**Improved survival prediction and comparison of prognostic models for patients with hepatocellular carcinoma treated with sorafenib**

Tim A. Labeur1-4, Sarah Berhane5, Julien Edeline6, Jean-Frederic Blanc7, Dominik Bettinger8, Tim Meyer9, Jeroen L.A. Van Vugt10, David W.G. Ten Cate10, Robert A. De Man11, Ferry A.L.M. Eskens12, Alessandro Cucchetti13, Laura J. Bonnett5, Otto M. Van Delden4, Heinz-Josef Klümpen2,
R. Bart Takkenberg3, Philip J. Johnson14

**Addresses and affiliations:**

1 Cancer Center Amsterdam, Amsterdam, The Netherlands.

2 Department of Medical Oncology, Amsterdam University Medical Centers, University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands. t.a.labeur@amsterdamumc.nl h.klumpen@amsterdamumc.nl

3 Department of Gastroenterology and Hepatology, Amsterdam University Medical Centers, University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands. r.b.takkenberg@amsterdamumc.nl

4 Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centers, University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands. o.m.vandelden@amsterdamumc.nl

5 Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom. S.Berhane@bham.ac.uk, L.J.Bonnett@liverpool.ac.uk
6 Department of Oncology, Centre Eugène Marquis, Rennes, France. j.edeline@rennes.unicancer.fr

7 Department of Hepatology, CHU Hôpital Saint André, Bordeaux, France. jean-frederic.blanc@chu-bordeaux.fr

8 Department of Medicine II, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Hustetter Str. 55 79106 Freiburg, Germany. dominik.bettinger@uniklinik-freiburg.de

9 UCL Cancer Institute, University College London, London, United Kingdom. t.meyer@ucl.ac.uk

10 Department of Surgery, Erasmus MC University Medical Center, Rotterdam, the Netherlands. j.l.a.vanvugt@erasmusmc.nl, d.w.g.tencate@erasmusmc.nl;

11 Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center, Rotterdam, the Netherlands. r.deman@erasmusmc.nl

12 Department of Medical Oncology, Erasmus MC University Medical Center, Rotterdam, the Netherlands. f.eskens@erasmusmc.nl

13Department of Medical and Surgical Sciences, Alma Mater Studiorum, University of Bologna, Bologna, Italy. alessandro.cucchett2@unibo.it

14 Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, United Kingdom. Philip.Johnson@liverpool.ac.uk

 **Corresponding author:**

Professor Dr. Philip J. Johnson

Department of Molecular and Clinical Cancer Medicine, University of Liverpool
Sherrington Building, Ashton Street, Liverpool L69 3GE, United Kingdom

Tel: +44 151 794 9821

Fax: +44 151 706 5826

E-mail: Philip.Johnson@liverpool.ac.uk

**Running head:** Prognostic models for sorafenib treatment of HCC **Words**: 239 (abstract) + 3986 (text, without references)

**Tables/Figures: 5 / 4**

**Abbreviations:** 95% CI, 95% confidence interval; AIC, Akaike Information Criterion; ALBI, albumin-bilirubin; APF, alpha-fetoprotein; AST, Asparatate Transaminase; BCLC, Barcelona Clinic Liver Cancer; CT, Computed Tomography; DCP, Des-gamma Carboxyprotrombin; ECOG PS, Eastern Cooperative Oncology Group performance status; HCC, Hepatocellular carcinoma; HAP, Hepatoma Arterial embolization Prognostic; JIS, Japan Integrated Staging (JIS) score; LR, likehood ratio; MRI, Magnetic Resonance Imaging; NLR, neutrophil-to-lymphocyt ratio; Overall survival (OS); PROSASH, Prediction Of Survival in Advanced Sorafenib-treated Hepatocellular carcinoma; SAP, Sorafenib Advanced hepatocellular carcinoma Prognostic; SD, standard deviation; SIRT, Selective Internal Radiation Therapy.

**Conflicts of interest:** DB receives teaching and speaking fees from Bayer Healthcare and from the Falk Foundation, Germany. HJK is member of the advisory board for Ipsen and Sirtex. JE receives speaking fees and travel grants from Bayer. JFB is member of the advisory board for Bayer, IPSEN, ESAI. RBT served as a speaker for Gore WL, Bayer, and Norgine. He is member of the advisory board for Gilead and Norgine. RdM served as a speaker for Norgine and as a consultant for Cook Medical. OvD served as consultant for Cook Medical. All other authors have declared no conflicts of interest. The study was designed and conducted by academic investigators.

 **Acknowledgments of research funding**: DB is supported by the Berta-Ottenstein-Programme, Faculty of Medicine, University of Freiburg. HJK has received an unrestricted research grant from Bayer. JE was partly funded by the Fondation de France for this work. RBT has received grants from the Netherlands Organisation for Health Research and Development (ZonMw) and The Netherlands Society for Gastroenterology (NVGE). SB acknowledges support from the UK EPSRC grant EP/N014499/1. TM is part funded by the NIHR UCLH Biomedical Research Centre.
These funding bodies had no role in the design of the study, collection, analysis, and interpretation of data, or in writing of the manuscript.

**Authors contributions:** Author contributions: TL lead the study, performed data acquisition and analysis, and wrote the manuscript. SB and LB performed and supervised data analysis. JE, JFB, DB, TM, JvV, DtC, RdM, FE, AC, OvD, HJK and RBT supervised the study, provided data and clinical input and provided mentorship for the study. PJ conceived the study design and is the guarantor of the article. All authors have reviewed and approved a final version of the manuscript.

