Statistical perspectives on using hepatocellular carcinoma risk models to inform surveillance decisions

Innes, Hamish; Nahon, Pierre

Published in:
Journal of Hepatology

DOI:
10.1016/j.jhep.2023.05.005

Publication date:
2023

Document Version
Author accepted manuscript

Statistical perspectives on using hepatocellular carcinoma risk models to inform surveillance decisions.

Hamish Innes[1,2,3] and Pierre Nahon [4]

Affiliations:

1. School of health and life sciences, Glasgow Caledonian University, Glasgow UK.
2. Lifespan and population health. University of Nottingham, Nottingham UK.
3. Public Health Scotland, Glasgow UK.
4. APHP, Liver Unit, Bobigny; Université Sorbonne Paris Nord, F-93000 Bobigny; Inserm, UMR-1138 “Functional Genomics of Solid Tumors”, Centre de recherche des Cordeliers, Université de Paris, Paris, France.

Word count: 2,770 (excluding references, abstract and figure legends).

Funding: HI is supported by a Viral Hepatitis Fellowship from the Medical Research Foundation (grant ID: C0825). PN has received research grants from AstraZeneca, AbbVie, Bristol–Myers Squibb and Eisai. Pierre Nahon’s research is funded in part by the European Union (GENIAL, Grant agreement ID: 101096312) and French Agence Nationale de la Recherche (France 2030 DELIVER ANR-21-RHUS-0001)

Author contributions: PN and HI have contributed equally in terms of the concept of the article and the writing and editing of the manuscript.

Correspondence: Address correspondence to: Dr Hamish Innes; Glasgow Caledonian University; Cowcaddens Road; G40BA; Glasgow, UK. Hamish.Innes@gcu.ac.uk

Conflicts of interest: Hi has no conflicts of interest to declare. PN has received honoraria from and/or consults for AstraZeneca, AbbVie, Bayer, Bristol-Myers Squibb, Eisai, Gilead,

Ipsen, MSD and Roche. He received research grants from AstraZeneca, AbbVie, Bristol–Myers Squibb and Eisai.

Summary:

More than 50,000 people are diagnosed with hepatocellular carcinoma (HCC) every year in Europe. Many cases are known to specialist liver centres years before they present with HCC. Despite this, HCC is usually detected at a late stage, when prognosis is very poor. For more than two decades, clinical guidelines have recommended uniform surveillance for all cirrhosis patients. However, studies continue to show that this broad-based approach is inefficient and poorly implemented in practice. A “personalised” approach, where the surveillance regimen is customised to the needs of the patient, is gaining growing support in the clinical community. The cornerstone of personalised surveillance is the HCC risk model – a mathematical equation predicting a patient’s individualised probability of developing HCC within a specific time window. However, although numerous risk models have now been published, few are being used in routine care to inform HCC surveillance decisions. In this article, we discuss methodological issues stymieing the use of HCC risk models in routine practice - highlighting biases, evidence gaps and misconceptions that future research must address.

Personalised surveillance:

For more than two decades, we have used crude decision rules to distribute HCC surveillance [1,2] (Figure 1). It is widely assumed that surveillance programmes will become more personalised in the years ahead, improving efficiency and patient outcomes. [3-5]

The basic idea behind personalised surveillance is to: a) consider the benefits, harms and costs of surveillance at the level of the individual patient; and then b) assign each patient to their optimal screening regimen on the basis of these considerations. In this way, one treats each patient as an individual, not as a group average.

In our opinion, opportunities to personalise surveillance are about to rapidly expand, as new technologies for early HCC detection emerge and as chronic liver diseases (CLDs) epidemiology evolves. [6] However, to truly leverage these opportunities, we must develop a framework for matching individuals to their optimal surveillance regimen. This is the topic of this article, and central to this is the concept of individualised HCC risk.

Individualised risk:

We are used to thinking about average HCC risks derived from a cohort of heterogeneous patients. However, the HCC risk for any specific patient can deviate starkly from the group average [7]; hence why average risks are not a good basis for making decisions about individuals.

An *individualised* risk is a risk estimate tailored to a specific individual. For example, an individualised 3-year HCC risk of 5% means a 5% chance a patient will develop HCC within the next 3 years (or conversely, a 95% chance they will not). Individualised HCC risk is estimated from a HCC risk model, which can be thought of as a simple input-process-output device (Figure 1). It takes information about a patient such as their age and platelet count (input) and through an explicit mathematical formula (process), converts this into an estimate of individualised risk (output).

