



Published in final edited form as:

Lancet Oncol. 2017 January ; 18(1): 122–131. doi:10.1016/S1470-2045(16)30598-8.

Risk-stratified staging in paediatric hepatoblastoma: a unified analysis from the Children's Hepatic tumors International Collaboration

Prof Rebecka L Meyers, MD*,

University of Utah School of Medicine, Salt Lake City, UT, USA

Rudolf Maibach, PhD*,

International Breast Cancer, Study Group Coordinating, Center, Bern, Switzerland

Prof Eiso Hiyama, MD*,

Hiroshima University, Hiroshima, Japan

Beate Häberle, MD*,

University of Munich, Munich, Germany

Mark Krailo, PhD*,

Children's Oncology Group, Monrovia, CA, USA

Prof Arun Rangaswami, MD*,

Stanford University, Palo Alto, CA, USA

Daniel C Aronson, MD,

Department of Paediatric Surgery, Noah's Ark Childrens' Hospital for Wales, University Hospital of Wales, Cardiff, UK

Prof Marcio H Malogolowkin, MD,

University of California Davis, Davis, CA, USA

Prof Giorgio Perilongo, MD,

University Hospital of Padua, Padua, Italy

Prof Dietrich von Schweinitz, MD,

Correspondence to: Prof Rebecka L Meyers, University of Utah School of Medicine, Salt Lake City, UT 84103, USA, rebecka.meyers@imail2.org.

*These authors contributed equally

Contributors

RLM, RM, EH, BH, MK, DCA, MHM, GP, DvS, and PC conceived the project. RLM, RM, EH, BH, MK, AR, DCA, MHM, GP, DvS, MA, DL-T, YT, RA, IL, TH, IS, KW, KY, ER, DS, MD, and PC participated in steering committee calls for the Children's Hepatic tumors International Collaboration (CHIC). RLM, RM, BH, MA, DL-T, and PC attended and organised CHIC steering committee retreats. EH, MK, AR, DCA, MHM, and GP attended CHIC steering committee retreats. YF, ER, DS, and MD managed the data. RLM, RM, EH, BH, MK, AR, DCA, YF, ER, DS, MD, and PC analysed the data. RLM, RM, EH, BH, MK, AR, DCA, and PC wrote the manuscript. RLM, RM, EH, BH, MK, AR, DCA, MHM, GP, DvS, MA, DL-T, YT, RA, IL, TH, IS, KW, KY, and PC edited and approved the final report.

Declaration of interests

We declare no competing interests.

For the **protocol** see <http://www.siopep.org>

See **Online** for appendix

University of Munich, Munich, Germany

Prof Marc Ansari, MD,
Geneva University Hospital, Geneva, Switzerland

Prof Dolores Lopez-Terrada, MD,
Baylor College of Medicine, Houston, TX, USA

Yukichi Tanaka, MD,
Kanagawa Children's Medical Center, Yokohama, Japan

Prof Rita Alaggio, MD,
University Hospital of Padua, Padua, Italy

Prof Ivo Leuschner, MD,
University of Kiel, Kiel, Germany

Tomoro Hishiki, MD,
Department of Pediatric Surgery, Chiba University, Hospital, Chiba, Japan

Irene Schmid, MD,
University of Munich, Munich, Germany

Kenichiro Watanabe, MD,
Shizuoka Children's Hospital, Shizuoka, Japan

Kenichi Yoshimura, PhD,
Innovative Clinical Research Center (iCREK), Kanazawa University Hospital, Kanazawa, Japan

Yurong Feng, MSe,
Children's Oncology Group, Monrovia, CA, USA

Eugenia Rinaldi,
CINECA, Bologna, Italy

Davide Saraceno,
CINECA, Bologna, Italy

Marisa Derosa, and
CINECA, Bologna, Italy

Prof Piotr Czauderna, MD*
Medical University of Gdansk, Gdansk, Poland

Summary

Background—Comparative assessment of treatment results in paediatric hepatoblastoma trials has been hampered by small patient numbers and the use of multiple disparate staging systems by the four major trial groups. To address this challenge, we formed a global coalition, the Children's Hepatic tumors International Collaboration (CHIC), with the aim of creating a common approach to staging and risk stratification in this rare cancer.

Methods—The CHIC steering committee—consisting of leadership from the four major cooperative trial groups (the International Childhood Liver Tumours Strategy Group, Children's

Oncology Group, the German Society for Paediatric Oncology and Haematology, and the Japanese Study Group for Paediatric Liver Tumours)—created a shared international database that includes comprehensive data from 1605 children treated in eight multicentre hepatoblastoma trials over 25 years. Diagnostic factors found to be most prognostic on initial analysis were PRETreatment EXTent of disease (PRETEXT) group; age younger than 3 years, 3–7 years, and 8 years or older; α fetoprotein (AFP) concentration of 100 ng/mL or lower and 101–1000 ng/mL; and the PRETEXT annotation factors metastatic disease (M), macrovascular involvement of all hepatic veins (V) or portal bifurcation (P), contiguous extrahepatic tumour (E), multifocal tumour (F), and spontaneous rupture (R). We defined five clinically relevant backbone groups on the basis of established prognostic factors: PRETEXT I/II, PRETEXT III, PRETEXT IV, metastatic disease, and AFP concentration of 100 ng/mL or lower at diagnosis. We then carried the additional factors into a hierarchical backwards elimination multivariable analysis and used the results to create a new international staging system.

Findings—Within each backbone group, we identified constellations of factors that were most predictive of outcome in that group. The robustness of candidate models was then interrogated using the bootstrapping procedure. Using the clinically established PRETEXT groups I, II, III, and IV as our stems, we created risk stratification trees based on 5 year event-free survival and clinical applicability. We defined and adopted four risk groups: very low, low, intermediate, and high.

Interpretation—We have created a unified global approach to risk stratification in children with hepatoblastoma on the basis of rigorous statistical interrogation of what is, to the best of our knowledge, the largest dataset ever assembled for this rare paediatric tumour. This achievement provides the structural framework for further collaboration and prospective international cooperative study, such as the Paediatric Hepatic International Tumour Trial (PHITT).

Funding—European Network for Cancer Research in Children and Adolescents, funded through the Framework Program 7 of the European Commission (grant number 261474); Children's Oncology Group CureSearch grant contributed by the Hepatoblastoma Foundation; Practical Research for Innovative Cancer Control and Project Promoting Clinical Trials for Development of New Drugs and Medical Devices, Japan Agency for Medical Research; and Swiss Cancer Research grant.