**Abstract**

**Background**: The ‘Prediction Of Survival in Advanced Sorafenib-treated HCC’ (PROSASH) model addressed the heterogeneous survival of patients with hepatocellular carcinoma (HCC) treated with sorafenib in clinical trials, but requires validation in daily clinical practice. This study aimed to validate, compare and optimize this model for survival prediction.

**Methods**: Patients treated with sorafenib for HCC at 5 tertiary European centres were retrospectively staged according to the PROSASH model. In addition, the optimized PROSASH-II model was developed using the data of 4 centres (training set) and tested in an independent dataset. These models for overall survival (OS) were then compared with existing prognostic models.
**Results**: The PROSASH model was validated in 445 patients, showing clear differences between the 4 risk groups (OS 16.9-4.6 months). A total of 920 patients (n=615 in training set, n=305 in validation set) were available to develop PROSASH-II. This optimized model incorporated fewer and less subjective parameters: the serum albumin, bilirubin and alpha-fetoprotein, and macrovascular invasion, extrahepatic spread and largest tumour size on imaging. Both PROSASH and PROSASH-II showed improved discrimination (C-index 0.62 and 0.63, respectively) compared with existing prognostic scores (C-index ≤0.59).
**Conclusions**: In HCC patients treated with sorafenib, individualized prediction of survival and risk group stratification using baseline prognostic and predictive parameters with the PROSASH model was validated. The refined PROSASH-II model performed at least as good with fewer and more objective parameters. This model can be used as a tool for tailored treatment of HCC in daily practice and to define pre-planned subgroups for future studies.

 **Keywords**: hepatocellular carcinoma; sorafenib; survival; prognosis; prediction; model.

**Lay summary:** Patients with incurable liver cancer (hepatocellular carcinoma) can be treated with sorafenib to expand their life expectancy, but the prognosis with this drug varies between patients. In this large international study, we tested and further improved a statistical method that allows clinicians to estimate the survival chances of an individual patient. This facilitates the communication with the patient when considering this treatment and will help further research to find better drugs.

**Introduction**

Hepatocellular carcinoma (HCC) is the most common primary liver cancer and the second leading cause of cancer-related death worldwide1. Most patients with HCC present with, or eventually progress to, advanced stage disease which bears a poor prognosis. Sorafenib, a multikinase inhibitor, was the first treatment to show a survival benefit in patients with advanced stage HCC. In two randomized-controlled trials, sorafenib improved the median overall survival (OS) by 2-3 months compared with placebo2,3. Since then, sorafenib has been the standard treatment for patients with advanced stage HCC who are ineligible for locoregional treatment and have preserved (Child-Pugh A) liver function.

However, there is significant heterogeneity in outcomes in patient treated with sorafenib with an OS ranging from <3 months to 2-3 years2-4. This indicates that the survival benefit offered by sorafenib varies between individual patients. Select subgroups may have similar or more benefit from alternative options such as lenvatinib5, best supportive care or clinical trials.
The variety in survival is inadequately captured by the currently available staging systems (i.e. Barcelona Clinic Liver Cancer [BCLC]). Therefore, guidelines have recommended exploration of further stratification of patients with intermediate (BCLC-B) and advanced stage HCC (BCLC-C)6. Previous studies have identified markers of liver function (i.e. albumin, bilirubin), clinical parameters (i.e. performance status, body composition), and tumour characteristics (i.e. alpha-fetoprotein [AFP], macrovascular invasion, tumour extent) that may aid in prognostic stratification prior to sorafenib treatment7-15. Predictive factors, i.e. those associated with improved survival benefit over placebo, included absence of extrahepatic spread, presence of hepatitis C virus and a low neutrophil-to-lymphocyt ratio (NLR)16. Based on a combination of baseline factors, several scoring systems have been proposed for survival stratification of patients with advanced HCC treated with sorafenib17-20. Limitations of these models include the use of factors that either have a degree of subjectivity (i.e. infiltrative tumour growth, ascites) or are not commonly available (i.e. Des-gamma carboxyprothrombin [DCP]). A recently proposed model, the ‘Prediction Of Survival in Advanced Sorafenib-treated HCC’ (PROSASH), provided individualized survival prediction with excellent risk group discrimination based on 9 parameters (age, macrovascular invasion, extrahepatic spread, performance status, disease aetiology, albumin, creatinine, Aspartate Transaminase (AST) and AFP)21. The PROSASH model was built and validated on the data from patients treated with sorafenib in 2 clinical trials22,23, but has not yet been validated in patients treated in routine clinical practice. Multiple studies in various tumour types have underlined the limited applicability of data from the strictly selected and homogeneous patients treated in clinical trials to the more heterogeneous population in routine clinical practice24-28. Moreover, the PROSASH model has not yet been compared with the currently existing prognostic scores (BCLC, Child-Pugh). Consequently, it remains unknown whether this new model outperforms the existing models and whether risk stratification of sorafenib treated patients might be further refined using data from ‘real-life’ patients.
Therefore, this study aimed to 1) validate the PROSASH model in HCC patients treated with sorafenib in daily clinical practice and 2) improve the PROSASH based on patients treated in clinical practice. Subsequently, PROSASH, the improved model (PROSASH-II) and existing prognostic models were compared to determine the utility for clinicians to predict the survival of these patients.

