1.0 Data Dictionary

1.1 Subject Data

1.1.1 Source.data.Subjects

The subjects table holds all diagnoses currently in the HMRN analysis dataset. One record in this table is for one diagnosis, a patient can have multiple diagnoses.

Variable	Data type	Definition
EGU_ID	varchar(10) PRIMARY KEY	Unique Identifier assigned by ECSG at time of case notification. The name EGU_ID reflects the group's previous name (Epidemiology & Genetics Unit).
Demographics		
Sex	varchar(1)	**M:** male **F:** female **U:** unknown
AgeDiagnosis	float	Age at diagnosis (year)
Diagnostics		
DateOfSample	date	Date of diagnostic sample
DateOfDiag	date	Date of diagnosis. This is the date the diagnosis was authorised by HMDS.
RegistryYear	int	The year of the Registry that a diagnoses was made. The registry years run from 1st September to 31st August with the first registry year starting 1st September, 2004.
Diagnosis	nvarchar(510)	Name of diagnosis assigned by HMDS
DiagNumber	tinyint	Diagnosis number in order of diagnosis date.
ICDO3Code	varchar(6)	International classification of diseases for oncology – 3rd edition (ICD-O-3) code assigned by HMDS. Please be aware, particularly for indolent diagnoses, this may be assigned where a code is not currently defined by the WHO classification. For example for code **9823/3**, covers the following diagnoses/subtypes: B-cell chronic lymphocytic leukaemia, Lymphoproliferative disorder NOS, Monoclonal B-cell lymphocytosis (CLL phenotype)
DiagnosticSubtypeICDO3	varchar(255)	This generally correponds to ICD-O-3 code, but aggregated or disaggrated where necessary and is the variable to be used for analyses at a diagnostic level. For example: **Anaplastic large cell lymphoma of T/null type ALK-** and **Peripheral T-cell lymphoma - common; unspecified** are both found under the ICD-O-3 code, 9702/3. **See Source.config.DiagnosticSubtypes.Subtype for full list of values**

Variable	Data type	Definition
DiagnosticGroupWHOBroad	varchar(200)	Broad WHO Diagnostic Category: **Myelodysplastic / Myeloproliferative neoplasms** **Myelodysplastic syndromes** **Myeloproliferative neoplasms** **Acute myeloid leukaemias** **Precursor lymphoid neoplasms (ALL)** **Mature lymphoid neoplasms**
DiagnosticGroupWHO	varchar(255)	WHO group for the diagnosis: **Chronic myeloid leukaemia** **Chronic myeloproliferative neoplasms** **Primary myelofibrosis** **Myeloid & lymphoid neoplasms with PDGFRA or PDGFRB rearrangement** **Myelodysplastic / Myeloproliferative neoplasms** **Myelodysplastic syndromes** **Acute myeloid leukaemia** **Acute leukaemia - mixed phenotype** **B-lymphoblastic leukaemia** **Monoclonal B-cell lymphocytosis** **Monoclonal gammopathy of undetermined significance** **Chronic lymphocytic leukaemia** **Hairy cell leukaemia** **Myeloma** **Plasmacytoma** **Extranodal marginal zone lymphoma** **Systemic marginal zone lymphoma** **Follicular lymphoma** **Mantle cell lymphoma** **Diffuse large B-cell lymphoma** **Burkitt lymphoma** **Lymphoproliferative disorder NOS** **B-cell lymphoma, intermediate between DLBCL and classical HL** **Lymphocyte predominant nodular Hodgkin lymphoma** **Classical Hodgkin lymphoma** **T-cell lymphoma** **T-cell leukaemias** **T-lymphoblastic leukaemia**
DiagnosticGroupWHOID	int	Current numeric code assigned to specific value of DiagnosticGroupWHO

