THE LANCET Oncology

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Meyers RL, Maibach R, Hiyama E, et al. Risk-stratified staging in paediatric hepatoblastoma: a unified analysis from the Children's Hepatic tumors International Collaboration. *Lancet Oncol* 2016; published online Nov 21. http://dx.doi.org/10.1016/S1470-2045(16)30598-8.

Web Extra Materials

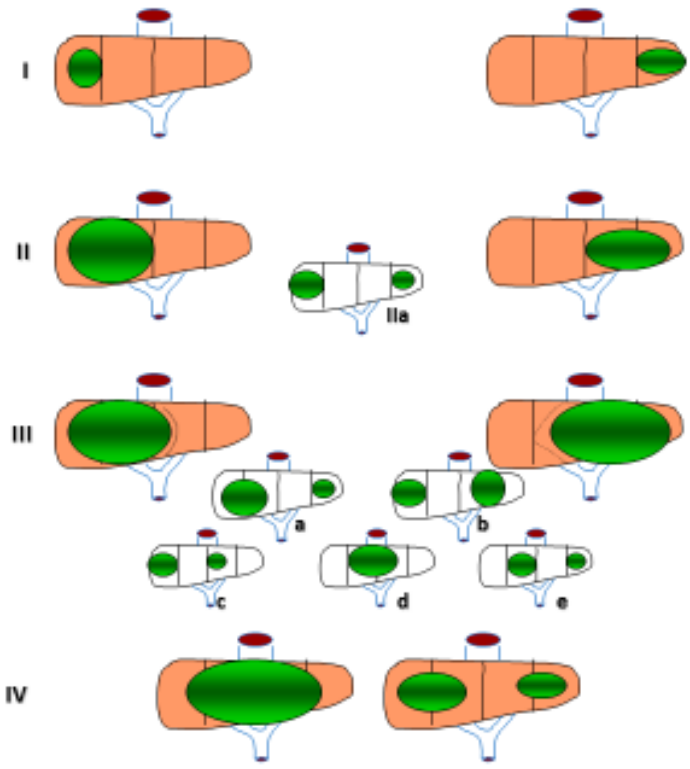
Manuscript reference number: THELANCETONCOLOGY-D-16-00970. Meyers RL, et al.

Risk stratified staging in paediatric hepatoblastoma: unified analysis from the Children's Hepatic tumor International Collaboration (CHIC)

Page 1

PRETEXT Group I, II, III, and IV. These four groups reflect hepatic parenchymal tumor involvement. When assessed at diagnosis they are PRETEXT. When assessed after neoadjuvant chemotherapy, but before surgical resection, they are called POST-TEXT. The group is defined by the number of contiguous whereas sections of liver that are tumor free. Left lateral section = Couinaud segments 2 and 3. Left medial section = Couinaud segment 4. Right anterior section = Couinaud segments 5 and 8. Right posterior section = Couinaud segments 6 and 7. Involvement of couinaud segment 1, the caudate lobe, is designated as an annotation factor "C".

PRETEXT Annotation Factors V, P, E, F, R, C, N, M denote extra-parenchymal tumor. The figure below details the radiographic definitions of these variables. Definitions of the annotation factors continue to evolve over time. For a detailed current assessment of these factors see Meyers et al, Reference 3.



PRETEXT

Pretreatment ***Ext***ent of Disease
Extent of parenchyma involvement at diagnosis

POST-TEXT

Postreatment ***Ext***ent of Disease,
extent of parenchyma involvement after pre-operative chemotherapy

I ... 3 contiguous sections tumor free
II ... 2 contiguous sections tumor free
III ... 1 contiguous sections tumor free
IV ...no contiguous sections tumor free

