We contacted the authors for this information, if these details were not clear in the original paper.

Our calculation of the 3-year observed HCC probability was based on the cumulative incidence function with non-HCC mortality treated as a competing risk.

Finally, we did not perform a calibration analysis for the genetic models, as they were not intended to estimate the probability of HCC at a particular point in time.

**RESULTS**

DERIVATION OF EXTERNAL VALIDATION COHORTS

2245 patients met our inclusion criteria from the Scottish cohort. We then excluded 106 patients with HCC prior to treatment completion. Thus, the final sample size was 2,139 (See Figure S1).

1019 participants from the STOPHCV study met our inclusion criteria. We then excluded 79 patients from Scotland and Wales. Secondly, 77 patients with HCC prior to SVR achievement were also excluded. Finally, a further 257 patients who achieved SVR before enrolling into STOPHCV were removed to avoid immortal time bias. Thus, the final sample size was 606 (see Figure S1).

PATIENT CHARACTERISTICS:

Patients in both cohorts were mainly middle-aged (i.e. between 40-65 years old), male (>70%) and of White ethnicity (>80%) (Table 1). However, there were notable differences between these two cohorts. Firstly, patients in the STOPHCV cohort were older than the Scottish cohort (mean age: 56.5 vs 50.2 years). Secondly, the proportion of patients that had achieved SVR through IFN-free therapies was higher in the STOPHCV that the Scottish cohort (92% achieved SVR via IFN-free therapies in STOPHCV versus 61% in Scottish cohort). Thirdly, the proportion of patients with past HCV genotype 3 infection was lower in the STOPHCV cohort (38% versus 50%). Finally, average values for the VHA, THRI and aMAP scores were all higher in the STOPHCV cohort versus the Scottish cohort, indicating that STOPHCV had higher predicted HCC risk.

The proportion of patients with missing predictions was generally <20%. However, missing data was more substantial in the Scottish cohort for the VHA model (24% missing) and the ANRS C012 model (60% missing).

CUMULATIVE INCIDENCE OF HCC AND NON-HCC MORTALITY:

In the Scottish cohort, participants were followed for a mean 3.9 years after SVR, during which time 118 incident HCC events and 214 non-HCC related deaths occurred (Table 2). The cumulative incidence of HCC and non-HCC mortality at 3 years was 3.3% (95% CI: 2.6-4.2) and 8.5% (95%CI: 7.2-9.8), respectively (Table 2 & Figure 1).

STOPHCV patients were followed for a mean of 3.4 years after SVR, during which time 40 incident HCCs and 36 non-HCC deaths occurred (Table 2). The cumulative incidence of HCC and non-HCC mortality at 3 years was 5.1% (95% CI: 3.5-7.0), and 5.0% (95% CI: 3.5-7.0), respectively (Figure 1).

Drug-related mortality and deaths from external causes were more common in the Scottish cohort versus STOPHCV. One-third of non-HCC mortality was from drug-related or external causes in the Scottish cohort, compared to only 10% in the STOPHCV study (Table 2).

PERFORMANCE OF HCC PREDICTION MODELS:

DISCRIMINATION

In cumulative incidence plots, higher predicted HCC risks were associated with a higher HCC cumulative incidence (Figures S2-3). However, the degree of discrimination varied considerably by prediction model and also by cohort.

In the Scottish cohort, the aMAP score exhibited the best discrimination (C-index: 0.771; 95% CI: 0.731-0.810), followed by the VHA model (0.715; 95% CI:0.668-0.761), THRI (0.719; 95% CI: 0.673-0.764), and ANRS CO12 (0.703; 95% CI: 0.656-0.749). (Figure 2)

HCC prediction models exhibited poorer discriminative performance (i.e. lower C-index values) in the STOPHCV cohort, but the general ranking was similar. For example, aMAP was also the top performing score in STOPHCV (C-index: 0.701; 95% CI: 0.638-0.764); followed by VHA (0.657; 95% CI: 0.576-0.737), followed by THRI (0.648; 95% CI:0.577-0.718). (Figure 4). The Dongiovanni GRS had a C-index value of 0.613 (95% CI:0.530-0.695), and the Gellert-Kristensen GRS C-index value was 0.559 (95% CI:0.473-0.645). (Figure 4). All C-index values were marginally higher when using the standard Harrell’s C-index as opposed to the Wolbers-modified C-index. (Table S3).