**Methods**

**Study population**

Patients with HCC treated with sorafenib were recruited consecutively at 5 tertiary European centres with specialist multidisciplinary services for HCC management: Bordeaux (n=306) and Rennes (n=129), France; Freiburg (n=183), Germany; Amsterdam (n=156) and Rotterdam (n=167), the Netherlands. The data was collected after obtaining the relevant authorization from the institutional review boards and this retrospective study was performed under ethically approved protocols (REC reference 12/LO/1088 and W17\_420#17.488). Patients were diagnosed with HCC by histological or radiological criteria in accordance with international guidelines6,29. Exclusion criteria included patients receiving combination treatments (i.e. selective internal radiation therapy [SIRT] with sorafenib) or those with an Eastern Cooperative Oncology Group performance status (ECOG PS) >2. Patients received sorafenib with a target dose of 400 mg BID, with toxicity-adjusted dosing and patient management according to the local practice.

**Data collection and outcomes**

Commonly available clinical, imaging and serum variables prior to sorafenib treatment were collected by members of the research team. Imaging parameters were obtained from the most recent radiological imaging prior to first dose of sorafenib. Radiological staging included a multiphasic contrast enhanced computed tomography (CT) or dynamic magnetic resonance imaging (MRI). The main outcome measure, overall survival (OS), was defined from the date of start of treatment to date of death or censored on the date of last follow-up.

Patients were staged according to the PROSASH model21. To assess whether improved prediction may be possible using data from daily practice, a new model was built and validated (PROSASH-II, detailed below). The utility of both models was compared with existing prognostic scores that could be assessed in the dataset, including the BCLC staging system, Child-Pugh classification, Albumin-Bilirubin (ALBI) grade30, Japan Integrated Staging (JIS) score31, Hepatoma Arterial embolization Prognostic (HAP)32 and the Sorafenib Advanced HCC Prognostic (SAP) score18. With the exception of BCLC stage and Child-Pugh classification, which are commonly used in daily practice and were coded by the individual investigators, all prognostic scores were calculated using the raw data.

**Statistical methods**

Continuous variables were described as means with standard deviation (SD) or medians with interquartile range in case of highly skewed distributions. Categorical variables were described as absolute and relative frequencies. The Kaplan-Meier method was used to generate and compare survival curves, and to estimate median OS with 95% confidence interval (95% CI). For all analyses, a two-tailed *p* < 0.05 was considered statistically significant. Statistical analysis was performed using SPSS Statistics for Windows Version 24.0 (IBM Corp., Armonk, NY, USA) and STATA/SE 14.1 (StataCorp, Texas, USA).

Model building, testing and external validation

For the building of a prognostic model from patients treated in daily practice, the data of 4 centres were clustered into a training dataset and the largest independent dataset (Bordeaux) was used as an external validation set. Baseline variables that were considered clinically relevant and available in both datasets were included in the model building process (Supplementary Table 1). Highly skewed variables were log-transformed. BCLC stage and Child-Pugh grade were excluded from the model building process due to multicollinearity with factors used in these scoring systems. Multiple imputations (10x) using chained equations were performed to account for missing key parameters that were missing at random in the training dataset33,34. Model performance, derived coefficients and *p*-values of imputed data were compared with complete case data.

In the training set, the association between OS and baseline variables was assessed in an exploratory univariable and subsequent multivariable flexible parametric survival analysis35-37. The advantages of a flexible parametric analysis over the more commonly used Cox proportional hazard analysis were previously described21,37. Risk factors were reported with Hazard Ratio (HR) and corresponding *p*-values. The multivariable model was built using a stepwise forward selection procedure of variables significant at the 5% level. The model was reported according to the TRIPOD guidelines38 as well as tested, optimized and validated using the methods described by Royston and Altman39. Any time-dependent effects and potential proportional hazard violations by variables in the model were examined using the likelihood ratio (LR) test37. The LR test was also used to optimize the degrees of freedom (number of knots) for the restricted cubic spline function37. Lastly, Martingale residuals were plotted against continuous variables to check the functional form and non-linearity.
A linear predictor was derived using the coefficients of the model variables. Four risk groups were generated by applying the previously suggested cut-offs at the 16th, 50th and 84th centiles of the training set’s linear predictor39. The model, including the linear predictor and the centile-based risk group stratification, was applied to the external validation set.

The calibration of survival prediction was visually assessed by comparing the similarity between the observed and predicted survival curves in both the training and validation set. The observed and predicted survival-percentage at 12 months were also compared. Model discrimination was visually inspected by examining the separation survival curves of the 4 risk groups. In addition, survival rates between the risk groups were compared using HRs or log-rank test and the accompanying *p*-values. Lastly, subgroup analyses of the new model were performed in patients with Child-Pugh A or Child-Pugh B, because current guidelines recommend selecting patients with Child-Pugh A patients only6,29.

Model comparison

The PROSASH model incorporates the variable ‘Aspartate Transaminase (AST)’ which was not available in the Rennes (training) and Bordeaux (validation) datasets.
Therefore, model comparisons were performed in 3 subgroups of patients:
1) the imputed training dataset,
2) the external validation set, with complete data for all prognostic models except for the PROSASH model and

3) patients with complete data for all prognostic scores.
For each prognostic model the utility and discriminative performance was quantified using the Akaike Information Criterion (AIC) Harrell’s c-index and Royston-Sauerbrei’s R2D40,41. A lower AIC indicates a better goodness of fit, whereas a higher Harrell’s c-index indicates a larger proportion of patient pairs has agreement between the survival prediction and observed survival outcome in terms of rank. A higher R2D reflects a better explained variation on the log relative hazard scale. Most prognostic models consist of a linear predictor or point-based system with a risk group categorization which can leads to loss of information (i.e. ALBI-score and ALBI grade 1, 2 and 3). To assess the difference, the performance of each model as a linear predictor or points and risk groups were assessed. Because of lacking data, the number of Child-Pugh points could not be calculated, thus only the Child-Pugh classes (A,B,C) were assessed.

