CrossMark

ORIGINAL ARTICLE

Recalibrating the Child–Turcotte–Pugh Score to Improve Prediction of Transplant-Free Survival in Patients with Cirrhosis

David E. Kaplan[1] · Feng Dai[2] · Melissa Skanderson[2] · Ayse Aytaman[3] ·
Michelle Baytarian[4] · Kathryn D'Addeo[2] · Rena Fox[5] · Kristel Hunt[6] ·
Astrid Knott[7] · Rajni Mehta[2] · Marcos Pedrosa[4] · Christine Pocha[7] ·
Adriana Valderrama[8] · Tamar Taddei[2] · for the VOCAL Study Group

Abstract

Background The Child–Turcotte–Pugh (CTP) score is a widely used and validated predictor of long-term survival in cirrhosis. However, the cutpoints for stratifying laboratory variables in CTP have never been validated.

Objective The objective of this study was to identify evidence-based cutpoints for the CTP laboratory subscores to improve its predictive capacity for transplant-free survival.

Design Retrospective observational study.

Data Source Using a cohort of 30,897 cirrhotic US Veteran patients with at least 5 years of follow-up, we performed Cox proportional hazard survival model iterations varying the upper and lower cutpoints for INR, total bilirubin and albumin CTP subscores. Cutpoints yielding the highest Harrell's *C*-statistics for concordance with transplant-free survival were incorporated into a modified CTP (mCTP) score. Validation of the mCTP was performed at multiple time frames within the follow-up period of the cohort and within subsets defined by disease etiology.

Results Modification of CTP cutpoints increased the Harrell's *C*-statistic for age- and gender-adjusted Cox proportional hazard models from 0.701 ± 0.002 to 0.709 ± 0.002 and the risk ratio per unit change from 1.49 (1.48–1.50) to 1.53 (1.52–1.54). The modified cutpoints showed superiority in predicting 5-year transplant-free survival in various disease etiology subgroups. A mCTP substituting serum creatinine for INR performed superiorly for predicting 5-year transplant-free survival.

Conclusion We propose an evidence-based recalibration of CTP score cutpoints that optimizes this model's capacity to predict transplant-free survival in patients with cirrhosis. The CTP score remains the best predictor of 5-year overall and transplant-free survival in patients with cirrhosis.

✉ David E. Kaplan
dakaplan@mail.med.upenn.edu

1 Gastroenterology Section, Department of Medicine, Corporal Michael J. Crescenz VA Medical Center, 3900 Woodland Avenue, Bldg. 21, Room A422, Philadelphia, PA 19104, USA

2 VA Connecticut-Healthcare System, 950 Campbell Avenue, West Haven, CT 06516, USA

3 VA New York Harbor Health Care System, 800 Poly Place, Brooklyn, NY 11209, USA

4 Boston VA Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130, USA

5 San Francisco VA Medical Center, 4150 Clement Street, San Francisco, CA 94121, USA

6 James J. Peters VA Medical Center, 130 West Kingsbridge Road, Bronx, NY 10468, USA

7 Minneapolis VA Health Care System, One Veterans Drive, Minneapolis, MN 55417, USA

8 Bayer HealthCare Pharmaceuticals, 100 Bayer Blvd, Whippany, NJ 07981, USA

Springer

Introduction

The Child–Turcotte–Pugh (CTP) score is a widely used and validated predictor of long-term survival in cirrhosis [1]. Described initially by Child and Turcotte [2] and modified by Pugh et al. [3], the CTP has undergone several additional modifications, most recently to change the cutpoint for serum albumin levels based on laboratory normal value ranges [4]. While highly predictive of surgical risks [5–9], hospital mortality [10, 11], post-embolization mortality [12, 13], transplantation waitlist mortality [14], and long-term survival [15] in cirrhosis, its major limitations are the dependence upon subjective variables and convenience-based cutpoints that have never been validated formally [4].

The Veterans Outcomes and Costs Associated with Liver Disease Study Group (VOCAL) recently developed and validated a method to calculate CTP scores [16] from administrative datasets, designated as the electronic CTP (eCTP). The VOCAL cohort includes over 59,000 Veterans with prevalent cirrhosis treated between the years 2008 and 2010 for whom baseline demographics, disease etiology, and overall and transplant-free survival data have been determined. By creating operational definitions for the subjective CTP variables (ascites and encephalopathy), we demonstrated that the eCTP score was superior to the model for end-stage liver disease (MELD) score, Charlson–Deyo index, VACS index [17], and CirCom score [18] for prediction of 5-year overall and transplant-free survival in this cohort.

Given the size of the cohort and comprehensive follow-up, the cohort appeared ideal for testing the hypothesis that the current CTP laboratory cutpoint values are not optimized for predicting survival in cirrhosis. We therefore performed serial modifications of the eCTP varying the upper and lower cutpoints for INR, total bilirubin and serum albumin to determine the impact of concordance with transplant-free survival in Cox proportional hazard survival models. We found that current cutpoints indeed do not optimally predict transplant-free survival, and we define cutpoints that provide better concordance with survival models. Compared to MELD, the current scoring system used for liver transplant allocation, both original and modified CTP (mCTP) are far superior predictors of intermediate- and long-term survival in this large cohort of cirrhotic patients.

Methods

Data Sources

The derivation of the VOCAL cohort has previously been described [16]. Using the Veterans Health Administration (VHA) Corporate Data Warehouse (CDW) with local IRB approvals at each site, data from all patients with cirrhosis using a validated algorithm (two outpatient or one inpatient ICD9-CM code for cirrhosis [571.2, 571.5]) [19] from the period of 1/1/2008 to 12/31/2010 were obtained. For these individuals, we obtained all in- and outpatient ICD9-CM codes, CPT codes, pharmacy data, and laboratory values from 1/1/2002 to 12/31/2013. Death was ascertained using the Vital Status File (censoring as of December 31, 2014). Liver transplantation status was obtained by cross-referencing United Network of Organ Sharing/Organ Procurement and Transplantation Network STAR-file data [20]. For non-laboratory CTP subscores, we utilized our previously validated algorithm incorporating pharmacy data, CPT codes, and ICD9 codes to determine the presence and severity of ascites and encephalopathy [16]. Laboratory values close to the end date of the specified quarter of analysis (e.g., 3/31/2008 for first quarter 2008) were used, and the duration of survival was calculated from the quarter end-date as previously described [16].

Modified Child–Turcotte–Pugh (mCTP) Score Iterations

Using R survival package [21, 22], and year 2008 quarter 1 data, mCTP scores were derived for serial modifications of the lower cutpoint of INR (from 1.0 to 4.0), the upper cutpoint of INR (from 2.0 to 5.0), the lower cutpoint of total bilirubin (from 1.0 to 4.0), the upper cutpoint of total bilirubin (from 2.0 to 5.0), the lower cutpoint of serum albumin (from 1.0 to 4.0), and the upper cutpoint of serum albumin (from 2.5 to 4.5). Ascites and encephalopathy eCTP subscores were held constant based on previous operational definitions [16]. The discriminative capability of each CTP score in Cox proportional hazards regression models adjusting for gender and age was evaluated using the concordance system of Harrell et al. [23] Optimal cutpoints were identified then validated in dataset from 2008 quarters 2–4 as well as evaluated for concordance in various patient subsets. The mCTP models were also compared to the original CTP (oCTP) score, a proposed modification of the CTP (Huo CTP) developed in Taiwan [24], and MELD score. Results were also confirmed by using competing risk models in R (mstate package

[21, 25]). Due to relatively lower LogWorth scores in multivariable models for the INR and encephalopathy subscores, a 5–13-point CTP model (mCTP [13]) in which the INR and hepatic encephalopathy subscores were dichotomized (1 = normal or below cutpoint, 2 = abnormal or above cutpoint) was also evaluated for its performance.