Introduction

Hepatoblastoma is rare with an annual incidence of 1.5 cases per million,¹ and advances in staging and risk stratification have lagged behind those now used for more common paediatric cancers. In 1990, the International Childhood Liver Tumours Strategy Group (SIOPEL) abandoned surgical exploration at diagnosis in favour of preoperative chemotherapy for all and introduced radiology-based staging called PRETEXT (PRETreatment EXTent of disease).² The Children's Oncology Group (COG) continued to advocate attempted surgical resection at diagnosis and therefore used the Evans' surgical staging based on exploratory surgery. In the 2000s, as radiographic imaging became increasingly sophisticated, hybrid systems that used varying aspects of PRETEXT, the adult TNM system, and Evans' surgical stage were introduced by the German Society for Paediatric Oncology and Haematology (GPOH) and by the Japanese Study Group for Pediatric Liver Tumors (JPLT).³ In its latest study,⁴ COG no longer advocated exploratory

surgery at diagnosis and introduced its own hybrid staging system using PRETEXT to define surgical resectability.⁴ Unsurprisingly, these disparate staging systems have made it difficult to compare results between the different groups.

Additionally, in the past decades, increasing amounts of data for many risk factors of hepatoblastoma have accumulated.^{5–8} Some factors present at diagnosis, such as PRETEXT group, tumour resectability, and metastatic disease, have become increasingly dominant in established treatment schema. Other factors—including α fetoprotein (AFP) expression, portal or hepatic venous macrovascular involvement, age, and fetal or small-cell-undifferentiated histology—have proven more subtle, achieving prognostic significance in some studies while remaining non-significant, or were not assessed, in other studies.^{5–8} Finally, a host of additional factors have been postulated to have prognostic value, but their effects have been difficult to quantify with statistical significance because of small patient numbers. These factors include spontaneous tumour rupture, tumour multifocality, and comorbidities such as Beckwith-Wiedemann syndrome, prematurity or low birthweight, congenital tumours, thrombocytosis, and macrotrabecular histology.⁹

In the past two decades, four major cooperative trial groups—SIOPEL, COG, GPOH, and JPLT—have undertaken prospective randomised trials of hepatoblastoma. In hopes of moving forward in a more cooperative manner, the Children’s Hepatic tumors International Collaboration (CHIC) was formed, with a primary objective of developing a common global approach to risk stratification in hepatoblastoma. CHIC formed a contract with CINECA, an Italian interuniversity consortium, to host and manage the dataset. Details of the formation of CHIC as a cooperative entity, including the challenges faced in the construction and cleaning of the dataset, have been discussed by Czauderna and colleagues.¹⁰ The CHIC steering committee set out in a memorandum of understanding the key variables available in each multicentre trial. We agreed on common definitions for all variables, and each trial group translated their unique datasets into the common format. Data in this common format were uploaded from eight separate trials: SIOPEL-2,¹¹ SIOPEL-3,^{12,13} COG-INT0098,¹⁴ COG-P9645,^{15,16} GPOH-HB89,^{17,18} GPOH-HB99,¹⁸ JPLT-1,¹⁹ and JPLT-2.²⁰

In a previous study,¹⁰ we applied univariable analysis to the CHIC database to identify the most important individual prognostic variables for event-free survival. In the current analysis, we aimed to use results from the prior univariable analysis to inform and construct a series of multivariable analyses to determine hierarchical levels of prognostic significance, and then assemble the factors that yielded significant prognostic associations into a series of progressively refined risk categories.

Methods

CHIC database and analytic strategy

We included data from 1605 patients treated in eight multicentre trials over 25 years in the CHIC database (figure 1; see protocol online).¹⁰ Potential prognostic factors present at diagnosis and their definitions are shown in table 1. Contemporary approaches to risk stratification and staging in hepatoblastoma are not uniform, but all study groups currently incorporate some version of the PRETEXT group, metastatic disease, and tumour AFP

expression. To make our proposed global stratification clinically appealing and applicable across all trial groups going forward, we chose to use, as a starting point, our accumulated clinical experience with these so-called common ground parameters, which we called backbone groups. These backbone groups not only define patient characteristics that are amenable to distinct and disparate approaches to treatment, but have also been repeatedly shown to identify groups that are homogeneous with respect to risk.⁶⁻¹⁰ After the collaborative creation of, to the best of our knowledge, the largest dataset ever compiled for this rare tumour, these backbone groups were first validated by univariate analysis.¹⁰ Validation of the backbone groups in the combined CHIC dataset is shown in the appendix (p 3), and initial univariable analysis of all potential risk factors present at diagnosis is shown in the appendix (p 5).

Inclusion and exclusion criteria

PRETEXT group (I, II, III, and IV) reflects hepatic parenchymal tumour involvement, whereas the PRETEXT annotation factors V (hepatic veins), P (portal veins), E (contiguous extrahepatic), F (multifocal), R (rupture), C (caudate), N (nodes), and M (metastatic) denote extraparenchymal tumour characteristics and have evolved over time (appendix p 1).^{4,21,22} Data for the final annotation factor (M) were collected rigorously and uniformly by all trials; therefore, in this risk analysis, we singled out metastatic disease as a separate stratification backbone. Of the remaining annotation factors, data for V, P, E, F, and R were collected by all studies with varied nomenclature and were included in this analysis. Data for C and N were not collected by all studies and were excluded. The nomenclature used to collect the V, P, E, F, and R variables was not identical across all study case report forms. For SIOPEL-2, SIOPEL-3, HB-99, and JPLT-2, the nomenclature used is defined by Roebuck and colleagues,²² but these variables were collected in different formats in JPLT-1, HB-89, INT-0098, and P9645. To unify the data, data from original case report forms were abstracted, centrally reviewed, cleaned, and re-coded according to the CHIC definitions to be then entered into the CHIC database. Where these annotation factors were not available, this variable was listed as missing and the patient excluded. In P9645, AFP concentration at diagnosis was reported without date of acquisition, which made it impossible to determine whether the value accurately reflected serum AFP concentration before chemotherapy. Since this biomarker is known to confer indispensable prognostic information,^{7,8} we chose not to enter data from P9645 into any analysis that concomitantly required AFP concentration at diagnosis. We assessed initial patient characteristics (age, sex, and stage) for participants of P9645 (n=277) and found no differences with respect to the other clinical trials in the dataset. Furthermore, overall survival and event-free survival were also highly congruent to outcomes of other included trials. 22 additional patients from other trials were excluded because of missing information on AFP. Another 43 patients were excluded for other missing variables, leaving 1263 patients for multivariable analyses.

