European Journal of Surgical Oncology 46 (2020) 560–571

Contents lists available at ScienceDirect

European Journal of Surgical Oncology

journal homepage: www.ejso.com

Preoperative risk score for prediction of long-term outcomes after hepatectomy for intrahepatic cholangiocarcinoma: Report of a collaborative, international-based, external validation study☆

Raffaele Brustia [a], Serena Langella [b], Takayuki Kawai [c], Gilton Marques Fonseca [d], Astrid Schielke [e], Fabio Colli [f], Vivian Resende [g], Francesco Fleres [h], Didier Roulin [i], Paul Leyman [j], Alessandro Giacomoni [k], Benjamin Granger [l], Laetitia Fartoux [a], Luciano De Carlis [k], Nicolas Demartines [i], Daniele Sommacale [h], Marcelo Dias Sanches [g], Damiano Patrono [f], Olivier Detry [e], Paulo Herman [d], Shinya Okumura [c], Alessandro Ferrero [b], Olivier Scatton [a,*], Study GROUP

[a] Department of Hepatobiliary and Liver Transplantation Surgery, Hôpital de la Pitié–Salpêtrière, Assistance Publique-Hôpitaux de Paris, France
[b] Department of General and Oncological Surgery, Ospedale Mauriziano, Torino, Italy
[c] Department of Surgery, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[d] Hepatobiliary Unit, Department of Gastroenterology, University of São Paulo Medical School, São Paulo, Brazil
[e] Department of Abdominal Surgery and Transplantation, CHU Liege (CHU-ULg), Liege, Belgium
[f] General Surgery 2U, Liver Transplant Unit, A.O.U. Città della Salute e della Scienza di Torino, Torino, Italy
[g] Department of Hepatobiliary and Liver Transplantation Surgery, Federal University Minas Gerais, Brazil
[h] Department of General, Digestive and Endocrine Surgery, Robert-Debré University Hospital, Reims, France
[i] Department of Hepatobiliary Surgery, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland
[j] Department of General, Dept of General & Digestive Surgery, GZA Hospitals, Antwerp, Belgium
[k] Division of General, HPB and Transplantation Surgery, Ospedale di Niguarda, Milano, Italy
[l] Department of Biostatistics, Public Health and Medical Information Hôpital de la Pitié–Salpêtrière, Assistance Publique-Hôpitaux de Paris, France

ARTICLE INFO

Article history:
Received 9 April 2019
Received in revised form
25 October 2019
Accepted 31 October 2019
Available online 5 November 2019

Keywords:
Liver surgery
Intra-hepatic cholangiocarcinoma
Prognostic score
External validation
Long-term outcomes

ABSTRACT

Purpose: A preoperative risk score (PRS) to predict outcome of patients with intrahepatic cholangiocarcinoma treated by liver surgery could be clinically relevant.To assess accuracy for broadly adoption, external validation of predictive models on independent datasets is crucial.
The objective of this study was to externally validate the score for prediction of long-term outcomes after liver surgery for intrahepatic cholangiocarcinoma proposed by Sasaki et al. and based on preoperative albumin, neutrophil-to-lymphocytes-ratio, CA19-9 and tumor size.
Methods: Patients treated by liver surgery for intrahepatic cholangiocarcinoma at 11 international HPB centers from 2001 to 2018 were included in the external validation cohort. Harrell's c-index and Hosmer-Lemeshow analyses were used to test PRS discrimination and calibration. Kaplan–Meier curve for risk groups as described in the original study were displayed.
Results: A total of 355 patients with 174 deaths during the follow-up period (median = 41.7 months, IQR 32.8–50.6) were included. The median PRS value was 14.7 (IQR 10.7–20.6), with normal distribution across the cohort. A Cox regression on PRS covariates found coefficients similar to those of the derivation cohort, except for tumor size. Measures of discrimination estimated by Harrell's c-index was 0.61(95% CI:0.56–0.67) and Hosmer-Lemeshow p = 0.175. The Kaplan-Meyer estimation showed reasonable discrimination across risk groups, with 5years survival rate ranging from 20.1% to 0%.

☆ Part of the results of this research study was presented as a poster to the ACHBT-SFCD 2018 French congresses in Paris, France. Moreover, it was accepted as oral presentation to the E-AHPBA 2019 European congress in Amsterdam, June 2019.

* Corresponding author. Hepatobiliary and Liver Transplantation Surgery Hôpital de la Pitié–Salpêtrière, Assistance Publique-Hôpitaux de Paris, France.
E-mail address: olivier.scatton@aphp.fr (O. Scatton).

https://doi.org/10.1016/j.ejso.2019.10.041

Conclusion: In this external validation cohort, the PRS had mild discrimination and poor calibration performance, similarly to the original publication. Nevertheless, its ability to identify different classes of risk is clinically useful, for a better tailoring of a therapeutic strategy.

Introduction

Intrahepatic cholangiocarcinoma (ICC) represents less than 10% of all cholangiocarcinomas [1], with age-adjusted incidence increasing in Western countries from 2.1 to 3.3 per 100000 [1,2]. The highly desmoplastic nature of ICC and genetic heterogeneity contribute to its therapeutic resistance and poor prognosis [1,3]. Surgery represents the mainstay of curative treatment [4], with 5 years overall survival (OS) varying from 15% to 50% [3,5–7]. Despite, given the high recurrence and the considerable morbidity rate, it is unclear which patients could really benefit of surgical resection [8–11].

Preoperative prediction models are used to estimate the probability of developing a particular outcome, stratifying patients according to their risk to develop – for example - recurrence or death from disease. Albeit their purpose is not to replace clinical judgment, they have a clear role in supporting clinical decisions. Evidence exists that their use provides more accurate estimates of risk as compared to subjective predictions [12]. To support its broad adoption, a clinical prediction model should be confirmed by applying this model to an independent, "external", dataset [13].

The aim of this study was to realize an external validation of a preoperative risk score (PRS) predicting long-term outcomes of patients treated for ICC, published by Sasaki et al. in 2018 [14].

We decided to select this score for validation because its composition is based on four easy-to-use clinical parameters, systematically collected in any standard preoperative workout: tumor size, carbohydrate antigen 19-9 (CA19-9) level, neutrophil-to-lymphocyte ratio (NLR) and albumin. Moreover, if validated, this PRS could be clinically relevant to help clinicians to draw tailored strategies, weighting the potential harms of extended surgery against the predicted prognosis.

Methods

This study was an international, multicenter independent cohort study for the external validation of a published ICC PRS [14].

The study was designed in February 2018: eleven international, tertiary hepato-pancreatic-biliary (HPB) centers from Europe, South America and Japan were proposed to participate to the study (complete list in Supp. Material). The study was approved by the ethical committee of each Institution.

The PRS derivation cohort in the original study [14] included 269 patients who underwent curative-intent liver surgery (LS) for ICC between 1990 and 2015 at 16 HPB centers.

To avoid any historical bias in the external validation cohort due to the evolution of clinical and surgical management, we decided to start inclusions of patients treated by LS from January 2001 up to June 2018, with a minimum follow-up period per patient of 6 months (so up-to December 2018).

Reporting of this study was based on the "transparent reporting of a multivariable prediction model for individual prognosis or diagnosis" (TRIPOD) Statement and guidelines [15] (supplementary_table_TRIPOD_1). More details for the Methods section in Supplementary_Material_2.

Study endpoints

The primary endpoint was overall patient survival, to validate the predictive discrimination value of the PRS score.

The secondary endpoints were the definition of preoperative, intraoperative and postoperative variables predicting survival.

Eligibility criteria

Adult patients (>18 y old) undergoing LS for ICC confirmed on the pathology report, were eligible to be included in the study cohort. The definition for LS included any procedure requiring the resection of one or more liver segments, by either an open or laparoscopic approach.

Patients undergoing local ablation procedures (radiofrequency or microwave ablation) were considered for inclusion only if this was part of a surgical strategy including the removal of at least one liver segment during the same intervention.

Patients who underwent R2 resection were excluded from the study, as well as patients for whom data allowing the calculation of the PRS were not available or who were lost to follow-up.

Primary outcome

To validate the PRS score, the primary outcome variables collected were the event (death) and time until the event (OS), this latter defined as the time from surgical intervention to death or to date of last follow-up.

PRS survival prediction model

The prognostic variables in the PRS multivariable Cox model were preoperative albumin level, preoperative leucocytes and neutrophils, CA 19-9 level and tumor diameter (cm) on the preoperative CT-scan [14]. With this PRS, the predicted risk of OS at 5 years as well as the predicted median OS in months can be calculated preoperatively for each patient. In the original study the PRS was divided in five class of risk: low-risk patients with a PRS between 0 and 5 had a predicted 5y OS of 66.1% (median OS "not reached") while high-risk patients with PRS >40 had a predicted 5y OS of 0% (median OS 5.1 months).

In this external validation study, we used the same four candidate prognostic variables defined in the original study.

Variables

The preoperative, intraoperative and postoperative variables used in this study were retrieved from each single center prospective or retrospective database. As far as possible, we used the same definitions, scoring system, tables and figures organization as in the original article [14].

Data from different Centers were harmonized and merged in a single dataset for analysis. Each patient was de-identified and assigned to an anonymized alphanumeric code. Data were regularly entered in a digital worksheet-database, hosted on a secured

computer (limited access, personalized username and password). The quality of data management was compliant to the reference methodology on personal data processing and protection (MR003), as stated by French data protection authority (*Commission Nationale de l'Informatique et des Libertés*, CNIL n°2208386 v 0).