Exploratory Analyses

Cox proportional hazard models were fit utilizing age, gender, three-point ascites subscore, three-point albumin subscore (using 3.6 and 3.3 g/dl cutpoints), and three-point total bilirubin subscore (2.9 and 3.3 mg/dl cutpoints) with several variations: three-point versus two-point INR stratification (1.8/4.5 vs. 3.3), three-point versus two-point encephalopathy stratification (standard vs. absent/present), and inclusion of a two-point serum creatinine (sCr) subscore using a cutpoint of 1.8 mg/dl derived from serial iteration of the models. A creatinine-mCTP model (mCTP-Cr) eliminating INR was then iterated in serial Cox models to identify cutpoints for all three laboratory variables to identify the model yielding the highest Harrell's *C*-statistic. To assess each model's concordance for predicting outcome across the range of clinical severity, oCTP, mCTP, mCTP-Cr, and MELD were assessed in serial subgroups of the cohort defined by three-level oCTP ranges (e.g., 5–7, 6–8, 7–9, …, 13–15) or defined by MELD ranges (6–10, 7–11, 8–12, …., 36–40).

Results

Demographics of the VOCAL cohort with prevalent cirrhosis during the first quarter 2008 have been previously described [16]. The cohort consists of predominantly well-compensated, white (72.3 %) male (97.2 %) cirrhotic patients. The dominant non-exclusive underlying liver diseases include hepatitis C infection (48.8 %) and alcohol abuse/dependence (62.6 %); 18.7 % have cirrhosis not attributable to alcohol or viral hepatitis. The median MELD score was 10. Median eCTP class/score was A6, and 43 % were eCTP B7 or higher. For this cohort, at least 5 years of follow-up for death or transplantation were available.

To screen for the directionality of changes in individual laboratory cutpoints, we first ran Cox proportional hazard models to estimate concordance of mCTP scores altering the upper and lower limits of each of the three laboratory variables INR, serum albumin, and total bilirubin while fixing the other two variable cutpoints at the original values. As shown in Fig. 1 (triangles), varying the lower INR cutpoint over a wide range identified optimal prediction at 1.8, very similar to the existing cutpoint of 1.7. By contrast, the

optimal value for the upper INR cutpoint was significantly greater than the existing 2.3, plateauing at approximately 4.5. For serum albumin, an upper cutpoint of 3.6 g/dl, similar to the existing 3.5 g/dl cutpoint, yielded optimal concordance. However, a lower cutpoint of 3.3 g/dl, significantly higher than the existing 2.8 g/dl, appeared optimal. Varying the lower cutpoint of total bilirubin over a wide range showed an initial peak of concordance at 1.3 mg/dl and second higher peak at 2.9 mg/dl, both significantly better than the current 2.0 mg/dl lower cutpoint. Marginally improved predictive capacity of the upper bilirubin cutpoint from the existing 3.0–3.3 mg/dl improved performance; further increases yielded no predictive improvement.

The optimal cutpoints were further refined by varying each single variable while fixing the other two variables at the newly identified optimal cutpoints. As shown in Fig. 1 (circles), estimates for optimal cutpoints for each of the laboratory components of the CTP score did not significantly change but the overall predictive capacity of the mCTP increased to approximately 0.709 ± 0.002 from 0.701 ± 0.002 with the oCTP model. The cutpoints for the final proposed mCTP and the oCTP are summarized in Table 1.

Using the optimized cutpoints in a multivariable Cox proportional hazard model including age, and the five CTP subscores with or without including gender in the model [Table 2(A)], serum albumin, total bilirubin, and ascites subscores had the greatest contribution (LogWorth) to the risk prediction of the CTP model. For these three parameters, incremental risk ratios for a one score (none) to two (mild or medically controlled) to three (severe or medically refractory) increased in a relatively linear fashion with relative risks (RR) of 1.5–1.6 for scores of 2 and 2.1–3.2 for scores of 3. Encephalopathy scores showed a modest increased RR for a score of 2 [RR 1.2 (95 % CI 1.1–1.3)] and a further modest increase for a score of 3 [RR 1.8 (95 % CI 1.7–1.9)]. INR showed an extremely modest increase in the RR of death or transplant for a score of 2 [RR 1.3 (95 % CI 1.2–1.4)] and no significant impact for a score of 3. Given the low RR for scores of 3 and overall small contribution to the CTP model, we designed a second mCTP model dichotomizing encephalopathy (1 = absent, 2 = present) and INR. To identify the appropriate single cutpoint for INR in a two-level model, we calculated Harrell's *C*-statistics varying INR over the range from 1 to 6.5 while varying albumin and total bilirubin cutpoints over narrower ranges (Fig. 4 in "Appendix"); we found that INR cutpoints from 3.1 to 4.0 optimized discriminative capacity. Our simplified mCTP (mCTP$_{13}$) using an INR cutpoint of 3.3 (for consistency with albumin and total bilirubin cutpoints) is shown in Table 1. The Harrell's *C*-statistic for age- and gender-adjusted mCTP$_{13}$ in the first quarter of 2008 cohort was identical to the mCTP (0.709 ± 0.002) with a significantly higher hazard ratio