VARIABILITY IN DISCRIMINATION

Our analysis of variability in model discrimination highlighted two patient factors of interest (Figures S4-S5). First, discrimination was better for younger patients versus older patients. This was apparent across both cohorts for all non-genetic models. In the Scottish cohort for example, the aMAP had a C-index of 0.59 (95%CI: 0.49-0.70) for those age>60 years at SVR achievement versus 0.80 (95%CI:0.75-0.84) for those aged<60 years (Figure 3). Secondly, GRS discrimination was better for patients with past genotype 3 infection. For example, the C-index of the Dongiovanni GRS was 0.78 (95%CI:0.70-0.87) in genotype 3 patients versus 0.50 (95%CI:0.39-0.62) in non-genotype 3 patients.

Otherwise, no major heterogeneity in model discrimination was observed according to IFN-free therapies, alcohol history, gender, or decompensated disease. (Figures S4-5).

CALBIRATION:

In the Scottish cohort, the observed 3-year probability of HCC was 3.3% (95% CI:2.6-4.2). This compared with predicted probabilities of 2.0% (THRI), 3.1% (VHA), 3.6% (aMAP) and 3.9% (ANRS CO12 model). In STOPHCV, the 3-year observed probability of HCC was 5.1% (95%CI: 3.5-7.0), compared to predicted probabilities of 2.5% (THRI model), 3.8% (VHA model) and 5.0% (aMAP model) (Figure S6).

When we examined calibration according to risk tertiles, we saw some instances of under-prediction in higher risk patients. For example, in the Scottish cohort, the observed 3-year risk for individuals whose THRI score was in tertile 3 (11.3% 95%CI:4.6-18.0) was almost twice the predicted risk (6.4%). This under-prediction affected the VHA model as well to some extent, but did not affect the aMAP or ANRS CO12 models (Figure 4).

**DISCUSSION**

HCC risk prediction models have the potential to support clinical decision making, but equally could cause harm if their predictions are not robust. In this study, we used external validation to quantify the performance of existing HCC prediction models for individuals with cirrhosis and cured HCV. There are three key findings to draw attention to. First, our data confirm that HCC prediction models are able to discriminate between patients who go onto develop HCC and those who do not. In other words, across all models, an increase in *predicted* HCC risk was mirrored by an increase in *observed* HCC risk (and vice versa). Nevertheless, not all models provide the same level of discrimination in a UK setting. Overall, the aMAP model exhibited the best discriminative ability with a C-index of 0.78 in the Scottish and 0.71 in STOPHCV. The aMAP model is derived entirely from routinely available prognostic factors – i.e. age, albumin, bilirubin, platelet count and gender – and thus this provides encouragement regarding opportunities for “real world” risk stratification in this growing patient group. Another corollary is that aMAP should be used as a benchmark for rival prognostic models to surpass. This will help the research community evaluate whether a proposed new model (of which many are likely to emerge in the years ahead) provides added-value over existing alternatives. A second novel aspect of this study is that it highlights the existence of heterogeneity in model performance – i.e. variability in model performance according to patient characteristics. For example, we found that most prediction models were more discriminating in younger patients versus older patients (Figure 3), although we are not clear exactly why this is the case. In a similar vein, we show that the Dongiovanni et al GRS exhibited far better discrimination for patients with past genotype 3 infection (C-index: 0.78) that for those with non-genotype 3 infection (C-index: 0.50). This could be because the Dongiovanni GRS was originally developed as a risk score for hepatic steatosis, which is well-known to be a more prominent histological feature of HCV genotype 3 infection versus genotype 2/3 infection.35,36 A third important observation from this study is to caution that some models may under predict 3-year HCC risk. This was most prominent for the THRI model among higher risk patients. Thus, re-calibration may be necessary before adopting this model in a UK setting (albeit, we acknowledge that underpredicting high risk patients would probably not alter HCC screening decisions).