Sample size

The only formal recommendation in the context of an external validation study is that a substantial validation sample is required [13]. For this reason, we decided to include at least the same number of patients used in the derivation cohort (n = 269): at the end of the accrual period, n = 355 patients were included.

However, there is no single rule based on predictor parameters that would guarantee an accurate estimation of logistic regression parameters. When dealing with mortality in external validation studies, a minimum number of 100 events has been recommended, or a minimum of 10 events *per* predictor parameter for proportional hazards regression [16,17] per variable. Nevertheless the "10 event rule" has generated much debate, with Authors suggesting numbers ranging from <10 up to 50 [18].

A complex statistical approach to fix this issue has been proposed by Riley et al. [18], suggesting that the minimum number of events per predictor parameter should be calculated to meet the following criteria:

- small optimism in predictor effect estimates as defined by a global shrinkage factor of $\geq$0.9.
- small absolute difference of $\leq$0.05 in the model's apparent and adjusted Nagelkerke's R2.
- precise estimation of the overall risk in the population.

Based on the data from the original publication from Sasaki et al. [14], and according to these steps, we calculated that the number of event *per* predictor should be 6.15. We observed 174 deaths in the external validation cohort, corresponding to 43.5 events *per* predictor parameter, satisfying the above reported calculation.

Missing data

No multiple imputation was used.

Statistical analysis methods

All analyses were performed using data from the external validation cohort, and the results were compared to those from the original derivation cohort [14]. In particular, to calculate the PRS score we used the same predictors described in the original derivation cohort: [9 + (−2.79 × albumin) + (0.50 × NLR) + (2.81 × natural logarithm CA 19-9) + (1.12 × tumor size)]. All variables with $p < 0.05$ were considered statistically significant.

Descriptive statistics

Categorical (qualitative) variables are reported as percentages, while quantitative continuous variables are summarized as means and standard deviation (SD) or median (range) for discrete variables, as appropriate. A Kaplan–Meier curve for the entire external validation cohort was created through survfit and ggsurvplot functions from survminer and survival packages.

Primary objective = external validation of PRS

The external validation of the PRS survival prediction model followed the methods described by Royston et al. [13].

- Regression on the Prognostic Index

The Prognostic Index (PI) is the weighted sum of the prognostic variables, where the weights are the regression coefficients from the derivation cohort. A Cox proportional hazards model was fit with the PI as the only prognostic variable. A calibration slope smaller than 1 indicates suboptimal discrimination. A score test was performed to test for if the slope was significantly different from 1.

- Model misspecification/fit

Model fit was defined as the agreement of the regression coefficients between the derivation and validation cohorts. It was assessed by fitting a Cox model that included the prognostic variables and the PI (using the original coefficients from the derivation cohort) as an 'offset' variable. The model is considered to fit well if the regression coefficients for the prognostic variables were not statistically significantly different from 0. This was tested jointly for significance using a pooled likelihood ratio (LR) test from each multiple imputation.

- Measures of discrimination

To determine the discriminative ability of the PRS survival prediction model, the Harrell's c-index of concordance was calculated in the validation cohort. Harrell's c-index reflects the proportion of all patient pairs in which the predicted and observed outcomes are accordant. An index value close to 1 is considered to reflect good performance of the model. The graphical expression was plot through a ROC curve from the pROC package.

- Measures of Calibration

Calibration is the agreement between prediction from the model and observed outcomes, reflecting the predictive accuracy of the model. The Hosmer-Lemeshow goodness of fit test can be calculated through the gof function from survMisc library, and represented through a calibration plot [12].

- Kaplan–Meier curve for risk groups

Kaplan–Meier curves for OS were created with five strata corresponding to the risk groups from the original study, in order to allow a visual evaluation of the discriminative ability of the PRS prediction model. Moreover, a comparison of Kaplan–Meier derivation and validation plots offers a rough assessment of the model calibration [13].

Secondary objectives = definition of preoperative, intraoperative and postoperative variables predicting survival.

Unadjusted hazard ratios (HRs) and 95% confidence intervals (95% CI) were calculated for pre-, per- and postoperative variables associated with death, by a Cox proportional hazard regression analysis. Variables with a p value > than 0.1 were entered in a multivariate Cox model to identify factors independently associated with death. The final model will express the adjusted HRs and 95%CI.

Statistical software

Data managing and statistical evaluation were performed with R software (version 3.5.2 and following. The R Foundation for Statistical Computing. www.cran.r-project.org, Vienna, Austria).

Results

Participants

During the seventeen years' study period (2001–2018), 355 patients undergoing to LS for ICC in 11 participating centers were included and represented the study cohort. Among them 52% (n = 185) were male with a median age of 68.0 (60.0–74) years. Obesity (BMI>30 kg/m^2) was observed in 17% of patients (n = 60) and diabetes, metabolic syndrome, and cirrhosis in 21% (n = 72), 20% (n = 57) and 10% (n = 32) of them, respectively. The median values of the four PRS predictors were 4.07 g/dL (3.70–4.30) serum albumin level, 2.5 (1.7–3.7) NLR, 30.0 IU/mL (10.9–106.5) serum carbohydrate antigen 19-9 level and 5.0 cm (3.4–8.0) for the maximum tumor diameter. Major hepatectomy was required in 69% (n = 170) of patients, with a laparoscopic approach in 12% (n = 43) of cases. An associated procedure was required in 22% (n = 58) of patients, as biliary or vascular reconstruction: 16% (n = 42) and 7% (n = 19), respectively. One, three, five and ten year's survival rate was 86%, 53%, 40% and 20%, respectively (supp_Fig. 1). The mean overall and disease-free survival for the entire cohort was 63.3 ± 4.0 and 56.4 ± 4.2 months, respectively.

More details on Table 1. In order to highlight the role of each predictor parameter of the PRS, in Table 2 are detailed their distribution *per* each PRS risk class, among the group of patients experiencing death (event).

Primary objective = external validation of PRS

The distribution of the PRS in the validation cohort follows a Normal distribution (One-sample Kolmogorov-Smirnov test D = 0.10357, p-value = 0.001513). The median observed PRS value in the original derivation cohort [14] and external validation cohort was 17 and 14.685 (IQR 10.722–20.634) (Fig. 1), respectively.

Regression on the Prognostic Index

The observed slope in the Cox proportional hazards model on the PRS in the external validation cohort was 0.02 (p = 0.01), suggesting a mild discrimination of the model (Table 3).

Model misspecification/fit

A Cox regression on the predictors covariates of the original PRS in the external validation cohort found similar coefficients, except for tumor size: 0.01324, 95%CI(0.995–1.0319)p = 0.1542 versus 0.112, 95%CI(1.06–1.18) p = 0.001 in the orginal derivation cohort [14]. (Table 3).

Measures of discrimination

In the original study the calculated c-index for PRS was 0.69 95% CI(0.65–0.74). In this external validation cohort, Harrell's c-index was 0.61 95%CI(0.56–0.67), which reflects modest discrimination, as well as in the original study (Table 3 and Fig. 2).

Measures of calibration

The Hosmer-Lemeshow goodness of fit test found 173.9 expected deaths against 174 deaths observed (p = 0.175) (Table 3; Calibration plot of the observed and predicted death depending on the severity of the PRS available on supp_fig_ 2).

Kaplan–Meier curves for risk groups

Fig. 3 displays the Kaplan–Meier survival estimation for the validation cohort, with the five strata corresponding to the PRS risk groups as described in the derivation cohort. Grossly, apart the first 6–12 months during which some overlap among PRS group 2 to 5 is observed, the five curves are well separated, similarly to what is observed in the derivation cohort.

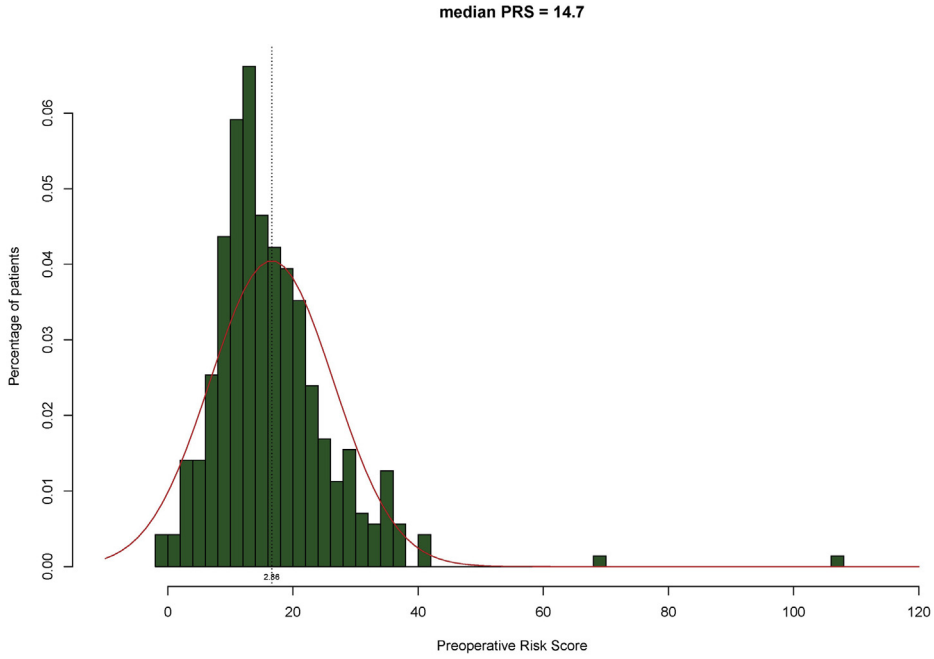

Fig. 1. The distribution of the preoperative risk scores (PRS) was normally distributed over the external derivation cohort (One-sample Kolmogorov-Smirnov test, D = 0.099055) with a median PRS value of 14.7. The vertical dotted line shows the PRS mean (16.6 ± 9.9) and the red line the theoretical Normal distribution. Two PRS outliers (69.2 and 107.8) were observed in the cohort.