Fig. 1 Impact of varying upper and lower cutoffs of CTP laboratory variables on concordance (Harrell's *C*) for prediction of 5-year transplant-free survival. **a–c** For each simulation, one variable was modified while fixing the other five variables at either original cutpoints (*triangle*), modified optimized cutpoints (*open circle*) or intermediate values (*diamond*). *Arrowhead* shows optimal cutpoint defined around original CTP cutpoints for other five variables. *Filled triangle* represents original CTP, and *filled circle* represents modified CTP. **d** Effect of iterating serum creatinine cutpoints in creatinine-modified CTP model. *Arrowhead* indicates point of optimal concordance
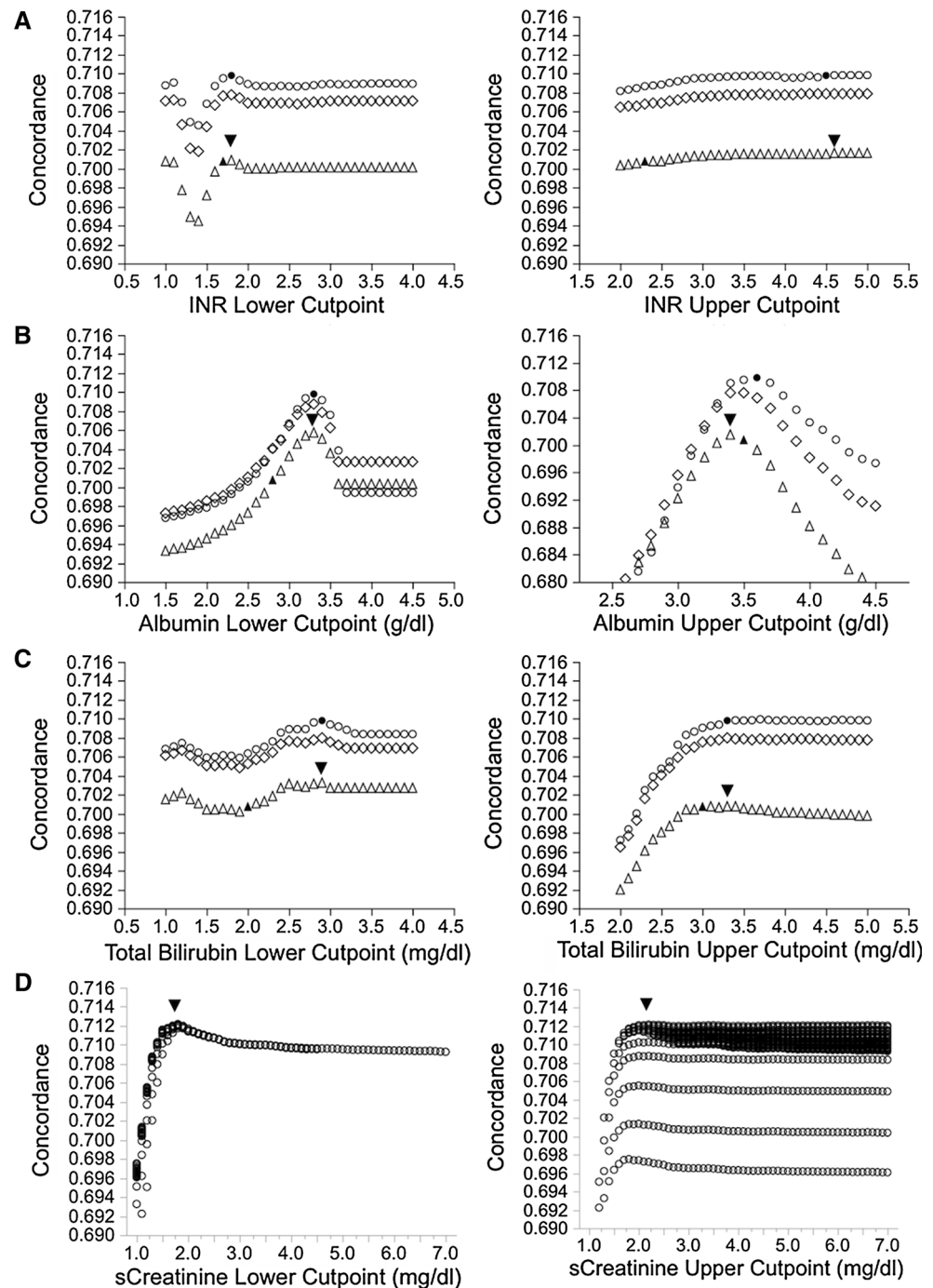
1.60 (95 % CI 1.58–1.61) than the mCTP [1.53 (95 % CI 1.52–1.54)].

We next investigated the utilization of sCr in mCTP models. Serial Cox proportional hazard models adding a lower and upper cutpoint for sCr factor were iterated over the range of sCr from 1.0 to 7.0 (Fig. 1d). These models showed that the optimal *C*-statistic was achieved using a lower sCr cutpoint of 1.8 and upper cutpoint of 2.3 mg/dl. After inclusion of sCr, the INR subscore no longer was statistically significantly associated with 5-year TFS (data

not shown), so a simplified model was created substituting INR with sCr. To evaluate the impact of this change on the other two laboratory cutpoints, we iterated upper and lower cutpoints for total bilirubin and serum albumin over a wide range to assess trends. We found that optimal cutpoints for albumin did not change (3.6 and 3.3 g/dl), the lower total bilirubin cutpoint did not change (2.9 mg/dl), but the optimal upper total bilirubin cutpoint increased up to 3.7 mg/dl. The final model (mCTP-Cr) is shown in Table 1. In multivariable Cox analysis [Table 2(B)], the

Table 1 Laboratory cutpoints for original CTP (oCTP), modified 5–15 CTP (mCTP), second modified CTP system dichotomizing INR and encephalopathy (mCTP$_{13}$) and creatinine-modified CTP (mCTP-Cr)

Subscore	oCTP				mCTP		
	INR	Albumin (g/dl)	Total bilirubin (mg/dl)		INR	Albumin (g/dl)	Total bilirubin (mg/dl)
1	<1.7	>3.5	<2.0		<1.8	>3.6	<2.9
2	1.7–2.3	2.8–3.5	2.0–3.0		1.8–4.5	3.3–3.6	2.9–3.3
3	>2.3	<2.8	>3.0		>4.5	<3.3	>3.3

Subscore	mCTP$_{13}^a$			mCTP-Cr		
	INR	Albumin (g/dl)	Total bilirubin (mg/dl)	Serum creatinine (mg/dl)	Albumin (g/dl)	Total bilirubin (mg/dl)
1	≤3.3	>3.6	<2.9	<1.8	>3.6	<2.9
2	>3.3	3.3–3.6	2.9–3.3	1.8–2.3	3.3–3.6	2.9–3.7
3		<3.3	>3.3	>2.3	<3.3	>3.7

[a] Encephalopathy subscore (1 = none, 2 = present)

HRs for creatinine subscores 2 and 3 were 1.6 (1.5–1.7) and 2.1 (1.9–2.2), respectively. The HR for a bilirubin score of 3 increased slightly with the new upper cutpoint. Using these modifications, the C-statistic for the mCTP-Cr model improved to 0.712 ± 0.002 (Table 3).

The performance of the oCTP, mCTP, and mCTP-Cr in predicting 5-year transplant-free survival was validated using datasets for patients in the cohort in the subsequent three quarters of 2008 (Table 3) and compared with MELD, VACS, and Huo CTP (Table 5 in "Appendix"). All CTP-based models were more predictive for 5-year transplant-free survival than MELD or VACS. The Huo CTP modification had a negligible difference from the oCTP. For cohorts drawn from each quarter, overall concordance of mCTP, mCTP$_{13}$, and mCTP-Cr exceeded that of oCTP and these were highly similar. Hazard ratios for mCTP$_{13}$ were significantly greater than mCTP due to the narrower range of scores (5–13 compared to 5–15). Kaplan–Meier survival curves for oCTP, mCTP, and mCTP-Cr (Fig. 4 in "Appendix") show that mCTP eliminates cross-over in survival curves for oCTP 12–13.

Across disease classes, mCTP and mCTP-Cr had statistically superior predictive capacity than oCTP particularly for individuals with alcoholic and non-viral cirrhosis (Table 4). mCTP and mCTP-Cr had statistically superior predictive capacity than oCTP and MELD in non-black men (Table 6 in "Appendix"), but was not statistically better in the relatively small African-American subset of patients. Due to limited numbers, no differences in performance in women were shown for mCTP and mCTP-Cr relative to oCTP, but all three models were superior to MELD. In addition, CTP-based predictive systems were markedly better at predicting 5-year transplant-free survival in individuals with MELD lower than 15, a group generally not considered for transplantation referral based on relatively low short-term mortality. Additionally, mCTP

and mCTP-Cr performed better than oCTP or MELD for predicting shorter term 1-, 2-, 3- and 4-year transplant-free survival (Table 7 in "Appendix").