An important question that this study doesn’t answer is how to translate a HCC risk prediction into a HCC screening decision for a given patient. There is agreement in the field that HCC risk prediction models will have most clinical utility for identifying patients whose risk of HCC is too low for screening to be of net benefit. However, there is considerable ambiguity regarding which patients are “low risk” and how this should be defined. In our view, the definition of low risk should reflect a compromise between multiple factors, such as: a) cost-effectiveness data; b) general population HCC incidence; c) patient preferences; d) clinical view and other clinical factors (e.g. likelihood of receiving curative treatment in the event of a HCC diagnosis); and e) resources available for HCC screening. In the real world, “low risk” is likely to represent a range of values rather than a hard threshold, and it is unlikely to be the same for all patients. It will also inevitably change as new surveillance technologies emerge with different performance characteristics to abdominal ultrasound. Thus, to support HCC screening decisions, versatile models are needed with good calibration across the risk spectrum. This is why we focused on calibration in this study. We deliberately avoided defining “low risk” based on what is optimal for a given model – i.e. which previous studies have done by identifying the risk threshold at which the sensitivity/specificity are optimised. This approach is statistically dubious37, but more to the point, it is equivalent to letting a statistical model dictate a clinical decision, as opposed to using a statistical model to help implement a clinical decision. Thus, auxiliary research to define “low risk” may be needed before models such as aMAP can be confidently deployed. Microsimulation Markov models38 may be useful for estimating the benefits of screening (i.e. in terms of life-years or quality-adjusted-life-years gained) according to 3-year HCC probability.

This study has several strengths. Firstly, our focus on externally validating competing models fills an important gap in the literature. i.e. most previous studies have opted to develop new risks models, rather than evaluate the performance of existing ones. Secondly, as previously discussed, we have assessed model performance not only in terms of discrimination, but calibration too. A third strength is that our estimates of model performance account for non-HCC mortality as a competing risk. This perspective is important because cirrhosis patients are at high risk of mortality from liver failure, and this may bias estimates of model performance.25 However, whilst we found that C-indexes were lower when accounting for competing risks, the differences were very modest. A fourth strength is the adoption of a dual cohort perspective, enabling us to perform the same analysis in two different cohorts and analyse variability. This has supported our investigation of heterogeneity in model performance. Another unique asset of this study is that we have collected data on genetic and non-genetic models, and have thus been able to compare the discriminative ability of these two model types. Our study has limitations too that warrant discussion. One of the main limitations is that predictions were missing for some patients. Whilst the proportion missing data was generally low (<20%), missing data was more substantial for the ANRS-CO12 (60% missing in Scottish cohort) and VHA models (24% missing in Scottish cohort). We used multiple imputation to maximise statistical power and correct potential bias from a complete-case analysis. Nevertheless, the performance of the ANRS-CO12 model in particular should be viewed with caution in light of the missing data. Second, we cannot exclude the possibility that some of the patients in our dataset may have been had HCC or been developing HCC before SVR was achieved. Third, we were not able to evaluate all models developed so far for HCV cirrhosis patients, including those proposed by Pons et al39, Audurea et al40 and Alonso Lopez et al.41 These scores were omitted from our analysis because data for factors such as liver stiffness and prothrombin time were unavailable in the Scottish and STOPHCV studies. This is also an inherent weakness of the scores themselves insofar as a model can only be useful if it can be calculated using “real world” data. Fourth, although patients were followed up from the point of SVR achievement, we did not have information on the specific date that the SVR test was performed. Thus, a conservative estimate of six months after treatment completion was used to ensure our analysis was not affected by immortal time bias28 (i.e. equivalent to SVR24).

In summary, this is the first study comparing the performance of competing HCC risk prediction models. Our findings highlight the opportunities for practical HCC risk-stratification in a UK setting for patients with cirrhosis and cured HCV. If models are to support HCC screening decisions however, then a consensus will ultimately be needed regarding the individualised probability of HCC at which screening should be avoided.

**REFERENCES:**

[1] Carrat F, Fontain H, Dorival C, Simony M, Diallo A, Hezode C, et al. Clinical outcomes in patients with chronic hepatitis C after direct-acting antiviral treatment: a prospective cohort study. Lancet 2019;393:1453-1464.

[2] McDonald SA, Pollock KG, Barclay ST, Goldberg DJ, Bathgate A, Bramley P, et al. Real-word impact following initiation of interferon-free hepatitis C regimens on liver-related outcomes and all-cause mortality among patients with compensated cirrhosis. J Viral Hepat. 2020;27:270-280.

[3] Sahakyan Y, Lee-Kim V, Bremner KE, Bielecki JM, Krahn MD. Impact of direct-acting antiviral regimens on mortality and morbidity outcomes in patients with chronic hepatitis c: Systematic review and meta-analysis. J Viral Hepat. 2021;28:739-754.

