**Using prognostic and predictive clinical features to make personalised survival prediction in advanced hepatocellular carcinoma patients undergoing sorafenib treatment**

Running title: survival prediction in sorafenib patients

*Sarah Berhane1,Richard Fox2, Marta García-Fiñana1, Alessandro Cucchetti3, Philip Johnson4*

***Addresses and affiliations:***

1Department of Biostatistics, University of Liverpool, Liverpool, UK

2Cancer Research UK Clinical Trials Unit, School of Cancer Sciences, University of Birmingham, Birmingham, B15 2TT, UK

3Department of Medical and Surgical Sciences, Alma Mater Studiorum, University of Bologna, Italy

4Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, UK

***Corresponding author:***

Professor Philip Johnson

Address: Department of Molecular and Clinical Cancer Medicine, The Sherrington Building, Ashton Street, Liverpool L69 3GE, UK

Tel:  0151 794 9821

Fax:  0151 706 5826

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

***Keywords:***

sorafenib; personalised survival prediction; advanced hepatocellular carcinoma; prognosis; risk stratification

**ABSTRACT**

**Background**

Sorafenib is the current standard of care for patients with advanced hepatocellular carcinoma (aHCC) and has been shown to improve survival by about 3 months compared to placebo. However survival varies widely from under three months to over two years. The aim of this study was to build a statistical model that allows personalised survival prediction following sorafenib treatment.

**Methods**

We had access to 1130 patients undergoing sorafenib treatment for aHCC as part of the control arm for two phase III randomised clinical trials (RCTs). A multivariable model was built that predicts survival based on baseline clinical features. The statistical approach permits both group level risk stratification and individual level survival prediction at any given time point. The model was calibrated and its discrimination assessed through Harrell’s c-index and Royston-Sauerbrei’s R2D.

**Results**

The variables influencing overall survival were vascular invasion, age, ECOG score, AFP, albumin, creatinine, AST, extra-hepatic spread and aetiology. The model predicted survival very similar to that observed. The Harrell’s c-indices for training and validation sets were 0.72 and 0.70 respectively indicating good prediction.

**Conclusions**

Our model (‘PROSASH’) predicts patient survival using baseline clinical features. However, it will require further validation in a routine clinical practice setting.

**BACKGROUND**

The multikinase inhibitor, sorafenib, was the first agent shown to offer survival benefit to patients with advanced hepatocellular carcinoma (aHCC) in a prospective placebo-controlled clinical trial (The SHARP trial).(1) Similar results were subsequently reported from the analogous Asia-Pacific study.(2) Sorafenib remains(3) the current standard of care for patients with aHCC in most countries offering a survival advantage of about three months compared to placebo. Although there are several potential new treatments for aHCC(4), sorafenib remains a first line treatment.(5)

Median overall survival (OS) for patients undergoing sorafenib treatment is about 10 months.(1, 6, 7) However, at the individual patient level, there is a wide variation ranging from 3 months to more than 2 years.(8, 9)As part of a personalised approach to sorafenib treatment, information on a patient’s baseline clinical features can potentially be utilised to make individual survival predictions. Previous studies looked at utility of current HCC staging systems in survival prediction but, none of them were found to be optimal.(10) Even within the Barcelona Clinic Liver Cancer staging system, where sorafenib is the recommended treatment (BCLC C), better survival varied considerably according to baseline clinical features.(9)

Subsequent studies have examined prognostic factors affecting OS among sorafenib patients as well as other variables predictive of sorafenib OS benefit. Factors found to influence prognosis included neutrophil-to-lymphocyte ratio (NLR), alpha-fetoprotein (AFP), tumour size/stage, extra-hepatic spread (EHD), Child-Pugh score, aspartate aminotransferase (AST), compensated cirrhosis, ascites, macroscopic vascular invasion (MVI), performance status (PS), albumin and bilirubin levels.(9, 11-16) Variables that predict greater sorafenib benefit compared to placebo included lower neutrophil-to-lymphocyte ratio, absence of extra-hepatic spread (EHS) and being hepatitis C positive.(8, 16-20)

In view of the aforementioned wide survival variation, we develop here for the first time a statistical model that predicts individual patient survival, using baseline clinical and laboratory features including those of prognostic and/or predictive significance.