To assess the predictive performance of mCTP, mCTP-Cr relative to oCTP and MELD, we analyzed serial subsets of the cohort over increasing MELD or oCTP score ranges (Fig. 2). We found that over MELD ranges 12–16 to 27–31, mCTP and mCTP-Cr exhibited higher concordance than oCTP. At MELD 28–32, mCTP and oCTP converge, but mCTP-Cr remains statistically superior until MELD 30–34, beyond which small cohort sizes limit the ability to differentiate model performance. Even at high MELD scores, CTP-based models better predict 5-year TFS (and 1–4-year TFS, data not shown). Over serial binning of the cohort by oCTP ranges (Fig. 2b), mCTP and mCTP-Cr maintain superior concordance for ranges from oCTP 6–8 to 9–11. For patients with oCTP 9–11, mCTP, oCTP, and MELD converge. Above an oCTP 11–13, mCTP-Cr has superior predictive capacity than mCTP or oCTP. For patients with oCTP 11–13, 12–14 or 13–15, MELD best predicts death or transplant events. Fewer than 2 % of cirrhotic patients were alive in these oCTP classes during the index quarter. Thus, among the models mCTP-Cr maintains the greatest degree of predictive consistency across the spectrum of liver disease severity.

Discussion

The CTP score, while developed to predict portosystemic shunt surgery outcomes in cirrhotic patients [2, 3], historically has remained the most widely used staging system to predict long-term survival in cirrhosis [1]. Despite widespread acceptance, several flaws of CTP as a prognostic staging system have been identified [26]. It relies on two subjective assessments, fostering inter-observer variation

Table 2 Multivariate Cox proportional hazard analysis of individual subscores in predicting 5-year transplant-free survival

Regressor	LogWorth	p value	Univariable score	RR (95 % CI)	p value	LogWorth	p value	Multivariable score	RR (95 % CI)	p value
(A) mCTP										
Age	185.6	<0.0001		1.02 (1.02–1.03)	<0.0001	227.5	<0.0001		1.028 (1.026–1.030)	<0.0001
Gender	14.8	<0.0001	M	1.0 (ref.)		2.8	0.001	M	1.0 (ref.)	
			F	0.65 (0.58–0.73)	<0.0001			F	0.83 (0.74–0.93)	0.001
Albumin	1186.8	<0.0001	1	1.0 (ref.)		714.4	<0.0001	1	1.0 (ref.)	
			2	1.8 (1.7–1.9)	<0.0001			2	1.6 (1.6–1.7)	<0.0001
			3	4.1 (3.9–4.2)	<0.0001			3	3.2 (3.1–3.3)	<0.0001
Ascites	447.9	<0.0001	1	1.0 (ref.)		155.6	<0.0001	1	1.0 (ref.)	
			2	2.0 (2.0–2.1)	<0.0001			2	1.5 (1.5–1.6)	0.0001
			3	4.9 (4.4–5.3)	<0.0001			3	2.5 (2.2–2.7)	<0.0001
Total bilirubin	577.7	<0.0001	1	1.0 (ref.)		144.4	<0.0001	1	1.0 (ref.)	
			2	2.0 (1.9–2.2)	<0.0001			2	1.5 (1.4–1.6)	<0.0001
			3	3.9 (3.6–4.1)	<0.0001			3	2.1 (2.0–2.2)	<0.0001
Encephalopathy	378.4	<0.0001	1	1.0 (ref.)		55.02	<0.0001	1	1.0 (ref.)	
			2	1.8 (1.7–1.9)	<0.0001			2	1.2 (1.1–1.3)	<0.0001
			3	4.0 (3.7–4.3)	<0.0001			3	1.8 (1.7–1.9)	<0.0001
INR	235.4	<0.0001	1	1.0 (ref.)		24.23	<0.0001	1	1.0 (ref.)	
			2	2.1 (2.0–2.2)	<0.0001			2	1.3 (1.2–1.4)	<0.0001
			3	3.5 (2.5–4.6)	<0.0001			3	1.3 (1.0–1.8)	0.07
(B) mCTP-Cr										
Age	185.6	<0.0001		1.02 (1.02–1.03)	<0.0001	266.7	<0.0001		1.03 (1.03–1.03)	<0.0001
Gender	14.8	<0.0001	M	1.0 (ref.)		2.4	0.003	M	1.0 (ref.)	
			F	0.65 (0.58–0.73)	<0.0001			F	0.84 (0.75–0.94)	0.003
Albumin	1186.8	<0.0001	1	1.0 (ref.)		697.4	<0.0001	1	1.0 (ref.)	
			2	1.8 (1.7–1.9)	<0.0001			2	1.6 (1.6–1.7)	<0.0001
			3	4.1 (3.9–4.2)	<0.0001			3	3.2 (3.0–3.3)	<0.0001
Ascites	447.9	<0.0001	1	1.0 (ref.)		146.1	<0.0001	1	1.0 (ref.)	
			2	2.0 (2.0–2.1)	<0.0001			2	1.5 (1.5–1.6)	<0.0001
			3	4.9 (4.4–5.3)	<0.0001			3	2.3 (2.1–2.5)	<0.0001
Total bilirubin	606.9	<0.0001	1	1.0 (ref.)		178.6	<0.0001	1	1.0 (ref.)	
			2	2.2 (2.1–2.3)	<0.0001			2	1.5 (1.4–1.6)	<0.0001
			3	4.6 (4.4–5.0)	<0.0001			3	2.2 (2.1–2.3)	<0.0001
Encephalopathy	378.4	<0.0001	1	1.0 (ref.)		55.2	<0.0001	1	1.0 (ref.)	
			2	1.8 (1.7–1.9)	<0.0001			2	1.2 (1.2–1.3)	<0.0001
			3	4.0 (3.7–4.3)	<0.0001			3	1.8 (1.7–1.9)	<0.0001
Creatinine	172.1	<0.0001	1	1.0 (ref.)		98.9	<0.0001	1	1.0 (ref.)	
			2	2.2 (2.1–2.4)	<0.0001			2	1.6 (1.5–1.7)	<0.0001
			3	2.7 (2.5–2.9)	<0.0001			3	2.1 (1.9–2.2)	<0.0001

Table 3 Performance of original CTP (oCTP), modified CTP (mCTP), simplified modified CTP (mCTP₁₃), and creatinine-modified CTP (mCTP-Cr) in predicting 5-year transplant-free survival in cohorts drawn from 2008 quarters 1–4

Quarter	N	Events	oCTP		mCTP	
			HR	Harrel's C	HR	Harrel's C
Q1	30,897	15,275	1.50 (1.49–1.52)	0.701 ± 0.002	1.53 (1.52–1.54)	0.709 ± 0.002
Q2	31,033	15,515	1.48 (1.47–1.49)	0.694 ± 0.002	1.52 (1.51–1.54)	0.705 ± 0.002
Q3	30,989	15,521	1.48 (1.47–1.49)	0.700 ± 0.002	1.51 (1.50–1.52)	0.708 ± 0.002
Q4	30,218	15,287	1.49 (1.48–1.50)	0.702 ± 0.002	1.53 (1.52–1.54)	0.712 ± 0.002

