

Supplementary Information(SI)

An individualized nomogram for survival prediction in
intrahepatic cholangiocarcinoma: A population-based
analysis

This file includes:

Supplementary Tables

Tables [S1](#) to [S8](#)

TOC

Supplementary tables-Tab.S1	3
Supplementary tables-Tab.S2	6
Supplementary tables-Tab.S3	8
Supplementary tables-Tab.S4	9
Supplementary tables-Tab.S5	10
Supplementary tables-Tab.S6	12
Supplementary tables-Tab.S7	13
Supplementary tables-Tab.S8	13

Supplementary tables-Tab.S1

Table S1. Comparison of representative previously published prognostic models for intrahepatic cholangiocarcinoma

Study	Prediction target	Study population / data source	Sample size	Main variables included	Validation method	Reported performance	Main features or limitations
Hahn et al. [10]	Survival prediction during chemotherapy	Non-resectable or recurrent ICC undergoing chemotherapy	121 patients	Baseline CA19-9, CA19-9 change, RECIST imaging response, tumor spread, metastasis, chemotherapy regimen	Retrospective cohort	AUC for CA19-9 change predicting PD: 0.73	Biomarker/imaging-based; single-center cohort.
Yuan et al. [16]	Overall survival	ICC patients from SEER database	1,933 patients	Age, surgery, radiotherapy, chemotherapy, T stage, N stage, M stage	single-institution external validation	C-index: 0.754; AUCs: 0.821, 0.828, and 0.836 for 1-, 3-, and 5-year OS	SEER-based model; small external cohort.
Hyder et al. [20]	Long-term survival	Patients with ICC undergoing surgical resection	514 patients	Age, Size, No. of Tumors Nodal Status, Vascular Invasion, Cirrhosis	Multicenter cohort validation	C-index: 0.692; bootstrap-corrected testing C statistic: 0.706; good calibration	Multicenter surgical cohort; external validation needed.
Shen et al. [35]	OS prediction in	SEER cohort aged 50–74 years	1,751 patients	Sex, diagnosis-to-treatment interval, grade, AJCC stage,	Random split internal	RSF C-index: 0.760; Brier score: 0.124;	ML-based model; age-specific cohort.

	high-incidence age group			surgery, sequence number, tumor size, chemotherapy	validation; hospital external validation	mean AUC: 0.882; internal/external C-index: 0.72/0.80	
Present study	1-, 3-, and 5-year overall survival	Population-based ICC patients from SEER database	Training cohort: 1,751; temporal validation cohort: 751	Age, sex, lymph node involvement, surgery, chemotherapy, radiotherapy, and tumor extension	Internal validation and temporal validation within SEER	Evaluated using C-index, time-dependent AUC, calibration curves, DCA, and IDI	Incorporates refined tumor-extension categories and systematically compares model performance with AJCC TNM staging

References in this supplementary table follow the numbering order of the main manuscript reference list.

[10] Hahn F, Müller L, Jungmann F, Mähringer-Kunz A, Tanyildizi Y, Düber C, et al. Survival prediction for patients with non-resectable intrahepatic cholangiocarcinoma undergoing chemotherapy: a retrospective analysis comparing the tumor marker CA 19-9 with cross-sectional imaging. *J Cancer Res and Clin Oncol*. 2020;146(7):1883-90.

[16] Yuan, C., Hu, Z., Wang, K., Zou, S.: Development and validation a nomogram for predicting overall survival in patients with intrahepatic cholangiocarcinoma. *Front Surgery*.2021;8:659422.

[20] Hyder O, Marques H, Pulitano C, Marsh JW, Alexandrescu S, Bauer TW, 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.

[35] Shen J, Yang D, Zhou Y, Pei J, Wu Z, Wang X, et al. Development of machine learning models for patients in the high intrahepatic cholangiocarcinoma incidence age group. *BMC Geriatr*. 2024;24(1):553.

Supplementary tables-Tab.S2

Table S2. Variable coding and cutoffs

In the SEER database, the age variable was grouped into 10-year intervals. Lung metastasis, bone metastasis, and brain metastasis were all recorded as binary variables (yes/no). Tumor size was categorized using a 5 cm cut-off value, based on established criteria from previous studies. Both tumor extension and lymph node metastasis were originally provided as numerical codes. These variables were reclassified according to the official SEER coding guidelines (<https://seer.cancer.gov/analysis/>). For instance, code 100 in the extension variable represents "Single lesion (one lobe)", while code 150 indicates "More than one lobe involved by contiguous growth". Lymph node metastasis was grouped following the same principles, in accordance with SEER coding rules.