Definition of AFP expression

AFP concentration of 100 ng/mL or less has previously been described as a risk factor in several published assessments of hepatoblastoma risk, and we considered it a surrogate marker for a biologically distinct subset of tumours that do not secrete AFP.^{2,5-9} However, in infants younger than 6 months, the range of so-called normal AFP values is wider and does

not begin to coalesce with adult normal values until age 6–12 months. Blohm and colleagues²³ detailed this decline by giving a range, with 95% CIs, of normal AFP values in the first 2 years of life (appendix p 2). To decide whether to use 100 ng/mL or less as a cutoff for normal AFP, or the upper 95.5% CI limit of normal defined by age in Blohm and colleagues' study,²³ we analysed data from 12 infants with AFP above 100 ng/mL but still within the normal range for their age. AFP was near the upper 95.5% CI limit in most of these patients and the number of patients in this category was small; hence, we decided to use 100 ng/mL as the cutoff for further analysis. We introduced a new category, AFP 101–1000 ng/mL, which might represent infant patients whose tumours do not express AFP, but whose serum AFP concentration remains above 100 ng/mL.

Backbone groups

The backbone factors—AFP 100 ng/mL or less, PRETEXT group (I, II, III, or IV), and the presence of metastases (yes or no)—have long been confirmed as being highly prognostic in hepatoblastoma,^{5–8} and so we called these traditional factors. The main focus of this analysis was to assess whether additional factors could add a substantial amount of information beyond these traditional factors. Therefore, we decided to analyse additional factors within groups defined by the traditional factors. Accordingly, a backbone classification was introduced, and five distinct backbone groups were defined (panel).

Patients with AFP concentration of 100 ng/mL or less at diagnosis were assigned to backbone 5; among the remaining patients, those with metastases at diagnosis were assigned to backbone 4. The remaining patients were assigned according to PRETEXT groups (I/II, III, or IV). Previous univariable analysis¹⁰ showed equivalent outcome for PRETEXT I and II; therefore, they were merged into a single backbone. The five backbones were validated in this dataset by initial univariable analysis and confirmed to differentiate the patients into groups that are distinctly different with respect to risk of event (appendix p 3).¹⁰ The log-rank *p* value for the comparison of all five groups was less than 0.0001.

Subset of PRETEXT annotation factors as a statistical variable

Before we assessed the interplay among the prognostic factors, the effect of the PRETEXT annotation factors V, P, E, F, and R was examined in the whole cohort. Each of these annotation factors taken by itself was significantly associated with event-free survival when analysed in the whole sample (table 2). However, addition of all five factors individually would have introduced too much variability into statistical models. Therefore, we decided that carrying them forward into the multivariable analysis as an aggregate factor would be most appropriate. This aggregate PRETEXT annotation factor was defined as positive (VPEFR+) if at least one of these five factors was present.

Multivariable backbone analysis

Within each backbone group, we put into our multi-variable analysis the factors that had been most predictive of outcome by univariable analysis.¹⁰ These factors were AFP 101–1000 ng/mL and greater than 1 000 000 ng/mL; age 0 to less than 1 year, 1–2 years, 3–7 years, and 8 years or above; and the presence or absence of a subgroup of the PRETEXT annotation factors V, P, E, F, and R. Sex, low birthweight, prematurity, and Beckwith-

Wiedemann syndrome were not prognostic in the univariable analysis, and were therefore not included in the multivariable analysis.

Statistical analysis

We expected that the potential contribution of additional factors to risk stratification would vary within each backbone group. Therefore, we sought to identify the combination of additional factors associated with either better or worse outcome compared with the mean outcome within each backbone. Event-free survival was defined as time of enrolment until first relapse, progression, second malignancy, or death for any reason. The correlation of event-free survival with potential factors was explored by multivariable Cox proportional hazards regression stratified by trial. Hazard ratios (HRs) above 1 indicate a higher risk for an event. Backward elimination was used to reach a parsimonious statistical model by stepwise elimination of factors from the model using a p value of 0.05 as a criterion of exclusion. We used the magnitude of the HR to guide the assignment to subgroups. Since we had only binary variables (0 and 1) in the models, the HR was appropriately interpretable. We decided that a higher HR should take precedence over a lower HR when guiding the assignment of factors to subgroups—ie, the factor with the higher HR took precedence. There were other approaches that we could have adopted, but we chose this method because it uses the concept of clinical utility for prognostic classification. Of note, the observed z scores led to the same conclusion as the HRs (data not shown but available upon request). Observations with missing values in one or more of the considered explanatory variables were excluded from the regression.

By choosing to proceed separately in the five clinically applicable backbone groups, we recognised the consequence that these groups were smaller in size than the aggregate dataset and did not lend themselves to further division into a training set and a validation set. In this situation, we felt that a more classical approach to validation (eg, a so-called leave k out of n method) could lead to overfitting of the model, which we wanted to avoid. Therefore, we interrogated the robustness of the model using the bootstrapping procedure.^{24–26} Each backbone group was sampled with replacement; 1000 samples of the backbone group size were taken and submitted to Cox regression with all explanatory factors (full model, stratified by trial). We determined the proportion of all 1000 replications in which a risk factor was significant at $p=0.05$. This percentage was complementary to the factor's prognostic importance. If one or more observations exercise a major leverage on the results of the backward elimination in the Cox regression, the resampling procedure tends to mitigate this effect and yield a more balanced result. The proportion of times each of the terms were identified for inclusion in the final model was tabulated to identify factors that were not selected in the entire data set but that might hold information regarding prognosis.²⁴

Final stratification

On the basis of results from the backward elimination model and the bootstrapping percentages, the next step was to form subgroups with distinct prognosis and clinical relevance. We took into account not only statistical significance but also the need to guide treatment in a clinically feasible way, the potential ease of application by clinicians of

disparate backgrounds, and the need to create treatment groups of a size that is amenable to study in clinical trials.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. Raw data were accessible only to the trial group statisticians (RM, EH, BH, MK, KW, and YF) and the database managers (ER, DS, and MD). RLM, RM, EH, BH, MK, AR, DCA, MHM, GP, and PC had final responsibility for the decision to submit for publication.

Results

The number and proportion of patients who contributed to the CHIC database (n=1605) from each collaborative group were 541 (34%) from SIOPEL, 447 (28%) from COG, 404 (25%) from JPLT, and 213 (13%) from GPOH (figure 1). Although survival tends to improve over time, all of these trials were based on some modification of cisplatin chemotherapy and complete surgical resection, a strategy introduced in the 1990s. Differences in outcome did not preclude aggregation of the dataset.⁷ Trial-specific HRs, p values, and Kaplan-Meier curves are shown in the appendix (p 4).