As a community, we expect that individualised HCC risk will be the cornerstone of personalised surveillance, directly influencing what surveillance regimen the patient receives. [3-5] As such, the development and validation of HCC risk models has become an active area of research with numerous models now available to clinicians. [8-11] In our view however, there are several barriers to using these models in an individualised screening context which are not widely recognised.

1. Low model quality and applicability

HCC risk models aim to improve patient care, yet they could equally cause harm if their predictions are biased. A key concern is overfitting, where the model is fitted to the noise in a dataset rather than the true signal. Overfitted models generate predictions that are too extreme – i.e. too low for lower risk patients and too high for higher risk patients [12].

The best defence against overfitting is to develop models from large longitudinal datasets with many incident HCC cases occurring over a sufficient long follow-up. Despite some exceptions however, most HCC risk models have been produced from small datasets, including several with <5 events per prognostic parameter (Figure 2). For these models, we should not expect their predictions to be reliable, even for a set of patients who closely resemble the dataset used to train the model.

Second, virtually all models to-date have been developed using Cox regression [8-11], which does not account for competing risk events which are particularly numerous in patients affected by CLDs and include non-HCC liver-related and extra-hepatic mortality. [13] A recent analysis suggests the impact of competing risk bias would be greatest if the model was being used to identify high risk patients – e.g. with a view to offering more costly surveillance modalities such as abbreviated MRI, as a more pronounced CLD impacts both HCC occurrence and liver failure/portal hypertension. [14]

Moreover, few models cater to patients with NAFLD or alcohol-related liver disease (Figure 2) – even though these aetiologies account for most HCC cases in the West, reflecting the lower investment in the past decades in the constitution of prospective observational cohorts as compared with viral hepatitis.

A comprehensive review is needed, but in our view, the methodological standard of HCC risk models must be improved if we want them to be used in the clinic.

2. Convenience external validation:

HCC risk models are normally developed from a just tiny fraction of the “at-risk” population. External validation (EV) means assessing model performance on “new” patients who are eligible to use the model but were not part of the development dataset. It can be a critical step towards clinical implementation providing EV cohorts are selected thoughtfully and strategically. However, in our experience, EV cohorts tend to be chosen more for reasons of convenience than with a clear implementation strategy in mind.

A crucial distinction is the difference between a model’s reproducibility and its transportability. [15] A model is reproducible if its performance can be replicated in cohorts which closely resemble the development cohort. Conversely, a model is transportable if the performance can be replicated in patients that deviate in some way from the patients on which the model was developed (e.g. patients with a different underlying chronic liver disease aetiology). In practice, if a HCC risk model were to be used successfully over a reasonably broad geographical area, it would need to exhibit good transportability as well as reproducibility. This is because a wide geographical area will typically include multiple health providers with diverse patient case-mixes – i.e. some that will closely resemble the development cohort (hence requiring reproducibility), but also others that may differ appreciably in terms of ethnicity, deprivation or aetiology for example (hence requiring transportability). Even if a model was intended for local use only (e.g. from a single provider/liver centre), transportability would still be required because patient case-mixes normally evolve over time (N.B. the Meld score is a case-in-point vis-à-vis how time trends in case-mix can gradually diminish model performance [16]).

Thus, when planning an EV study, we believe the following points are paramount:

a) be explicit about the target population of intended use.
b) consider the epidemiology of HCC in that target population, including regional heterogeneity and secular trends in “at risk” patients.

c) approach EV strategically, choosing EV cohorts that test those aspects of transportability relevant to the target population of interest.

Stratifying model performance by fundamental variables (e.g. age, sex, ethnicity) can also give insight into transportability if sample size permits. A good example relates to cured HCV cirrhosis, where HCC risk models discriminate much better in younger patients than older patients [17] This is valuable information from an implementation perspective – e.g. it suggests the performance of these models would gradually decline over time as the HCV cirrhosis population ages.

3. Model discrimination: letting the perfect be the enemy of the good

A HCC risk model must be able to discriminate between individuals who go onto develop HCC from those who do not. A model with good discrimination will assign higher predicted risks to patients who develop HCC versus patients who do not, whereas in a poorly discriminating model, the predicted risks will be similar between these two groups.

One of the most widely used metrics to quantify discriminative ability is Harrel's concordance index (C-index), which has a minimum possible value of 0.5 (zero discrimination) and a maximum possible value of 1.0 (perfect discrimination). To many, the C-index is a "black box" metric that is therefore difficult to interpret. But its clinical relevance can be easily illustrated through simulation analyses (Figure 4).