In addition, any group may have one or more

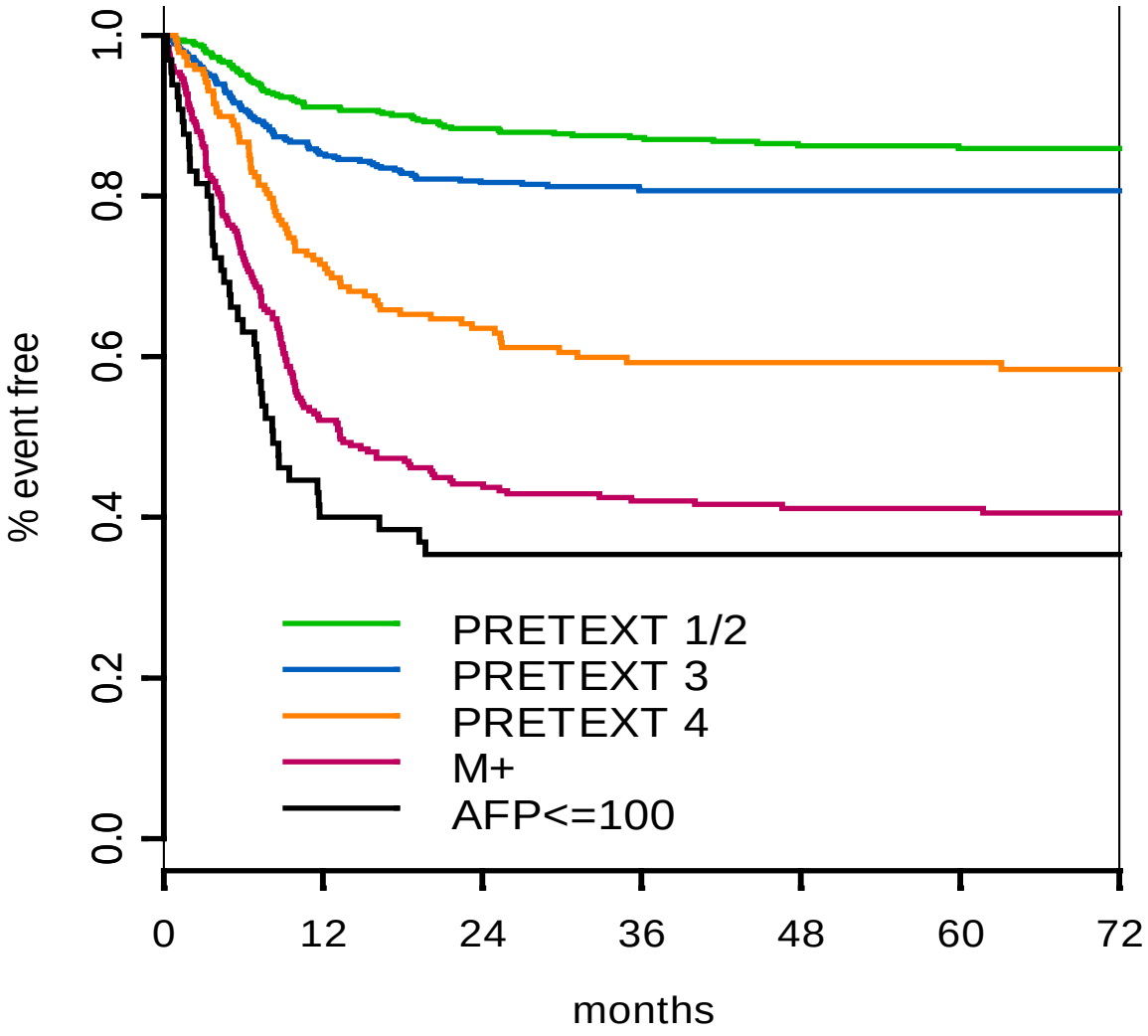
PRETEXT Annotation Factors:

V ...involvement vena cava, all 3 hepatic veins
P ...involvement portal bifurcation, both R and L
E ...contiguous extrahepatic tumor
F ....mutifocal tumor
R ... tumor rupture prior to diagnosis
C ...caudate lobe
N ... lymph node involvement
M ...metastasis, distant extrahepatic tumor

AFP in infants less than two years of age.