Table 1
Characteristics of the external validation cohort.

	PRS 1	PRS 2	PRS 3	PRS 4	PRS 5	Overall
	(n = 72)	(n = 183)	(n = 73)	(n = 22)	(n = 5)	(n = 355)
PREOPERATIVE CHARACTERISTICS						
Gender						
F	30 (41.7%)	89 (48.6%)	38 (52.1%)	9 (40.9%)	4 (80.0%)	170 (47.9%)
M	42 (58.3%)	94 (51.4%)	35 (47.9%)	13 (59.1%)	1 (20.0%)	185 (52.1%)
Age, years						
Mean (SD)	67.7 (9.11)	64.8 (11.7)	67.9 (9.53)	64.3 (8.79)	74.3 (3.68)	66.2 (10.6)
Median [Min, Max]	70.0 [46.0, 86.5]	66.0 [32.0, 84.0]	69.0 [43.0, 84.0]	65.5 [43.0, 80.0]	74.0 [69.0, 79.0]	68.0 [32.0, 86.5]
BMI, Kg/m^2						
Mean (SD)	26.3 (6.05)	24.3 (5.60)	25.2 (6.37)	26.9 (5.19)	24.8 (1.34)	25.0 (5.85)
Median [Min, Max]	26.7 [13.4, 43.0]	24.0 [10.5, 45.6]	24.7 [7.72, 43.0]	25.6 [19.3, 39.0]	25.3 [23.0, 26.0]	24.6 [7.72, 45.6]
BMI, class						
<18	4 (5.6%)	19 (10.4%)	7 (9.6%)	0 (0%)	0 (0%)	30 (8.5%)
18_25	29 (40.3%)	89 (48.6%)	36 (49.3%)	10 (45.5%)	3 (60.0%)	167 (47.0%)
26_30	19 (26.4%)	49 (26.8%)	14 (19.2%)	5 (22.7%)	2 (40.0%)	89 (25.1%)
31_35	13 (18.1%)	16 (8.7%)	11 (15.1%)	5 (22.7%)	0 (0%)	45 (12.7%)
>35	4 (5.6%)	4 (2.2%)	4 (5.5%)	1 (4.5%)	0 (0%)	13 (3.7%)
Obesity (BMI>30)						
0	52 (72.2%)	157 (85.8%)	57 (78.1%)	15 (68.2%)	5 (100%)	286 (80.6%)
1	18 (25.0%)	20 (10.9%)	15 (20.5%)	7 (31.8%)	0 (0%)	60 (16.9%)
Diabetes						
0	54 (75.0%)	143 (78.1%)	58 (79.5%)	17 (77.3%)	4 (80.0%)	276 (77.7%)
1	17 (23.6%)	35 (19.1%)	14 (19.2%)	5 (22.7%)	1 (20.0%)	72 (20.3%)
Arterial Hypertension						
0	39 (54.2%)	102 (55.7%)	36 (49.3%)	10 (45.5%)	1 (20.0%)	188 (53.0%)
1	33 (45.8%)	76 (41.5%)	35 (47.9%)	12 (54.5%)	4 (80.0%)	160 (45.1%)
Dyslipidemia						
0	40 (55.6%)	103 (56.3%)	44 (60.3%)	15 (68.2%)	3 (60.0%)	205 (57.7%)
1	19 (26.4%)	21 (11.5%)	11 (15.1%)	4 (18.2%)	2 (40.0%)	57 (16.1%)
Metabolic syndrome						
0	46 (63.9%)	121 (66.1%)	47 (64.4%)	14 (63.6%)	4 (80.0%)	232 (65.4%)
1	20 (27.8%)	20 (10.9%)	11 (15.1%)	5 (22.7%)	1 (20.0%)	57 (16.1%)
Viral Hepatitis						
0	56 (77.8%)	147 (80.3%)	60 (82.2%)	20 (90.9%)	4 (80.0%)	287 (80.8%)
1	16 (22.2%)	33 (18.0%)	12 (16.4%)	2 (9.1%)	1 (20.0%)	64 (18.0%)
HIV						
0	58 (80.6%)	125 (68.3%)	57 (78.1%)	19 (86.4%)	5 (100%)	264 (74.4%)
1	1 (1.4%)	1 (0.5%)	0 (0%)	0 (0%)	0 (0%)	2 (0.6%)
Alcohol						
0	51 (70.8%)	135 (73.8%)	59 (80.8%)	15 (68.2%)	4 (80.0%)	264 (74.4%)
1	21 (29.2%)	44 (24.0%)	12 (16.4%)	7 (31.8%)	1 (20.0%)	85 (23.9%)
Biliary disease						
0	58 (80.6%)	123 (67.2%)	51 (69.9%)	17 (77.3%)	5 (100%)	254 (71.5%)
1	1 (1.4%)	4 (2.2%)	4 (5.5%)	2 (9.1%)	0 (0%)	11 (3.1%)
Digestive disease						
0	59 (81.9%)	122 (66.7%)	51 (69.9%)	18 (81.8%)	5 (100%)	255 (71.8%)
1	0 (0%)	5 (2.7%)	4 (5.5%)	1 (4.5%)	0 (0%)	10 (2.8%)
Hemochromatosis						
0	58 (80.6%)	122 (66.7%)	54 (74.0%)	18 (81.8%)	5 (100%)	257 (72.4%)
1	1 (1.4%)	4 (2.2%)	1 (1.4%)	1 (4.5%)	0 (0%)	7 (2.0%)
Smoke						
0	43 (59.7%)	83 (45.4%)	38 (52.1%)	16 (72.7%)	5 (100%)	185 (52.1%)
1	15 (20.8%)	40 (21.9%)	17 (23.3%)	3 (13.6%)	0 (0%)	75 (21.1%)
ASA score						
I	3 (4.2%)	7 (3.8%)	4 (5.5%)	2 (9.1%)	0 (0%)	16 (4.5%)
II	30 (41.7%)	63 (34.4%)	29 (39.7%)	9 (40.9%)	2 (40.0%)	133 (37.5%)
III	20 (27.8%)	39 (21.3%)	13 (17.8%)	5 (22.7%)	2 (40.0%)	79 (22.3%)
IV	2 (2.8%)	6 (3.3%)	4 (5.5%)	0 (0%)	0 (0%)	12 (3.4%)
Weight loss						
0	43 (59.7%)	72 (39.3%)	30 (41.1%)	10 (45.5%)	2 (40.0%)	157 (44.2%)
1	4 (5.6%)	22 (12.0%)	14 (19.2%)	5 (22.7%)	0 (0%)	45 (12.7%)
Preoperative PVE						
0	55 (76.4%)	114 (62.3%)	52 (71.2%)	18 (81.8%)	3 (60.0%)	242 (68.2%)
1	4 (5.6%)	12 (6.6%)	5 (6.8%)	1 (4.5%)	1 (20.0%)	23 (6.5%)
Preoperative Chemotherapy						
0	54 (75.0%)	113 (61.7%)	53 (72.6%)	19 (86.4%)	5 (100%)	244 (68.7%)
1	4 (5.6%)	15 (8.2%)	4 (5.5%)	0 (0%)	0 (0%)	23 (6.5%)
Serum albumin level, g/dL						
Mean (SD)	4.18 (0.443)	4.02 (0.537)	3.83 (0.571)	3.75 (0.553)	3.62 (0.680)	3.99 (0.543)
Median [Min, Max]	4.20 [3.10, 5.50]	4.10 [2.50, 7.20]	3.95 [2.00, 5.30]	3.75 [2.80, 4.90]	3.60 [3.00, 4.70]	4.07 [2.00, 7.20]
Neutrophil to lymphocyte ration						
Mean (SD)	2.50 (1.83)	2.85 (1.83)	3.72 (2.29)	4.52 (5.71)	15.6 (22.3)	3.24 (3.67)
Median [Min, Max]	2.22 [0.567, 12.1]	2.31 [0.436, 13.1]	3.37 [0.512, 14.6]	3.04 [0.705, 28.5]	5.20 [0.858, 54.5]	2.53 [0.436, 54.5]
Serum carbohydrate antigen 19–9 level, IU/mL						

Table 1 (*continued*)