The initial step of backward elimination yielded different models within each backbone. For each backbone, a hierarchy of HRs was variably based on patient age, VPEFR positivity, and AFP concentration (table 3). The bootstrapping analysis (table 4) confirmed or muted the results qualitatively. As qualitative observations, the bootstrapping percentages had to be interpreted in conjunction with the prognostic effect of significant risk factors (table 3) and clinical inference. The use of observed 5 year event-free survival and varying bootstrapping percentages to declare a certain factor as essential for the risk classification was clearly a subjective collaborative decision, as shown in the next step, which consisted of forming relevant prognostic subgroups within each backbone. This step was straightforward for both PRETEXT I/II (backbone 1) and metastatic disease (backbone 4), in which only one factor was of hierarchical statistical significance. For PRETEXT III (backbone 2) and PRETEXT IV (backbone 3), we decided to use the magnitude of the HR to guide the assignment of factors to subgroups: the factor with the higher HR took precedence. For PRETEXT III, the HRs were of the same order of magnitude; therefore, we formed all possible subgroups from the two factors. The PRETEXT IV backbone group had only 161 patients, and the subgroup of age older than 3 years (n=34) could not be further subdivided by VPEFR. The subgroups were then rearranged from best to worst by 5 year event-free survival (table 5). On the basis of results and judgments that led to these subgroups, we labelled event-free survival greater than or equal to 89% as very low risk and low risk, 50–88% as intermediate risk, and less than 50% as high risk. The distinction between very low risk and low risk was defined by the treatment event that defines the potential for surgical resection at diagnosis for PRETEXT I and II tumours. A detailed step-by-step analysis of the stepwise process used to identify subgroups within each backbone is shown in the appendix (p 6).

Finally, after multiple iterations of expressing the model in grids and tables, the classification trees shown in figure 2 were deemed the best balance between clinical utility

and risk assignment precision. PRETEXT I/II was split into separate trees to make clear the clinically driven decision to add patients with low AFP concentration and small resectable tumours (ie, PRETEXT I and AFP < 100 ng/mL) to the low-risk group. Similarly, two of the backbone groups (AFP < 100 ng/mL and metastatic disease [M+]) were incorporated into each of the PRETEXT group trees. Therefore, although the statistical analysis was done within the five backbone groups, for clinical clarity the results are shown as stratification trees structured to conform to the four clinically familiar PRETEXT groups.

Discussion

In this analysis, we started with a platform of clinically relevant historic data, informed the platforms with a multivariable analysis of significant risk factors, identified the constellations of variables that were most predictive of outcome, and finally used these constellations of factors to build a new risk-stratified staging system that would be used by all. We propose to call this new staging system the Children's Hepatic tumors International Collaboration—Hepatoblastoma Stratification (CHIC-HS). We believe that the new risk stratification system is crucial for two reasons. First, it will allow comparison of treatment results of past and future multicentre studies. Second, and most importantly, it builds the foundation for a new global collaborative trial in paediatric hepatoblastoma. Our initial univariable analysis confirmed the prognostic importance of established factors such as metastatic disease, low AFP concentration (< 100 ng/mL), and PRETEXT group.¹⁰ The increased statistical power provided by this collaboration has allowed us to clarify the previously suspected, but unproven, importance of PRETEXT annotation factors (VPEFR) and age of patient. We now have a method for us to match, in future trials of hepatoblastoma, the assigned therapeutic approach to the estimated risk of relapse. Moreover, these results allow us to interpret data presented by other investigators, making them comparable by accounting for prognostic factors.

Although evidence shows that prognosis of tumours with well differentiated fetal histology is good and that prognosis of small-cell undifferentiated tumours might be inferior, the prognostic significance of histological subtypes has not been uniform across all study groups.^{2,5,7,8,20,27} This non-uniformity might be due to disparate histological definitions and reporting categories. Therefore, we are undertaking a separate analysis that includes a comprehensive retrospective review of all available histological material, with histological subtype reassignment for patients in the database, using an international consensus histology classification developed in 2014.²⁸

Conversion of the statistically generated model into a tool that not only predicts outcome, but also has utility in treatment assignment in future clinical trials, was challenging in some special cases. Clearly, treatment in these situations might call for individualised clinical reflection and will need to be a focus of future prospective validation. First, in older patients with small tumours (PRETEXT I/II), our model suggests a relatively poor prognosis—ie, the older age seems to override the importance of low PRETEXT. Because many of these tumours will be surgically resectable at diagnosis, we elected to keep the 3–7 year age group in the lower-risk group. The ability to surgically resect tumours that could have limited chemosensitivity supports this approach. Second, few patients had low PRETEXT and

positive VPEFR in our analysis, with intermediate results in the bootstrapping analysis. In support of overall model uniformity, we elected to place these patients in the intermediate-risk group. Third, in patients with PRETEXT I and low AFP (< 100 ng/mL), low AFP was not as predictive of a poor outcome as for patients in backbone 5. Therefore, stratification of these patients to a high-risk treatment group with intensified chemotherapy might be unjustified where the tumour is very small and surgically resectable. Fourth, in patients with PRETEXT III tumour who were younger than 8 years and had no metastasis (M-), although only a small subset had AFP 100–1000 ng/mL, their 5 year event-free survival was much lower than the others (table 5). Those with AFP higher than 1000 ng/mL had comparatively better survival regardless of VPEFR positivity. Finally, age has a variable effect in each backbone group. In the PRETEXT I/II and PRETEXT IV backbone groups, age 3–7 years and 8 years or older were predictors of worse prognosis. Curiously, this effect was not seen in PRETEXT III. Therefore, we are left with a clinical challenge of having disparate effects of age across different PRETEXT and backbone groups, rendering it difficult to make a universal treatment recommendation based on age that applies across all groups (appendix p 6). We made the decision to simplify these age groups into two categories (< 8 years or ≥ 8 years) for all backbone groups, except for PRETEXT IV in which prognosis for patients aged 3–7 years was significantly worse and equivalent to that seen in those aged 8 years or older. Future age cut-offs might be further adjusted on the basis of trial design.

