

Epidemiological Data Structure

Core Variable	Variable Definition	Variable values and codes
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+original ID
Status	Case-Control Status	1=HCC Case 2=Healthy Control 3=HCV-induced cirrhosis control
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 For Future Collaboration EUROPE (300s) SOUTH AMERICA (400s) ASIA (500s) AUSTRALIA (600s)
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 or 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
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
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
HX_cancer	Prior History of Cancer	0=No 1=Yes
Smoke	Status of Cigarette Smoking	0=Never Cigarette Smoking 1=Current Cigarette Smokers with ≤20 pkys (including ex-smoker <1 year since quit) 2=Current Cigarette Smokers with >20 pkys (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_hepatitis	Past history of hepatitis	0=No 1=HAV 2=HBV 3=HCV 4=HBV+HCV
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)
Metformin	Ever treated by metformin	0=No 1=Yes
Statin	Ever received Statin	0=No 1=Yes
FHX_cancer	First degree relative with Cancer	0=No 1=Yes
FHX_LC	First degree relative with liver cancer	0=No 1=Yes
Virus_test	Ever blood testing for HCV and HBV	0=No 1= Yes (HCV) 2=Yes (HBV) 3=Yes (HCV and HBV)
Virus_markers	HCV and HBV markers if blood was tested	0=HCV-/HBV- 1=HCV+/HBV- 2=HCV+/missing HBV 3=HCV-/HBV+ 4=HCV-/HBV missing

		5=HCV missing/HBV- 6=HCV missing/HBV+ 7=HCV+HBV+
--	--	--

Clinical Data Structure

Core Variable	Variable Definition	Variable values and codes
DxType	Type of diagnosis	1=Radiologically 2=Pathologically
DXDate	Date of HCC diagnosis	Month/day/year
Cirrhosis	History of Cirrhosis at time of HCC diagnosis	0=1 1=2
DXCirrhosis	Method of cirrhosis diagnosis at time of HCC diagnosis	1=Pathological 2=Radiological (no pathology) 3=Clinical (no path or image)
Tumor Differentiation	HCC Pathological Differentiation	1=Well diff. 2=Well-moderately 3=Moderately 4=Moderately-poorly 5=Poorly 9=Not documented
Ascites	Clinical, pathological, radiological evidence of Ascites	1=Absent 2=Slight 3=Moderate
Bilirubin	Baseline level of bilirubin at time of diagnosis	Continuous variable (mg/dL)
Bilirubin_Cat	Baseline categorical level of bilirubin at time of diagnosis	1= \leq 2 mg/dL 2=2-3 mg/dL 3= $>$ 3 mg/dL
Albumin	Baseline level of albumin at time of diagnosis	Continuous variable
Albumin_Cat	Baseline categorical level of albumin at time of diagnosis	1= $>$ 3.5 g/dL 2=2.8-3.5 g/dL 3= $<$ 2.8 g/dL
Prothrombin	Baseline Prothrombin Time (at time of diagnosis)	1=1-3 2=4-6 3= $>$ 6
INR	Baseline INR level (at time of diagnosis)	1= $<$ 1.7 2=1.8-2.3 3= $>$ 2.3
Encephalopathy	Evidence of encephalopathy (at time of diagnosis)	1=None 2=Grade 1-2 3=Grade 3-4
Child_Pugh	Child-Pugh score (computational variable) =Ascites + Bilirubin_cat + Albumin_cat + INR or Prothrombin + Encephalopathy	Continuous variable
CP_Score	Child-Pugh score Child_Pugh of 5-6= A Child_Pugh of 7-9=B Child_Pugh 10-15=C	0=A 1=B 2=C
TNM_T	Baseline HCC tumor status for TNM Staging	1= T1 (Single tumor without vascular invasion) 2=T2 (Single tumor with vascular invasion or multiple tumors with none > 5 cm) 3=T3 (multiple tumors > 5 cm or tumor involving a major branch of portal or hepatic vein(s)) 4=T4 (Perforation of visceral peritoneum)
TNM_N	Baseline lymph node involvement	0=N0 (No) 1=N1 (Yes)
TNM_M	Baseline distance metastasis	0=M0 (No) 1=M1 (Yes)
TNM Staging	Baseline 6 th Edition of TNM staging system	1=I (T1, N0, M0) 2=II (T2, N0, M0) 3=IIIA (T3, N0, M0)

		4=IIIB (T4, N0, M0) 5=IIIC (Any T, N1, M0) 6=IV (Any T, Any M, M1)
HCC_Nod	Baseline Tumor Nodularity	0=Uninodular 1=Multinodular
HCC_Ext	Baseline HCC Extension	0= ≤50% extension 1 > 50% extension
HCC_Morph	Baseline HCC (Tumor) morphology	0=uninodular and ≤50% extension 1=multinodular and ≤50% extension 2=Massive or extension >50%
AFP	Baseline Alpha-feto protein (ng/dL)	Continuous variable
AFP_cat	Baseline categorical level of AFP (ng/dL)	0= < 400 1=≥ 400
PVT	Baseline radiological evidence of Portal venous thrombosis	0=No 1=Yes
Clip_score	Baseline score of the CLIP (computational) Clip Score=CP_score + HCC_Morph + AFP_cat + PVT	0 1 2 3 4 5 6
HCC_Size	Baseline Tumor Size (Cm)	Continuous variable
HCC_Feature	Baseline Tumor Feature	0= single ≤2cm 1=Single ≤ 5cm or 3 nodules <3 cm 2=single > 5 cm or multinodular 3=portal invasion
ECOC	Baseline Performance Status of the HCC patient (at Diagnosis)	0= Asymptomatic (Fully active, able to carry on all pre-disease activities without restriction) 1=Symptomatic but completely ambulatory (Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature) 2=Symptomatic, <50% in bed during the day (Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours) 3= Symptomatic, >50% in bed, but not bedbound (Capable of only limited self-care, confined to bed or chair 50% or more of waking hours) 4= Bedbound (Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair)
BCLC Staging	Baseline BCLC staging system	Stage 0=0 Stage A=1 Stage B=2 Stage C=3 Stage D=4
Okuda_TS	Baseline Recoding of Tumor Size (Okuda Classification)	0= <5% (negative)) 1=>5% (Positive)
Okuda_Ascites	Baseline recoding of Ascites(Okuda Classification)	0=undetectable (negative) 1=detectable (positive)
Okuda_Alb	Baseline recoding of Albumin (Okuda Classification)	0= ≥ 3 g/dL (negative) 1=< 3 g/dL (positive)
Okuda_Bilirub	Baseline recoding of bilirubin (Okuda Classification)	0=< 3 mg/dl (negative) 1=>3 mg/dL (negative)
Okuda	Baseline Okuda Stage System	0= Stage I (No positive) 1= Stage II (one or two positives) 2= Stage III (three or four positives)
Vascular Invasion	Vascular Invasion	0=1 1=Yes

Treatment	Type of received HCC treatment	0=No treatment 1=Chemotherapy 2= Chemotherapy and others 3=Surgical Resection and Transplant 4=Surgery and others 5=Local (TAC)
Sorafenib Intake	History of Sorafenib Intake	0=No 1=Yes
Sorafenib_PTT	Treatment intake prior to sorafenib	0=No 1=Yes
Death Date	Death Date	Month/Day/Year (XX/XX/XXXX)
Sorafenib Date	Date of Sorafenib Intake	Month/Year
Sorafenib Response	Response to Sorafenib	0=Response 1=Progression
Sorafenib progression	Date of Sorafenib Progression	Month/Year
Death Status	Survival Status	0= Alive or Censor 1=Dead
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)