

Upper gastrointestinal cancer dataset using UKACR cancer repository dataset and HES data 1998-2007

Data range

All gastroesophageal cancers (OG) (ICD10 C15 and C16) from UK cancer repository dataset for diagnosis years 1998-2007

All hepatic, pancreatic and biliary cancers (HPB) (ICD10 C17.0 and C22-C25) from UK cancer repository dataset for diagnosis years 1998-2007

		Final numbers in dataset	
		OG	HPB
Data are from the UK cancer repository dataset unless specified as HES data fields		136,241	101,336
Removed from dataset	Not resident in England	4	2
	Date of death before date of diagnosis	5	5
	Duplicates removed (IARC)	2,416	1,924
	Registry duplicate	12	26
Entered to analysis		133,804	99,379
	DCOs removed	4,990	9,377
Entered to analysis (excl DCOs)		128,814	90,002

Heading	Explanation
ID	Unique anonymised row identifier in the NCDR - one row per patient tumour
patient_number_map_to2	This allows mapping of known duplicates. One patient number is assigned to all cases which appear to be duplicates within or
tumournumbermerge	Tumour number supplied by the registries in the NCDR
patientnumbermerge	Patient number supplied by the registries in the NCDR
source	Registry supplying and year data supplied
nhsno	NHS number
dob	Date of birth
sex	Sex
site3	Three digit ICD10 code
site4	Four digit ICD10 code - NB no "."
type5	Morphology code - NB no leading "M"
charlson_score_1year	Charlson score at 1 year. -1 = no match in the HES data
charlson_score_3year	Charlson score at 3 years. -1 = no match in the HES data
dodeath	Date of death - this is an amalgam of the registries dates of death and the ONS date of death
docensor	Date of censor = Date of death or 31st December 2007 if alive or dead after this date
cod_1a	Cause of death Ia
cod_1b	Cause of death Ib
cod_1c	Cause of death Ic
cod_2	Cause of death II
dodiag	Date of diagnosis supplied by the cancer registries
ONS_Stage	ONS stage - stage 1 - 4 supplied by NYCRIS only
original_stage	This is the stage supplied by the cancer registries
Recorded_Dukes	This is a derived field only used for colorectal cancer at present but will come for the other cancers
description_stage	This field relates to the ONS_Stage field and is supplied by NYCRIS only
finaldukes	This is a derived field only used for colorectal cancer at present but will come for the other cancers
tnm	TNM - supplied by ECRIC only
t_clin	Clinical T where available
n_clin	Clinical N where available
m_clin	Clinical M where available
t_path	Pathological T where available
n_path	Pathological N where available
m_path	Pathological M where available
dukesfromtnmclin	This is a derived field only used for colorectal cancer at present but will come for the other cancers
dukesfromtnmpath	This is a derived field only used for colorectal cancer at present but will come for the other cancers
met	Did the patient have metastases
grade	Grade of the tumour
grade_description	Description of the grade by each registry
place_of_death	Place of death - seems very variable
basis_of_dagnosis	Basis of diagnosis
dco	Death certificate only case
Ethnicity	Ethnicity as recorded in the HES data. T= no links in HES and U is no valid ethnicity recorded in HES
postcode_unformatted	Postcode with no spaces
year_of_diagnosis	Year of diagnosis
year_of_death	Year of death
age	Age at diagnosis
census_soa_lower	Super lower output area code
2004Quintile	2004 IMD income domain quintile
2007Quintile	2007 IMD income domain quintile
cancer_gor	Government Office code
GoR_name	Government Office text
cancer_network_code	Cancer Network code
CN_name	Cancer Network text
cancer_pct_current	PCT code
PCT_name	PCT text
cancer_registry_code	Cancer Registry code
CR_name	Cancer Registry text
HESsurg	Did the patient have surgery in HES, 1 = yes
HESsurg hosp	Hospital code of place giving this surgery
HESOPSC4	Operation code for this surgery
HESopdate	Date of the operation for this surgery
HESdiagtosurg	Time from diagnosis to operation (-30 to 183 days)
HESopseverity	The severity of this operation, 1 = most severe
HESchemo	Did the patient have chemotherapy in HES, 1 = yes
HESchemodate	Date of the first mentioned chemotherapy
HESdiagtochemo	Time from diagnosis to the first chemotherapy (-30 to 183 days)
CRsurg	Did the patient have surgery in the registries data
CRsurg hosp	Hospital code of place giving this surgery
CROPSC4	Operation code for this surgery
CRsurgdate	Date of the operation for this surgery
CRdiagtosurg	Time from diagnosis to operation (-30 to 183 days)
CRopseverity	The severity of this operation, 1 = most severe
CRchemo	Did the patient have chemotherapy in the registries data
CRchemodate	Date of the first mentioned chemotherapy
CRdiagtochemo	Time from diagnosis to the first chemotherapy (-30 to 183 days)