	PRS 1 (n = 72)	PRS 2 (n = 183)	PRS 3 (n = 73)	PRS 4 (n = 22)	PRS 5 (n = 5)	Overall (n = 355)
Mean (SD)	7.41 (8.09)	61.2 (97.5)	715 (1430)	12800 (14900)	3610 (4580)	1020 (4820)
Median [Min, Max]	2.10 [0.200, 25.5]	30.4 [1.00, 805]	142 [5.00, 8110]	5580 [10.7, 50000]	1800 [28.0, 10900]	30.0 [0.200, 50000]
Maximum diameter of the tumor, cm						
Mean (SD)	3.95 (1.93)	5.14 (2.55)	7.78 (3.81)	9.94 (6.20)	30.8 (33.7)	6.10 (5.80)
Median [Min, Max]	3.50 [0.800, 10.4]	4.80 [1.00, 13.0]	8.00 [1.30, 16.0]	8.00 [4.00, 27.0]	16.0 [7.50, 90.0]	5.00 [0.800, 90.0]
Preoperative risk score (PRS)						
Mean (SD)	6.50 (2.84)	14.4 (2.85)	23.9 (2.93)	34.0 (2.13)	60.1 (29.3)	16.6 (9.86)
Median [Min, Max]	7.18 [-1.09, 9.94]	14.0 [8.08, 20.5]	22.9 [20.0, 29.7]	34.6 [30.2, 37.0]	41.8 [40.3, 108]	14.7 [-1.09, 108]
PERI OPERATIVE CHARACTERISTICS						
Surgical approach						
Laparoscopy	18 (25.0%)	20 (10.9%)	5 (6.8%)	0 (0%)	0 (0%)	43 (12.1%)
Open	54 (75.0%)	163 (89.1%)	68 (93.2%)	22 (100%)	5 (100%)	312 (87.9%)
Conversion to open						
0	14 (19.4%)	16 (8.7%)	2 (2.7%)	0 (0%)	0 (0%)	32 (9.0%)
1	3 (4.2%)	1 (0.5%)	2 (2.7%)	0 (0%)	0 (0%)	6 (1.7%)
Number of removed segments						
I	11 (15.3%)	8 (4.4%)	1 (1.4%)	1 (4.5%)	0 (0%)	21 (5.9%)
II	14 (19.4%)	24 (13.1%)	5 (6.8%)	2 (9.1%)	0 (0%)	45 (12.7%)
III	8 (11.1%)	21 (11.5%)	6 (8.2%)	2 (9.1%)	2 (40.0%)	39 (11.0%)
IV	8 (11.1%)	35 (19.1%)	21 (28.8%)	8 (36.4%)	1 (20.0%)	73 (20.6%)
V	7 (9.7%)	16 (8.7%)	8 (11.0%)	2 (9.1%)	0 (0%)	33 (9.3%)
VI	4 (5.6%)	12 (6.6%)	13 (17.8%)	3 (13.6%)	1 (20.0%)	33 (9.3%)
Major hepatectomy						
0	29 (40.3%)	37 (20.2%)	6 (8.2%)	3 (13.6%)	1 (20.0%)	76 (21.4%)
1	24 (33.3%)	80 (43.7%)	48 (65.8%)	15 (68.2%)	3 (60.0%)	170 (47.9%)
Assocated procedures						
0	55 (76.4%)	105 (57.4%)	35 (47.9%)	12 (54.5%)	4 (80.0%)	211 (59.4%)
1	4 (5.6%)	24 (13.1%)	22 (30.1%)	7 (31.8%)	1 (20.0%)	58 (16.3%)
Biliary procedures						
0	57 (79.2%)	112 (61.2%)	39 (53.4%)	14 (63.6%)	3 (60.0%)	225 (63.4%)
1	2 (2.8%)	16 (8.7%)	18 (24.7%)	5 (22.7%)	1 (20.0%)	42 (11.8%)
Vascular procedures						
0	56 (77.8%)	122 (66.7%)	52 (71.2%)	14 (63.6%)	4 (80.0%)	248 (69.9%)
1	3 (4.2%)	6 (3.3%)	5 (6.8%)	5 (22.7%)	0 (0%)	19 (5.4%)
Total vascular exclusio						
0	58 (80.6%)	120 (65.6%)	52 (71.2%)	17 (77.3%)	4 (80.0%)	251 (70.7%)
1	1 (1.4%)	8 (4.4%)	5 (6.8%)	2 (9.1%)	0 (0%)	16 (4.5%)
Extracorporeal circulation						
0	58 (80.6%)	128 (69.9%)	57 (78.1%)	18 (81.8%)	4 (80.0%)	265 (74.6%)
1	1 (1.4%)	0 (0%)	0 (0%)	1 (4.5%)	0 (0%)	2 (0.6%)
Liver cooling						
0	58 (80.6%)	128 (69.9%)	56 (76.7%)	18 (81.8%)	4 (80.0%)	264 (74.4%)
1	1 (1.4%)	0 (0%)	1 (1.4%)	1 (4.5%)	0 (0%)	3 (0.8%)
Blood transfusion						
0	43 (59.7%)	88 (48.1%)	33 (45.2%)	9 (40.9%)	2 (40.0%)	175 (49.3%)
1	7 (9.7%)	20 (10.9%)	13 (17.8%)	6 (27.3%)	1 (20.0%)	47 (13.2%)
N° RBC						
0	43 (59.7%)	88 (48.1%)	33 (45.2%)	8 (36.4%)	2 (40.0%)	174 (49.0%)
1	0 (0%)	2 (1.1%)	3 (4.1%)	0 (0%)	1 (20.0%)	6 (1.7%)
2	6 (8.3%)	10 (5.5%)	6 (8.2%)	2 (9.1%)	0 (0%)	24 (6.8%)
3	0 (0%)	2 (1.1%)	0 (0%)	1 (4.5%)	0 (0%)	3 (0.8%)
4	1 (1.4%)	3 (1.6%)	1 (1.4%)	1 (4.5%)	0 (0%)	6 (1.7%)
5	0 (0%)	0 (0%)	0 (0%)	1 (4.5%)	0 (0%)	1 (0.3%)
6	0 (0%)	1 (0.5%)	1 (1.4%)	1 (4.5%)	0 (0%)	3 (0.8%)
8	0 (0%)	1 (0.5%)	1 (1.4%)	0 (0%)	0 (0%)	2 (0.6%)
Blood loss, mL						
Mean (SD)	440 (767)	631 (837)	621 (553)	2030 (5150)	222 (38.7)	676 (1500)
Median [Min, Max]	250 [0.00, 5400]	445 [0.00, 6000]	500 [20.0, 2990]	586 [0.00, 21700]	200 [200, 267]	400 [0.00, 21700]
Blood loss, class						
0_200	23 (31.9%)	36 (19.7%)	14 (19.2%)	4 (18.2%)	2 (40.0%)	79 (22.3%)
200_400	11 (15.3%)	32 (17.5%)	10 (13.7%)	4 (18.2%)	1 (20.0%)	58 (16.3%)
400_600	10 (13.9%)	25 (13.7%)	12 (16.4%)	2 (9.1%)	0 (0%)	49 (13.8%)
600_800	2 (2.8%)	19 (10.4%)	8 (11.0%)	1 (4.5%)	0 (0%)	30 (8.5%)
>800	6 (8.3%)	28 (15.3%)	14 (19.2%)	6 (27.3%)	0 (0%)	54 (15.2%)
PATHOLOGY CHARACTERISTICS						
Mass forming type						
0	2 (2.8%)	14 (7.7%)	2 (2.7%)	2 (9.1%)	0 (0%)	20 (5.6%)
1	45 (62.5%)	94 (51.4%)	52 (71.2%)	16 (72.7%)	4 (80.0%)	211 (59.4%)
Maximum diameter of the tumor, mm						
Mean (SD)	41.1 (21.1)	52.0 (25.5)	74.7 (39.0)	102 (60.6)	100 (57.2)	58.2 (35.6)
Median [Min, Max]	35.0 [8.00, 104]	50.0 [10.0, 140]	80.0 [10.0, 160]	80.0 [40.0, 270]	85.0 [25.0, 160]	50.0 [8.00, 270]
Diameter of the tumor, classes, cm						
0_1	1 (1.4%)	2 (1.1%)	1 (1.4%)	0 (0%)	0 (0%)	4 (1.1%)

(*continued on next page*)

Table 1 (*continued*)