Variable	Data type	Definition
DiagnosticGroup	varchar(50)	Diagnostic group acronym: used for convenience in quick selections: **Acute-Mixed** **AML** **APML** **B-ALL** **B-CLL** **Burkitt** **CML** **DLBCL** **DLBCL-HL** **Follicular** **Hairy** **HDCN** **Hodgkin** **LPD-NOS** **Mantle** **MBL** **MDS** **MDS/MPN** **MGUS** **MPN** **Myeloma** **MZL-E** **MZL-S** **PDGFR** **Plasmacytoma** **T-ALL** **T-Leukaemia** **T-Lymphoma**
DiagnosticGroupICD10	varchar(50)	ICD10 Group for the diagnosis It is frequently not possible to bridge code from ICD-O-3 to ICD 10 and clinically these groups are generally meaningless with the exception of Hodgkin lymphoma and plasma cell neoplasms. However, we recognise that in order to make comparisons with other published data it is sometimes necessary to aggregate the diagnoses to these groups **Hodgkin lymphoma** **Leukaemia** **Non-Hodgkin lymphoma** **Plasma cell neoplasms** **Other disorders**
ICD10	varchar(100)	ICD 10 Code, please refer to theDiagnosticGroupICD10 definition **Hodgkin's disease (C81)** **Non-Hodgkin lymphoma (C82–C85)** **Myeloma (C90)** **Lymphoid leukaemia (C91)** **Myeloid leukaemia (C92-C94)** **Myelodysplastic syndromes (D46)** **Other neoplams of uncertain or unknown behaviour (D47)**

Variable	Data type	Definition
Lineage	varchar(15)	Lineage of the diagnosis **Myeloid** **Lymphoid**
Area based measures		
CancerNetwork	varchar(50)	These areas are now offically defunct and have been replaced by Strategic Clinical Networks, however, clinically the Network still operates at this level. **N06:** Yorkshire **N07:** Humber & Yorkshire Coast
LowerSOA_2001	varchar(50)	2001 Census Geography Lower Super output area
IMDIncome	float	Index of Multiple Deprivation - Income Domain (2007). Continuous score ranging from **0 (affluent)** to **1 (deprived)**. http://data.gov.uk/dataset/index_of_multiple_deprivation_imd_2007 IMD 2004 and 2010 are also available on request.
IMDIncomeCat5	tinyint	Defined by ranking every Lower SOA in England by their score **1** (affluent), **2**, **3**, **4**, **5** (deprived)
IMDIncomeCat7	tinyint	**1** (affluent), **2**, **3**, **4**, **5**, **6**, **7** (deprived)
Mortalilty		
EventType	varchar(10)	From the mortailty linkage: **D:** death **CAN:** cancelled cipher **AF/ SMO:** under care of armed forces/service medical officer **E / EMB:** embarkation from the UK **SCT:** Registered in Scotland **NULL:** No event recorded
EventDate	date	Date of event (EventType) and this date should be used as the censor date. For patients still alive (EventType = NULL) the date will be the last date information on flagging was known. If EventDate = NULL, this means the subjects has not been flagged for mortaility data.
AgeEvent	float	Age at event date (death, cancellation, censor)
Data Collection		