**METHODS**

**Sorafenib clinical trials**

We had access to the sorafenib control arms of two multinational phase III randomised clinical trials (RCT) that compared sorafenib with each of brivanib(6) and sunitinib(7). The brivanib trial sorafenib arm consists of 588 patients with aHCC accrued between 2009 and 2011. The sunitinib trial sorafenib arm comprised 542 patients with aHCC, recruited between July 2008 and May 2010. Inclusion criteria (see supplementary table 1) in both trials were very similar. All data were obtained with permission from Bristol-Myers Squibb and Pfizer respectively.

**Variables**

Baseline variables available for analysis that were common to both trials are shown in Table 1. All variables were measured at baseline before the start of the treatment. Those used for model building were: age, gender, race, Eastern Cooperative Oncology Group (ECOG) score, creatinine, bilirubin, AST, AFP, albumin, international normalised ratio (INR), aetiology, tumour type (solitary/multiple), tumour size, presence of extra-hepatic spread and vascular invasion. Aetiology was categorised as HCV-related, hepatitis B virus (HBV)-related and “other”. The “other” group of aetiology also included patients with no known risk factors. Extra-hepatic spread included patients with lymph node involvement. Tumour number was presented as a binary variable (“solitary” or “multiple”) rather than discrete because there was more missing data in the latter (6.1% versus 11.2%). This discrepancy in missingness is due to some patients being marked as having multiple tumours rather than the actual number being given. Child-Pugh grade was excluded from the model building and replaced instead by its derivation components albumin, bilirubin and INR. The amount of missingness for each variable is summarised in Supplementary Table 2. Patients with missing data in any of the listed variables above were excluded from the modelling analysis.

**Statistical methods**

Analysis was carried out using Stata/SE 14.1 (StataCorp, Texas, USA). Continuous variables were reported as mean (with standard deviation [SD]) or median (with interquartile range [IQR]), the latter for variables with highly skewed distributions. Categorical variables were presented as counts and percentages. Continuous variables which exhibited extreme skewness were log transformed. Overall survival (OS) was measured from date of randomisation until date of death (any cause). Patients who were still alive were censored at their date of last follow up. Overall survival curves for each dataset were plotted using the Kaplan-Meier (KM) method and median survival with 95% confidence interval (95% CI) were reported. Survival distributions were compared using hazard ratio (HR) and corresponding p-values.

Univariable and multivariable analyses were undertaken using a flexible parametric survival model (see below). A multivariable model that predicts survival based on the baseline clinical features of patients was built using the sorafenib arm of the brivanib trial (training set). The sorafenib arm from the sunitinib trial was used as a validation set. The model was validated and tested according to the methodology suggested by Royston and Altman.(21)

Flexible parametric survival model

Flexible parametric survival regression was proposed by Royston and Parmar(22) in 2002. It is an extension of the Weibull model, and models the log baseline cumulative hazard using restricted cubic splines. The mathematical basis of the model and its advantages over the traditional Cox regression is described in the supplementary appendix titled “*The flexible parametric survival model*” and Supplementary figure 1.

Model building and validation

Univariable analysis was undertaken to examine the prognostic influence of each individual variable. The hazard ratio, 95% CI and p-values were reported. A multivariable model was then built using a stepwise backward selection of variables significant at the 5% level. Any strong interactions between the variables in the model were also examined. The hazard ratio, 95% CI and p-values for the multivariable model were reported.

A time-dependent (TD) effect for each variable in the final model was sequentially added and tested using the likelihood ratio (LR) test to inspect for any proportion hazards violation. The optimal degrees of freedom (d.f.) or knots for the restricted cubic spline function was chosen by testing and comparing different d.f. using the LR test. The functional forms of the continuous variables were examined by plotting a smoothed curve through Martingale residuals estimates, with zero gradient signalling an appropriate form.