**Results**

**Study population**
In total, 941 patients who received sorafenib for advanced HCC between February 2003 and December 2016 were identified for this study. Of these, 21 patients (2%) were excluded because they received a combination of sorafenib with locoregional treatment (n=20) or due an ECOG PS >2 (n=1). Subsequently, 920 patients were included in this study, of whom 615 (67%) patients were included in the training cohort and 305 patients (33%) in the external validation cohort. The baseline characteristics of both cohorts are summarized in Table 1. Both cohorts had similar baseline features except that in the external validation cohort, more patients had ECOG PS 0 (65% vs 45%, *p*<0.001) and alcohol-induced liver disease was more common (64% vs 35%, *p*<0.001) compared with the training cohort, respectively. The median OS was 8.3 months (95% CI 7.6-9.2) in all patients. There was no statistically significant difference in survival between the training and validation cohort (HR 1.05, 95% 0.91-1.21, *p*=0.128)(Supplementary Figure 1).

**Validation of the PROSASH model in routine clinical practice**
The PROSASH model could be applied to 445/615 (73%) of patients from the training set who had a median OS of 8.0 months (95% CI 6.7-9.1). None of the patients from the external validation set were available due to missing AST (Supplementary Table 1). With the exception of risk group 2 vs 1 (HR 1.35, 0.94-1.92, *p*=0.102), there were clear survival differences between the 4 risk groups with a median OS ranging from 16.9 months to 4.6 months (Figure 1) in risk group 1 and 4, respectively.

**Prognostic factors and improved model: PROSASH-II**
First, multiple imputation was performed on the training set to account for missing data (Supplementary Table 1). An exploratory univariable analysis showed that albumin, Ln(bilirubin), ECOG PS, macrovascular invasion, extrahepatic spread, largest tumour size, number of liver lesions, Ln(AFP) and receiving prior HCC treatments were associated with OS (Supplementary Table 2).
The stepwise multivariable regression identified albumin, Ln(bilirubin), macrovascular invasion, extrahepatic spread, largest tumour size and Ln(AFP) as statistically significant prognostic factors (Table 2). These 6 baseline variables and their coefficients were incorporated in a multivariable model, named the PROSASH-II (Prediction Of Survival in Advanced Sorafenib-treated HCC v2):
Linear predictor: (-0.0337 x albumin in g/l) +
 (0.315 x Ln(bilirubin in µmol/l)) +
 (0.295 x macrovascular invasion, where 0=No and 1=Yes) +
 (0.181 x extrahepatic spread, where 0=No and 1=Yes) +
 (0.0336 x Largest tumour size in cm) +
 (0.0703 x Ln(AFP U/L))

A comparison of the model variables using complete case data (Supplementary Table 3) and imputed data showed very similar coefficients and *p*-values, indicating that the model was not greatly impacted by the imputation of missing data.
Using the centile-based cut-points, four risk groups were created: ≤-0.0760 (risk group 1), >-0.0760 to ≤0.355 (risk group 2), >0.355 to ≤0.858 (risk group 3) and >0.858 (risk group 4).
To simplify individual survival prediction, the calculation for the linear predictor and risk groups was incorporated in an online calculator (<https://jscalc.io/calc/qXgkZNB1h6B1jEfq>). This calculator can be used to determine the risk group and chance of survival at 3, 6, 12 and 24 months for each patient. For example, a patient with an albumin of 45 g/l, a bilirubin of 7 µmol/l, an AFP of 5789 U/L, the largest tumour measuring 5.9 cm with macrovascular invasion, but without extrahepatic spread, will have a predicted survival of 87% , 70%, 44% and 19% and 9% at 3, 6, 12, 24 and 36 months, respectively. The equations for these predictions are detailed in Supplementary Appendix A.

 **PROSASH-II performance in training and validation set**There were clear and statistically significant survival differences between the PROSASH-II risk groups in the training set (Figure 2A), with a median OS ranging from 19.6 months (risk group 1) to 3.9 months (risk group 4). The PROSASH-II model could be applied to 292 (93%) patients from the validation set. With the exception of risk group 1, which had fewer patients (n=36, 12%) and showed overlap in 95% CI with risk group 2 (HR 1.32, 95% CI 0.85-2.05, *p*=0.220), the risk groups showed evenly good discrimination in the validation set (Figure 2B).
Indicated by the concordance in the observed and predicted survival curves of both the training and validation sets (Figure 3A/B), the model showed good overall calibration. Similarly, the predicted and observed median OS and survival at 12 months closely matched in both datasets (Table 3). Although the model slightly underestimated the OS of risk group 1 in the training set, this was not the case in the validation set.
Given the similarities of baseline characteristics and model performance in the training and validation sets, all patients were clustered together then model-based stratification was re-applied. The median OS was 19.0, 11.2, 7.2 and 3.4 months with a 12-month survival of 65.6%, 45.6%, 31.2% and 10.1%, in risk group 1-4 respectively. There was no overlap in hazard ratios (Table 3), indicating good discrimination. Similar to the training set, there was a trend towards a slight survival underestimation of patients in risk group 1 (Figure 4), however overall the predicted and observed survival were closely matched.