Variable	Data type	Definition
FormStatus	varchar(20)	Status of data collection (first abstraction) : **FormNotPrinted**: data collection process not yet initiated (~8 months after date of diagnosis) **Input**: data collection completed and input onto HILIS **FormNotCompleted**: data collection not possible due to reason stated by variable FormNotCompleted **FormNotReturned**: data collection still in process **Returned**: form returned to HMDS awaiting input onto HILIS **Scunthorpe/Grimsby**: data collection was not initiated for patients diagnosed at these hospital in 2004
FormNotCompleted	varchar(20)	The reason for the data collection form not being fully completed; specifically treatment and prognostic data will be unavailable **DERMATOLOGY**: treated by dermatology & not refered to haematology (cutaneous T-cell lymphomas) **NO_HAEM_REFERRAL**: not referred to haematology. Usually because either the patient has serious competing co-morbidity or the diagnoses was incidental and does not require further investigations/treatment **NO_HOSP_REFERRAL**: diagnoses and management being undertaken in primary care only **NO_TREATMENT**: this option is now obsolete as this is recorded as part of the data collection procedure **NON_YHHN_HOSP**: infrequently a patient resident in the study area chooses to be treated outside the Network. Either in one of the hospitals nearby such as Middlesborough or the patient may choose to be nearer relatives. Yorkshire Humberside Haematology Network (YHHN) refers to the original name for HMRN and is still in use specifically for patient contact. **NOTES_UNOBTAINABLE**: despite repeated attempts it has not been possible to obtain the medical records. **POST_MORTEM_SAMPLE**: occasionally a diagnoses made at HMDS is from a post mortem sample so no data are available **PRIVATE_PATIENT**: the patient is being treated privately **SHEFF_CHILDREN**: specifically for paediatric cases diagnosed in the region but being treated at Sheffield Childrens' Hospital
DateOfFollowUp	date	The last date medical records were accessed by the nurses
DataCollectionDOD	date	Date of death recorded at data collection from the medical records. This is used to signify the data collection process is complete and should not be used for survival analyses

Variable	Data type	Definition
DiagnosticHospital	varchar(50)	Hospital where the first diagnostic sample originated from. **Airedale** **Bradford Royal** **Calderdale** **Leeds Teaching Hospitals** **Harrogate** **York** **Dewsbury & District** **Pinderfields** **Pontefract** **Hull & East Yorkshire Hospitals** **Scarborough** **Diana Princess of Wales** **Scunthorpe** **Private** **Pathlinks** **Surgery/Medical Centre** **Lincoln County** **Nottingham** **Middlesbrough** **Newcastle**
MDT	varchar(50)	Multi Disciplinary Team (Diagnostic Hospital). **Airedale/Bradford** **Leeds** **Harrogate/York** **Mid Yorkshire** **HYCCN (Humber Yorkshire Coast Cancer Network)**
TreatingHospital	varchar(50)	Hospital where the first treatment was given. **Airedale** **Bradford** **Calderdale** **St James's, Leeds** **Harrogate** **York** **Mid-Yorks** **Hull & East Yorkshire Hospitals** **Scarborough** **Diana Princess of Wales** **Scunthorpe** **Private** **Pathlinks** **Surgery/Medical Centre** **Middlesbrough**

Variable	Data type	Definition
AntecedentEvent	nvarchar(30)	Details of antecedent or concurrent event the patient has been diagnosed with. Always recorded **Immunodeficiency (HIV)** **Solid organ transplant** **Down syndrome** Recorded if treated with chemotherapy and/or radiotherapy **Autoimmune disorder** **Haematological malignancy** **Non-haematological malignancy** **Non-haematological malignancy and haematological malignancy** **Other**
NationalDataLinkage	int	Quality of the national data linkage (Mortality, HES & cancer registration) **1: Linked** **2: Potentially missing data** **3: Not linked**

1.1.2 Source.data.SubjectsWide

This table contains the same variables as Source.data.Subjects except all of the fields that relate to the diagnosis are repeated for each diagnosis a patient has using the DiagNumber. This means that each patient has only one row in the table as opposed to multiple rows in the Subjects view.

1.2 Treatment Data

1.2.1 Source.data.Treatment

This table contains treatment data for all patients up to the last date their medical records were abstracted as indicated by the field DateOfFollowUp in Subjects and SubjectsWide.