Suppose we have a cohort of cirrhosis patients with cured hepatitis C. Let us assume the average five-year risk of HCC in this cohort is 10.5%, based loosely on a recent meta-analysis. [18] Imagine our goal was to use a HCC risk model to differentiate patients in this cohort with a 5-year risk HCC risk <15% versus ≥15%. Possibly, this could relate to offering abbreviated MRI surveillance to the latter group but not to the former – a decision rule suggested by a recent health economics analysis [19]. Figure 1 shows that if a model has zero discrimination (i.e. C-index =0.50), all patients are assigned the risk of the group

average (i.e. 10.5%), and as a result, everyone falls into the <15% group. However, as the C-index increases, the model gets better at separating these two risk profiles, and stark differences in HCC incidence emerge. Thus, a greater C-index means greater prognostic separation between risk strata, and from a personalised surveillance perspective, prognostic separation is key because it is the premise one has for treating the two groups differently. After all, if the incidence of HCC was comparable between the <15% and >15% risk groups, then there would be no justification for providing abbreviated MRI to the latter whilst withholding it from the former.

Another point implied by Figure 4 is that discrimination is a question of degree, rather than an attribute a model either does or doesn't have. This begs the question: what *degree* of discrimination does a HCC risk model need to have to be clinically useful and improve the status quo? EASL guidelines suggest a C-index of ≥0.80 is required [20], implying that existing models -which generally exhibit C-indexes of 0.70 to 0.80 [8-11] - are inadequate for clinical use. However, Figure 4 suggest ≥0.80 is essentially an arbitrary threshold. In our view, there is no good reason why a model must have a minimum C-index of 0.80 to support personalised surveillance. We must remember that a model does not need to be perfect to be useful, and that clinical utility should be judged from impact studies (discussed later), not from the C-index alone.

4. Conflating predictions and decisions:

Even if we are satisfied with our ability to estimate individualised HCC risk reliably, we forget that individualised risk is only a means to an end, not an end in itself. The next step is to develop a decision rule, mapping a patient's individualised risk to their optimal surveillance regimen (Figure 1).

Most HCC risk models group patients into "low"; "moderate" and "high" risk categories using statistical criteria (e.g. optimising sensitivity and specificity via the Youden index). Usually

there is an implication that patients assigned to the “low” risk group do not need screening, which raises controversy from both clinical and ethical standpoints. In reality, optimal cut-offs are simply not an appropriate basis for making surveillance decisions. [21] For starters, they assume sensitivity and specificity are equally important (even though we know sensitivity trumps specificity when surveying cirrhosis patients for HCC). Moreover, they tend to have very wide uncertainty intervals which are rarely acknowledged.

Health economics can provide valuable information about decision rules for personalised surveillance. For example, a previous cost-effectiveness analysis (CEA) suggests biannual ultrasound becomes cost-effective once the HCC incidence rate exceeds 1.3% per year. [21], implying ultrasound should only be offered to patients exceeding this threshold. [3] However, we should remember that CEAs were developed to inform population-level decision-making; caution is required if extrapolating their results to an individualised context. For example, CEAs generally factor low adherence into their estimates, which will augment the minimum HCC incidence required for cost effectiveness. [22, 23] Although accounting for low adherence is appropriate for population-level analyses, it is more contentious if extrapolated to the individual level – i.e. it could lead to motivated patients being denied surveillance on the tacit assumption their adherence will be poor. Another issue is that the data on quality-of-life used in CEAs are sparse or are out-of-date and not generalisable to contemporary patients. [22, 23] Without better input data, CEA-based thresholds may not have the face-validity to win clinician and patient support.

In our view, we should be thinking about developing decision rules via stakeholder consensus (i.e. between patients, clinicians and commissioners). One approach is to use the delphi technique – weighing up the trade-offs, cost effectiveness, affordability, opportunity costs, resource implications etc of various candidate decision rules. Transparency is crucial; the process taken to develop the decision rule will be as important as the decision rule itself.

5. Unchallenged assumptions about the centrality of individualised risk:

Individualised HCC risk is expected to be the cornerstone of personalised surveillance algorithms, directly pinpointing a patients' optimal surveillance regimen. In general, lower risk patients are expected to be assigned cheaper and less effective surveillance, whereas higher risk patients will be assigned more costly and more effective regimens.[3] The rationale is that screening higher risk patients will yield a greater absolute number of early stage HCCs, and therefore more patients treated with curative intent. In this way, the higher costs associated with more effective screening modalities can be offset by the greater benefit accrued. However, note that we are tacitly assuming the benefit of early HCC detection will be the same for high and low risk patients. This may not be correct. A recent study suggests that patient characteristics associated with higher individualised HCC risk (i.e. older age and advanced cirrhosis) are also associated with reduced odds of curative-intent HCC treatment due to higher rates of comorbidities or impaired liver function precluding allocation to liver resection or ablation. [24] Thus, detecting more HCCs early, may not necessarily translate into treating more patients with curative-intent.