The linear predictor of the final model was then derived using its coefficients. In order to generate four risk categories, previously suggested cut-offs(21) at the 15th, 50th and 85th centiles were applied to the linear predictor of the training set. Subsequent model predictions were grouped according to this classification. Individual level survival prediction was undertaken by calculating the survival function at time t (i.e. probability of a patient surviving past time t). The method for deriving the survival function formula is described in the supplementary appendix.

The derived model was validated on the sorafenib arm of the second RCT (sunitinib trial). Kaplan-Meier (KM) survival curves according to the risk categories were plotted and visually inspected for both the training and validation sets. Median OS and HR were also calculated for each risk category.

Model calibration

Model predictions according to the risk categories were visually inspected by overlaying the observed KM and predicted mean survival curves into one graph and examining how closely they agree. In addition to this, the corresponding observed versus predicted median survival as well as observed versus predicted percentage survival at 12 months were derived and reported. This was carried out for both training and validation sets.

Model discrimination

Model discriminative performance was measured using Harrell’s c-index(23) and Royston-Sauerbrei’s R2D(24). Harrell’s c-index measures the proportion of patient pairs where the survival predictions and observed outcomes are in agreement with respect to rank. R2D assesses the level of explained variation on the log relative hazard scale. Model parameters derived from the training set were first applied to the validation set before calculation of Harrell’s c-index and R2D. A higher value of Harrell’s c-index and R2D is indicative of better model discrimination.

Missing data

In order to investigate the nature of missingness and its effect on the final model parameters, multiple imputation of missing data using chained equations(25-27) was undertaken and coefficients and p-values between the complete case final model and the one with the imputed data were compared for any divergence.

**RESULTS**

Both sorafenib arms had similar baseline features (Table 1) with the exception of the presence of Child-Pugh B patients (n=46) in the brivanib trial patients (none in the sunitinib trial). Comparing the KM plots between the two datasets showed that there was no evidence of a statistical difference in survival (HR=1.11, 95% CI 0.97, 1.28, p=0.128) (Supplementary figure 2).

**Univariable and multivariable analyses**

Supplementary Table 3 reports the hazard ratios, 95% CI and p-values of the univariable flexible parametric survival models. They show that age, ECOG, aetiology, tumour size, extra-hepatic spread, vascular invasion, log(bilirubin), log(AST), log(AFP), albumin and INR were statistically significant prognostic factors.

Table 2 shows the variables that were selected for the final multivariable model, along with the hazard ratios, 95% CI and p-values. The variables in the final model were vascular invasion, age, ECOG, log(AFP), albumin, log(creatinine), log(AST), extra-hepatic spread and aetiology, along with an interaction between age and vascular invasion.

Partitioning the linear prediction using the prescribed cut-offs produced four distinct risk categories (1-4) in both the training and validation sets (Figure 1A and 1B). The observed group-level median OS was comparable in both the training and validation sets, ranging approximately from 4 months in risk category 4 to 30 months in risk group 1.

The curves in Figure 1 were generated as follows. First the linear predictor was derived:

Linear predictor, η = (0.327 x vascular invasion\*) +

 (-0.0231 x (Age – 60)) +

 (0.0303 x [(Age – 60) x vascular invasion\*]) +

 (0.455 x ECOG\*\*) +

 (0.0831 x ln(AFP)) +

 (-0.0553 x albumin) +

 (0.709 x ln(creatinine)) +

 (0.349 x ln(AST)) +

 (0.298 x extra-hepatic spread\*) +

 *(0.526 x HBV\*) +*

 *(0.507 x other aetiology if not HCV/HBV\*)*

Where

\*0=no and 1=yes, \*\*0=ECOG0 and 1=ECOG1

HCV is a reference group

The following cut-offs were then applied to the linear predictor, η, to generate the four risk categories: ≤2.898 (risk category 1), >2.898 to ≤3.666 (risk category 2), >3.666 to ≤4.559 (risk category 3) and >4.559 (risk category 4).