**Subgroup analysis according to Child-Pugh class**In a subgroup analysis of Child-Pugh A patients (n=767), who had a median OS of 9.1 months, there were clear survival differences between the various PROSASH-II risk groups (Supplementary Figure 2A). The median OS was 19.0, 10.8, 7.6 and 4.5 months across risk group 1-4 respectively. For the subgroup analysis of patients with Child-Pugh B liver function, 136 patients were available with a median OS of 4.3 months (Supplementary Figure 2B). None of these patients were assigned to risk group 1 and only 10 (13.4%) to risk group 2. There was a trend towards a poorer survival across risk group 2 to 4 with a median OS of 13.4, 5.4 and 3.1 months, respectively. The difference between risk group 2 and 3 was not significant due to limited patient numbers (HR 1.98, 0.97-4.04, *p*=0.062). There were statistically significant survival differences between risk group 3 and 4 (log-rank *p*=0.002).

**PROSASH-II Model performance and comparison**The performance of the different prognostic models were compared and summarized in Table 4. Comparisons were performed in the training set with imputed missing data (n=615), the validation set with complete data (n=290) and a subgroup of 438 patients with complete data for all prognostic models. Across the various prognostic models, there was a slight loss in discriminative power when patients were categorized in risk groups or prognostic classes. Moreover, there was a trend towards a higher C-index and R2D and lower AIC across all assessed prognostic models in the validation set compared with the training set. In all different subsets, the models with the lowest predictive performance in terms of AIC, C-index and R2D were the BCLC, Child-Pugh and JIS. The HAP and SAP score performed very similarly in the different subsets.
In the training set, the higher C-index (0.65, IQR 0.64-0.65) and R2D (0.12, 95% CI 0.08-0.17) of the PROSASH-II indicated improved discriminative performance and explained variation compared with the currently available models. Likewise, the PROSASH-II had a lower AIC (1684) which indicated a better goodness of fit.
In the validation set, the PROSASH-II model had a higher C-index (0.68, 95% CI 0.65-0.72) and lower AIC (828) than commonly used scores such as BCLC and Child-Pugh. It also had the highest R2D (0.16, 95% CI 0.08-0.24) of all tested models, reflecting better explained variation. However, the model appeared to have a similar prognostic performance as the HAP and SAP scores, the latter showing a slightly higher C-index (0.69, 95% CI 0.66-0.72) and lower AIC (817) than the PROSASH-II model.
In the complete case subset for all models (n=438), the PROSASH-II model had the highest C-index (0.63, 95% CI 0.60-0.66) and R2D (0.10, 95% CI 0.06-0.15) and lowest AIC (1260). The slightly lower values for AIC (1278), C-index (0.62, 95% CI 0.59-0.65) and R2D (0.07, 95% CI 0.04-0.11) of the original PROSASH model indicated a very comparable predictive performance.