[4] El-Serag HB, Kanwal F, Richardson P, Kramer J. Risk of hepatocellular carcinoma after sustained virological response in veterans with hepatitis C virus infection. Hepatology. 2016;64:130-7.

[5] Innes H, Barclay ST, Hayes PC, Fraser A, Dillon JF, Stanley A, et al. The risk of hepatocellular carcinoma in cirrhotic patients with hepatitis C and sustained viral response: Role of the treatment regimen. J hepatol 2018;68:646-654

 [6] van der Meer AJ, Feld JJ, Hofer H, Almasio PL, Calvaruso V, Fernandez-Rodriguez CM, et al. Risk of cirrhosis-related complications in patients with advanced fibrosis following hepatitis C virus eradication. J Hepatol. 2017; 66:485-493.

[7] Ioannou GN, Beste LA, Green PK, Singal AG, Tapper EB, Waljee AK, et al. Increased risk for hepatocellular carcinoma persists up to 10 years after HCV eradication in patients with baseline cirrhosis or high FIB-4 scores. Gastroenterology. 2019;157:1264-1278.

[8] Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Niksic M, et al. Global surveillance of trends in cancer survival 2000-2014 (CONCORD-3): analysis of individual records for 37 513025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet 2018;391:1023-1075.

[9] EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma.. J hepatol. 2012;56:908-943.

[10] Heimbach JK, Kulik LM, Finn RS, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology 2018;67:358–380.

[11] Kanwal F, Singal AG. Surveillance for hepatocellular carcinoma: current best practice and future direction. Gastroenterology. 2019;157:54-64

[12] Fujiwara N, Friedman SL, Goossens N, Hoshida Y. Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine. J Hepatol. 2018;68:526-549

[13] Fan R, Papatheodoridis G, Sun J, Innes H, Toyoda H, Xie Q, et al. aMAP risk score predicts hepatocellular carcinoma development in patients with chronic hepatitis. J Hepatol. 2020;73:1368-1378.

[14] Sharma SA, Kowgier M, Hansen BE, Brouwer WP, Maan R, Wong D, et al. Toronto HCC risk index: A validated scoring system to predict 10-year risk of HCC in patients with cirrhosis. J Hepatol. 2018;68:92-99.

[15] Ioannou GN, Green PK, Beste LA, Mun EJ, Kerr KF, Berry K. Development of models estimating the risk of hepatocellular carcinoma after antiviral treatment for hepatitis C. J Hepatol. 2018;69:1088-1098.

[16] Ganne-Carrie N, Layese R, Bourcier V, Cagnot C, Marcellin P, Guyader D, et al. Nomogram for individualised prediction of hepatocellular carcinoma occurrence in hepatitis C virus cirrhosis (ANRS CO12 CirVir). Hepatology. 2016;64:1136-1147.

[17] Dongiovanni P, Stender S, Pietrelli A, Mancina RM, Cespiati A, Petta S, et al. Causal relationship of hepatic fat with liver damage and inuslin resistance in non-alcoholic fatty liver. Journal of Internal Medicine. 2018;283:356-370.

[18] Degasperi E, Galmozzi E, Pelusi S, D’ambrosio R, Soffredini R, Borghi M. et al. Hepatic fat – genetic risk score predicts hepatocellular carcinoma in patients with cirrhotic HCV treated with DAAs. Hepatology. 2020;72:1912-1923.

[19] Gellert-Kristensen H, Richardson TG, Davey-Smith G, Nordestgaard BG, Tybjaerg-Hansen A, Stender S. Combined effect of PNPLA3, TM6SF2 and HSD17B13 variants on risk of cirrhosis and hepatocellular carcinoma in the general population. Hepatology. 2020;72:845-856.

[20] Trepo E, Romeo S, Zucman-Rossi J, Nahon P. PNPLA3 gene in liver diseases. J Hepatol. 2016;399-412.

[21] Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162:W1-73.

[22] Collin GS, de Groot JA; Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol. 2014;14:40. doi: 10.1186/1471-2288-14-40.

[23] D’Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. J Hepatol. 2006;44:217-31.

[24] Hong CY, Sinn DH, Kang D, Paik SW, Guallar E, Cho J, Gwak GY. Incidence of extrahepatic cancers among individuals with chronic hepatitis B or C virus infection: A nationwide cohort study. J viral Hepat. 2020;27:896-903.