This analysis was bridled by the retrospective nature of the data analysis. To include such a large number of patients with a rare tumour in the database, data by necessity were abstracted from clinical trials done over two decades. Each trial had unique objectives, overall treatment outcomes have improved, and surgical strategies have evolved. The trial outcomes were sufficiently homogenous to allow aggregation as a dataset (appendix p 4). As in all of modern medicine, supportive care, diagnostic imaging, and surgical technique have refined across treatment eras, which have resulted in improved outcomes over time, especially for some of the lower-risk subgroups. Importantly, however, two cornerstones of therapy—cisplatin and attempted complete surgical resection—were at the core of all of these trials, and this has not changed in present trials. Although all of the trials have prescribed treatment with variations on this theme, the CHIC analysis was not designed to capture information on alternative therapeutic strategies (eg, preoperative *vs* postoperative chemotherapy, or alternative combinations of cisplatin chemotherapy). In the end, it became a question of feasibility because of the rarity of this tumour and the small numbers of patients receiving individual therapeutic regimens. Nevertheless, we are confident that the constellation of relative risk within these evolving strategies remains.

The relevance of this analysis to present patients derives from the continued use of treatment based on cisplatin and complete surgical resection, the increasing incidence of this rare childhood malignant disease, and the imminent initiation of the Paediatric Hepatic International Tumour Trial (PHITT), the first international trial of childhood hepatic malignancies in which a large cohort of patients will be treated in a uniform manner. Although outcomes in some cohorts of patients have improved, other groups (eg, patients with very large tumours, older patients, and those with metastatic disease or low AFP concentrations at diagnosis) continue to have a relatively poor prognosis. Substantial room for improvement in outcomes for these patients remains. This model represents the first

iteration of an approach to risk stratification that will undoubtedly be refined by advances in our understanding of biological variables and in a uniform approach across international cooperative groups to diagnostic staging and treatment response assessment.

Finally, for some combinations of the investigated factors the sample sizes were small, resulting in large 95% CIs around the estimated HRs. Additional ongoing efforts to further refine risk stratification in hepatoblastoma importantly include CHIC histological subtype review, which is currently underway and aimed at prognostic assessment of histology; validation of the proposed model with contemporary datasets comprised of patients treated in SIOPEL-4, SIOPEL-6, AHEP-0731, and JPLT-3 once data from these latest trials become available; prospective validation built into PHITT; and expansion of the database and analysis to interrogate the potential importance of treatment-defined factors, such as quality of response to chemotherapy and timing or type of surgical resection.

CHIC-HS is the most refined system so far for risk stratification in paediatric hepatoblastoma. As a unified international approach, it allows, for the first time, the comparison of emerging study results from different trial groups by defining a common risk stratification. Ongoing efforts will incorporate histological subtypes and data from our current studies. This international cooperative effort has been a crucial element in the planning for PHITT, which will prospectively validate this proposed model and also make possible the parallel collection and interrogation of biological material and molecular markers. Thus, CHIC-HS is an important next step towards an individualised treatment approach for hepatoblastoma.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by the European Network for Cancer Research in Children and Adolescents, funded through the Framework Program 7 of the European Commission (grant number 261474); Children's Oncology Group CureSearch grant contributed by the Hepatoblastoma Foundation; Practical Research for Innovative Cancer Control and Project Promoting Clinical Trials for Development of New Drugs and Medical Devices, Japan Agency for Medical Research; and the Swiss Cancer Research grant. This has truly been an international collaborative effort.

References

1. Spector LG, Birch J. The epidemiology of hepatoblastoma. *Pediatr Blood Cancer*. 2012; 59:776–79. [PubMed: 22692949]
2. Czauderna P, Lopez-Terrada D, Hiyama E, et al. Hepatoblastoma state of the art: pathology, genetics, risk stratification, and chemotherapy. *Curr Opin Pediatr*. 2014; 26:19–28. [PubMed: 24322718]
3. Perilongo G, Malogolowkin M, Feusner J. Hepatoblastoma clinical research: lessons learned and future challenges. *Pediatr Blood Cancer*. 2012; 59:818–21. [PubMed: 22678761]
4. Meyers RL, Tiao G, de ville de Goyet, et al. Hepatoblastoma state of the art: PRETEXT, surgical resection guidelines and the role of liver transplantation. *Curr Opin Pediatr*. 2014; 26:29–36. [PubMed: 24362406]

5. Fuchs J, Rydzynski J, von Schweinitz D, et al. Pretreatment prognostic factors and treatment results in children with hepatoblastoma: a report from the German Cooperative Pediatric Liver Tumor Study HB94. *Cancer*. 2002; 95:172–82. [PubMed: 12115331]
6. Aronson DC, Schnater JM, Staalman CR, et al. Predictive value of the pre-treatment extent of disease system in hepatoblastoma: results from the International Society of Pediatric Oncology Liver Tumor Study Group SIOPEL-1 study. *J Clin Oncol*. 2005; 23:1245–52. [PubMed: 15718322]
7. Meyers RL, Rowland JH, Krailo M, et al. Pretreatment prognostic factors in hepatoblastoma: a report of the Children's Oncology Group. *Pediatr Blood Cancer*. 2009; 53:1016–22. [PubMed: 19588519]
8. Maibach R, Roebuck D, Brugieres L, et al. Prognostic stratification for children with hepatoblastoma: the SIOPEL experience. *Eur J Cancer*. 2012; 48:1543–49. [PubMed: 22244829]
9. von Schweinitz D. Hepatoblastoma: recent developments in research and treatment. *Semin Pediatr Surg*. 2012; 21:21–30. [PubMed: 22248967]
10. Czauderna P, Haeberle B, Hiyama E, et al. The Children's Hepatic tumors International Collaboration (CHIC): novel global rare tumor database yields new prognostic factors in hepatoblastoma and becomes a research model. *Eur J Cancer*. 2016; 52:92–101. [PubMed: 26655560]
11. Perilongo G, Shafford E, Maibach R, et al. Risk adapted treatment for childhood hepatoblastoma: final report of the second study of the Internal Society Of Pediatric Oncology, SIOPEL 2. *Eur J Cancer*. 2004; 40:411–21. [PubMed: 14746860]
12. Perilongo G, Maibach R, Shafford E, et al. Cisplatin versus cisplatin plus doxorubicin for standard risk hepatoblastoma. *N Engl J Med*. 2009; 361:1662–70. [PubMed: 19846851]
13. Zsiros J, Maibach R, Shafford E, et al. Successful treatment of childhood high-risk hepatoblastoma with dose-intensive multiagent chemotherapy and surgery: final results of the SIOPEL-3HR study. *J Clin Oncol*. 2010; 28:2584–90. [PubMed: 20406943]
14. Ortega JA, Douglass EC, Feusner JH, et al. Randomized comparison of cisplatin/vincristin/5-fluorouracil and cisplatin/doxorubicin for the treatment of pediatric hepatoblastoma (HB): a report from the Children's Cancer Group and the Pediatric Oncology Group. *J Clin Oncol*. 2000; 18:2665–75. [PubMed: 10894865]
15. Malogolowkin MH, Katzenstein HM, Krailo M, et al. Intensified platinum therapy is an ineffective strategy for improving outcome in pediatric patients with advanced hepatoblastoma. *J Clin Oncol*. 2006; 24:2879–84. [PubMed: 16782927]
16. Katzenstein HM, Chang KW, Krailo MD, et al. Amifostine does not prevent platinum-induced hearing loss associated with treatment of children with hepatoblastoma; a report of the Intergroup Hepatoblastoma Study P9645 as part of Children's Oncology Group. *Cancer*. 2009; 115:5828–35. [PubMed: 19813275]
17. von Schweinitz D, Hecker H, Schmidt-von-Arndt G, et al. Prognostic factors and staging systems in childhood hepatoblastoma. *Int J Cancer*. 1997; 74:593–99. [PubMed: 9421354]
18. Haeberle B, von Schweinitz D. Treatment of hepatoblastoma in German cooperative pediatric liver tumor studies. *Front Biosci*. 2012; 1:493–98.
19. Sasaki F, Matsunaga T, Iwafuchi M, et al. Outcome of hepatoblastoma treatment with JPLT-1 Protocol-1: a report from the Japanese study group for pediatric liver tumor. *J Pediatr Surg*. 2002; 37:851–56. [PubMed: 12037748]
20. Hishiki T, Matsunaga T, Sasaki F, et al. Outcome of hepatoblastoma treated using the Japanese Study Group for Pediatric Liver Tumor (JPLT) protocol-2: report from the JPLT. *Pediatr Surg Int*. 2011; 27:1–8. [PubMed: 20922397]
21. Pritchard J, Plaschkes J, Shafford EA, et al. SIOPEL I: the first SIOP hepatoblastoma (hepatoblastoma) and hepatocellular carcinoma (HCC) study. Preliminary results. *Med Ped Oncol*. 1992; 20:389.
22. Roebuck DJ, Aronson D, Clapuyt P, et al. 2005 PRETEXT: a revised staging system for primary malignant liver tumours of childhood developed by the SIOPEL group. *Pediatr Radiol*. 2007; 37:123–32. [PubMed: 17186233]