**Discussion**
In this large multicentre study of patients treated with sorafenib for HCC, the clinical trial based PROSASH model was successfully validated and optimized (PROSASH-II) in routine clinical practice. The PROSASH-II model, which uses fewer and more objective parameters and performed at least as good as PROSASH, offers individualized survival prediction and performs better than frequently used prognostic models (i.e. BCLC and Child-Pugh).
In light of the modest survival benefit (2-3 months) and significant costs and toxicity of sorafenib in advanced HCC, various studies have raised concerns on the cost-effectiveness of sorafenib in daily practice42-44. The BCLC staging system and Child-Pugh score are the mostly used prognostic models, but they have clear limitations: Child-Pugh incorporates subjective parameters which can lead to misclassification and inter-observer variability9, whereas the prognostic value of BCLC staging for patients treated with the same modality is low. To optimize cost-effectiveness and aid clinicians in survival prediction and clinical decision making, several other prognostic models have been proposed to stratify these patients (Table 5). Interestingly, most of these models were not specifically built for sorafenib treated HCC patients and none of them performed optimal9,18,19,45-47. Lack of consensus, easy applicability and external validation have hampered implementation of these prognostic scores in clinical practice.
We were able to compare 8 different prognostic models: ALBI, Child-Pugh, BCLC, HAP, SAP, JIS, PROSASH and the newly proposed PROSASH-II model (Table 5). All tested models included parameters for liver function (i.e. albumin, bilirubin, AST), most of them included tumour-related parameters (i.e. AFP, tumour size, macrovascular invasion) and some included ‘other’ baseline parameters (age, HCC aetiology, ECOG PS). Only a few scores have incorporated *predictive* parameters that were associated with increased benefit of sorafenib over placebo (extrahepatic spread, NLR and hepatitis C virus infection)16. This may reflect the modest impact of sorafenib on the natural history of advanced HCC. The well-known prognostic impact of the severity of the underlying liver disease was confirmed in this study, reflected by multivariable significance and incorporation of albumin and bilirubin in the PROSASH and PROSASH-II models. In accordance with prior studies9,48,49, we showed that despite using less parameters, ALBI has a better discrimination than the Child-Pugh classification.
Although initially developed to stratify HCC patients treated with TACE, the HAP-score showed that a further improvement of predictive accuracy is possible by combining liver function (albumin, bilirubin) and tumour-related (AFP, tumour size) parameters18. The highly comparable SAP score, which adds ECOG PS, performed similarly in our study. Depending on the subgroup of patients, the HAP and SAP scores performed slightly worse or similar to the PROSASH and PROSASH-II models. Given the overlap of 4 prognostic parameters (albumin, bilirubin, AFP, tumour size) which are dichotomized in the HAP and SAP scores, this is not unexpected. However, neither the SAP nor HAP score offer individualized survival prediction and do not incorporate predictors of improved sorafenib benefit.
Both the PROSASH and PROSASH-II model offer individualized survival prediction and propose an externally validated 4-tier subgroup classification with a median survival of 17-10-7-5 months and 19-11-7-3 months, for risk group 1-4 respectively. The PROSASH incorporated albumin, AFP, AST, creatinine, age, extrahepatic spread, macrovascular invasion, ECOG PS and disease aetiology (9 parameters in total), whereas the PROSASH-II incorporated albumin, AFP, extrahepatic spread, macrovascular invasion, tumour size and bilirubin (6 parameters in total). It is inevitable that different studies with different datasets lead to (slightly) different prognostic models. However, despite the different origins (clinical trial vs daily practice), there is significant overlap in the PROSASH-I and -II variables which suggests that these variables are stable and clinically relevant. As pointed out by several statistical experts, there is no widely agreed approach to build a multivariable prognostic model from a set of candidate predictors50,51. In this study, we aimed to report on the optimized statistical associations in daily clinical practice of sorafenib treated patients guided by 2 main principles in prognostic model building. Firstly, the parameters should be commonly available in centres treating patients with HCC. Secondly, models should be widely validated and universally applicable. For this purpose, we used large international datasets which have inevitable differences in data availability. As suggested by Royston et al., this was handled by multiple imputation of randomly missing data (Supplementary Table 1) and by balancing data availability (i.e. parameter selection) and analytic power (i.e. patient numbers)50. Using this approach, we were able to build the PROSASH-II model which required fewer and only highly reproducible parameters while it performed better in terms of C-index, AIC, and R2D than its predecessor. Disease aetiology and ECOG PS are less objective parameters which may lead to inter- and intra-user variability in daily practice, favouring PROSASH-II as a tool that can aid clinicians in providing patient tailored treatment. Moreover, PROSASH-II was build and tested on a daily clinical practice population in which it will be applied. Currently, guidelines recommend to consider all patients with well-preserved liver function (Child-Pugh-A) who are unsuitable for loco-regional therapy for sorafenib treatment. The clear subgroup survival differences of PROSASH-II risk groups in Child-Pugh A patients shows that even in ‘guideline concordant patients’ a more individualized decision is possible. Patients within risk group 3 (median OS 7-8 months) may have more benefit from alternative treatments (lenvantinib, clinical trials i.e. with PD1/PD-L1 blockers), whereas patients within risk group 4 (median OS 3-5 months) could be counselled to receive best-supportive care only. A similar stratification was seen in patients classified as Child-Pugh B who are currently not recommended to be treated with sorafenib and have a poor prognosis (median OS of 4.3 months). Still, a small subgroup of these patients (risk group 2, <10%) had a better prognosis (risk group 2, median OS 13.4 months) and could be considered for treatment with sorafenib.
In addition, the PROSASH-II stratification could be used for pre-planned or post-hoc subgroup analyses of ongoing and finalized phase-III studies comparing sorafenib with alternative treatments. Another application would be to generate survival curves of patients with advanced HCC treated with new agents in phase I-II studies. A quantitative comparison between the observed survival outcomes of tested agent and the predicted sorafenib survival remains difficult in these ‘in silico’ clinical trials, but it could aid in deciding whether these agents can proceed to be tested in a phase III trial.
This study has several limitations, foremost the retrospective design and its inherent limitations. Due to missing parameters some previously proposed prognostic factors (i.e. NLR16,52-55, body composition13,56) could not be taken into account and not all previously proposed models could be included in the comparison (CLIP, NIACE). Secondly, the study was performed in patients treated in European countries and should be validated in other geographical areas (i.e. Asia).
Despite over a decade of sorafenib usage and extensive studies, no molecular markers with a strong association with mechanism of sorafenib action have been identified, reflecting the complexity of advanced stage HCC and the difficulty of simplifying this into easily applicable biomarkers8. Our calculator provides an clinically applicable and validated model for the unmet need of outcome prediction prior to sorafenib treatment. Future studies could improve the risk stratification, survival prediction and clinical decision making by not only taking into account baseline factors (pre-sorafenib), but also parameters that can be monitored and could be of potential prognostic influence during treatment (i.e. sorafenib dose, dynamics in liver function, AFP,radiological response or pattern of progression). The more recently approved second-line treatments for advanced HCC (i.e. regorafenib [2017], cabozantinib [2019]) most likely did not have a major impact on current model, because the included patients were treated with sorafenib prior to FDA/EMEA approval of these treatments and the strict inclusion criteria of these landmark trials . Future studies aiming to implement these variables into robust tools and validated models will require large collaborations with detailed and high quality (prospective) datasets. To avoid statistical bias (overfitting), it remains important to externally validate novel prognostic models.
In conclusion, our study validated the PROSASH model in routine daily practice and proposed an improved model (PROSASH-II) which uses less baseline clinical features. The PROSASH-II model outperforms the currently available models and offers risk group stratification and individualized survival prediction that can be used for tailored treatment of HCC in daily practice and pre-planned subgroups analyses of future studies.

**Figure legends**

**Figure 1.** Overall survival according to the PROSASH risk groups with 95% confidence intervals.

**Figure 2A/B**. Overall survival according to the PROSASH-II risk groups in the training (A) and validation (B) set with 95% confidence intervals.

**Figure 3A/B**. Calibration plot of the predicted (dotted line) and observed (solid line) of the overall survival according to PROSASH-II risk groups in the training (A) and validation (B) set.

**Figure 4**. Calibration plot of the predicted (dotted line) and observed (solid line) of the overall survival according to the PROSASH-II risk groups (1-4) in all patients.

**References**

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: a cancer journal for clinicians.* 2015;65(2):87-108.