To calculate the survival function for an individual patient at time t, the following three steps were undertaken. The derivation of these equations is explained in more detail in the supplementary.

(1) The log cumulative baseline hazard (spline function) at time t was derived as follows:

s(log t) = -8.167 + 2.465 log t + 0.113 z2

where z2 = $(log t – 1.833)\_{+}^{3}$ – $0.361(log t+1.017)\_{+}^{3}$ – $0.639(log t – 3.443)\_{+}^{3}$

The “+” notation denotes (x)+=max(0, x)

(2) Baseline survival function, at time t, was expressed as:

S0(t) = exp(-exp(s(log t)))

(3) Survival function, S(t), at time t for an individual subject can then be calculated by:

S(t) = S0(t)exp(η)

where η is the linear predictor.

The values for S0(t) at time points 3, 6, 12, 18 and 24 months were 0.997, 0.991, 0.977, 0.965 and 0.955 respectively. For other time points, S0(t) can be calculated by following Steps 1 and 2.

For example, a patient with the following baseline features: No vascular invasion, 71 years of age, ECOG 1, AFP 850.3ng/ml, albumin 46g/l, creatinine 35.36μmol/l, AST 29U/l, no EHD and “other” aetiology (non-HBV, non-HCV) will have a survival function of 97%, 90%, 77%, 67% and 59% at times 3, 6, 12, 18 and 24 months respectively.

An online calculator to generate the survival predictions at a group and individual level is available at: <https://jscalc.io/calc/oGSDLHDsDg9g2XBF>

**Model calibration**

Observed KM and model-predicted survival curves according to the risk categories were closely matched in both the training and validation sets (Figure 2A and 2B). This was also reflected by the similarities between the observed and predicted median OS as well as observed and predicted percentage survival at 12 months (Table 3) in both datasets. There was some discrepancy, however, between the observed and predicted median OS in the lowest risk category of the training set although the percentage survival at 12 months was almost identical (78 vs 76%). Table 3 shows the median OS, percentage survival at 12 months, hazard ratio and p-value according to each risk category.

**Model discrimination**

The Harrell’s c-index figures (Table 3) show that there was a slight fall in the Harrell’s c-index C (from 0.72 to 0.70) and R2D (from 0.27 to 0.18) in the validation set compared to the training set; signalling a small deterioration in predictive power of the model. However both figures were indicative of good prediction.

**All patients combined**

Since both the training and validation sets showed similar survival both overall and within each risk category, they were merged and KM survival curves (Figure 3A) and calibration plots (Figure 3B) were generated.

Observed percentage survival at 12 months for patients within risk categories 1 to 4 were 76%, 55%, 28% and 6% respectively, with corresponding HR (in comparison to risk category 1) being 1.93 (95% CI 1.46, 2.54; p<0.0001), 3.68 (95% CI 2.81, 4.81; p<0.0001) and 8.16 (95% CI 6.03, 11.04; p<0.0001). Model-predicted and observed survival were very similar, with the exception of some discrepancy in median OS in the lowest risk category, as mirrored in the individual training and validation set results. Table 3 shows the median OS, hazard ratio and p-value when the model is applied to all the patients combined.

Finally, comparing the parameters between the complete case model and the one using imputed data showed very similar coefficients and p-values (Supplementary table 4), therefore indicating that the final model is not greatly affected by missing data.

**DISCUSSION**

Although the SHARP trial and the Asia-Pacific study(1, 2) clearly demonstrated a significant improvement in survival over placebo in patients with advanced HCC, the absolute improvement in median survival was less than 3 months. Furthermore, as we show here, there is heterogeneity around the median survival figures. In this study, where patient inclusion criteria was analogous to that of the SHARP trial, survival ranged from less than one month to more than two years. Others have found similar heterogeneity.(8, 9) At present the recommended indication for sorafenib treatment is patients with well-preserved liver function and being unsuitable for loco-regional therapies,(5) and this does not take into account likely survival. These figures are put into further context by a recent review of sorafenib amongst over 1000 Medicare beneficiaries(28) noting that ‘survival is exceptionally short…..and downsides of sorafenib use - high drug-related symptom burden and high drug cost - must be considered in light of this minimal benefit’. To overcome this issue, in this paper, we have developed a statistical model [henceforth known as PROSASH (**PR**ediction **O**f **S**urvival in **A**dvanced **S**orafenib-treated **H**CC] that allows personalised survival predictions with a view to aiding patient counselling and trial design.