23. Blohm MEG, Vesterling-Hörner D, Calaminus G, Göbel U. Alpha₁-fetoprotein (AFP) reference values in infants up to 2 years of age. *Pediatr Hematol Oncol.* 1998; 15:135–42. [PubMed: 9592840]
24. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema DF. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* 2001; 54:774–81. [PubMed: 11470385]
25. Braga-Neto UM, Dougherty ER. Is cross validation valid for small-sample microarray classification? *Bioinformatics.* 2004; 20:374–80. [PubMed: 14960464]
26. Frazier LA, Hale JP, Rodriguez-Galindo C, et al. Revised risk classification for paediatric extracranial germ cell tumours based on 25 years of clinical trial data from the United Kingdom and United States. *J Clin Oncol.* 2014; 33:195–201. [PubMed: 25452439]
27. Malogolowkin MH, Katzenstein HM, Meyers RL, et al. Complete surgical resection is curative for children with hepatoblastoma with pure fetal histology: a report from the Children’s Oncology Group. *J Clin Oncol.* 2011; 29:3301–06. [PubMed: 21768450]
28. Lopez-Terrada D, Alaggio R, DeDavila MT, et al. Towards an international paediatric liver tumour consensus classification: proceedings of the Los Angeles COG International Pathology Paediatric Liver Tumours Symposium. *Mod Pathol.* 2014; 26:19–28.

Research in context

Evidence before this study

The challenges of clinical research in paediatric rare cancers have been detailed and, importantly, include disparate staging systems. In the past decades, we have accumulated increasing data for many risk factors of hepatoblastoma, and a host of additional factors have been postulated to have prognostic value, but their effects have been difficult to quantify with statistical significance because of small numbers of patients.

Added value of this study

The Children's Hepatic tumors International Collaboration (CHIC) represents a decade-long effort in data collection, cleaning, and analysis, as well as collaboration and bridge-building. The creation of this new staging system was a true global effort, taking into account not only statistical significance in a large database, but also the need to guide treatment in a clinically feasible way, the potential ease of application by clinicians of disparate backgrounds, and the need to create treatment groups of sizes that are amenable to study in clinical trials. These new risk cohorts will now become international targets for biological interrogation, therapy intensification, and therapy reduction.

Implications of all the available evidence

This effort is an essential step in the development of a unified risk stratification that will allow international collaboration in the study of hepatoblastoma. Importantly, it creates an opportunity to launch a truly global clinical trial in this rare paediatric cancer.

Panel: Backbone groups

- Backbone 1: PRETEXT I/II, not metastatic, AFP >100 ng/mL
- Backbone 2: PRETEXT III, not metastatic, AFP >100 ng/mL
- Backbone 3: PRETEXT IV, not metastatic, AFP >100 ng/mL
- Backbone 4: Metastatic disease at diagnosis, AFP >100 ng/mL, any PRETEXT group
- Backbone 5: AFP \leq 100 ng/mL at diagnosis, metastatic or not, any PRETEXT group

PRETEXT=PRETreatment EXTent of disease. AFP= α fetoprotein.