2. Cheng A-L, Kang Y-K, Chen Z, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. *The Lancet Oncology.* 2009;10(1):25-34.

3. Llovet J, Ricci S, Mazzaferro V. Sorafenib in advanced hepatocellular carcinoma [J]. *N Engl J Med.* 2008;359.

4. Marrero JA, Kudo M, Venook AP, et al. Observational registry of sorafenib use in clinical practice across Child-Pugh subgroups: The GIDEON study. *Journal of hepatology.* 2016;65(6):1140-1147.

5. Kudo M, Finn RS, Qin S, et al. Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. *Lancet (London, England).* 2018;391(10126):1163-1173.

6. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *Journal of hepatology.* 2018;69(1):182-236.

7. Bruix J, Raoul JL, Sherman M, et al. Efficacy and safety of sorafenib in patients with advanced hepatocellular carcinoma: subanalyses of a phase III trial. *Journal of hepatology.* 2012;57(4):821-829.

8. Marisi G, Cucchetti A, Ulivi P, et al. Ten years of sorafenib in hepatocellular carcinoma: Are there any predictive and/or prognostic markers? *World journal of gastroenterology.* 2018;24(36):4152-4163.

9. Edeline J, Blanc JF, Johnson P, et al. A multicentre comparison between Child Pugh and Albumin-Bilirubin scores in patients treated with sorafenib for Hepatocellular Carcinoma. *Liver international : official journal of the International Association for the Study of the Liver.* 2016;36(12):1821-1828.

10. Giannini EG, Bucci L, Garuti F, et al. Patients with advanced hepatocellular carcinoma need a personalized management: A lesson from clinical practice. *Hepatology (Baltimore, Md).* 2018;67(5):1784-1796.

11. Iavarone M, Cabibbo G, Piscaglia F, et al. Field-practice study of sorafenib therapy for hepatocellular carcinoma: a prospective multicenter study in Italy. *Hepatology (Baltimore, Md).* 2011;54(6):2055-2063.

12. Jackson R, Psarelli EE, Berhane S, Khan H, Johnson P. Impact of Viral Status on Survival in Patients Receiving Sorafenib for Advanced Hepatocellular Cancer: A Meta-Analysis of Randomized Phase III Trials. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2017;35(6):622-628.

13. Labeur TA, van Vugt JLA, Ten Cate DWG, et al. Body Composition Is an Independent Predictor of Outcome in Patients with Hepatocellular Carcinoma Treated with Sorafenib. *Liver Cancer.* 2018.

14. Llovet JM, Pena CE, Lathia CD, Shan M, Meinhardt G, Bruix J. Plasma biomarkers as predictors of outcome in patients with advanced hepatocellular carcinoma. *Clin Cancer Res.* 2012;18(8):2290-2300.

15. Pinter M, Sieghart W, Graziadei I, et al. Sorafenib in unresectable hepatocellular carcinoma from mild to advanced stage liver cirrhosis. *The oncologist.* 2009;14(1):70-76.

16. Bruix J, Cheng AL, Meinhardt G, Nakajima K, De Sanctis Y, Llovet J. Prognostic factors and predictors of sorafenib benefit in patients with hepatocellular carcinoma: Analysis of two phase III studies. *Journal of hepatology.* 2017;67(5):999-1008.

17. Adhoute X, Penaranda G, Raoul JL, et al. Prognosis of advanced hepatocellular carcinoma: a new stratification of Barcelona Clinic Liver Cancer stage C: results from a French multicenter study. *European journal of gastroenterology & hepatology.* 2016;28(4):433-440.

18. Edeline J, Blanc JF, Campillo-Gimenez B, et al. Prognostic scores for sorafenib-treated hepatocellular carcinoma patients: A new application for the hepatoma arterial embolisation prognostic score. *European journal of cancer (Oxford, England : 1990).* 2017;86:135-142.

19. Takeda H, Nishikawa H, Osaki Y, et al. Proposal of Japan Red Cross score for sorafenib therapy in hepatocellular carcinoma. *Hepatology research : the official journal of the Japan Society of Hepatology.* 2015;45(10):E130-140.

20. Yoo JJ, Chung GE, Lee JH, et al. Sub-classification of Advanced-Stage Hepatocellular Carcinoma: A Cohort Study Including 612 Patients Treated with Sorafenib. *Cancer research and treatment : official journal of Korean Cancer Association.* 2018;50(2):366-373.

21. Berhane S, Fox R, Garcia-Finana M, Cucchetti A, Johnson P. Using prognostic and predictive clinical features to make personalised survival prediction in advanced hepatocellular carcinoma patients undergoing sorafenib treatment. *Br J Cancer.* 2019.

22. Cheng AL, Kang YK, Lin DY, et al. Sunitinib versus sorafenib in advanced hepatocellular cancer: results of a randomized phase III trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2013;31(32):4067-4075.

23. Johnson PJ, Qin S, Park JW, et al. Brivanib versus sorafenib as first-line therapy in patients with unresectable, advanced hepatocellular carcinoma: results from the randomized phase III BRISK-FL study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2013;31(28):3517-3524.

24. Dierks J, Gaspersz MP, Belkouz A, et al. Translating the ABC-02 trial into daily practice: outcome of palliative treatment in patients with unresectable biliary tract cancer treated with gemcitabine and cisplatin. *Acta oncologica (Stockholm, Sweden).* 2017:1-6.

25. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *Br J Cancer.* 2014;110(3):551-555.

26. Labeur TA, Ten Cate DWG, Bart Takkenberg R, et al. Are we SHARP enough? The importance of adequate patient selection in sorafenib treatment for hepatocellular carcinoma. *Acta oncologica (Stockholm, Sweden).* 2018;57(11):1467-1474.