This leads to broader questions about the adequacy of using individualised HCC risk alone to match patients to their optimal surveillance regimen. For example, it has been suggested that HCC risk could be used to assign patients to their optimal surveillance *interval* (i.e. more frequent surveillance for higher risk patients). [4]. However, the optimal screening interval depends on the tumour doubling time, and to our knowledge, there is no reason to assume a HCC arising in a low risk patient will grow at a different speed to one arising in a high risk patient. Thus, the validity of using individualised HCC risk to infer a patient's optimal screening interval requires further justification. We also know that specific surveillance modalities are substandard in some patient groups (e.g. lower performance of ultrasound due to impaired liver visualization in obese patients), and these subgroups are not entirely distinguishable by their individualised risk. This again suggests that in some cases, factors

beyond individualised risk may need to be considered to identify a patient's optimal surveillance regimen.

6. Lack of impact data:

In general, we assume personalised surveillance will be non-inferior (at minimum) to current practice in terms of patient outcomes. Supporting evidence from impact studies may be required to substantiate this expectation. The purpose of an impact study is to determine the effect of a decision rule on clinician behaviour, patient behaviour, and subsequent patient outcomes. [25]

Evaluating the impact of a HCC surveillance decision rule in a prospective study in terms of surrogate measures of HCC mortality – e.g. early HCC detection rates, allocation to curative-intent treatment– would be challenging, but arguably achievable. Decision modelling analyses may also be useful as a first port of call to estimate the impact of a surveillance decision rule and what assumptions impact is sensitive to.[26]

We can also learn from prominent impact studies being conducted outside of liver disease, such as the WISDOM study, evaluating risk stratified breast cancer surveillance to usual care in the United States. [27]

7. Scalability challenges:

To date, most HCC risk models are available simply as web-based calculators, requiring clinicians to gather data for each prognostic factor manually. To be used at scale however, risk calculation would need to be automated, drawing on routine data collected in electronic health records. Yet, routine data is often incomplete and unstructured (e.g. free text data). For some scores, complete automation may not even be possible or may require advanced data extraction techniques. [28] Moreover, data may be missing for one or more prognostic

factor(s), thereby precluding an estimate of individualised risk from being made. How should these cases be managed in practice? Ideally, imputation of missing predictor values should be part of the automation process, but there is still uncertainty about how this can be done in real time.

It is right that we have prioritised understanding the validity, performance and clinical impact of HCC risk models – however, given the large number of at-risk patients, we must begin now to also think about the operational challenges of using HCC risk models at scale.

Conclusions:

The rise of the HCC risk model represents a major step towards personalising HCC surveillance. At the same time, our article draws attention to factors that will hinder their implementation in routine practice. Many of these challenges are statistical issues as much as clinical ones and they tend to be under-recognised or misunderstood by the research community. Box 1 lists some key recommendations for future research; by channelling our efforts strategically into these areas today, we will be better placed to harness the opportunities of personalised surveillance tomorrow.

REFERENCES:

[1] Bruix J, Sherman M, Llovet JM, Beaugrand M, Lencioni R, Burroughs AK, et al. Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona. J Hepatol. 2001;35:421-430.

[2] European Association for the Study of the Liver. EASL Clinical practice guidelines: management of hepatocellular carcinoma. J Hepatol. 2018; 69:182-236.

[3] Ioannou GN. HCC surveillance after SVR in patients with F3/F4 fibrosis. J Hepatol. 2021;74:458-465.

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

[5] Nahon P, Quang WV, Ganne-Carrie N. Stratification of hepatocellular carcinoma risk following HCV eradication or HBV control. J Clin Med. 2021;10:353. Doi: 10.3390/jcm10020353.

[6] Singal AG, Reig M, Villanueva A. Emerging tools for hepatocellular carcinoma surveillance. Am J Gastroenterol. 2022. *In press*

[7] Innes H, Hamill V, McDonald SA, Hayes PC, Johnson P, Dillon JF, et al. Comparing predicted probability of hepatocellular carcinoma in patients with cirrhosis with the general population: an opportunity to improve risk communication. Am J Gastroenterol. 20122;117:1454-1461.

[8] 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;733:1368-1378.