[25] Jepsen P, Vilstrup V, Andersen PK. The clinical course of cirrhosis: the importance of multistate models and competing risks analysis. Hepatology. 2015;62:2922-302.

[26] McDonald SA, Hutchinson SJ, Innes HA, Allen S, Bramley P, Bhattacharyya, et al. Attendance at specialist hepatitis clinics and initiation of antiviral treatment among persons chronically infected with hepatitis C: examining the early impact of Scotland’s hepatitis C action plan. J Viral Hepat. 2014;21:366-76.

[27] McDonald SA, Innes HA, Hayes PC, Dillon JF, Mills PR, Goldberg DJ, et al. What is the impact of a country-wide scale-up in antiviral therapy on the characteristics and sustained viral response rates of patients treated for hepatitis C? J Hepatol. 2015;62:262-268.

[28] Suissa S. Immortal time bias in pharmaco-epidemiology. Am J Epidemiol. 2008;167:492-9.

[29] Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:B2393.

[30] Rubin D. Multiple imputation for nonresponse in surveys. New York: Wiley, 1987.

[31] Enzo Coviello, 2003. "[STCOMPET: Stata module to generate cumulative incidence in presence of competing events](https://ideas.repec.org/c/boc/bocode/s431301.html)," [Statistical Software Components](https://ideas.repec.org/s/boc/bocode.html) S431301, Boston College Department of Economics, revised 11 Nov 2012.

[32] Pencina MJ, D’Agostino RB, Song L. Quantifying discrimination of Framingham risk functions with different survival C statistics. Stat Med. 2012;31:1543-53.

[33] Wolbers M, Koller MT, Witteman JCM, Steyerberg EW. Prognostic models with competing risks. Methods and application to coronary risk prediction. Epidemiology. 2009;20:555-561.

[34] Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol. 2013;13:33.

[35] Asselah T, Rubia-Brandt L, Marcellin P, Negro F. Steatosis in chronic hepatitis C: why does it really matter. Gut 2006;55:123-130.

[36] Mihm S, Fayyazi, Hartmann H, Ramadori G. Analysis of histopathological manifestations of chronic hepatitis C virus infection with respect to virus genotype. Hepatology. 1997;25:735-9.

[37] Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ. 2006;332:1080.

[38] Innes H, Goldberg D, Dusheiko G, Hayes P, Mills PR, Dillon JF, et al. Patient-important benefits of clearing the hepatitis C virus through treatment: a simulation model. J Hepatol. 2014;60:1118-26.

[39] Pons M, Rodriguez-Tajes S, Esteban JI, Marino Z; Vargas V; Lens S; et al. Non-invasive prediction of liver-related events in patients with HCV-associated compensated advanced chronic liver disease after oral antivirals. J Hepatol. 2020;72:472-480.

[40] Audureau E, Carrat F, Layese R, Cagnot C, Asselah T, Guyader D, et al. Personalised surveillance for hepatocellular carcinoma in cirrhosis- using machine learning adapted to HCV status. J Hepatol. 2020;73;1434-1445

[41] Lopez SA, Manzano ML, Gea F, Gutierrez ML, Ahumada AM, Devesa MJ, et al. A model based on noninvasvie markers predicts very low hepatocellular carcinoma risk after viral response in hepatitis C virus-advanced fibrosis. Hepatology. 2020;72:1924-1934.

**TABLE AND FIGURE LEGENDS**

Table 1: Description of the Scottish and STOP-HCV cohorts at treatment initiation.

Table 2: Description of follow-up data and outcome events observed in the Scottish and STOP-HCV cohorts.

FIGURE 1: Stacked cumulative incidence (CI) curves for HCC and non-HCC mortality.

Cumulative incidence curves are generated non-parametrically – i.e. without any modelling assumptions. For the blue line, non-HCC mortality is treated as a competing risk event. Vice versa, for the red line, HCC outcome is treated as a competing risk event

FIGURE 2: Discriminative ability of HCC prediction models in Scottish and STOPHCV cohorts, in terms of the Concordance Index (C-index).

Concordance index refers specifically to the Wolbers concordance index, which takes account of competing risk events. Here, non-HCC mortality is treated as a competing risk. The dashed line represents the point of zero discrimination.

FIGURE 3; Comparison of the discriminative ability for HCC incidence, according to age.