Variables	variable coding and cutoffs
Sex	
Male	1
Female	0
Age	
Under 50	0
50-60	1
60-70	2
70-80	3
80-90	4
Bone Mets	
Yes	1
No	0
Brain Mets	
Yes	1
No	0
Liver Mets	
Yes	1
No	0

Lung Mets		
Yes		1
No		0
Tumor Size		
<5 cm		0
≥5cm		1
Tumor Extension		
Single lesion one lobe		0
Multiple tumors one lobe		1
Single lesion More than one lobe		2
Multiple tumors More than one lobe		3
Lymph nodes		
No		0
Yes		1
Chemotherapy		
No		0
Yes		1
Surgery		
No		0
Yes		1
Radiation		
No		0
Yes		1

Supplementary tables-Tab.S3

Table S3. The proportional hazards assumption

Variable	χ^2	df	P-value
Gender	2.74	1	0.0978
Age	21.47	1	<0.001
Extension	6.39	1	0.0115
Lymph nodes	14.97	1	<0.001
Chemotherapy	252.38	1	<0.001
Radiation	46.35	1	<0.001
Surgery	25.84	1	<0.001
Global test	371.66	8	<0.001

Note: The proportional hazards assumption was assessed using Schoenfeld residuals. A significant P-value ($P < 0.05$) indicates violation of the proportional hazards assumption for that variable. The global test evaluates the overall model assumption. Variables including age, extension, lymph node status, chemotherapy, radiation, and surgery showed significant time-varying effects, while gender and tumor size satisfied the proportional hazards assumption.

Supplementary tables-Tab.S4

Table S4. Missingness of key variables in the original SEER extraction cohort

Variable / exclusion criterion	Definition used for missingness	Missing / blank / unknown, n	Proportion, %
Bone metastasis	Unknown or unavailable	379	3.75
Brain metastasis	Unknown or unavailable	387	3.83
Liver metastasis	Unknown or unavailable	362	3.58
Lung metastasis	Unknown or unavailable	418	4.14
Any metastatic-site variable	Any of bone, brain, liver, or lung metastasis recorded as unknown	523	5.18
	or unavailable		
Tumor size	Blank or coded as 999	7,785	77.06
Tumor extension	Blank or coded as 999	7,133	70.60
Chemotherapy or radiation	Blank information for chemotherapy or radiation status	0	0.00
Survival months	Unknown survival time	14	0.14
Lymph node status	Blank or coded as 999	7,090	70.18

Note: Percentages were calculated using the original SEER extraction cohort as the denominator (n = 10,104). For metastatic-site variables, unknown or unavailable records were treated as missing. For tumor size, tumor extension, and lymph node status, blank values and code 999 were treated as missing.

Supplementary tables-Tab.S5

Table S5. Comparison of available survival-related characteristics between included and excluded patients in the original SEER extraction cohort

Characteristic	Included patients, (n = 1,751)	Excluded patients (n = 8,353)	P value	SMD
Age, years			<0.001	0.136
<50	177 (10.1%)	638 (7.6%)		
50–60	355 (20.3%)	1,462 (17.5%)		
60–70	565 (32.3%)	2,760 (33.0%)		
70–80	472 (27.0%)	2,396 (28.7%)		
>80	182 (10.4%)	1,097 (13.1%)		
Sex			0.747	0.008
Male	889 (50.8%)	4,281 (51.2%)		
Female	862 (49.2%)	4,072 (48.8%)		
status			<0.001	0.363
Alive	189 (10.8%)	2,037 (24.4%)		
Dead	1562 (89.2%)	6,316 (75.6%)		
Survival months, median (IQR)	14 (5–35)	7 (2–16)	<0.001	0.569
Surgery			<0.001	0.274
No	1189 (67.9%)	6,639 (79.5%)		
Yes	562 (32.1%)	1,714 (20.5%)		
Radiation			0.470	0.018
No	1462 (83.5%)	7,031 (84.2%)		
Yes	289 (16.5%)	1,322 (15.8%)		
Chemotherapy			<0.001	0.110
No	838 (47.9%)	3,540 (42.4%)		
Yes	913 (52.1%)	4,813 (57.6%)		
Bone metastasis			<0.001	0.169
No	1657 (94.6%)	7,206 (86.3%)		
Yes	94 (5.4%)	768 (9.2%)		
Brain metastasis			0.692	0.015
No	1742 (99.5%)	7,933 (95.0%)		
Yes	9 (0.5%)	33 (0.4%)		
Liver metastasis			<0.001	0.408
No	1598 (91.3%)	6,090 (72.9%)		
Yes	153 (8.7%)	1,900 (22.7%)		
Lung metastasis			<0.001	0.181
No	1,597 (91.2%)	6,777 (81.1%)		
Yes	154 (8.8%)	1,157 (13.9%)		
Tumor size			-	-
<5 cm	704 (40.2%)	-		
≥5 cm	1047 (59.8%)	-		
Tumor extension			-	-
Single lesion, one lobe	710 (40.5%)	-		
Multiple tumors, one lobe	187 (10.7%)	-		

Characteristic	Included patients, (n = 1,751)	Excluded patients (n = 8,353)	P value	SMD
Single lesion, >1 lobe	209 (11.9%)	-		
Multiple tumors, >1 lobe	645 (36.8%)	-		
Lymph node Involvement			-	-
No	1242 (70.9%)	-		
Yes	509 (29.1%)	-		