[9] Ioannou GN, Green P, Kerr KF, Berry K. Models estimating risk of hepatocellular carcinoma in patients with alcohol or NAFLD-related cirrhosis for risk stratification. J Hepatol. 2019;71:523-533.

[10] Semmler G, Meyer EL, Kozbial K, Schwabl P, Hametner-Schreil S, Zanetto A, et al. HCC risk stratification after cure of hepatitis C in patients with compensated advanced chronic liver disease. J Hepatol. 2022;76:812-821.

[11] 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.

[12] Riley RD, Snell KIE, Harrel Jr FE, Martin GP, Reitsma JB, Moons KGM, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;18:368. Doi: 10.1136/bmj.m441.

[13] 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.

[14] Innes H, Johnson P, McDonald SA, Hamill V, Yeung A, et al. Competing risk bias in prognostic models predicting hepatocellular carcinoma occurrence: impact on clinical decision-making. Gastro Hep Advances. 2022;1:129-136. Doi: https://doi.org/10.1016/j.gastha.2021.11.008

[15] Debray TPA, Vergouwe Y, Koffijberg H, Nieboier D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol. 2015;68:279-289.

[16] Godfrey EL, Malik TH, Lai JC, Mindikoglu AL, Galvan NTN, Cotton RT, et al. The decreasing predictive power of MELD in an era of changing etiology of liver disease. Am J Transplant.2019;19:3299-3307.

[17] Innes H, Jepsen P, McDonald S, Dillon J, Hamill V, Yeung A, et al. Performance of models to predict hepatocellular carcinoma risk among UK patients with cirrhosis and cured HCV infection. JHEP Rep. 2021;3:100384. Doi: 10.1016/j.jhepr.2021.100384

[18] Lockart I, Yeo MGH, Hajarizadeh B, Dore GJ, Danta M. HCC incidence after hepatitis Ca cure among patients with advanced fibrosis or cirrhosis: a meta-analysis. Hepatology. 2022. doi: 10.1002/hep.32341

[19] Nahon P, Najean M, Layese, R, Zarca K, Segar LB, Cagnot C, et al. Early hepatocellular carcinoma detection using magnetic resonance imaging is cost-effective in high-risk patients with cirrhosis. Jhep Rep. 2022;4: doi: https://doi.org/10.1016/j.jhepr.2021.100390

[20] European Association for the Study of the Liver. EASL Clinical practice guidelines on non-invasive tests for evaluation of liver disease severity and prognosis- 2021 update. J Hepatol. 2021;75:659-689.

[21] Wyants L, van Smeden M, McLernon DJ, Timmerman D, Steyberg EW, Calster BV. Three myths about risk thresholds for prediction models. BMC Med. 2019;17:192.

[22] Zangneh HF, Wong WWL, Sander B, Bell CM, Mumtaz K, Kowgier M, et al. Cost effectiveness of hepatocellular carcinoma surveillance after a sustained virologic response to therapy in patients with hepatitis C virus infection and advanced fibrosis. Clin Gastroenterol Hepatol. 2019;17:1840-1849.

[23] Parikh ND, Singal AG, Hutton DW, Tapper EB. Cost-effectiveness of hepatocellular carcinoima: An assessment of benefits and harms. Am J Gastroenterol. 2020;115:1642-1649.

[24] Hamill V, Gelson W, MacDonald D, Richardson P, Ryder SD, Aldersley M, et al. Delivery of biannual ultrasound surveillance for individuals with cirrhosis and cured hepatitis C in the UK. Liver Int. 2023. Doi 10.1111/liv.15528.

