

DataSchema

Core Variable	Variable Definition	Value Labeling
Study_ID	Identification Number for the participated Institution	1=MD Anderson Cancer Center 2=Mayo Clinic 3= Veteran Administration Houston 4=NCI_EDRN 5= Baylor College of Medicine (BCM) 6= Columbia University 7= St Luke's Hospital 8= Beth Israel Deaconess Medical Center 9= Massachusetts General Hospital 10= Memorial Sloan Kettering 11= Houston Methodist Hospital 12= Moffitt Cancer Center 13= University of Pittsburgh Cancer Institute 14= Toronto University 15= University of California San Francisco 16= University of Texas Southwest Dallas 17= University of Texas Medical School 18= University of California San Diego 19= University of California Los Angeles 20=HEDS
Subject_ID	Identification Number of participated subject	Study_ID*1000000+orginal ID

DataSchema

Core Variable	Variable Definition	Value Labeling
Status	Case-Control Status	1=HCC Case 2=Healthy Control 3=HCV-induced cirrhosis control
HCC Diagnosis	Method of HCC diagnosis	1=Pathological 2=Radiological 3=Clinical 4=Undetermined
Site	Study Site	USA (100s) 101=Houston, Texas 102=Dallas, Texas 103= Rochester, Minnesota 104=Los Angeles, California 105=San Francisco, California 106=San Diego, California 107=Pittsburgh, Pennsylvania 108=Manhattan , New York 109= New York, New York 110=Baltimore, Maryland 111= Boston, Massachusetts 112=Tampa, Florida CANADA (200s) 201 = Toronto, Canada

DataSchema

Core Variable	Variable Definition	Value Labeling
Sex	Gender of the participated subject	1= Male 2=Female
Age	Calculated Age (years)	For Cases: Age at Diagnosis For Healthy Controls: Age at interview For HCV controls: Age of cirrhosis evidence
Race	Race of the subject	1 = White, Anglo-Caucasian (non-Hispanic) 2 = Spanish origin (Hispanic) only 3 = White, Spanish origin 4 = Black, Spanish origin 5 = Black / African American (non-Hispanic) 6 = Asian / Pacific Islander 7 = American Indian / Alaskan Native 8 = Other
Edu	Level of Education of the participant	1 = 8 th grade of less 2 = 9 th - 11 th grade 3 = 12 th grade or high school graduate 4 = some college or college graduate 5 = one or more years of graduate school
Data_Collection	Source of data collection	1 = telephone 2 = self-administered 3 = in person 4 = other

DataSchema

Core Variable	Variable Definition	Value Labeling
Birth_date	Date of birth of the participants	Month/Day/Year (XX/XX/XXXX)
Enrollment_date	Date of participant enrollment in the study	Month/Day/Year (XX/XX/XXXX)
Case_DX	HCC method of diagnosis	1=Pathology 2=Histology
DX_date	Date of radiological or pathological diagnosis	Month/Day/Year (XX/XX/XXXX)
Wt_current	Current Weight (Kg)	kg=lbs* 0.45
Ht_current	Usual adult height (cm)	cm=inches* 2.54
Wt_past	Average Weight prior diagnosis or enrollment (Kg)	kg=lbs* 0.45
Ht_past	Average Height prior diagnosis or enrollment (cm)	cm=inches* 2.54
BMI_current	Current Body Mass Index	kg/m ² . Formula: wt/(ht/100) ²
BMI_past	Past BMI	kg/m ² . Formula: wt/(ht/100) ²
HX_obesity	History of obesity or overweight	1= Slim 2=Normal 3=Overweight 4=Obese

DataSchema

Core Variable	Variable Definition	Value Labeling
Virus_test	Ever blood testing for HCV and HBV	0=No 1= Yes (HCV) 2=Yes (HBV) 3=Yes (HCV and HBV)
HX_hepatitis	Past history of hepatitis	0=No 1=HAV 2=HBV 3=HCV 4=HBV+HCV
HCV	Anti-HCV	0=Negative 1=Positive 9=Missing
HCV_Subtype	HCV Genotypes	
HBsAg	HBsAg	0=Negative 1=Positive 9=Missing
HBC	Anti-HBc	0=Negative 1=Positive 9=Missing

DataSchema

Core Variable	Variable Definition	Value Labeling
Virus_markers	HCV and HBV markers if blood was tested	1=HCV-/HBsAg-/Anti-HBc 2=HCV-/HBsAg+/Anti-HBc+ 3=HCV-/HBsAg-/Anti-HBc+ 4=HCV+/HBsAg-/Anti-HBc 5=HCV+/HBsAg+/Anti-HBc+ 6=HCV+/HBsAg-/Anti-HBc+
Smoke	Status of Cigarette Smoking	0=Never Cigarette Smoking 1=Current Cigarette Smokers with ≤ 20 pkyrs (including ex-smoker <1 year since quit) 2=Current Cigarette Smokers with >20 pkyrs (including ex-smoker <1 year since quit)
Alcohol	Status of Alcohol Use	0=Never drink alcoholic beverage 1=Ever use with < 60 ml ethanol /day 2=Ever use with > 60 ml ethanol/day
HX_cancer	Prior History of Cancer	0=No 1=Yes 2=Nonmelanoma
FHX_cancer	First degree relative with Cancer	0=No 1=Yes
FHX_LC	First degree relative with liver cancer	0=No 1=Yes
CLD	History of chronic or genetic liver disease	0=No 1=Yes