Note: Values are presented as n (%) unless otherwise indicated. Percentages were calculated using the corresponding cohort size as the denominator. Missing or unknown values among excluded patients are not displayed in this table and are summarized separately in Supplementary Table S4. Tumor size, tumor extension, and lymph node involvement were not formally compared because missing or unknown values in these variables were major determinants of exclusion from the complete-case cohort. P values were calculated using the chi-square test for categorical variables and the Mann–Whitney U test for survival months. SMD values are absolute standardized mean differences

Supplementary tables-Tab.S6

Table S6. Complete Cox model coefficients

Variable	Category / contrast	β coefficient	SE	HR	95% CI	P value
Sex	Male vs Female	0.1144	0.0511	1.121	1.026–1.229	0.0251
Age	50–60 vs <50	0.0945	0.1007	1.099	0.902–1.339	0.3479
Age	60–70 vs <50	0.1663	0.0962	1.181	0.978–1.425	0.0838
Age	70–80 vs <50	0.3790	0.0983	1.461	1.205–1.772	0.0001
Age	>80 vs <50	0.3072	0.1188	1.360	1.077–1.716	0.0097
Lymph node involvement	Yes vs No	0.3567	0.0600	1.429	1.270–1.607	<0.0001
Tumor extension	Multiple tumors, one lobe vs single lesion,	0.1059	0.0898	1.112	0.932–1.326	0.2385

Variable	Category / contrast	β coefficient	SE	HR	95% CI	P value
	one lobe					
Tumor extension	Single lesion, >1 lobe vs single lesion, one lobe	0.4066	0.0841	1.502	1.274–1.771	<0.0001
	one lobe					
Tumor extension	Multiple tumors, >1 lobe vs single lesion, one lobe	0.4400	0.0631	1.553	1.372–1.757	<0.0001
Radiation	Yes vs No	-0.2887	0.0684	0.749	0.655–0.857	<0.0001
Surgery	Yes vs No	-1.5134	0.0670	0.220	0.193–0.251	<0.0001
Chemotherapy	Yes vs No	-0.5849	0.0592	0.557	0.497–0.624	<0.0001

The linear predictor is calculated as:

$$\begin{aligned}
LP = & 0.1144 \times I(\text{Male}) + 0.0945 \times I(\text{Age}50 -- 60) \\
& + 0.1663 \times I(\text{Age}60 -- 70) + 0.3790 \times I(\text{Age}70 -- 80) \\
& + 0.3072 \times I(\text{Age} > 80) \\
& + 0.3567 \times I(\text{Lymphnodeinvolvement}) \\
& + 0.1059 \times I(\text{Multiple tumors, one lobe}) \\
& + 0.4066 \times I(\text{Singlelesion, } > 1\text{lobe}) \\
& + 0.4400 \times I(\text{Multiple tumors, } > 1\text{lobe}) \\
& - 0.2887 \times I(\text{Radiotherapyyes}) \\
& - 1.5134 \times I(\text{Surgeryyes}) \\
& - 0.5849 \times I(\text{Chemotherapy yes})
\end{aligned}$$

The predicted survival probability is then calculated as:

$$S(t|X) = S_0(t)^{\exp(LP)}$$

Supplementary tables-Tab.S7

Table S7. Baseline survival estimates

Time point	Baseline survival ($S_0(t)$)
12 months	0.4103
36 months	0.0844
60 months	0.0293

Note: β coefficients, standard errors, hazard ratios, 95% confidence intervals, and P values were derived from the final Cox proportional hazards model. Reference categories were female sex, age <50 years, no lymph node involvement, single lesion confined to one lobe, no radiotherapy, no surgery, and no chemotherapy. The baseline survival estimates $S_0(t)$ correspond to the reference covariate pattern. HR, hazard ratio; CI, confidence interval; SE, standard error.

Therefore:

$$S(1\text{-year} | X) = 0.4103^{\exp(LP)}$$

$$S(3\text{-year} | X) = 0.0844^{\exp(LP)}$$

$$S(5\text{-year} | X) = 0.0293^{\exp(LP)}$$

Worked example

For example, consider a male patient aged 60–70 years, with multiple tumors involving more than one lobe, lymph node involvement, no surgery, chemotherapy, and no radiotherapy.

Total linear predictor LP 0.4925

Thus:

$$\exp(LP) = \exp(0.4925) = 1.6364$$

The predicted survival probabilities are:

Time point	Formula	Predicted survival probability
1-year OS	$0.4103^{1.6364}$	0.2327
3-year OS	$0.0844^{1.6364}$	0.0175
5-year OS	$0.0293^{1.6364}$	0.0031

Supplementary tables-Tab.S8

Table S8. Sensitivity analyses comparing alternative age specifications

Model	Age specification	Tumor size specification	C-index
Main model	10-year age categories	<5 cm / $\geq$ 5 cm	0.760
Sensitivity model 1	Continuous age variable	<5 cm / $\geq$ 5 cm	0.747
Sensitivity model 2	Restricted cubic spline age model	<5 cm / $\geq$ 5 cm	0.746

Note: Restricted cubic spline analysis for age showed no statistically significant nonlinear component (P for nonlinearity = 0.681).