Based on the Wolbers concordance index, taking account of non-HCC mortality as a competing risk.

FIGURE 4: Agreement between observed and predicted 3-year HCC probability, by risk tertile.

T1, T2 and T3 denote risk tertiles 1, 2 and 3, respectively. Risk tertiles refer to three groups: a) T1: those whose prediction is in the 33rd percentile or lower; b) T2: those in the 33rd -67th percentile and c) T3: those in the 68th-100th percentile. The green line indicates perfect agreement between observed and predicted risk. Values above the green line indicate that the observed risk is higher than the predicted risk (and vice versa).

**FUNDING AND ACKNOWLEDGEMENTS:**

We would like to acknowledge funding from Public Health Scotland.

We would like to thank the participants of the STOP-HCV cirrhosis study.

HCV Research UK participating sites and Principal Investigators are as follows:

Agarwal K. (Kings College Hospital, London), Aldersley M. (St James’s University Hospital, Leeds),

Aspinall R. (Queen Alexandra Hospital, Portsmouth), Barclay S. (Glasgow Royal Infirmary), Barnes E. (John Radcliffe Hospital, Oxford), Benselin J. (University of Nottingham), Brown A. (St Mary’s Hospital, London), Ch’ng C. (Singleton Hospital, Swansea), Corless L. (Hull Royal Infirmary), Cramp M. (Derriford Hospital, Plymouth), Dillon J. (Ninewells Hospital, Dundee), Shirley English (Aberdeen Royal Infirmary); Forton D. (St George’s Hospital, London), Foster G. (The London Hospital, Fraser A. (Aberdeen Royal Infirmary), Gelson W. (Addenbrookes Hospital, Cambridge), Gorard D. (Wycombe Hospital), Gordon F. (Bristol Royal Infirmary), Kennedy N. (Monklands

Hospital), Knowles J. (James Cook University Hospital, Middlesbrough), Leen C. (Western General Hospital, Edinburgh McPherson S. (Freeman Hospital, Newcastle), Moreea S. (Bradford Teaching Hospitals NHS Foundation Trust), Mutimer D. (Queen Elizabeth Hospital, Birmingham), Prince M. (Manchester Royal Infirmary, Richardson P. (Royal Liverpool University Hospital), Rosenberg W. (Royal Free Hospital & University College Hospital), Ryder S. (Queen’s Medical Centre, Nottingham) Kara Rye (Royal Shrewsbury Hospital), Stone B. (Royal Hallamshire Hospital, Sheffield), Thursz M. (St Mary’s Hospital, London),, Ustianowski A. (North Manchester General Hospital), Verma S. (Royal Sussex County Hospital, Brighton), ,Wiselka M. (Leicester Royal Infirmary).

**STOP-HCV Consortium**: Barnes E (University of Oxford), Ball JK (University of Nottingham), Brainard D (Gilead Sciences), Burgess G (Conatus Pharmaceuticals), Cooke G (Imperial College, London), Dillon J (University of Dundee), Foster GR (Queen Mary University of London), Gore C (Hepatitis C Trust), Guha N (University of Nottingham), Halford R (Hepatitis C Trust), Whitby K (Gilead Sciences), Holmes C(University of Oxford), Howe A (British Columbia Centre for Excellence), Hudson E (University of Oxford), Hutchinson S (Glasgow Caledonian University), Irving WL (University of Nottingham), Khakoo S (University of Southampton), Klenerman P (University of Oxford), Martin N (UC San Diego), Massetto B (Gilead Sciences), Mbisa T (Public Health England), McHutchinson J (Gilead Sciences), McKeating J (University of Oxford), McLauchlan J (Centre for Virus Research, Glasgow), Miners A (London School of Hygiene and Tropical Medicine), Murray A (OncImmune Limited), Shaw P (Merck & Co), Simmonds P (University of Oxford), Spencer C (Wellcome Trust), Thomson E (Centre for Virus Research, Glasgow), Vickerman P (University of Bristol), Zitzmann N (University of Oxford).

HCV Research UK was established by a grant from the Medical Research Foundation (award no: C0365).

The STOP-HCV study was funded by a grant from the Medical Research Council, United Kingdom (grant MR/K01532X/1). EB is an NIHR Senior Investigator.

Hamish Innes is supported by a viral hepatitis fellowship from the Medical Research Foundation (award no: C0825).

The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.











