

Data Structure for Clinical Data Set

Variable Name	Label	Variable Type	Coding Values
PATIENT IDENTIFIERS			
MDACC	Medical Record Number	Numeric	---
LIVTX	Liver Case Control ID	Numeric	---
Birthdate	Date of birth	Date	---
dxdate	Date of diagnosis at MDACC	Date	---
mdaccregistration	Tumor registry date of Anderson first registration	Date	---
age	Age at HCC diagnosis	Numeric	---
sex	Gender	Categorical	1 = Male 2 = Female 9 = Missing
Race	Race/ethnicity of patient	Categorical	1 = White 2 = Hispanic 3 = Black 4 = Asian 5 = Other 8 = Not available 9 = Missing
ResCity	Patient's city of residence	String	---
ResCounty	Patient's county of residence	String	---
Zipcode	Patient's zip code for city of residence	Numeric	---
ResCountry	Country of Residence	String	---
ResSt	State of Residency	String	---
ResState	USA State of Residency or Nationality	Categorical	1 = Texas 2 = Louisiana 3 = Florida 4 = Mississippi 5 = New Mexico 6 = Georgia 7 = Oklahoma 8 = Non-US 9 = Tennessee 10 = Missouri 11 = Illinois 12 = Kentucky 13 = Alabama 14 = California 15 = Pennsylvania 16 = New York 17 = Kansas 18 = Michigan 19 = Arizona 20 = Nevada 21 = Minnesota 22 = South Carolina 23 = Oregon 24 = Indiana

- 25 = Montana
- 26 = Colorado
- 27 = North Carolina
- 28 = West Virginia
- 29 = Washington
- 30 = Arkansas
- 31 = Idaho
- 32 = Hawaii
- 33 = Nebraska
- 34 = New Jersey
- 35 = Wisconsin
- 36 = Wyoming
- 37 = Iowa
- 38 = South Dakota
- 39 = Maryland
- 40 = Massachusetts
- 41 = Virginia
- 42 = Maine
- 43 = Vermont
- 44 = North Dakota
- 45 = Alaska
- 46 = Utah
- 47 = Ohio
- 48 = Connecticut
- 49 = Rhode Island
- 50 = Delaware
- 51 = New Hampshire

referral	Referral location	Categorical	1 = Texas and adjacent states (OK=7, NM=5, AR, and LA=2) 2 = Other States 3 = International
Edu	Education level	Categorical	1 = Less than 8 th grade 2 = 9 th to 11 th grade 3 = 12 th grade or high school equivalent 4 = Some college or college graduate 5 = Graduate Study 9 = Not available
maritalST	Marital Status	Categorical	1 = Married 2 = Widowed 3 = Divorced 4 = Living as married 5 = Separated 6 = Never married 9 = Missing

PATHOLOGY

pathology_date	Date of initial pathology diagnosis	Date	---
accession	Accession number on pathology report	Numeric	---

Path_loc	Location of pathology	Categorical	1 = Pathology performed at MDACC 2 = Pathology performed outside and reviewed at MDACC 3 = Missing
path	Pathology differentiation	Categorical	1 = Well differentiated 2 = Well-moderately differentiated 2 = Moderately differentiated 3 = Moderately-Poorly differentiated 3 = Poorly differentiated 4 = Fibrolamellar 5 = Clear cell type 9 = Not Stated/Missing 10 = No biopsy 11 = Neuroendocrine 12 = Other 13 = Non differentiated
Liver_Steatosis	Presence of liver steatosis (pathology)	Categorical	1 = Yes 2 = No 3 = Not available in medical records/Missing

LAB VALUES

AFP	Alpha feto-protein at diagnosis	Numeric	---
albumin	Albumin at diagnosis	Numeric	---
ALK	Alkaline phosphatase at diagnosis	Numeric	---
ALT	Alanine aminotransferase at diagnosis	Numeric	---
AST	Aspartate aminotransferase at diagnosis	Numeric	---
bilirubin	Bilirubin at diagnosis	Numeric	---
Creatinine	Serum creatinine at diagnosis	Numeric	---
INR	International Normalized Ratio at diagnosis	Numeric	---
LDH	Lactate dehydrogenase at diagnosis	Numeric	---
LDL	Low-density lipoprotein at diagnosis	Numeric	---
Triglyceride	Triglyceride at diagnosis	Numeric	---
PT	Prothrombin time	Numeric	---
HCV	Antibody to hepatitis C virus	Categorical	1 = Yes 2 = No 9 = Missing
HBsAg	Hepatitis B surface antigen	Categorical	1 = Yes 2 = No 9 = Missing