Using data collected for regulatory purposes with similar criteria to that used for the SHARP trial and Asia-Pacific study(1, 2), we show that it is possible to predict survival on the basis of clinical features available at the time of diagnosis. Within the entire cohort of patients, we identified four risk categories (Figure 1) whose median overall survival ranged from 4 months in the highest risk category to over 20 months in the lowest risk category (Table 3). The corresponding percentage survival at 12 months for risk categories 1 to 4 were approximately 8%, 28%, 55%, and 76% respectively (Table 3). This stratification in addition to the individual patient predictions, permits ‘personalisation’ of sorafenib therapy and may be useful in clinical trials to optimally stratify patients where sorafenib is the appropriate control arm. Thus it would be possible to ensure that patients in a randomised phase III trial would have equivalent and matching prognostic features. Furthermore, the median survival figures in the higher risk groups are, in fact, worse than in the control (placebo) group of the SHARP study and this may lead clinicians to consider if the toxicity consequent on sorafenib therapy is worth any small potential survival benefit.

The modelalso offers insight into some of the factors that influence survival (Table 2). Notably, using HCV as the reference aetiology, the prognosis is clearly much worse in the HBV and ‘other’ groups. This is consistent with recent findings both from a retrospective review of the SHARP trial and Asia-Pacific study(1, 2, 16) and meta-analysis studies(8, 20), which identify HCV positivity as a key *predictive* factor for benefit after sorafenib. Our model thuscontains the major factors that have been found to be either predictive of sorafenib benefit such as extra-hepatic spread, or prognostic, such as vascular invasion and AFP in the combined analysis of the SHARP and Asia-Pacific trials(16) apart from NLR, which was not a recorded in our datasets. Although such data permits optimisation of the patient groups in whom sorafenib is administered, molecular markers related to the mechanism of action of targeted agents remain an important and, as yet unfulfilled, goal.(29)

The inclusion of aetiological factors in our model is clearly justified on the basis of the previously mentioned evidence that HCV positivity is a predictor of survival benefit compared to placebo. However, the inclusion of aetiological factors other than HBV and HCV, which are relatively ‘objective’ is problematical. Thus the lifetime consumption of alcohol is very difficult to record accurately and when both alcohol and a type of viral hepatitis are recorded, attribution to one specific aetiology becomes highly speculative. The diagnosis of NAFLD is equally difficult in the setting of HCC since evidence of NAFLD has often disappeared by the time the patient develops cirrhosis and HCC(30, 31). In the light of these observations we believe that categorisation aetiology as HCV, HBV or ‘other’ is the fairest option.

Another limitation of our study is that we did not have sufficient data to take into account pre- and post- study treatments. Since these are not predicated in the trial protocol there is a wide variation in the treatment options, and their duration, which makes it difficult to model statistically.

Since the guidelines for sorafenib treatment are based on clinical trial data it seems reasonable to build our prognostic model on similar datasets. Nonetheless, it will be important to validate the performance of the model in larger datasets and in patients treated in the routine clinical practice setting. Although the predicted survival in our model was very close to that observed at each risk category (Figure 2), there was some discrepancy between the observed and predicted median survival in the low risk category of the training set towards the end of the follow-up period (after about 15 months). Possible explanations for this discrepancy may be due to the small number of patients within that group surviving beyond 20 months such that (a) there was not enough information for the model to correctly extrapolate survival for and (b) such patients may have different features (compared to others) and factors that affect their survival may not be accounted for by the model. The observed and predicted percentage survival at 12 months in this category was, however, very close (78 vs 76%).