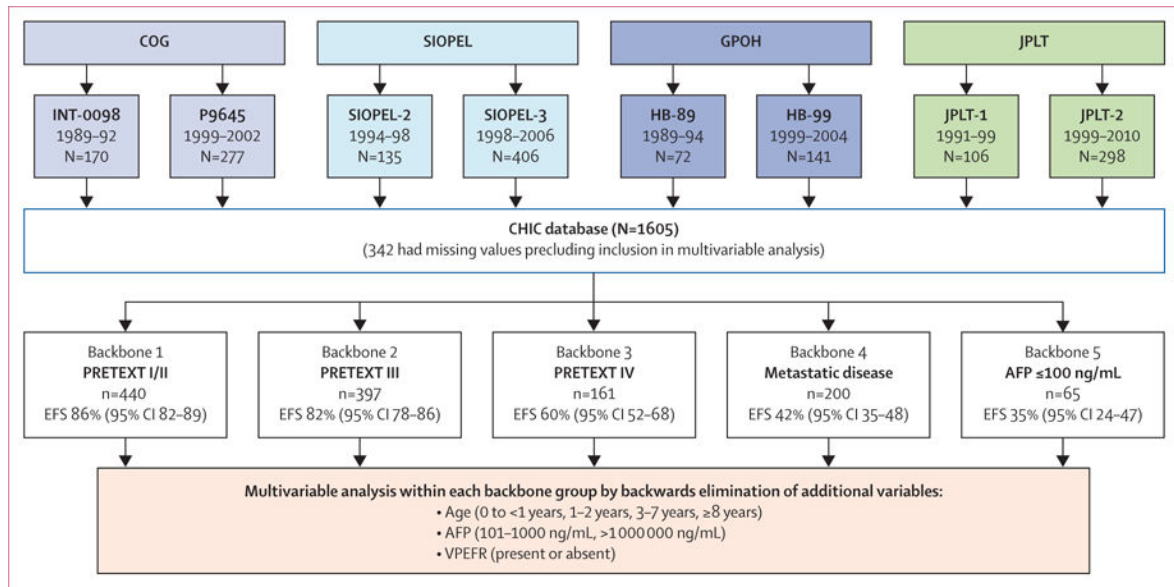


Figure 1. Construction and analysis of the CHIC database

Eight trials from four major groups were used to construct the CHIC database; dates of patient enrolment in these trials, missing values, and EFS in the five backbone groups are shown. Detailed definitions of the PRETEXT annotation factors VPEFR are shown in table 2 and appendix p 1. AFP=α fetoprotein. CHIC=Children's Hepatic tumors International Collaboration. COG=Children's Oncology Group. EFS=event-free survival. GPOH=German Society for Paediatric Oncology and Haematology. JPLT=Japanese Study Group for Pediatric Liver Tumors. PRETEXT=PRETreatment EXTent of disease. SIOPEL=International Childhood Liver Tumours Strategy Group.

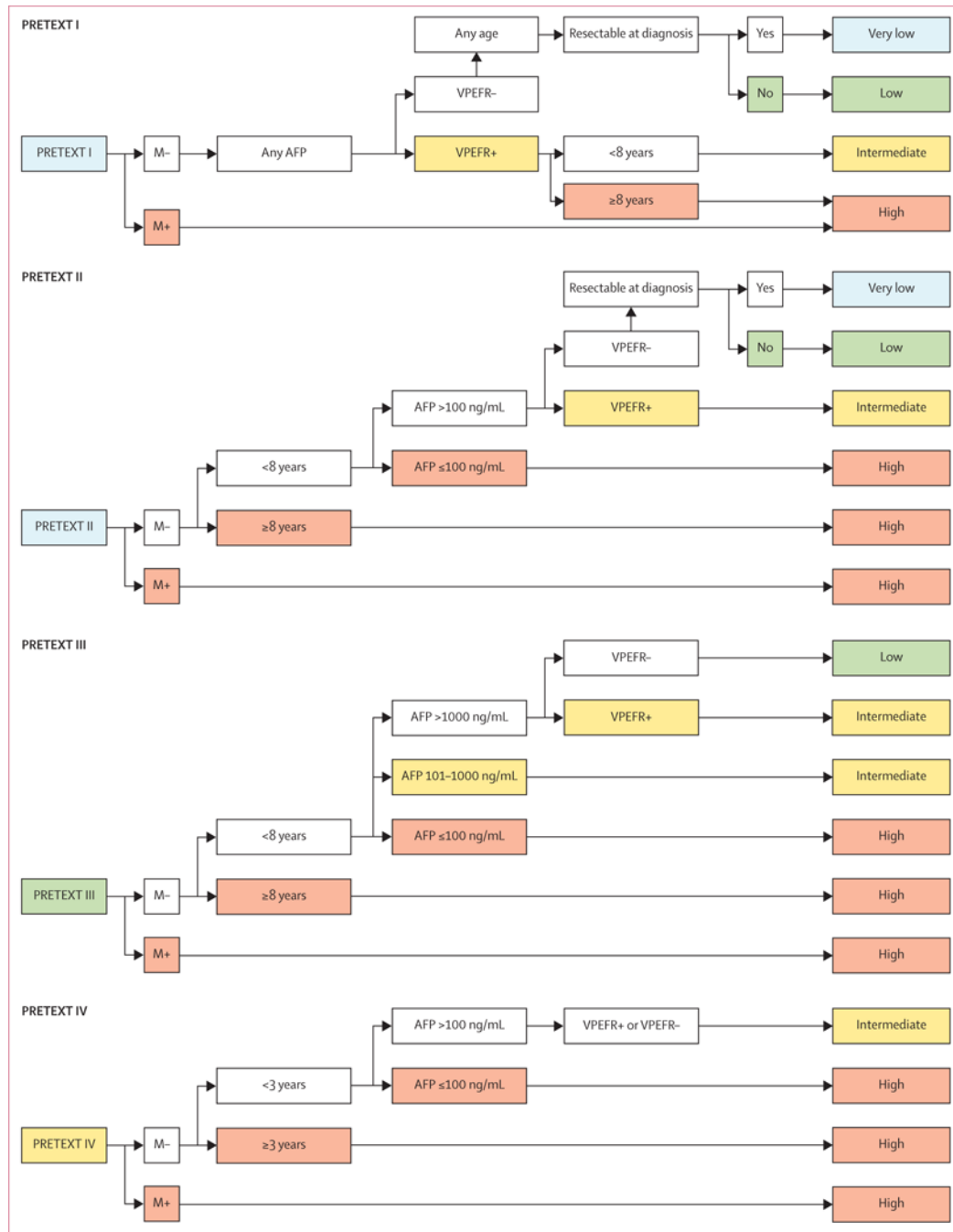


Figure 2. Risk stratification trees for the Children’s Hepatic tumors International Collaboration—Hepatoblastoma Stratification (CHIC-HS)

Very-low-risk group and low-risk group are separated only by their resectability at diagnosis, which has been defined by international consensus as part of the surgical guidelines for the upcoming collaborative trial, Paediatric Hepatic International Tumour Trial (PHITT).

Separate risk stratification trees are used for each of the four PRETEXT groups. Detailed definitions of the PRETEXT annotation factors VPEFR are shown in table 2 and appendix p 1. AFP=α fetoprotein. M=metastatic disease. PRETEXT=PRETreatment EXTent of disease.