AntiHBc	Antibody to hepatitis B core antigen	Categorical	1 = Yes 2 = No 9 = Missing
RISK FACTORS			
smok	Cigarette smoking status	Categorical	1 = Yes 2 = No 3 = Not available in medical records/Missing
drink	Drinking alcohol status	Categorical	1 = Yes 2 = No 3 = Not available in medical records/Missing
estrogen_replacement	Ever Use of Estrogen intake	Categorical	1 = Yes 2 = No 9 = Missing
OralContraceptive	Ever use of Oral Contraceptive Pills	Categorical	1 = Yes 2 = No 9 = Missing
Diabetes	Diabetes mellitus	Categorical	1=Yes 2= No 9 = Missing
weight	Patient's weight (in kg)	Numeric	---
height	Patient's height (in cm)	Numeric	---
BMI	Body Mass Index	Numeric	---
SYMPTOMS			
weight_loss	Patient experiencing weight loss as a symptom	Categorical	1 = Yes 2 = No 3 = Not available in medical records/Missing
Bleeding	Patient experiencing bleeding as a symptom	Categorical	1 = Yes 2 = No 3 = Not available in medical records/Missing
Fatigue	Patient experiencing fatigue as a symptom	Categorical	1 = Yes 2 = No 3 = Not available in medical records/Missing
Fever	Patient experiencing fever as a symptom	Categorical	1 = Yes 2 = No 3 = Not available in medical records/Missing
Pain	Patient experiencing pain as a symptom	Categorical	1 = Yes 2 = No 3 = Not available in medical records/Missing
FAMILY HISTORY			
fhx_can	Family history of cancer (First or Second Degree)	Categorical	1 = Yes 2 = No 3 = Not available in medical records/Missing

fhx_livc	Family history of liver cancer (First or Second Degree)	Categorical	1 = Yes 2 = No 3 = Not available in medical records/Missing
SURVIVAL			
deathdate	Death date or date of last visit	Date	---
death	Death Status	Categorical	0 = Alive 1 = Death
RADIOLOGY			
Evidence_of_cirh	Presence of cirrhosis (Radiology or Pathology)	Categorical	1 = Yes 2 = No 3 = Not available in medical records/Missing
Cirr_dx	Method of Cirrhosis Diagnosis	Categorical	1 = Pathology 2 = Radiology 3 = Clinical
Tr_Size	Tumor Size	Numeric	---
T_involvement	Tumor % of involvement for CLIP calculation	Categorical	1 = Less than or equal to 50% 2 = Greater than 50% 3 = Unavailable radiology report 4 = Unavailable due to resection 5 = Unavailable due to post treatment effect
vascular_inv	Vascular invasion	Categorical	1 = Yes 2 = No 3 = Could not be determined 9 = Unavailable
portal_throm	Vascular thrombosis	Categorical	1 = Yes 2 = No 3 = Unavailable
Major_branch_HV_PV_invasion	Invasion of the major branch of portal vein or hepatic vein	Categorical	0 = No 1 = Yes 3 = No vascular invasion 9 = Missing
lymphnode	Lymph node involvement	Categorical	1 = None 2 = Present 3 = Unavailable 10 = MUO
tumor_nodul	Tumor nodularity	Categorical	1 = Solitary 2 = Multinodular 3 = Unavailable due to liver transplantation 4 = Unavailable due to partial hepatectomy 5 = Unavailable due to local therapy

			6 = Unavailable documents
Num_nodules	BCLC tumor nodularity	Categorical	1 = 1 2 = 2 – 3 3 = Greater than 3 4 = Unavailable due to liver transplantation 5 = Unavailable due to partial hepatectomy 6 = Unavailable due to local therapy 7 = Unavailable documents
mets	Metastasis	Categorical	1 = None 2 = Present 3 = Unavailable 10 = MUO
GBD	Status of gallbladder disease	Categorical	1 = Cholecystectomy prior to HCC 2 = Cholecystectomy concurrent with HCC resection 3 = Evidence of gallstones 4 = Evidence of cholecystitis 5 = Normal 6 = Gallbladder sludge 9 = Unknown or missing comments by radiologist 10 = Benign changes or Adenomyosis 99 = Unknown for lack of imaging
OrganInvasion	Adjacent organ invasion by the tumor	Categorical	1 = Peritoneal invasion 2 = Gallbladder invasion 3 = Other organ invasion 4 = None or no comment by radiologists 9 = Missing
ascites_	Presence of ascites	Categorical	1 = None 2 = Slight 3 = Moderate 4 = Massive
ecog	Continuous values of performance status	Categorical	0 = PS0 1 = PS1 2 = PS2 3 = PS3 4 = PS4 9 = No available data
enceph_cpg	Encephalopathy	Categorical	1 = None 2 = Minimal 3 = Advanced