We believe that the statistical approach adopted here could be used to generate a ‘virtual control group’ for phase II, single arm, trials of new agents. Thus, we can generate survival curves that predict the outcome of patients in such trials *had they received sorafenib* although quantitative estimation of differences between the trial arms (sorafenib predicted vs actual new agent) remains methodologically challenging. Such an approach might be a useful preliminary screen for new agents when a go /no go decision concerning progression from phase II to phase III has to be made.

**ADDITIONAL INFORMATION**

**Ethical approval and consent to participate:** The subjects in this study were part of two large trials that were previously published(6, 7). As such, ethical approval has been acquired and study performed in accordance with the Declaration of Helsinki, with all patients giving informed consent.

**Consent to publish:** not applicable

**Availability of data and material:** All data remain the property of the sponsors of the trials from which they were extracted, as noted in the acknowledgments section.

**Conflict of interest:** none

**Funding:** SB and MGF acknowledge support from the UK EPSRC grant EP/N014499/1

**Authors' contributions:**

Concept and design: PJ, SB

Statistical analysis: SB, RF, MGF

Drafting of the manuscript: all

**Acknowledgements:** We are grateful to Professor Patrick Royston for his advice on the statistical methodology applied in this paper and to Bristol Myers Squibb and Pfizer for providing the data. This study has been previously presented at The International Liver Congress, 53rd annual meeting of the European Association for the Study of the Liver (EASL), Paris, France, 2018(32)

**REFERENCES**

1. Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med. 2008;359(4):378-90.

2. Cheng AL, Kang YK, Chen Z, Tsao CJ, Qin S, Kim JS, 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. Lancet Oncol. 2009;10(1):25-34.

3. Finn RS, Zhu AX, Farah W, Almasri J, Zaiem F, Prokop LJ, et al. Therapies for advanced stage hepatocellular carcinoma with macrovascular invasion or metastatic disease: A systematic review and meta‐analysis. Hepatology. 2018;67(1):422-35.

4. Raoul JL, Kudo M, Finn RS, Edeline J, Reig M, Galle PR. Systemic therapy for intermediate and advanced hepatocellular carcinoma: Sorafenib and beyond. Cancer Treat Rev. 2018;68:16-24.

5. European Association for the Study of the Liver. Electronic address eee, European Association for the Study of the L. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. J Hepatol. 2018;69(1):182-236.

6. Johnson PJ, Qin S, Park JW, Poon RT, Raoul JL, Philip PA, 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. J Clin Oncol. 2013;31(28):3517-24.

7. Cheng AL, Kang YK, Lin DY, Park JW, Kudo M, Qin S, et al. Sunitinib versus sorafenib in advanced hepatocellular cancer: results of a randomized phase III trial. J Clin Oncol. 2013;31(32):4067-75.

8. 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. J Clin Oncol. 2017;35(6):622-8.

9. Giannini EG, Bucci L, Garuti F, Brunacci M, Lenzi B, Valente M, et al. Patients with advanced hepatocellular carcinoma need a personalized management: A lesson from clinical practice. Hepatology. 2018;67(5):1784-96.

10. Samawi HH, Sim HW, Chan KK, Alghamdi MA, Lee-Ying RM, Knox JJ, et al. Prognosis of patients with hepatocellular carcinoma treated with sorafenib: a comparison of five models in a large Canadian database. Cancer Med. 2018.

11. Yuan J, Liang H, Li J, Li M, Tang B, Ma H, et al. Peripheral blood neutrophil count as a prognostic factor for patients with hepatocellular carcinoma treated with sorafenib. Molecular and Clinical Oncology. 2017;7(5):837-42.

12. Pinter M, Sieghart W, Hucke F, Graziadei I, Vogel W, Maieron A, et al. Prognostic factors in patients with advanced hepatocellular carcinoma treated with sorafenib. Aliment Pharmacol Ther. 2011;34(8):949-59.

13. Li J, Wang L, Cong N, Shi C, Bu W, Song J, et al. Efficacy of Sorafenib for Advanced Hepatocellular Carcinoma and Prognostic Factors. Hepatogastroenterology. 2014;61(132):954-7.