Quarter	N	Events	mCTP₁₃		mCTP-Cr	
			HR	Harrel's C	HR	Harrel's C
Q1	30,897	15,275	1.60 (1.58–1.61)	0.709 ± 0.002	1.57 (1.56–1.59)	0.712 ± 0.002
Q2	31,033	15,515	1.59 (1.57–1.60)	0.705 ± 0.002	1.57 (1.55–1.58)	0.708 ± 0.002
Q3	30,989	15,521	1.57 (1.56–1.59)	0.707 ± 0.002	1.56 (1.54–1.57)	0.711 ± 0.002
Q4	30,218	15,287	1.60 (1.58–1.61)	0.712 ± 0.002	1.57 (1.55–1.58)	0.714 ± 0.002

and hampering application to large datasets. The cutpoints for the objective laboratory variables have never been validated. Additionally, the dynamic range of the CTP with only 11 total strata and three classes theoretically reduces its utility for prioritization in organ allocation. We have previously shown that by creating operational definitions for ascites and encephalopathy based on ICD9-CM diagnosis codes, CPT procedure codes, laboratory and pharmacy refill data, and the subjective variables of the CTP score can be estimated (eCTP) with a high degree of precision [16]. In this study, which represents the largest dataset used to evaluate CTP to date, the electronic oCTP was significantly more accurate than MELD or VACS scores for predicting 1–5-year transplant-free survival. Addition of comorbidity indices such as Charlson–Deyo and CirCom [18] had minimal impact on prediction.

Our cohort of 30,897 cirrhotic subjects with prevalent cirrhosis in the first quarter of 2008 who have at least 5 years of follow-up allows for robust modeling to determine evidence-based, optimal CTP cutpoints. One striking finding on multivariable Cox proportional hazard modeling was that cutpoint estimation for INR, albumin, bilirubin and creatinine were highly independent suggesting that these three measures truly reflect different aspects of the pathophysiology of end-stage liver disease associated with outcome.

The lower bound of INR associated with increased mortality risk was adequately estimated by the oCTP score. The similarity of the lower cutpoint we identified as optimal and the original value is notable, approximating a prothrombin time >4 s longer than control. However, the original upper cutpoint of 2.3 (based on a prothrombin time 6 s longer than control) does not appear to have intrinsic validity; increased risk of death was more closely associated with significantly higher degrees of coagulopathy. INR only

modestly contributes to CTP in multivariable Cox proportional hazard models as a three-level factor. Indeed, a single cutpoint at INR between 3.1 and 3.5 better discriminated the degree of coagulopathy associated with increased risk of death or transplantation. Eliminating patients on warfarin did not impact these findings (data not shown). Dichotomizing the INR subscore (1 = INR <3.3, 2 = INR ≥3.3) also did not negatively impact predictive capacity. Notably, this cutpoint is nearly identical to that identified by Asrani and Kim [27] as the point at which increasing INR no longer is associated with increasing mortality. Our data strongly support the clinical gestalt that by the time in the natural history of cirrhosis at which significant INR increases are observed, the prognosis is so poor that further increases have little impact on observed survival.

Many clinicians consider declining albumin to be the most sensitive marker of progressive liver dysfunction, both as a reflection of impaired synthetic reserve and as a precursor of the neurohumoral response that results in ascites and hepatorenal syndrome. Indeed, relatively modest reductions in albumin (3.7–3.8) from normal markedly increased risk of mortality. Nonetheless, the optimal upper cutpoint of 3.6 g/dl from our cohort approximated the originally chosen value of 3.5 g/dl. Strikingly however, high risk was imparted at serum albumin values much higher than the original 2.8 g/dl, occurring at a lower cutpoint of 3.3 g/dl. Seventeen percent of the cohort had serum albumin values of 2.8–3.2 g/dl and were given an additional point by the mCTP. Median survival for individuals with albumin 2.8–3.2 was significantly shorter [907 days (95 % CI 872–943)] than for those with albumin 3.3–3.5 [1629 days (95 % CI 1569–1700)]. Given the marked disparity in survival of individuals with albumin levels in the lower half of the oCTP albumin score 2 range (2.8–3.2 vs. 3.3–3.5), raising the lower cutpoint of albumin to 3.3 g/dl

Table 4 Performance of original CTP (oCTP) and modified CTP (mCTP) in predicting 5-year transplant-free survival in specific disease etiology subsets

Disease	N	Events	oCTP		mCTP		mCTP-Cr	
			HR	Harrel's C	HR	Harrel's C	HR	Harrel's C
HCV, no EtOH	5549	2075	1.68 (1.64–1.73)	0.735 ± 0.006	1.77 (1.72–1.81)	0.740 ± 0.006	1.80 (1.75–1.86)	0.741 ± 0.006
EtOH, no HCV	9755	5165	1.39 (1.36–1.41)	0.668 ± 0.004	1.42 (1.40–1.44)	0.680 ± 0.004	1.46 (1.44–1.48)	0.682 ± 0.004
HCV + EtOH	9533	5066	1.50 (1.48–1.52)	0.707 ± 0.004	1.51 (1.49–1.53)	0.710 ± 0.004	1.56 (1.54–1.59)	0.712 ± 0.004
HBV	589	240	1.64 (1.52–1.76)	0.727 ± 0.019	1.77 (1.63–1.91)	0.740 ± 0.019	1.82 (1.67–1.99)	0.738 ± 0.019
Other	5777	2872	1.45 (1.42–1.49)	0.711 ± 0.005	1.57 (1.53–1.61)	0.726 ± 0.005	1.63 (1.60–1.68)	0.732 ± 0.005
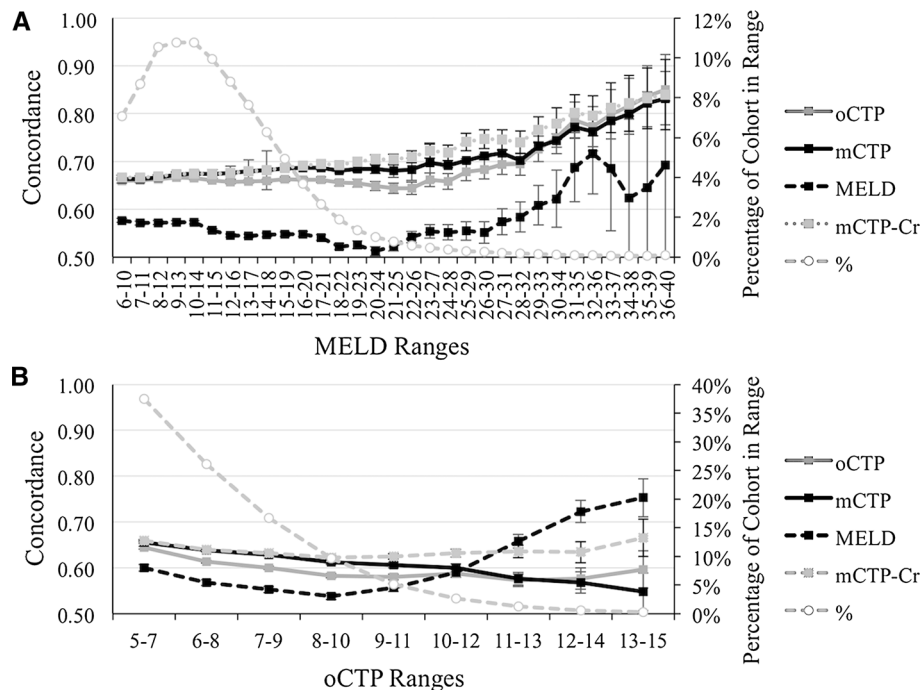
Fig. 2 Impact of varying clinical severity on concordance of oCTP, mCTP, mCTP-Cr, and MELD on 5-year transplant-free survival. **a** The 2008 quarter 1 cohort was serially subsetted by five-point MELD ranges and the C-statistics for oCTP (*solid gray*), mCTP (*solid lack*), mCTP-Cr (*dashed gray*, *solid marker*), and MELD (*dashed black*) plotted. The percentage of patients in each subset is plotted on the secondary axis (*gray dotted line*, *round marker*). **b** The 2008 quarter one cohort was serially subsetted by three-point oCTP ranges and the C-statistics for oCTP (*solid gray*), mCTP (*solid lack*), mCTP-Cr (*dashed gray*, *solid marker*), and MELD (*dashed black*) plotted. The percentage of patients in each subset is plotted on the secondary axis (*gray dotted line*, *round marker*)