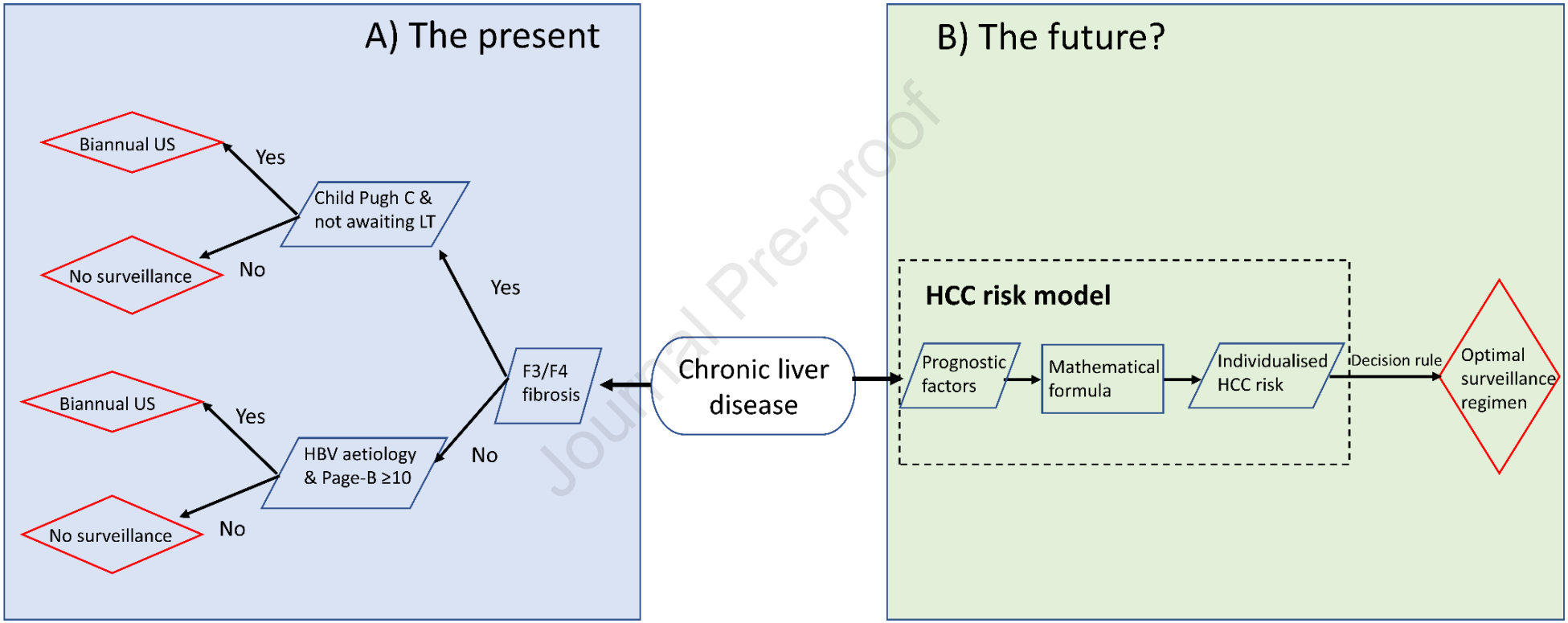
[25] Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ. 2009;338:B606. Doi: 10.1136/bmj.b606.

[26] Jenniskens K, Lagerweij GR, Naaktgeboren CA, Hooft L Moons KGM. Poldervaart JM, et al. Decision analytical modelling was useful to assess the impact of a prediction model on health outcomes before a randomised trial. J Clin Epidemiol. 2019;115:106-115.

[27] Esserman LJ, Wisdom Study and Athena investigators. The WISDOM study: breaking the deadlock in the breast cancer screening debate. NPJ Breast Cancer. 2017;3.34. doi: 10.1038/s41523-017-0035-5.