DataSchema

Core Variable	Variable Definition	Value Labeling
Diabetes	History of diabetes	0=No 1=Yes
Type_Dib	Type of diabetes	1=Type 1 diabetes (insulin dependent) 2=Type 2 (non-insulin dependent)
Diabetes_Dur	Duration of Diabetes prior to diagnosis or enrollment	0=No Diabetes 1=< 3 years 2=3-10 years 3=>10 years
Diabetes_Date	Year of diabetes diagnosis	Continuous (years)
Hypertension	History of Hypertension	0=No 1=Yes
Wilson	History of Wilson Disease	0=No 1=Yes
AAD	History of Alpha Anti-Trypsin Deficiency	0=No 1=Yes
AIH	History of Autoimmune Hepatitis	0=No 1=Yes
HHC	History of Hemochromatosis	0=No 1=Yes
BC	History of Biliary Cirrhosis	0=No 1=Yes

DataSchema

Core Variable	Variable Definition	Value Labeling
Cirrhosis	History of Cirrhosis	0=1 1=2
Tumor Differentiation	Pathological Differentiation	1=Well diff. 2=Well-moderately 3=Moderately 4=Moderately-poorly 5=Poorly
Ascites	Baseline Ascites	1=None 2=Mild 3=Moderate to Sever
Enceph	Hepatic encephalopathy	1=None 2=Grade1-!! 3=Grade III-IV (refractory)
Albumin	Baseline level of Albumin	Continuous Variable (g/dL)
Bilirubin	Baseline bilirubin level	Continuous Variable (mg/dL)
INR	Baseline International Normalized Ratio level	Continuous Variable
ALT	Baseline ALT level	Continuous Variable (U/L)
AST	Baseline AST level	Continuous Variable (U/L)

DataSchema

Core Variable	Variable Definition	Value Labeling
Creatinine	Baseline Creatinine	Continuous Variable (mg/DL)
PT	Baseline Prothrombin Level	Continuous Variable
Child-Pugh	Child-Pugh Actual Score (5-15)	Continuous Variable
Child-Pugh	Child-Pugh score categories	1=A 2=B 3=C
Child-Pugh_Time	Time of Child-Pugh Classification	1=At time of diagnosis 2=At time of patients' recruitment (Study enrollment)
MEDLD	Model for End Stage Liver Disease MELD Score= $(0.957 * \ln(\text{Serum Cr}) + 0.378 * \ln(\text{Serum Bilirubin}) + 1.120 * \ln(\text{INR}) + 0.643) * 10$ (if hemodialysis, value for Creatinine is automatically set to 4.0)	Continuous Variable If any score is <1, the MELD assumes the score is equal to 1.
AFP	Baseline Alpha-feto protein	Continuous values
AFP-Class	AFP Classification	0 =<400 1= ≥400

DataSchema

Core Variable	Variable Definition	Value Labeling
Platelets	Baseline Platelets	Continuous Variable
TumorN	Tumor nodularity	1= Single <2cm 2=Single >2-≤5cm 3=Single >5cm 4=Multiple (2-3 nodules) each ≤3 cm 5= Multiple (2-3 nodules) each >3 cm
TumorS	Tumor Size	0=≤50% of the liver 1= >50% of the liver
TNodN	Number of tumor nodularity	Continuous Variable
TMorph	Tumor Morphology	0= Uniodular and extension ≤50% 1=Multinodular and extension ≤50% 2=Massive or extension >50%
VasInv	Vascular Invasion	0=Negative 1=Positive
VasInvType	Type of Vascular Invasion	1=Microvascular 2=Macrovascular
Porthrombosis	Portal Vein Thrombosis	0=No 1=Yes
Metastasis	Extra Hepatic Distant Disease	0=No 1=Yes

DataSchema

Core Variable	Variable Definition	Value Labeling
Lymph nodes	Lymphatic involvement	0=No 1=Yes
ECOG	Performance Status (0-5)	0 1 2 3 4 5
TNM Staging	Baseline TNM staging system	1=I 2=II 3=IIIA 4=IIIB 5=IIIC 6=IVA 7=IVB
CLIP Staging	Baseline CLIP staging system	0 1 2 3 4 5 6

DataSchema

Core Variable	Variable Definition	Value Labeling
BCLC Staging	Baseline BCLC staging system	0=0 A=1 B=2 C=3 D=4
Okuda	Baseline Okuda Score	0 1
Baseline Treatment (At time of Diagnosis)	Type of received HCC treatment: Currently accepted treatment classification approaches including BCLC treatment algorithm, adopted by AASLD, EASL and NCCS. For Survival Analysis baseline treatment (at time of diagnosis) is the most predictive factor	0=No treatment 1= Surgery and liver transplant 2= Ablative therapies (RFA, ETOH, cryoablation) 3=Local Therapy (TACE, TARE, radiation) 4= Systemic Therapy including Sorafenib 5=Surgery with systemic therapy
Sorafenib	History of Sorafenib Intake	0=No 1=Yes
Sorafenib Date	Date of Sorafenib Intake	Month/Year

DataSchema

Core Variable	Variable Definition	Value Labeling
Sorafenib_R	Response to Sorafenib	0=Response 1=Progression
Sorafenib_PD	Date of Sorafenib Progression	Month/Year
Death Status	Survival Status	0= Alive or Censor 1=Dead
Death Date	Death Date	Month/Day/Year (XX/XX/XXXX)
Overall Survival	Duration between diagnosis and death or lost to follow	Continuous variable (Months)
PFS	Progression Free Survival duration between initiation of sorafenib to disease progression	Continuous variable (Months)