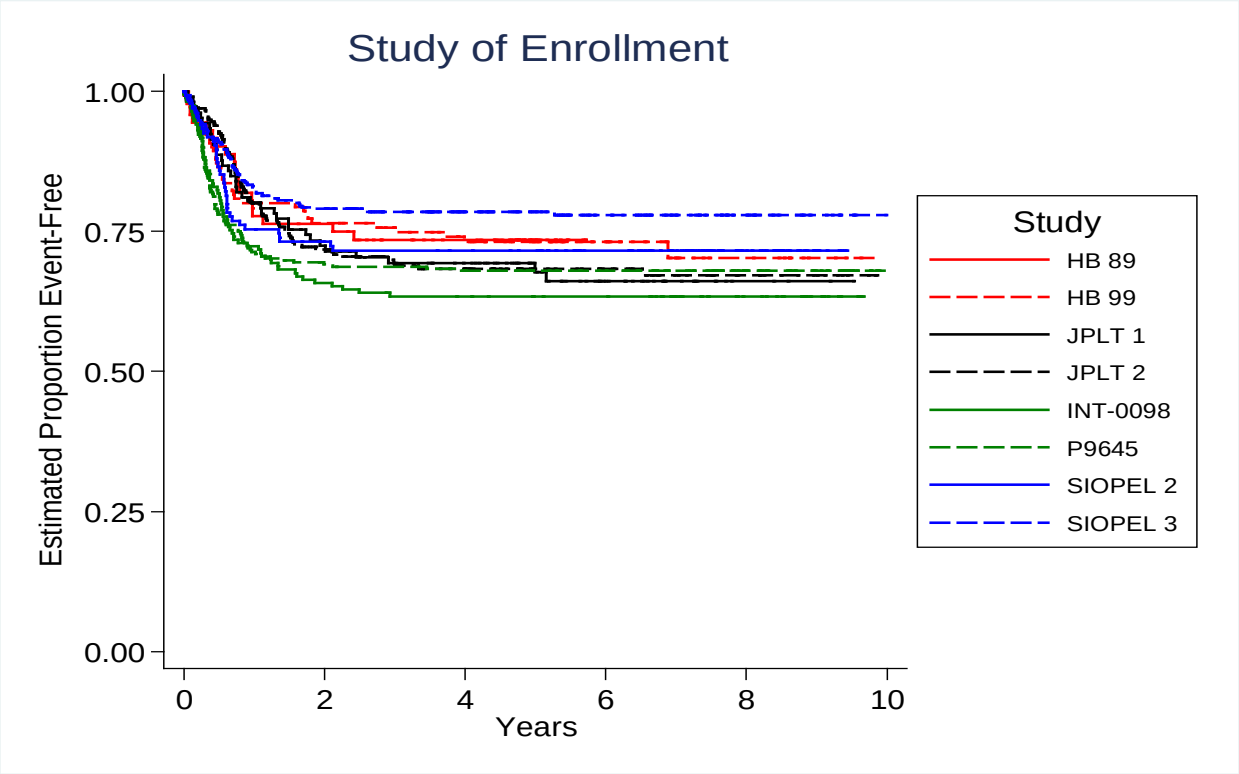
Blohm et al (reference 22) establish reference values and factors associated with serum AFP elevation in infants. Five hundred twenty-four samples were collected from infants up to the age of 2 years at the University Hospital Düsseldorf (Germany). At birth mean serum AFP levels were 41,687 ng/ml in 256 term babies and 158,125 ng/ml in 90 premature babies born before the 37th gestational week, excluding samples from children with factors known to be associated with AFP elevation. In the first 4 weeks of life, AFP levels decreased by 50% in 5.1 days in term babies. Between day 180 and 720 of life, AFP levels up to 87 ng/ml were within the 95.5% interval (assumed logarithmic normal distribution) with a mean of 8 ng/ml without a further decline. By the age of 2 years the infants of this study had not reached adult serum AFP levels (0-6 ng/ml).

Kaplan Meier. Event Free Survival (EFS) for the five backbone groups.



Univariable results comparing outcome by multicenter trial.
For each of the eight multicenter trials included in the CHIC database, see table below for number of patients, and comparative Hazard Ratio's and P values. These comparisons use HB 89 as the reference study. Kaplan Meier curves for the eight studies immediately following the table.

	HB89	HB 99	JPLT 1	JPLT 2	INT 0098	P9645	SIOPEL 2	SIOPEL 3
#patients	72	141	106	298	169	187	135	406
Hazard Ratio	-	1.02	1.22	1.15	1.46	1.05	1.09	0.81
P Value	-	0.918	0.489	0.572	0.152	0.867	0.739	0.396



Page 5. Univariable Results CHIC database analysis, see Czauderna P et al[10]. Univariate analysis has been previously published and are shown in the table below. Because of its importance in the new stratification schema, the univariate analysis results for the PRETEXT annotation factors V, P, E, F, and R is also shown in the main manuscript in Table 2.

Patient Feature	Characteristic	Value	Percent of Total (Percent of Non-Missing Values)	Relative Risk	P-Value
Age At Start Treatment (Years)	Mean	2.1			
	Median	1.4			
	Range	0-15.5			
	≤ 2 Year	1318	82	1.0	< 0.0001
	3-7	281	14	1.6	
	≥8	69	4.3	3.3	
Patient Sex	Female	639	40	1.0	0.41
	Male	966	60	1.1	
Prematurity	No	818	79	1.0	0.46
	Yes	217	21	1.1	
	Missing	570		-	
Birth Weight	1500 g or more	719	97	1.0	0.83
	Less than 1500 g	25	3.4	1.1	
	Missing	861		-	
Beckwith-Wiedemann	No	999	97	1.0	
	Yes	29	2.8	0.97	0.93
	Missing	577			
Metastatic Disease at Diagnosis (M)	No	1320	83	1.0	
	Yes	277	17	3.8	< 0.0001
	Missing	8			
AFP (ng/nl)	1000-1.2x10^6	998	78	1.0	< 0.0001
	>1.2x10^6	113	8.8	1.5	
	100-999	110	8.6	1.7	
	-99	65	5.1	4.3	
	Missing	319			
PRETEXT Group	I	97	6	1.0	< 0.0001
	II	529	34	1.4	
	III	621	40	2.4	
	IV	297	19	4.8	
	Missing	61		-	
Multifocal Primary Tumor (F)	No	1295	18	1.0	< 0.0001
	Yes	280	82	2.3	
	Missing	30			
Tumor Rupture At Enrollment (R)	No	1440	95	1.0	< 0.0001
	Yes	69	5	2.1	
	Missing	96			
Hepatic Venous Involvement (V)	No	1386	90	1.0	< 0.0001
	Yes	147	9.5	2.2	
	Missing	72			
Portal Venous Involvement (P)	No	1387	90	1.0	<0.001
	Yes	146	10	2.3	
	Missing	72			
Extrahepatic Tumor Extension (E)	No	1529	96	1.0	0.0013
	Yes	71	4	1.9	
	Missing	5			