14. Lee S, Kim BK, Kim SU, Park SY, Kim JK, Lee HW, et al. Clinical outcomes and prognostic factors of patients with advanced hepatocellular carcinoma treated with sorafenib as first-line therapy: a Korean multicenter study. J Gastroenterol Hepatol. 2014;29(7):1463-9.

15. Baek KK, Kim JH, Uhm JE, Park SH, Lee J, Park JO, et al. Prognostic factors in patients with advanced hepatocellular carcinoma treated with sorafenib: a retrospective comparison with previously known prognostic models. Oncology. 2011;80(3-4):167-74.

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. J Hepatol. 2017;67(5):999-1008.

17. Kelley RK, Venook AP. Sorafenib in hepatocellular carcinoma: separating the hype from the hope. J Clin Oncol. 2008;26(36):5845-8.

18. Harding JJ, Abou-Alfa GK. Treating advanced hepatocellular carcinoma: How to get out of first gear. Cancer. 2014;120(20):3122-30.

19. Huitzil-Melendez F, Saltz L, Song J, Capanu M, Jacobs G, Humphrey J, editors. Retrospective analysis of outcome in hepatocellular carcinoma (HCC) patients (pts) with hepatitis C (C+) versus B (B+) treated with sorafenib (S). Gastrointestinal Cancers Symposium; 2007.

20. Shao YY, Shau WY, Chan SY, Lu LC, Hsu CH, Cheng AL. Treatment efficacy differences of sorafenib for advanced hepatocellular carcinoma: a meta-analysis of randomized clinical trials. Oncology. 2015;88(6):345-52.

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

22. 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. Stat Med. 2002;21(15):2175-97.

23. Newson RB. Comparing the predictive powers of survival models using Harrell's C or Somers' D. Stata Journal. 2010;10(3):339.

24. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med. 2004;23(5):723-48.

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

26. Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Statistics in medicine. 1999;18(6):681-94.

27. Van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Statistical methods in medical research. 2007;16(3):219-42.

28. Sanoff HK, Chang Y, Lund JL, O'Neil BH, Dusetzina SB. Sorafenib Effectiveness in Advanced Hepatocellular Carcinoma. Oncologist. 2016;21(9):1113-20.

29. Marisi G, Cucchetti A, Ulivi P, Canale M, Cabibbo G, Solaini L, et al. Ten years of sorafenib in hepatocellular carcinoma: Are there any predictive and/or prognostic markers? World J Gastroenterol. 2018;24(36):4152-63.

30. Caldwell SH, Oelsner DH, Iezzoni JC, Hespenheide EE, Battle EH, Driscoll CJ. Cryptogenic cirrhosis: clinical characterization and risk factors for underlying disease. Hepatology. 1999;29(3):664-9.

31. Poonawala A, Nair SP, Thuluvath PJ. Prevalence of obesity and diabetes in patients with cryptogenic cirrhosis: a case-control study. Hepatology. 2000;32(4 Pt 1):689-92.

32. Berhane S, Fox R, Chan S, Yau T, Johnson P. A statistical model for survival risk prediction in patients with advanced hepatocellular carcinoma undergoing sorafenib treatment. Journal of Hepatology. 2018;68:S197-S8.

**FIGURE LEGENDS**

**Figure 1A-B. Survival according to the risk categories as defined by the sorafenib model.** Kaplan-Meier survival curves and the corresponding risk table for (A) training and (B) validation set.

**Figure 2A-B. Calibration plots.** Comparing observed KM curves (solid line) and model-predicted mean survival curve (dashed line) for each risk category in the (A) training and (B) validation set.

**Figure 3A-B. Survival and calibration plots for all patients combined.** (A) Kaplan-Meier survival curves according to the risk categories as defined by the sorafenib model and (B) calibration plots of observed KM curves (solid line) and model-predicted mean survival curve (dashed line). Risk tables for both graphs shown below each figure.