appears to be alteration of the oCTP that yields the greatest magnitude of improvement in model performance.

We observed a bimodal behavior of the lower bilirubin cutoff, in which a cutpoint of 1.3 mg/dl was superior to the current 2.0 mg/dl cutpoint, but an even stronger concordance was generated using a higher cutpoint of 2.9 mg/dl. By contrast with the MELD, there was a relatively low transition to a high-risk bilirubin value at 3.3 mg/dl [27]. We could not replicate an improvement of prediction by addition of points to the CTP for very high bilirubin levels (>8 mg/dl) as proposed by Huo et al. [24] either in the complete cohort or those with decompensated cirrhosis (Table 7 in "Appendix"). These data are in concordance with the relatively low weighting of total bilirubin in the

MELD score, but differ from the linear association with mortality observed over a large dynamic range of total bilirubin when utilized in MELD models [27]. We ultimately found that in mCTP-Cr the optimal cutpoints were 2.9 and 3.7 mg/dl.

The inclusion of "subjective" assessments of HE and ascites has been a consistent critique of the CTP scoring system [26]. However, both ascites and HE have been shown to have a strong impact on survival [28–30] and quality of life in end-stage liver disease [31] but correlate poorly with MELD score [30–32]. Encephalopathy is very difficult to quantify administratively [33], and the lack of impact of a higher score could reflect misclassification or could reflect lack of internal validity for a three-level

stratification. Dichotomizing the CTP encephalopathy subscore (1 = absent, 2 = present) based on hospitalizations and/or treatment for HE might simplify and reduce the subjectivity of this assessment. We found that doing so had no significant impact on the predictive capacity of CTP models (C-statistic 0.709 ± 0.002 for both three- and two-level models). Similarly, the Mayo group found that there was no statistical difference in Kaplan–Meier survival rates for HE grades of 2 or 3 [30]. However, in mCTP-Cr a three-level encephalopathy score proved superior (data not shown). The presence of encephalopathy ICD9 codes or medications remains critical to CTP for predictive capacity and partially explains the improved performance of CTP over MELD in predicting survival in this and other cohorts [16, 30]. A similar simplification of ascites scoring worsened CTP predictive capacity; however, the strict operational definition we used for severe or refractory ascites, which required >1 large-volume paracentesis in the preceding 90 days or history of TIPSS placement, could strongly reduce the subjectivity of this assessment in practice.

Strengths of our study include large cohort size, completeness of follow-up for outcomes, and variety of disease etiologies. The limitations of this study, similar to the original derivation cohort, include the predominance of male patients and complexity of the eCTP algorithm, that requires access to claims, laboratory and pharmacy data. Therefore, replicating this study requires comprehensive administrative datasets. External validation of the mCTP and mCTP-Cr score on non-VA administrative datasets would be important before recommending a wholesale change of the 50-year-old CTP scoring system.

Conclusion

The mCTP and creatinine-modified mCTP-Cr models show superiority over oCTP and MELD in predicting 1-, 2-, 3-, 4- and 5-year transplant-free survival in a large cohort of cirrhotic individuals ranging from well to poorly compensated. Augmenting the clinical gestalt of highly skilled and experienced liver surgeons that has stood the test of time with modern evidence-based estimates based on survival models can improve the capacity of CTP to predict short- and long-term survival in multiple cirrhosis etiologies and demographic subsets. We recommend modifying the CTP score used in clinical practice to incorporate the current evidence-based cutpoints.

Compliance with ethical standards

Conflict of interest None of the authors have any conflicts of interest to disclose.

Appendix

See Tables 5, 6 and 7, Figs. 3 and 4.

Table 5 Performance of original CTP (oCTP), modified CTP (mCTP), simplified modified CTP (mCTP$_{13}$), Taiwanese-proposed modification of CTP (Huo CTP), model for end-stage liver disease (MELD), Veteran's Aging Cohort Study (VACS) in predicting 5-year transplant-free survival in specific disease etiology subsets

Regressor	Unit	HR	Harrell's C	R^2	LR test	Wald	Logrank
eCTP	per unit change	1.49 (1.48–1.50)	0.701 ± 0.002	0.213	7394	8711	9188
mCTP	per unit change	1.53 (1.52–1.54)	0.709 ± 0.002	0.232	8172[a]	9842	10,413
mCTP$_{13}$	per unit change	1.60 (1.58–1.61)	0.709 ± 0.002	0.231	8110[a]	9330	9939
mCTP-Cr	per unit change	1.57 (1.56–1.59)	0.712 ± 0.002	0.242	8563[a]	10,225	10,687
Huo CTP	per unit change	1.47 (1.46–1.48)	0.695 ± 0.002	0.211	7306	8716	8974
MELD	per unit change	1.10 (1.09–1.10)	0.639 ± 0.002	0.110	3608[a]	4230	4240
VACS	per 10 unit change	1.32 (1.31–1.33)	0.670 ± 0.002	0.156	5246[a]	5602	5605
eCTP class	B versus A	2.48 (2.40–2.57)	0.682 ± 0.002	0.180	6132	6925	8112
	C versus A	7.35 (6.96–7.77)					
mCTP class	B versus A	2.80 (2.71–2.90)	0.693 ± 0.002	0.205	7090	8004	9564
	C versus A	8.91 (8.42–9.43)					
mCTP-Cr class	B versus A	2.88 (2.79–2.98)	0.690 ± 0.002	0.196	6733	7514	8845
	C versus A	9.84 (9.21–10.50)					