SCORING SYSTEM

CPS	Child Pugh Score	Numeric	---
CPS_cat	Child Pugh Score Category	Categorical	1 = A 2 = B 3 = C
CLIP	CLIP Staging System	Categorical	1 = Stage 0 – 2 2 = Stage 3 3 = Stage 4 – 6 9 = Missing
TNIM	TNM Staging System	Categorical	1 = Stage I 2 = Stage II 3 = Stage IIIA 4 = Stage IIIB 5 = Stage IIIC 6 = Stage IVA 7 = Stage IVB
BCLC	BCLC Staging System	Categorical	0 = Stage 0 1 = Stage A 2 = Stage B 3 = Stage C 4 = Stage D 9 = Missing

TREATMENT

Baseline_ttt	Treatment at Baseline	Categorical	1 = Surgical 2 = Ablation 3 = Local 4 = Systemic 5 = Supportive care 9 = Missing records
Baseline_TTT_date	Date of baseline treatment response	Date	---
Baseline_TTT_Res	Response of baseline treatment	Categorical	1 = Post-surgical recurrence 2 = Progressed/death 3 = Respond 9 = Missing
Baseline_TTT_Res_Date	Date of baseline treatment response	Date	---
Ever_surgery_TTT	Follow-up surgery	Categorical	1 = Yes 2 = No 9 = Missing
Second_line_ttt	Second line therapy	Categorical	1 = Yes 2 = No 9 = Missing
Second_line_ttt_SD	Second line treatment start date	Date	---
Second_line_ttt_LD	Second line treatment end date	Date	---
Second_line_ttt_Res	Second line treatment response	Categorical	1 = Post surgical recurrence 2 = Progressed/death 3 = Respond 9 = Missing

Third_line_ttt	Third line therapy	Categorical	1 = Yes 2 = No 9 = Missing
Third_line_ttt_SD	Third line treatment start date	Date	---
Third_line_ttt_LD	Third line treatment end date	Date	---
Third_line_ttt_Res	Third line treatment response	Categorical	1 = Post surgical recurrence 2 = Progressed/death 3 = Respond 9 = Missing
Fourth_line_ttt	Fourth line therapy	Categorical	1 = Yes 2 = No 9 = Missing
Fourth_line_ttt_SD	Fourth line treatment start date	Date	---
Fourth_line_ttt_LD	Fourth line treatment end date	Date	---
Fourth_line_ttt_Res	Fourth line treatment response	Categorical	1 = Post surgical recurrence 2 = Progressed/death 3 = Respond 9 = Missing
Fifth_line_ttt	Fifth line therapy	Categorical	1 = Yes 2 = No 9 = Missing
Fifth_line_ttt_SD	Fifth line treatment start date	Date	---
Fifth_line_ttt_LD	Fifth line treatment end date	Date	---
Fifth_line_ttt_Res	Fifth line treatment response	Categorical	1 = Post surgical recurrence 2 = Progressed/death 3 = Respond 9 = Missing
Sixth_line_ttt	Sixth line therapy	Categorical	1 = Yes 2 = No 9 = Missing
Sixth_line_ttt_SD	Sixth line treatment start date	Date	---
Sixth_line_ttt_LD	Sixth line treatment end date	Date	---
Sixth_line_ttt_Res	Sixth line treatment response	Categorical	1 = Post surgical recurrence 2 = Progressed/death 3 = Respond 9 = Missing
Surgery1	Patient ever had surgery	Categorical	1 = Yes 2 = No
Surgdate1	Date of surgery 1	Date	---
Recurr1	Presence of recurrence	Categorical	1 = Yes 2 = No
Daterecurr1	Date of recurrence	Date	---
Surgery2	Patient ever had second surgery	Categorical	1 = Yes 2 = No
Surgdate2	Date of surgery 2	Date	---
Recurr2	Second recurrence	Categorical	1 = Yes 2 = No

Daterecurr2	Date of second recurrence	Date	---
Margin	Safe margin during surgery	Categorical	1 = Yes 2 = No