	PRS 1 (n = 72)	PRS 2 (n = 183)	PRS 3 (n = 73)	PRS 4 (n = 22)	PRS 5 (n = 5)	Overall (n = 355)
1_2	8 (11.1%)	15 (8.2%)	5 (6.8%)	0 (0%)	0 (0%)	28 (7.9%)
2_3	19 (26.4%)	28 (15.3%)	7 (9.6%)	0 (0%)	1 (20.0%)	55 (15.5%)
3_4	13 (18.1%)	31 (16.9%)	8 (11.0%)	2 (9.1%)	0 (0%)	54 (15.2%)
4_5	12 (16.7%)	23 (12.6%)	5 (6.8%)	0 (0%)	0 (0%)	40 (11.3%)
5_6	7 (9.7%)	24 (13.1%)	3 (4.1%)	3 (13.6%)	0 (0%)	37 (10.4%)
6_7	5 (6.9%)	21 (11.5%)	4 (5.5%)	4 (18.2%)	0 (0%)	34 (9.6%)
7_8	2 (2.8%)	16 (8.7%)	9 (12.3%)	3 (13.6%)	1 (20.0%)	31 (8.7%)
>8	4 (5.6%)	22 (12.0%)	31 (42.5%)	10 (45.5%)	3 (60.0%)	70 (19.7%)
Number of lesions						
1	49 (68.1%)	99 (54.1%)	50 (68.5%)	15 (68.2%)	3 (60.0%)	216 (60.8%)
2	2 (2.8%)	9 (4.9%)	2 (2.7%)	1 (4.5%)	1 (20.0%)	15 (4.2%)
3	1 (1.4%)	7 (3.8%)	1 (1.4%)	0 (0%)	0 (0%)	9 (2.5%)
4	0 (0%)	0 (0%)	1 (1.4%)	1 (4.5%)	0 (0%)	2 (0.6%)
5	0 (0%)	1 (0.5%)	0 (0%)	0 (0%)	0 (0%)	1 (0.3%)
6	1 (1.4%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0.3%)
12	0 (0%)	0 (0%)	0 (0%)	1 (4.5%)	0 (0%)	1 (0.3%)
Single or multiple nodules						
Multiple	4 (5.6%)	18 (9.8%)	4 (5.5%)	3 (13.6%)	1 (20.0%)	30 (8.5%)
Single	49 (68.1%)	99 (54.1%)	50 (68.5%)	15 (68.2%)	3 (60.0%)	216 (60.8%)
Pathology-proven satellite lesions						
0	45 (62.5%)	86 (47.0%)	33 (45.2%)	9 (40.9%)	3 (60.0%)	176 (49.6%)
1	8 (11.1%)	31 (16.9%)	21 (28.8%)	9 (40.9%)	1 (20.0%)	70 (19.7%)
American Joint Committee on Cancer						
Tumor stage (T)						
I	27 (37.5%)	35 (19.1%)	14 (19.2%)	5 (22.7%)	1 (20.0%)	82 (23.1%)
II	32 (44.4%)	93 (50.8%)	39 (53.4%)	7 (31.8%)	3 (60.0%)	174 (49.0%)
III	9 (12.5%)	39 (21.3%)	14 (19.2%)	6 (27.3%)	0 (0%)	68 (19.2%)
IV	4 (5.6%)	14 (7.7%)	6 (8.2%)	4 (18.2%)	1 (20.0%)	29 (8.2%)
Lymph node status (N)						
N0	40 (55.6%)	114 (62.3%)	37 (50.7%)	11 (50.0%)	4 (80.0%)	206 (58.0%)
I	7 (9.7%)	37 (20.2%)	28 (38.4%)	7 (31.8%)	0 (0%)	79 (22.3%)
X	25 (34.7%)	32 (17.5%)	8 (11.0%)	4 (18.2%)	1 (20.0%)	70 (19.7%)
Metastasis (M)						
0	52 (72.2%)	139 (76.0%)	59 (80.8%)	17 (77.3%)	3 (60.0%)	270 (76.1%)
I	0 (0%)	4 (2.2%)	2 (2.7%)	0 (0%)	0 (0%)	6 (1.7%)
Histologic differentiation						
Well	9 (12.5%)	17 (9.3%)	4 (5.5%)	4 (18.2%)	1 (20.0%)	35 (9.9%)
Moderate	34 (47.2%)	61 (33.3%)	27 (37.0%)	11 (50.0%)	2 (40.0%)	135 (38.0%)
Poor	15 (20.8%)	50 (27.3%)	26 (35.6%)	4 (18.2%)	2 (40.0%)	97 (27.3%)
Biliary invasion						
0	35 (48.6%)	73 (39.9%)	31 (42.5%)	7 (31.8%)	3 (60.0%)	149 (42.0%)
1	9 (12.5%)	23 (12.6%)	15 (20.5%)	6 (27.3%)	0 (0%)	53 (14.9%)
Vascular invasion						
0	24 (33.3%)	57 (31.1%)	16 (21.9%)	5 (22.7%)	1 (20.0%)	103 (29.0%)
1	32 (44.4%)	68 (37.2%)	39 (53.4%)	13 (59.1%)	3 (60.0%)	155 (43.7%)
Nervous engainment						
0	37 (51.4%)	78 (42.6%)	28 (38.4%)	6 (27.3%)	3 (60.0%)	152 (42.8%)
1	17 (23.6%)	38 (20.8%)	28 (38.4%)	12 (54.5%)	1 (20.0%)	96 (27.0%)
Adjacent liver						
Fibrosis (F)						
0	31 (43.1%)	79 (43.2%)	37 (50.7%)	14 (63.6%)	4 (80.0%)	165 (46.5%)
I	19 (26.4%)	41 (22.4%)	22 (30.1%)	3 (13.6%)	0 (0%)	85 (23.9%)
II	1 (1.4%)	12 (6.6%)	2 (2.7%)	0 (0%)	0 (0%)	15 (4.2%)
III	3 (4.2%)	8 (4.4%)	1 (1.4%)	2 (9.1%)	0 (0%)	14 (3.9%)
IV (Cirrhosis)	9 (12.5%)	18 (9.8%)	3 (4.1%)	2 (9.1%)	0 (0%)	32 (9.0%)
Steatosis						
0	24 (33.3%)	76 (41.5%)	30 (41.1%)	13 (59.1%)	3 (60.0%)	146 (41.1%)
1	27 (37.5%)	37 (20.2%)	24 (32.9%)	5 (22.7%)	1 (20.0%)	94 (26.5%)
Resection margin						
R 0	58 (80.6%)	153 (83.6%)	58 (79.5%)	16 (72.7%)	3 (60.0%)	288 (81.1%)
R 1	14 (19.4%)	28 (15.3%)	14 (19.2%)	5 (22.7%)	1 (20.0%)	62 (17.5%)
Resection margin, mm						
Mean (SD)	6.57 (8.72)	7.71 (11.2)	7.62 (10.9)	2.87 (4.21)	4.00 (4.08)	7.02 (10.2)
Median [Min, Max]	3.00 [0.00, 40.0]	5.00 [0.00, 90.0]	3.00 [0.00, 50.0]	0.900 [0.00, 15.0]	4.00 [0.00, 8.00]	4.00 [0.00, 90.0]
Missing	19 (26.4%)	67 (36.6%)	19 (26.0%)	4 (18.2%)	1 (20.0%)	110 (31.0%)
Resection margin, mm classes						
0_5	32 (44.4%)	71 (38.8%)	35 (47.9%)	15 (68.2%)	2 (40.0%)	155 (43.7%)
5_10	11 (15.3%)	21 (11.5%)	8 (11.0%)	2 (9.1%)	2 (40.0%)	44 (12.4%)
10_15	4 (5.6%)	8 (4.4%)	4 (5.5%)	1 (4.5%)	0 (0%)	17 (4.8%)
15_20	2 (2.8%)	9 (4.9%)	2 (2.7%)	0 (0%)	0 (0%)	13 (3.7%)
>20	4 (5.6%)	7 (3.8%)	5 (6.8%)	0 (0%)	0 (0%)	16 (4.5%)
POSTOPERATIVE CHARACTERISTICS						
Postoeprative complications						
0	46 (63.9%)	101 (55.2%)	37 (50.7%)	6 (27.3%)	0 (0%)	190 (53.5%)

Table 1 (*continued*)

	PRS 1 (n = 72)	PRS 2 (n = 183)	PRS 3 (n = 73)	PRS 4 (n = 22)	PRS 5 (n = 5)	Overall (n = 355)
1	19 (26.4%)	62 (33.9%)	31 (42.5%)	14 (63.6%)	3 (60.0%)	129 (36.3%)
Clavien-Dindo						
0	38 (52.8%)	66 (36.1%)	26 (35.6%)	4 (18.2%)	0 (0%)	134 (37.7%)
I	4 (5.6%)	10 (5.5%)	4 (5.5%)	2 (9.1%)	1 (20.0%)	21 (5.9%)
II	11 (15.3%)	30 (16.4%)	13 (17.8%)	5 (22.7%)	0 (0%)	59 (16.6%)
III	3 (4.2%)	18 (9.8%)	10 (13.7%)	5 (22.7%)	1 (20.0%)	37 (10.4%)
IV	0 (0%)	2 (1.1%)	1 (1.4%)	0 (0%)	1 (20.0%)	4 (1.1%)
V	1 (1.4%)	2 (1.1%)	3 (4.1%)	2 (9.1%)	0 (0%)	8 (2.3%)
ICU stay						
Mean (SD)	1.00 (1.82)	1.63 (4.41)	2.45 (9.23)	1.40 (2.29)	1.50 (0.707)	1.66 (5.42)
Median [Min, Max]	0.00 [0.00, 7.00]	0.00 [0.00, 36.0]	0.00 [0.00, 63.0]	0.00 [0.00, 8.00]	1.50 [1.00, 2.00]	0.00 [0.00, 63.0]
Hospital stay						
Mean (SD)	10.8 (8.84)	14.8 (14.8)	17.3 (19.4)	22.2 (18.4)	14.5 (3.42)	15.0 (15.4)
Median [Min, Max]	9.00 [3.00, 50.0]	10.0 [2.00, 114]	10.0 [3.00, 104]	16.0 [2.00, 73.0]	15.0 [10.0, 18.0]	10.0 [2.00, 114]
Missing	20 (27.8%)	68 (37.2%)	19 (26.0%)	4 (18.2%)	1 (20.0%)	112 (31.5%)
Postoperative chemotherapy						
0	39 (54.2%)	84 (45.9%)	28 (38.4%)	11 (50.0%)	2 (40.0%)	164 (46.2%)
1	12 (16.7%)	31 (16.9%)	26 (35.6%)	6 (27.3%)	1 (20.0%)	76 (21.4%)
Recurrence						
0	41 (56.9%)	71 (38.8%)	27 (37.0%)	7 (31.8%)	2 (40.0%)	148 (41.7%)
1	29 (40.3%)	101 (55.2%)	44 (60.3%)	15 (68.2%)	2 (40.0%)	191 (53.8%)
Disease free survival, moths						
Mean (SD)	30.3 (30.6)	28.3 (31.3)	18.2 (20.5)	15.9 (24.7)	5.63 (5.04)	25.6 (29.1)
Median [Min, Max]	17.6 [0.100, 122]	14.5 [0.0667, 157]	11.6 [0.00, 107]	6.28 [0.0667, 92.0]	4.18 [1.71, 14.2]	13.6 [0.00, 157]
Death						
0	53 (73.6%)	87 (47.5%)	30 (41.1%)	9 (40.9%)	2 (40.0%)	181 (51.0%)
1	19 (26.4%)	96 (52.5%)	43 (58.9%)	13 (59.1%)	3 (60.0%)	174 (49.0%)
Overall survival, moths						
Mean (SD)	36.8 (29.8)	37.5 (31.7)	30.4 (25.9)	21.6 (23.7)	11.7 (13.1)	34.6 (29.9)
Median [Min, Max]	27.8 [0.100, 122]	27.9 [0.0667, 157]	24.4 [0.197, 131]	12.8 [0.0667, 92.0]	5.00 [2.43, 34.1]	25.9 [0.0667, 157]