[a] $p < 0.0001$ for difference of LR test by ANOVA with eCTP

Table 6 Performance of original CTP (oCTP), modified CTP (mCTP), simplified modified CTP (mCTP$_{13}$), creatinine-modified CTP (mCTP-Cr), and model for end-stage liver disease (MELD) in predicting 5-year transplant-free survival in specific demographic subgroups
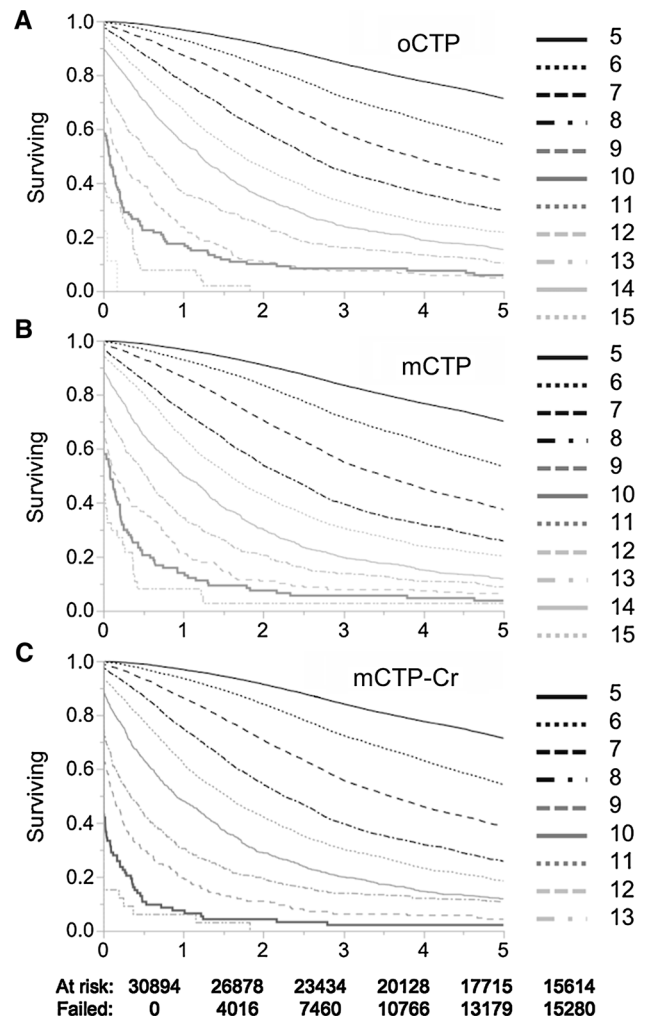
Subgroup	Regressor	HR	Harrell's C	R^2	LR test
Race = white	oCTP	1.47 (1.46–1.49)	0.695 ± 0.003	0.207	5182
n = 22,325	mCTP	1.51(1.50–1.53)	0.705 ± 0.003	0.229	5820
Events = 11,361	mCTP-Cr	1.56 (1.55–1.58)	0.708 ± 0.003	0.240	6118
	MELD	1.10 (1.10–1.10)	0.635 ± 0.003	0.107	2535
Race = black	oCTP	1.61 (1.57–1.65)	0.723 ± 0.007	0.231	1164
n = 4434	mCTP	1.64 (1.60–1.69)	0.726 ± 0.007	0.236	1196
Events = 1835	mCTP-Cr	1.69 (1.64–1.73)	0.730 ± 0.007	0.250	1275
	MELD	1.11 (1.10–1.12)	0.661 ± 0.007	0.130	519
Race = other	oCTP	1.47 (1.44–1.51)	0.703 ± 0.006	0.214	997
n = 4138	mCTP	1.51 (1.47–1.55)	0.715 ± 0.006	0.237	1118
Events = 2079	mCTP-Cr	1.56 (1.53–1.60)	0.717 ± 0.002	0.247	1171
	MELD	1.09 (1.08–1.10)	0.641 ± 0.006	0.108	475
Males	oCTP	1.49 (1.47–1.50)	0.699 ± 0.002	0.211	7136
n = 30,060	mCTP	1.53 (1.51–1.54)	0.708 ± 0.002	0.231	7900
Events = 14,974	mCTP-Cr	1.57 (1.56–1.59)	0.711 ± 0.002	0.240	8265
	MELD	1.10 (1.09–1.10)	0.637 ± 0.002	0.108	3428
Females	oCTP	1.62 (1.52–1.72)	0.726 ± 0.017	0.214	201
n = 837	mCTP	1.66 (1.56–1.77)	0.725 ± 0.017	0.226	214
Events = 301	mCTP-Cr	1.77 (1.66–1.90)	0.729 ± 0.017	0.247	237
	MELD	1.17 (1.14–1.19)	0.683 ± 0.017	0.153	139
eCTP >7	oCTP	1.41 (1.38–1.44)	0.627 ± 0.004	0.125	936
n = 7035	mCTP	1.39 (1.36–1.42)	0.648 ± 0.004	0.152	1157
Events = 5418	mCTP-Cr	1.42 (1.40–1.45)	0.658 ± 0.004	0.175	1355
	MELD	1.05 (1.04–1.06)	0.574 ± 0.004	0.156	142
	Huo CTP	1.38 (1.35–1.40)	0.639 ± 0.004	0.147	1122
MELD <15	oCTP	1.54 (1.52–1.57)	0.669 ± 0.003	0.134	3121
n = 23,739	mCTP	1.55 (1.53–1.58)	0.676 ± 0.003	0.145	3396
Events = 10,251	mCTP-Cr	1.58 (1.56–1.59)	0.680 ± 0.003	0.155	4010
	MELD	1.10 (1.09–1.11)	0.588 ± 0.003	0.040	891
	Huo CTP	1.49 (1.47–1.51)	0.662 ± 0.003	0.125	2907

Table 7 Performance of original CTP (oCTP), modified CTP (mCTP), simplified modified CTP (mCTP$_{13}$), creatinine-modified CTP (mCTP-Cr), and model for end-stage liver disease (MELD) in predicting 1-, 2-, 3-, 4- and 5-year transplant-free survival

Timeframe	Model	HR	Harrell's C
5-year	oCTP	1.49 (1.48–1.50)	0.701 ± 0.002
	mCTP	1.53 (1.52–1.54)	0.709 ± 0.002
	mCTP-Cr	1.57 (1.56–1.59)	0.712 ± 0.002
	MELD	1.10 (1.09–1.10)	0.639 ± 0.002
4-year	oCTP	1.52 (1.51–1.54)	0.709 ± 0.003
	mCTP	1.56 (1.55–1.58)	0.719 ± 0.003
	mCTP-Cr	1.61 (1.60–1.63)	0.721 ± 0.003
	MELD	1.10 (1.10–1.11)	0.645 ± 0.003
3-year	oCTP	1.57 (1.55–1.58)	0.723 ± 0.003
	mCTP	1.61 (1.59–1.62)	0.733 ± 0.003
	mCTP-Cr	1.66 (1.64–1.68)	0.734 ± 0.003
	MELD	1.11 (1.11–1.12)	0.653 ± 0.003
2-year	oCTP	1.63 (1.61–1.65)	0.744 ± 0.003
	mCTP	1.67 (1.65–1.69)	0.755 ± 0.003
	mCTP-Cr	1.73 (1.71–1.75)	0.757 ± 0.003
	MELD	1.13 (1.12–1.13)	0.664 ± 0.003
1-year	oCTP	1.73 (1.71–1.75)	0.781 ± 0.005
	mCTP	1.76 (1.74–1.79)	0.791 ± 0.005
	mCTP-Cr	1.84 (1.81–1.87)	0.795 ± 0.005
	MELD	1.15 (1.15–1.16)	0.691 ± 0.005