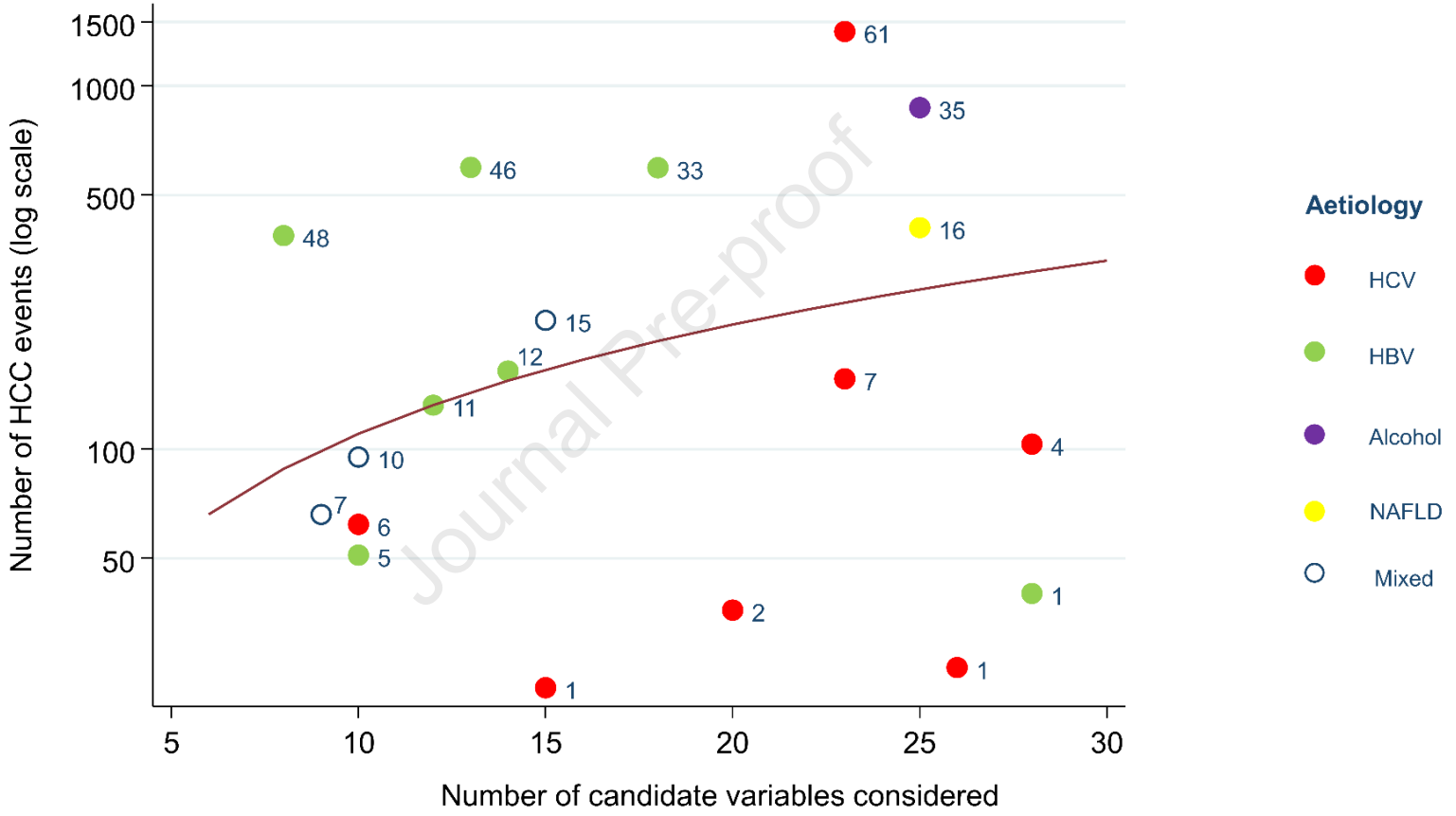
[28] Aakre C, Dziadzko M, Keegan MT, Herasevich V. Automating clinical score calculation within the electronic health record. Appl Clin Inform. 2017;8:369-380.

Fig 1. Current and future approaches to HCC surveillance



Currently, surveillance decisions recommended in guidelines are centred on fibrosis stage (panel A), even though this is difficult to measure and is just one of many risk factors influencing HCC risk. As new surveillance modalities emerge beyond ultrasound, an alternative approach will be to base screening decisions directly on individualised HCC risk (panel B). However, despite growing support, there are several methodological challenges with this approach which are not widely recognised and are not being addressed.