Detailed explanation of the process used to create risk categories within each backbone.

a) **Backbone PRETEXT I/II.** In Backbone PRETEXT I/II only age proved to be of additional statistical significance for risk classification. The bootstrap results show that the age group 1-2 years differs only in 6.1% of replications from age group <1 year. We therefore decided to merge these two groups into just one (age $\leq$ 3) for risk classification. For age 3 years and above, there is a clear gradient in the hazard ratios (see table 3), but the two groups are rather small. Age 0-<3 (n=365) show an EFS of 91%. Age 3-7 (56 patients) and age $\geq$ 8 (19 patients) showed 5-year EFS rates of 72% and 40% respectively (see TABLE 5). We decided to keep age 3-7 with age <3, retaining just one age cut-off of $\geq$ 8 years, and therefore achieving a uniform age cut throughout PRETEXT I/II/III groups.

b) **Backbone PRETEXT III.** Within backbone PRETEXT III two factors proved to be significantly prognostic, 101<AFP$\leq$1000 and VPEFR (see tables 3 and 4). We did not formally test the interaction between the two factors, but we formed 3 groups by cross-classification, see Table 5, which shows that AFP has a stronger impact on prognosis than VPEFR, and also shows a strong effect of VPEFR within the AFP>1000 group. Age did not show a significant impact and there were only 4 pts aged 8 or older and therefore this factor could not be tested; we kept the factor age cut of $\geq$ 8 in the classification tree to be consistent with the PRETEXT I/II/III groups.

c) **Backbone PRETEXT IV**. Within backbone PRETEXT IV two factors proved to be significantly prognostic, age and VPEFR (see TABLE 3). In this cohort of 161 patients, the cross classification of 4 age categories with 2 VPEFR categories would yield 8 subgroups, some of very small size and therefore not amenable to a solid statistical evaluation. Age groups 0 to <1 and 1 to 2 yielded similar outcomes (data not shown) and therefore we merged the two groups. This is supported by the only 32.0% significant associations of age 1 to 2 with EFS in TABLE 4. Age 3-7 (20 patients) and age $\geq$ 8 (14 patients) both show a similarly elevated risk for an EFS event (see TABLE 5); their relevance is supported by the 65.5% and 76.8% significant associations in TABLE 4. Therefore we decided that in this backbone, the cut-off for risk classification by age should be at 3 years. Furthermore, the distinction between VPEFR present *vs* absent is important as shown by the results in both TABLE 3 and 4, but not practicable any more for age 3-7 or 8 and higher because of reduced precision. This effect of small patient numbers is illustrated in TABLE 5 by the very wide 95% confidence intervals around the 5 year EFS rates of 40% (CI: 19-61%) and 31% (10-65%). We therefore decided to use the VPEFR categories only in the risk classification for patients below 3 years where they distinguish between a low risk and an intermediate risk group, whereas patients with age>3 all fall into the high risk group.

d) **Backbone M + (metastatic).** The outcome of metastatic patients with an AFP >100 is much worse. The presence of metastases dominates other risk factors which accordingly contribute only marginally to the prognosis. One remarkable exception is an AFP between 100 and 1000, which still confers a significantly worse prognosis, as shown in tables 3 and 4. Since the presence of metastases alone prognosticates a 5-year EFS of less than 50% (table 5), this factor is used in the classification trees to assign patients to the high risk stratum, regardless of any other factor.

e) **Backbone AFP $\leq$100.** Within backbone AFP $\leq$100 no factors added additional prognostic significance. Therefore an AFP$\leq$100 means that the patient is classified as high risk, with one exception: in PRETEXT I patients the tumor is resectable, and in the absence of any other risk factor, the patient is classified as very low risk.