Fig. 4 Kaplan–Meier survival curves for original CTP (**a**, oCTP), modified CTP (**b**, mCTP), and Creatinine-modified CTP (**c**, mCTP-Cr)

Fig. 3 Impact of varying INR cutpoint of a modified CTP score dichotomizing encephalopathy and INR on unadjusted concordance (Harrell's C) for prediction of 5-year transplant-free survival. INR cutpoints were varied between 1.0 and 6.5, while varying albumin and bilirubin cutpoints over narrower ranges (±0.5 from optimal)

References

1. D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *J Hepatol*. 2006;44:217–231.
2. Child CG, Turcotte JG. Surgery and portal hypertension. In: Child CG, ed. *The liver and portal hypertension*. Philadelphia: Saunders; 1964:50–64.
3. Pugh RN, Murray-Lyon IM, Dawson JL, et al. Transection of the oesophagus for bleeding oesophageal varices. *Br J Surg*. 1973;60:646–649.
4. Cholongitas E, Papatheodoridis GV, Vangeli M, et al. Systematic review: the model for end-stage liver disease—should it replace Child–Pugh's classification for assessing prognosis in cirrhosis? *Aliment Pharmacol Ther*. 2005;22:1079–1089.
5. Farnsworth N, Fagan SP, Berger DH, et al. Child–Turcotte–Pugh versus MELD score as a predictor of outcome after elective and

emergent surgery in cirrhotic patients. *Am J Surg*. 2004;188: 580–583.

6. Befeler AS, Palmer DE, Hoffman M, et al. The safety of intra-abdominal surgery in patients with cirrhosis: model for end-stage liver disease score is superior to Child–Turcotte–Pugh classification in predicting outcome. *Arch Surg*. 2005;140:650–654; **discussion 55**.

7. Schroeder RA, Marroquin CE, Bute BP, et al. Predictive indices of morbidity and mortality after liver resection. *Ann Surg*. 2006;243:373–379.

8. Hoteit MA, Ghazale AH, Bain AJ, et al. Model for end-stage liver disease score versus child score in predicting the outcome of surgical procedures in patients with cirrhosis. *World J Gastroenterol*. 2008;14:1774–1780.

9. Delis S, Bakoyiannis A, Madariaga J, et al. Laparoscopic cholecystectomy in cirrhotic patients: the value of MELD score and Child–Pugh classification in predicting outcome. *Surg Endosc*. 2010;24:407–412.

10. Butt AK, Khan AA, Alam A, et al. Predicting hospital mortality in cirrhotic patients: comparison of Child–Pugh and acute physiology, age and chronic health evaluation (APACHE III) scoring systems. *Am J Gastroenterol*. 1998;93:2469–2475.

11. Ho YP, Chen YC, Yang C, et al. Outcome prediction for critically ill cirrhotic patients: a comparison of APACHE II and Child–Pugh scoring systems. *J Intensive Care Med*. 2004;19:105–110.

12. Brown DB, Fundakowski CE, Lisker-Melman M, et al. Comparison of MELD and Child-Pugh scores to predict survival after chemoembolization for hepatocellular carcinoma. *J Vasc Interv Radiol*. 2004;15:1209–1218.

13. Reichman TW, Bahramipour P, Barone A, et al. Hepatitis status, Child–Pugh classification, and serum AFP levels predict survival in patients treated with transarterial embolization for unresectable hepatocellular carcinoma. *J Gastrointest Surg*. 2005;9:638–645.

14. Boin IF, Leonardi MI, Pinto AO, et al. Liver transplant recipients mortality on the waiting list: long-term comparison to Child–Pugh classification and MELD. *Transplant Proc*. 2004;36:920–922.

15. Said A, Williams J, Holden J, et al. Model for end stage liver disease score predicts mortality across a broad spectrum of liver disease. *J Hepatol*. 2004;40:897–903.

16. Kaplan DE, Dai F, Aytaman A, et al. Development and performance of an algorithm to estimate the Child–Turcotte–Pugh score from a National Electronic Healthcare database. *Clin Gastroenterol Hepatol*. 2015;13:2333-41.e1–6.

17. Justice AC, McGinnis KA, Skanderson M, et al. Towards a combined prognostic index for survival in HIV infection: the role of 'non-HIV' biomarkers. *HIV Med*. 2010;11:143–151.

18. Jepsen P, Vilstrup H, Lash TL. Development and validation of a comorbidity scoring system for patients with cirrhosis. *Gastroenterology*. 2014;146:147–156; **quiz e15–e16**.

19. Kramer JR, Davila JA, Miller ED, et al. The validity of viral hepatitis and chronic liver disease diagnoses in Veterans Affairs administrative databases. *Aliment Pharmacol Ther*. 2008;27: 274–282.

20. UNOS STAR file. *Based on OPTN Data as of June 1, 2015*.

21. Therneau T. *A Package for Survival Analysis in S*, Version 2.38; 2015. http://CRAN.R-project.org/package=survival.

22. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria; 2013. http://www.R-project.org/.

23. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–387.

24. Huo TI, Lin HC, Wu JC, et al. Proposal of a modified Child–Turcotte–Pugh scoring system and comparison with the model for end-stage liver disease for outcome prediction in patients with cirrhosis. *Liver Transplant*. 2006;12:65–71.

25. de Wreede LCFM, Putter H. mstate: an R package for the analysis of competing risks and multi-state models. *J Stat Softw*. 2011;38:1–30.

26. Wiesner RH, McDiarmid SV, Kamath PS, et al. MELD and PELD: application of survival models to liver allocation. *Liver Transplant*. 2001;7:567–580.

27. Asrani SK, Kim WR. Model for end-stage liver disease: end of the first decade. *Clin Liver Dis*. 2011;15:685–698.

28. Wong RJ, Gish RG, Ahmed A. Hepatic encephalopathy is associated with significantly increased mortality among patients awaiting liver transplantation. *Liver Transplant*. 2014;20:1454–1461.

29. Patidar KR, Thacker LR, Wade JB, et al. Covert hepatic encephalopathy is independently associated with poor survival and increased risk of hospitalization. *Am J Gastroenterol*. 2014;109:1757–1763.

30. Stewart CA, Malinchoc M, Kim WR, et al. Hepatic encephalopathy as a predictor of survival in patients with end-stage liver disease. *Liver Transplant*. 2007;13:1366–1371.

31. Saab S, Ibrahim AB, Shpaner A, et al. MELD fails to measure quality of life in liver transplant candidates. *Liver Transplant*. 2005;11:218–223.

32. Yoo HY, Edwin D, Thuluvath PJ. Relationship of the model for end-stage liver disease (MELD) scale to hepatic encephalopathy, as defined by electroencephalography and neuropsychometric testing, and ascites. *Am J Gastroenterol*. 2003;98:1395–1399.

33. Lo Re V III, Lim JK, Goetz MB, et al. Validity of diagnostic codes and liver-related laboratory abnormalities to identify hepatic decompensation events in the Veterans Aging Cohort Study. *Pharmacoepidemiol Drug Saf*. 2011;20:689–699.