F=Female, M = men, BMI = body mass index, HIV=Human Immunodeficientia Virus, ASA = American Society of Anesthesiologists, PVE=Portal Vein Embolisation, PRS=Preoperative Risk Score, IU=International Unit, RBC = Red Blood Cells, ICU=Intensive care unit.

Secondary objectives = definition of preoperative, intraoperative and postoperative variables predicting survival

Preoperative variables (supplementary_Table 1)

A multivariate analysis was adjusted on history of digestive disease, cirrhosis, weight loss, serum albumin level, NLR and CA 19-9 preoperative value. After adjusting, NLR (HR 1.08, 95%CI(1.00,1.16) p = 0.035) and preoperative CA 19-9 (HR 1.00, 95%CI(1.00,1.01) p = 0.007) were independently associated with death.

Peri and postoperative variables (supplementary_Table 2)

A multivariate analysis was adjusted on open/laparoscopic approach, vascular reconstruction, blood loss (classes), blood transfusion (Y/N), postoperative complication (Y/N), length of hospital stay and adjuvant chemotherapy. After adjusting only blood transfusion (HR 1.90, 95%CI(1.01,3.59) p = 0.046) resulted independently associated with death.

Pathology variables (supplementary_Table 3)

A multivariate analysis was adjusted on the tumor size > 7 cm (Y/N), presence of pathology-proven satellite lesions, T and N stage, histologic differentiation, positive resection margin (R1), nervous invasion, presence of steatosis or cirrhosis. After adjusting for confounding variables the T IV stage (HR 2.57, 95%CI(1.027,6.441) p = 0.044), NI status (HR 2.67, 95%CI(1.553,4.606) p = 0.001), positive resection margins (R1) (HR 2.12, 95%CI(1.205,3.740) p = 0.009) and the presence of steatosis (HR 1.71, 95%CI(1.117,2.605) p = 0.013) were independently associated with death.

Table 2
Distribution of each predictor parameter of the PRS among the patients experiencing death (event).

PREDICTOR PARAMETER	PRS 1 (n = 19)	PRS 2 (n = 96)	PRS 3 (n = 43)	PRS 4 (n = 13)	PRS 5 (n = 3)	Overall (n = 174)
Serum Albumin						
Mean (SD)	4.14 (0.622)	4.02 (0.459)	3.72 (0.602)	3.92 (0.601)	3.59 (0.959)	3.95 (0.549)
Median [Min, Max]	4.20 [3.10, 5.50]	4.10 [2.76, 5.00]	3.90 [2.00, 4.77]	4.10 [2.80, 4.90]	3.08 [3.00, 4.70]	4.00 [2.00, 5.50]
NLR						
Mean (SD)	2.61 (2.40)	2.91 (1.93)	3.81 (2.52)	5.15 (7.13)	23.9 (26.9)	3.63 (4.88)
Median [Min, Max]	2.18 [0.995, 12.1]	2.30 [0.553, 13.1]	3.03 [0.683, 14.6]	3.02 [1.50, 28.5]	13.8 [3.52, 54.5]	2.60 [0.553, 54.5]
CA 19–9 IU/mL						
Mean (SD)	5.81 (6.30)	68.2 (110)	915 (1780)	14600 (17800)	5370 (5450)	1450 (6150)
Median [Min, Max]	3.00 [0.600, 18.5]	30.0 [1.00, 805]	142 [10.0, 8110]	5590 [39.8, 50000]	5140 [28.0, 10900]	39.5 [0.600, 50000]
Tumor size (cm)						
Mean (SD)	4.23 (1.80)	4.93 (2.55)	8.39 (3.58)	8.78 (4.83)	37.7 (45.5)	6.56 (7.25)
Median [Min, Max]	3.50 [1.50, 8.20]	4.00 [1.00, 13.0]	9.00 [2.00, 14.5]	7.00 [4.00, 20.0]	15.5 [7.50, 90.0]	5.00 [1.00, 90.0]

PRS=Preoperative Risk Score, NLR= Neutrophil to lymphocyte ration, SD=Strandard Deviation, IU=International Unit.

Table 3
External validation.

Regression on the Prognostic Index (PRS) in the external validation cohort.					
		Slope		SE	p
Preoperative Risk Score		0.001		0.0007	0.076
Model misspecification/fit in the derivation and external validation cohort.					
Derivation cohort n = 269[a]	HR	Coefficient	SE	95% CI	p
Serum albumin	0.76	−0.279	0.150	0.55–0.99	0.047
NLR	1.05	0.050	0.018	1.02–1.09	0.009
LogN CA 19–9	1.33	0.281	0.041	1.22–1.45	<0.001
Tumr size (per cm)	1.12	0.112	0.026	1.06–1.18	<0.001
Validation cohort n=355	**HR**	**Coefficient**	**SE**	**95% CI**	**p**
Serum albumin	0.7003	−0.35623	0.14900	0.523–0.9378	0.0168
NLR	1.0418	0.04099	0.01610	1.009–1.0753	0.0109
LogN CA 19–9	1.1584	0.14700	0.03440	1.083–1.2391	<0.0001
Tumr size (per cm)	0.0133	0.01324	0.00929	0.995–1.0319	0.1542
Likelihood ratio test= 32.69 on 4 df, p=0.000001					
Measures of discrimination (Harrel's c-index).					
		Harrell's c-index		95% CI	
Preoperative Risk Score		0.6142853		0.5558–0.6723	
Calibration function (Hosmer-Lemeshow goodness of fit test).					
	Observation	Events (death)	Expected	p	
n	355	174	173.9	0.175	

NLR= Neutrophil to lymphocyte ration, SE=Strandard Error, HR=Hazare Ratio, CI = confidence interval, PRS=Preoperative Risk Score, df = degrees of freedom.
[a] Data copied from the original publication [14].

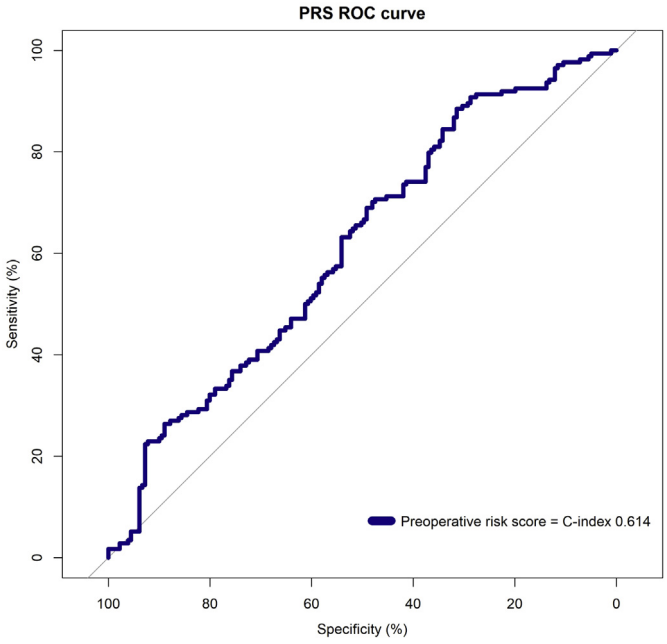

Fig. 2. Receiver Operator Characteristics (ROC) curve for PRS and mortality. The Harrell's c-index, corresponding to the area under the ROC (AuROC) was 0.61 95% CI(0.56–0.67).

Model improvement

Within the previous analysis, the presence of steatosis was independently associated with death. Given the increasing diffusion of the assessment of steatosis on radiological preoperative imaging (MRI or CT), we included this predictor parameter within a new PRS to test its performance compared to the PRS alone. Considering steatosis as binary (Y/N), the new PRS + Steatosis is calculated as follows = [9 + (−2.79 × albumin) + (0.50 × NLR) + (2.81 × natural logarithm CA 19-9) + (1.12 × tumor size) + (0.53 × steatosis)]. No difference was observed between the auROC for PRS and PRS + Steatosis, displayed in Fig. 4: 0.614 (95%CI:0.53–0.67) vs 0.606 (95%CI:0.55–0.67) respectively, p = 0.66.