27. Mol L, Koopman M, van Gils CW, Ottevanger PB, Punt CJ. Comparison of treatment outcome in metastatic colorectal cancer patients included in a clinical trial versus daily practice in The Netherlands. *Acta oncologica (Stockholm, Sweden).* 2013;52(5):950-955.

28. Templeton AJ, Vera-Badillo FE, Wang L, et al. Translating clinical trials to clinical practice: outcomes of men with metastatic castration resistant prostate cancer treated with docetaxel and prednisone in and out of clinical trials. *Annals of oncology : official journal of the European Society for Medical Oncology.* 2013;24(12):2972-2977.

29. Heimbach JK, Kulik LM, Finn RS, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology (Baltimore, Md).* 2018;67(1):358-380.

30. Johnson PJ, Berhane S, Kagebayashi C, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2015;33(6):550-558.

31. Kudo M, Chung H, Osaki Y. Prognostic staging system for hepatocellular carcinoma (CLIP score): its value and limitations, and a proposal for a new staging system, the Japan Integrated Staging Score (JIS score). *Journal of gastroenterology.* 2003;38(3):207-215.

32. Kadalayil L, Benini R, Pallan L, et al. A simple prognostic scoring system for patients receiving transarterial embolisation for hepatocellular cancer. *Annals of oncology : official journal of the European Society for Medical Oncology.* 2013;24(10):2565-2570.

33. Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *J Survey methodology.* 2001;27(1):85-96.

34. Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in medicine.* 1999;18(6):681-694.

35. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in medicine.* 2002;21(15):2175-2197.

36. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research.* 2007;16(3):219-242.

37. Royston P, Lambert PC. *Flexible parametric survival analysis using Stata: beyond the Cox model.* Stata College Station, Texas; 2011.

38. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine.* 2015;162(1):W1-73.

39. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC medical research methodology.* 2013;13:33.

40. Newson RB. Comparing the Predictive Powers of Survival Models Using Harrell's C or Somers’ D. *The Stata Journal.* 2010;10(3):339-358.

41. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Statistics in medicine.* 2004;23(5):723-748.

42. Camma C, Cabibbo G, Petta S, et al. Cost-effectiveness of sorafenib treatment in field practice for patients with hepatocellular carcinoma. *Hepatology (Baltimore, Md).* 2013;57(3):1046-1054.

43. Parikh ND, Marshall VD, Singal AG, et al. Survival and cost-effectiveness of sorafenib therapy in advanced hepatocellular carcinoma: An analysis of the SEER-Medicare database. *Hepatology (Baltimore, Md).* 2017;65(1):122-133.

44. Sanoff HK, Chang Y, Lund JL, O'Neil BH, Dusetzina SB. Sorafenib Effectiveness in Advanced Hepatocellular Carcinoma. *The oncologist.* 2016;21(9):1113-1120.

45. Adhoute X, Penaranda G, Raoul JL, Bourliere M. Staging of hepatocellular carcinoma: BCLC system, what else! *Liver international : official journal of the International Association for the Study of the Liver.* 2016;36(9):1395-1396.

46. Samawi HH, Sim HW, Chan KK, et al. Prognosis of patients with hepatocellular carcinoma treated with sorafenib: a comparison of five models in a large Canadian database. *Cancer medicine.* 2018.

47. Endo M, Nishikawa H, Kita R, et al. Comparison of five staging systems in hepatocellular carcinoma treated with sorafenib: A single-center experience. *Molecular and clinical oncology.* 2016;4(4):515-522.

48. Kuo YH, Wang JH, Hung CH, et al. Albumin-Bilirubin grade predicts prognosis of HCC patients with sorafenib use. *Journal of gastroenterology and hepatology.* 2017;32(12):1975-1981.

49. Tada T, Kumada T, Toyoda H, et al. Impact of albumin-bilirubin grade on survival in patients with hepatocellular carcinoma who received sorafenib: An analysis using time-dependent receiver operating characteristic. *Journal of gastroenterology and hepatology.* 2018.

50. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ (Clinical research ed).* 2009;338:b604.

51. Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the practicing statistician. *Biometrical journal Biometrische Zeitschrift.* 2018;60(3):431-449.

52. Casadei Gardini A, Scarpi E, Faloppi L, et al. Immune inflammation indicators and implication for immune modulation strategies in advanced hepatocellular carcinoma patients receiving sorafenib. *Oncotarget.* 2016;7(41):67142-67149.

53. da Fonseca LG, Barroso-Sousa R, Bento Ada S, et al. Pre-treatment neutrophil-to-lymphocyte ratio affects survival in patients with advanced hepatocellular carcinoma treated with sorafenib. *Medical oncology (Northwood, London, England).* 2014;31(11):264.

54. Howell J, Pinato DJ, Ramaswami R, et al. Integration of the cancer-related inflammatory response as a stratifying biomarker of survival in hepatocellular carcinoma treated with sorafenib. *Oncotarget.* 2017;8(22):36161-36170.

55. Lue A, Serrano MT, Bustamante FJ, et al. Neutrophil-to-lymphocyte ratio predicts survival in European patients with hepatocellular carcinoma administered sorafenib. *Oncotarget.* 2017;8(61):103077-103086.

56. Fujiwara N, Nakagawa H, Kudo Y, et al. Sarcopenia, intramuscular fat deposition, and visceral adiposity independently predict the outcomes of hepatocellular carcinoma. *Journal of hepatology.* 2015;63(1):131-140.