Table 1

Prognostic factors and their classification

	Classification
PRETEXT group	I=one section involved, three sections tumour free; II=one or two sections involved, two sections tumour free; III=two or three sections involved, one section tumour free; IV=four sections involved [*]
PRETEXT annotation factors	
V: involvement of vena cava—all three hepatic veins or the intrahepatic inferior vena cava, or both	Yes vs no
P: involvement of portal vein—both left and right portal vein, or portal bifurcation, or both	Yes vs no
E: contiguous extrahepatic intra-abdominal tumour extension—contiguous involvement of adjacent organs (eg, diaphragm and bowel)	Yes vs no
F: multifocal liver tumour—two or more tumour nodules separated by normal hepatic parenchyma	Yes vs no
R: tumour rupture at diagnosis	Yes vs no
M: metastasis—non-contiguous tumour spread, usually to the lungs	Yes vs no
AFP concentration, ng/mL	100 vs 101–1000 vs 1001–1 000 000 vs 1 000 000
Age, years	0 to <1 vs 1–2 vs 3–7 vs 8

AFP=α fetoprotein. PRETEXT=PRETreatment EXTent of disease.

^{*} See appendix p 1.

Table 2

Prognostic effect of PRETEXT annotation factors

	Number of patients/ total number of assessable patients	Hazard ratio (95% CI)	p value
V: involvement of vena cava or all three hepatic veins, or both	147/1533	2.20 (1.67–2.79)	<0.0001
P: involvement of portal bifurcation or both right and left portal veins, or both	146/1533	2.26 (1.76–2.90)	<0.0001
E: extrahepatic contiguous tumour extension	71/1600	1.91 (1.33–2.73)	0.0004
F: multifocal liver tumour	280/1575	2.34 (1.90–2.88)	<0.0001
R: tumour rupture at diagnosis	69/1509	2.05 (1.43–2.95)	<0.0001
VPEFR+: one or more of V, P, E, F, or R present	533/1605	2.51 (2.08–3.02)	<0.0001
PRETEXT=PRETreatment EXTent of disease.			

PRETEXT=PRETreatment EXTent of disease.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Prognostic effect of significant risk factors, stratified by backbone group

	p value	Hazard ratio
PRETEXT I/II (n=440, 7 excluded)		
Age <3 years *		1 (reference)
Age 3–7 years	0.0005	3.08
Age 8 years	<0.0001	6.50
PRETEXT III (n=397, 20 excluded)		
AFP		
>1000 ng/mL		1 (reference)
101–1000 ng/mL	0.0008	3.29
PRETEXT annotation factors (VPEFR)		
None		1 (reference)
One or more	0.0038	2.08
PRETEXT IV (n=161, 2 excluded)		
Age		
<3 years *	..	1 (reference)
3 years †	0.0069	2.21
PRETEXT annotation factors (VPEFR)		
None	..	1 (reference)
One or more	0.0037	2.44
Metastatic disease (n=200, 14 excluded)		
AFP >1000 ng/mL	..	1 (reference)
AFP 100–1000 ng/mL	<0.0001	4.42
AFP 100 ng/mL (n=65, none excluded) ‡		
None added additional prognostic significance

PRETEXT=PRETreatment EXTent of disease. AFP=α fetoprotein.

* Age 1–3 years was not significantly different from age 0 to less than 1 years, and therefore these age groups were merged into one baseline group (age <3 years).

† Age 3–7 years and age 8 years were merged because both showed a similar risk in the model.

Table 4

Proportion of replicates that had a significant association with event-free survival in bootstrapping, by backbone group

Proportion of replicates with significant association*	
PRETEXT I/II	
Age 1–2 years	6.1%
Age 3–7 years	77.3%
Age 8 years	94.2%
AFP 101–1000 ng/mL	8.0%
AFP >1 000 000 ng/mL	4.4%
VPEFR+	29.6%
PRETEXT III	
Age 1–2 years	55.2%
Age 3–7 years	18.0%
Age 8 years	22.3%
AFP 101–1000 ng/mL	78.1%
AFP >1 000 000 ng/mL	6.2%
VPEFR+	69.6%
PRETEXT IV	
Age 1–2 years	32.0%
Age 3–7 years	65.5%
Age 8 years	76.8%
AFP 101–1000 ng/mL	21.9%
AFP >1 000 000 ng/mL	5.6%
VPEFR+	67.3%
Metastatic	
Age 1–2 years	6.6%
Age 3–7 years	16.5%
Age 8 years	41.0%
AFP 101–1000 ng/mL	98.1%
AFP >1 000 000 ng/mL	10.1%
VPEFR+	9.3%
PRETEXT III	4.1%
PRETEXT IV	28.1%
AFP <100 ng/mL	
Age 1–2 years	9.1%
Age 3–7 years	35.5%
Age 8 years	26.7%
VPEFR+	24.2%
M+	38.9%
PRETEXT III	9.6%

Proportion of replicates with significant association[*]	
PRETEXT IV	28.8%

PRETEXT=PRETreatment EXTent of disease. AFP=a fetoprotein. M=metastatic disease.

*
p=005; each backbone group underwent 1000 replications in the bootstrap procedure.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Kaplan-Meier estimates of event-free survival in risk categories defined by statistical analysis, by backbone group

	Number of patients in subgroup	Number of events*	Observed 5 year event-free survival (95% CI)
PRETEXT I/II[†]			
Age <3 years	365	33	91% (87–93)
Age 3–7 years	56	14	72% (57–83)
Age 8 years	19	11	40% (18–61)
PRETEXT III			
AFP >1000 ng/mL, negative VPEFR	260	27	89% (85–92)
AFP >1000 ng/mL, positive VPEFR	109	29	73% (64–80)
AFP 101–1000 ng/mL, positive or negative VPEFR	28	11	61% (40–76)
PRETEXT IV			
Age <3 years, negative VPEFR	51	9	84% (70–92)
Age <3 years, positive VPEFR	76	33	56% (44–67)
Age 3–7 years, positive or negative VPEFR	20	11	40% (19–61)
Age 8 years, positive or negative VPEFR	14	9	31% (10–65)
Metastatic disease[‡]			
AFP >1000 ng/mL	183	95	47% (40–55)
AFP 10–1000 ng/mL	17	14	18% (4–38)
AFP <100 ng/mL [‡]	65	42	35% (24–47)

PRETEXT=PRETreatment EXTent of disease. AFP=α fetoprotein.

* First relapse, progression, second malignancy, or death for any reason.

[†] PRETEXT I and II are merged here but not in the classification trees where PRETEXT I appears as its own tree, because all patients with PRETEXT I are amenable to resection and diagnosis; hence, AFP < 100 ng/mL is not a risk factor for chemoresistance (see Discussion).

[‡] To maximise clinical utility of the classification trees, the metastatic disease and AFP < 100 ng/mL backbone groups are included within the tree of the corresponding PRETEXT group.