Given the underrepresentation of the PRS class-risk 4 (n = 22) and 5 (n = 5), with significant overlap in Kaplan-Meier *strata*, we merged these two groups in a single one PRS >30 (n = 27). The Kaplan-Meier survival estimation with 4 class-risk is displayed in Fig. 5.

Discussion

ICC is a rare disease, and we were challenged to organize a large international cohort for external validation of a preoperative risk score (PRS) of survival after LS for ICC, proposed by Sasaki et al. [14]. The PRS was originally obtained after random splitting a data frame of 538 patients in a training and a validation cohort: besides a good discrimination, the calibration of the original PRS was fair [14]. One of the reasons explaining such results may be found in the absence of a preemptive sample size calculation [18,19].

Our study allowed the external validation of the PRS and its ability to class five risk groups of patients, based on a pre-operative assessment of the disease. According to our analysis, this score seems useful in clinical practice, and may help to decide in the future whose patients could be considered or not for upfront surgery. To be clear, we don't feel that clinicians (surgeons or oncologists) will refuse to offer a minor hepatic resection (segmentectomy or left lateral section) even for high-risk class IV-V patients: the debate is more likely to be on complex liver surgery (major right or left hepatectomy, with or without associated biliary or vascular procedures), or in case of repeated hepatectomy. The PRS is intended to be a decision aid during multi-disciplinary team meetings.

The methodological strength of our study relies in the fact that the external validation was based on a larger independent cohort (n = 355) than the derivation one (n = 269), and did not include any

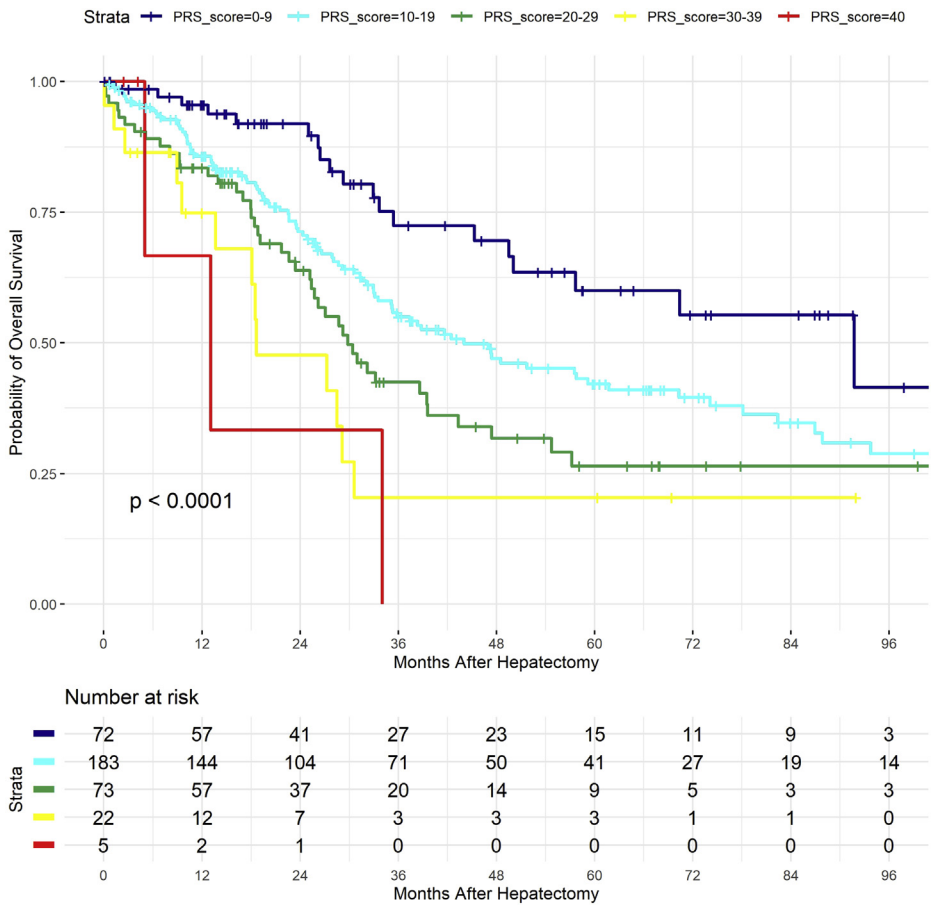

Fig. 3. Kaplan-Meier survival estimation for the different PRS classes: PRS 0–9 (PRS_score = 1), PRS 10–19 (PRS_score = 2), PRS 20–29 (PRS_score = 3), PRS 30–39 (PRS_score = 4), PRS 40 (PRS_score = 5).

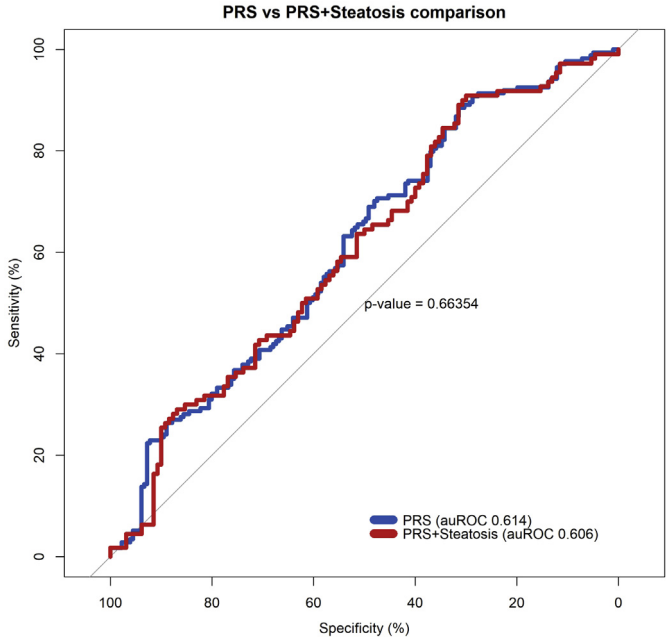

Fig. 4. Receiver Operator Characteristics (ROC) curve comparing the PRS and the improved version of PRS integrating steatosis (PRS + Steatosis). No difference was observed between the area under the ROC (AuROC) for PRS and PRS + Steatosis: 0.614 (95%CI:0.53–0.67) vs 0.606 (95%CI:0.55–0.67) respectively, p = 0.66.

center involved in the original study. For this type 4 analysis (predictive performance of a published prediction model on a separate dataset), we followed the methodology from Royston et al. [13] and the reporting recommendations of the TRIPOD statement [15]. Taken altogether, these points reinforce the quality and usefulness of the score validation process. The slope of the PI, Harrell's c-index and the five separate *strata* in the Kaplan–Meier curves suggested poor but acceptable discrimination: this can be expected in such in validation studies realized on large, multicenter international retrospective cohorts, because of the different case-mix.

Although five classes of risk were defined, only three of them (PRS class 1–3) appear clearly separated (in both derivation and validation dataset), and potentially useful in clinical practice. In particular, the observed survival rates at one year for PRS classes 1–3 were of 95.5%, 85.7% and 83.5%, while in PRS classes 4–5 were 74.8% and 66.7%, respectively. This discrimination is stable over the time: at 3 years, with 72.4%, 55.0% and 42.5% for PRS classes 1–3 against 20.4% and 0% for PRS classes 4–5, respectively, and at five years, with 55.3%, 42.1% and 26.4% for PRS classes 1–3 against 20.4% and 0% for PRS classes 4–5, respectively. Given the underrepresentation and *strata* overlap of the two latter classes, we merged them together (Fig. 5): a reduced model with 4 classes may probably be more useful than the original one.

When considering the four prognostic variables included in the PRS score, half of them (CA 19-9 and tumor size) are directly related to the tumoral features, while the two others (albumin, NLR) are dependent from the patient condition or to the underlying liver disease. Indeed, 20% of the included patients had a significant liver