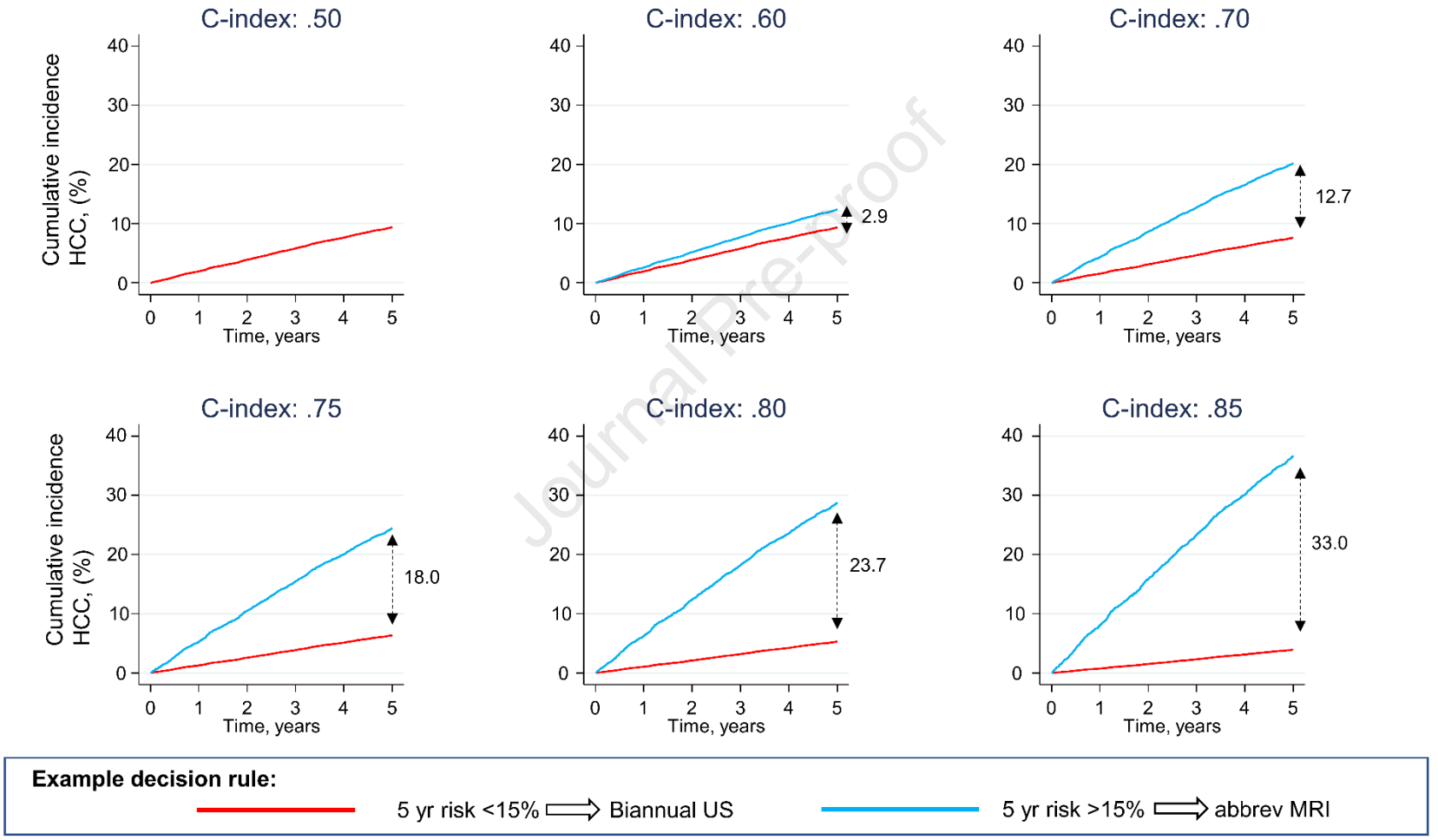
Fig 2. Event-Per-Predictor parameter (EPP) ratio for current HCC risk models



We undertook a brief search for HCC risk models published in the four top liver disease journals (J Hepatol; Gastroenterol; Hepatology, Gut) in the last ten years. 19 models were identified in total– each represented by a data point in this figure. The number adjacent to each data point is the model's Events Per Predictor parameter (EPP) ratio, calculated by dividing the number of HCC events in the model development dataset (y axis) by the number of prognostic parameters considered (x axis). Based on a recent sample size framework [12], we would estimate the

minimum EPP for a HCC risk model in cirrhosis patients to be at least 11 (red line). Several models are well below this threshold and can be considered at high risk of overfitting - i.e. fitting to idiosyncrasies rather than generalisable patterns.

Figure 3: Clinical relevance of the C-index.



Figures are generated from a simulated cohort, designed to resemble cirrhosis patients with cured hepatitis C. The HCC incidence rate is 2.1% per year, based on data from a recent meta analysis.[18] The rate of non-HCC mortality is 4% per year, which serves as a competing risk event. We assumed the rate of non-HCC mortality was higher in patients with HCC risk >15%. The 15% risk threshold for abbreviated MRI is derived from a recent cost effectiveness study. [19] Analyses were performed in stata version 16, using survsim and stcrreg commands.

BOX 1: Research recommendations.

1. Perform living systematic reviews to summarise the performance, applicability and biases of available HCC risk models.
2. Adopt a competing risk perspective when developing HCC risk models and ensure sample size is sufficient.
3. Address the shortage of HCC risk models catering to patients with alcohol-related liver disease and non-alcohol fatty liver disease.
4. Conduct external validation with the population of intended use in mind.
5. Collect health-related quality-of-life data from contemporary cirrhosis and HCC patients to inform CEA-based decision thresholds.
6. Think carefully about how individualised HCC risk should be converted into a surveillance decision. Consider developing a decision-rule via stakeholder consensus.
7. Avoid deriving decision rules using optimal cut-point methods.
8. Justify the validity of inferring surveillance regimen from a patient's individualised HCC risk. Under what circumstances may additional factors need to be considered?
9. Do not judge the clinical utility of a HCC risk model via the C-index alone (e.g. avoid the >0.80 heuristic). Instead, quantify clinical utility via impact studies.
10. Consider potential barriers to scalability (e.g. automated risk calculation and missing prognostic factor data).