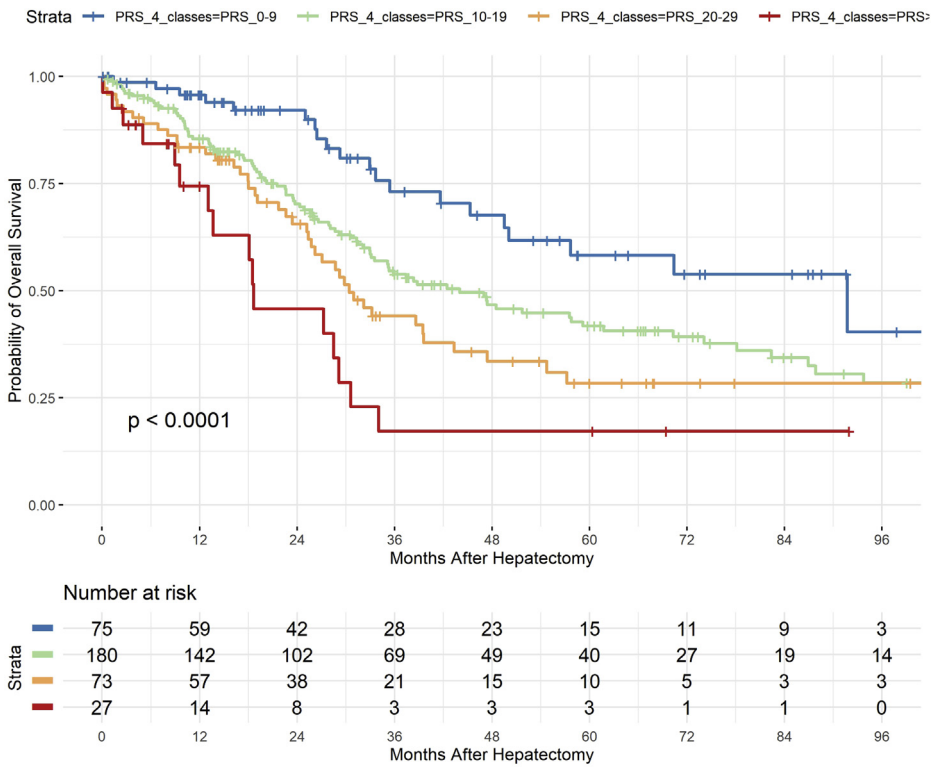

Fig. 5. Kaplan-Meier survival estimation after merging the *strata* of the previous class 4 and 5 in a single one (Class 4, PRS>30): PRS 0–9 (PRS_score = 1), PRS 10–19 (PRS_score = 2), PRS 20–29 (PRS_score = 3), PRS > 30 (PRS_score = 4).

fibrosis (F I-III) or cirrhosis (F IV). These latter values in particular can fluctuate according the patient's general condition, and are probably less relevant than those - more objective - related to the tumoral features. In future studies, it could be interesting to focus on the dynamic evolution of the PRS score: at the time of the diagnosis, and after the introduction of a neo-adjuvant treatment and/or a prehabilitation program. Similar to a "test of time", if there is a change of PRS class after such treatment, it would possible to consider a more aggressive treatment for this category of patients.

When focusing on preoperative variables, univariate and multivariate Cox analyses showed significant correlation with survival for NLR and CA 19-9 in this validation cohort. In contrast with the original derivation cohort [14] and to previously published studies, tumor size [20] and albumine [21] were not significantly associated with survival. The lack of statistical significance for these two predictors may be due to the different historical period (1990–2015 derivation cohort, versus 2001–2018 validation cohort) and case-mix: even in the absence of statistical analyses, a simple comparison reveals how patients in the validation cohort were older, had lower CA 19-9 level and higher rate of T stage II-IV, R1 resection and poor histological differentiation as compared to the derivation cohort. These differences represent classical limitations and biases of any retrospective multicenter cohort study realized over a long time-period.

Anyhow, the observation of very similar c-indexes in the original (0.69, 95%CI:0.65–0.74) and external cohort (0.61, 95% CI:0.56–0.67) despite the different characteristics of both cohorts, allows to speculate the reproducibility – and therefore the usefulness – of such a PRS.

Last, there are still two potential ways to improve the performance of the PRS. We tried to integrate steatosis into the score (PRS + Steatosis, Fig. 4), but the results were not different from the PRS alone. Recently Lunsford et al. [22] showed some promising results of liver transplantation after neo-adjuvant treatment for advanced ICC, and this is probably the room for improvement of the PRS. However neo-adjuvant regiment was administered only in some 6% of our patient's cohort, with heterogeneity in molecule, cycles and duration: results of prospective trials are urgently needed, to include or not neo-adj treatment in the PRS. Another way to test the usefulness of the PRS – as it is – might be its dynamic evolution after a neo-adjuvant treatment: a patient jumping from a mid-class risk to a low-class risk after neo-adjuvant treatment could be considered as a good prognostic sign, authorizing some heavier treatments or surgical resection.

To resume, the PRS model has acceptable performance, is generalizable among different ICC populations and moreover it is easy-to-use through any digital spreadsheet. As a perspective, PRS paves the way to a tailored strategy, avoiding upfront surgery for class 4–5 patients and proposing aggressive surgery for class 1 patients. Further studies are needed to ascertain if class 2–3 patients may benefit of an induction treatment, including preoperative neoadjuvant chemotherapy, to seek for an improvement of the parameters before any potential surgery.

Declaration of competing interest

The authors declare no conflicts of interest.

Acknowledgements

We wish to thank two Reviewers and the Editor for their constructive comments which helped improve the article upon revision.

For this study research, neither grants nor founding were perceived by any of the participating centers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejso.2019.10.041.

References

[1] Razumilava N, Gores GJ. Cholangiocarcinoma. Lancet (London, England) 2014;383(9935):2168–79. https://doi.org/10.1016/S0140-6736(13)61903-0.

[2] Shaib Y, El-Serag H. The epidemiology of cholangiocarcinoma. Semin Liver Dis 2004;24(02):115–25. https://doi.org/10.1055/s-2004-828889.

[3] Mavros MN, Economopoulos KP, Alexiou VG, Pawlik TM. Treatment and prognosis for patients with intrahepatic cholangiocarcinoma. JAMA Surg 2014;149(6):565. https://doi.org/10.1001/jamasurg.2013.5137.

[4] DeOliveira ML, Cunningham SC, Cameron JL, et al. Cholangiocarcinoma: thirty-one-year experience with 564 patients at a single institution. Ann Surg 2007;245(5):755–62. https://doi.org/10.1097/01.sla.0000251366.62632.d3.

[5] Li D-Y, Zhang H-B, Yang N, Quan Y, Yang G-S. Routine lymph node dissection may be not suitable for all intrahepatic cholangiocarcinoma patients: results of a monocentric series. World J Gastroenterol 2013;19(47):9084–91. https://doi.org/10.3748/wjg.v19.i47.9084.

[6] Bektas H, Yeyrek C, Kleine M, et al. Surgical treatment for intrahepatic cholangiocarcinoma in Europe: a single center experience. J Hepatobiliary Pancreat Sci 2015;22(2):131–7. https://doi.org/10.1002/jhbp.158.

[7] Adachi T, Eguchi S, Beppu T, et al. Prognostic impact of preoperative lymph node enlargement in intrahepatic cholangiocarcinoma: a multi-institutional study by the kyushu study group of liver surgery. Ann Surg Oncol 2015;22(7):2269–78. https://doi.org/10.1245/s10434-014-4239-8.

[8] Spolverato G, Yakoob MY, Kim Y, et al. The impact of surgical margin status on long-term outcome after resection for intrahepatic cholangiocarcinoma. Ann Surg Oncol 2015;22(12):4020–8. https://doi.org/10.1245/s10434-015-4472-9.

[9] Bagante F, Spolverato G, Cucchetti A, et al. Defining when to offer operative treatment for intrahepatic cholangiocarcinoma: a regret-based decision curves analysis. Surgery 2016;160(1):106–17. https://doi.org/10.1016/j.surg.2016.01.023.

[10] Hyder O, Marques H, Pulitano C, et al. A nomogram to predict long-term survival after resection for intrahepatic cholangiocarcinoma: an Eastern and Western experience. JAMA Surg 2014;149(5):432–8. https://doi.org/10.1001/jamasurg.2013.5168.

[11] Spolverato G, Kim Y, Alexandrescu S, et al. Is hepatic resection for large or multifocal intrahepatic cholangiocarcinoma justified? Results from a multi-institutional collaboration. Ann Surg Oncol 2015;22(7):2218–25. https://doi.org/10.1245/s10434-014-4223-3.

[12] Collins GS, De Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. 2014. https://doi.org/10.1186/1471-2288-14-40.

[13] Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;13:33. https://doi.org/10.1186/1471-2288-13-33.

[14] Sasaki K, Margonis GA, Andreatos N, et al. Preoperative risk score and prediction of long-term outcomes after hepatectomy for intrahepatic cholangiocarcinoma. J Am Coll Surg 2018;226(4):393–403. https://doi.org/10.1016/j.jamcollsurg.2017.12.011.

[15] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594. https://doi.org/10.1136/BMJ.G7594.

[16] Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol 2005;58(5):475–83. https://doi.org/10.1016/j.jclinepi.2004.06.017.

[17] Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995;48(12):1503–10. http://www.ncbi.nlm.nih.gov/pubmed/8543964. [Accessed 9 January 2019].

[18] Riley RD, Snell KIE, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 2019;38(7):1276–96. https://doi.org/10.1002/sim.7992.

[19] Riley RD, Snell KIE, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: Part I – continuous outcomes. Stat Med 2019;38(7):1262–75. https://doi.org/10.1002/sim.7993.

[20] Bagante F, Spolverato G, Weiss M, et al. Defining long-term survivors following resection of intrahepatic cholangiocarcinoma. J Gastrointest Surg 2017;21(11):1888–97. https://doi.org/10.1007/s11605-017-3550-7.

[21] Shen J, Wen T, Li C, Yan L, Li B, Yang J. The prognostic prediction role of preoperative serum albumin level in patients with intahepatic cholangiocarcinoma following hepatectomy. Dig Dis 2018;36(4):306–13. https://doi.org/10.1159/000487479.

[22] Lunsford KE, Javle M, Heyne K, et al. Liver transplantation for locally advanced intrahepatic cholangiocarcinoma treated with neoadjuvant therapy: a prospective case-series. lancet Gastroenterol Hepatol 2018;3(5):337–48. https://doi.org/10.1016/S2468-1253(18)30045-1.