Variable	Data type	Definition
tx_id	int PRIMARY KEY	Primary key HILIS
EGU_ID	nvarchar(16)	ECSG Patient Identifier
tx_pid	Int	HILIS PID
hospital	nvarchar(max)	Treating hospital
type	nvarchar(max)	Treatment type: **Bisphosphonates** **Care of dying pathway** **chemotherapy** **clinical trial** **combination** **community monitoring** **died pre-treatment** **discharged from haematology** **discharged to GP** **erythropoietin** **H. pylori eradication** **immunosuppressive** **Liverpool care pathway** **Lost to hospital** **non-haematological** **observation** **palliative** **radiotherapy** **refused treatment** **resection** **splenectomy** **stem cell harvest** **stem cell transplant** **steroids** **supportive care** **telephone clinic** **venesection** **vertebroplasty**
detail	nvarchar(max)	Regimen string as received from HILIS
start_date	date	Date treatment was first started. If date is not known the missing value code 01/01/1000 is used.
end_date	date	Last date the treatment was given. If date is not known the missing value code 01/01/1000 is used.
response	nvarchar(100)	**CR**: Complete Remission **MR**: Minimal Residual Disease

		NE: Non Evaluable **PD**: Progressive Disease **Plateau**: Plateau Phase **PR**: Partial Response **sCR**: Strict Complete Remission (Myeloma Only) **SD**: Stable Disease **VGPR**: Very Good Partial Response (Myeloma Only) **NULL**: Not Applicable
Regimen	nvarchar(200)	Regimen string as received from HILIS, if Regimen = **NULL** and detail <> **NULL** then the regimen has not been linked to the config.Treatments table.
Fields assigned after linkage to config.Treatment Table		
RegimenAgents	nvarchar(max)	Agents used in the regimen
RegimenGroup	nvarchar(200)	Collapsed groups for regimens, mostly affecting clinical trials e.g. for all arms of AML17, RegimenGroup = **AML17**
RegimenName	nvarchar(200)	Homogenised name for different versions of regimen names, i.e. where Regimen = **COP or CVP**, RegimenName = **CVP**
RegimenType	nvarchar(200)	Corrected regimen type for treatments such as Aspirin or Bisphosphonates, which HILIS classifies under chemotherapy but for analysis purposed this would be unhelpful. For Regimen = **Aspirin** or **G-CSF**, RegimenType = **supportive care.**
Intensive	int	Indicates intensive / non-intensive chemotherapy, only applicable for some diagnoses, needs to be requested for additional diagnoses. **0**: non-intensive **1**: intensive **NULL**: not assigned

1.2.2 Source.data.TreatmentWide

This table contains the same data as 'Source.data.Treatment' but reshaped into a wide format with one row per patient.

1.3 Presentation data

1.3.1 Source.data.Presentation_Clinical

This contains all data with which the subject presented at the time of ***initial*** diagnosis; it is recorded on the data collection forms and is disease specific. It should be linked to subjects on both EGU_ID and DiagNumber, as all data relates to only one diagnosis, usually, but not exclusively, the first.

Variable	Data type	Applicable diagnostic group	Definition
EGU_ID	nvarchar(10), PRIMARY KEY		
DiagNumber	tinyint		Indicates diagnosis to which data relates.
Acute_cns	varchar(1)	Acute	Central nervous system involvement **Y**: Yes **N**: No **U**: Unknown
ecog	int	All groups	ECOG performance status **0** to **4** **8**: Not known **9**: Not done
ldh	nvarchar(10)	Lymphoproliferative	Lactate dehydrogenase level **low** **normal** **raised** **unknown** **not done**
hb	float	Myeloproliferative Myelodysplastic Plasma-cell Lymphoproliferative	Haemoglobin b level (g/dL) at the time of diagnosis. **88.8**: Not known **99.9**: Not done
wbc	float	Myeloproliferative Myelodysplastic Lymphoproliferative	White blood cell (10^9/L) count at time of diagnosis. **8888**: Not known **9999**: Not done
lymphs	float	Lymphoproliferative	Lymphocytes count (10^9/L) at time of diagnosis. **8888**: Not known **9999**: Not done
albumin	float	Plasma-cell Lymphoproliferative	Albumin level (g/L) at the time of diagnosis. **888**: Not known **999**: Not done

Variable	Data type	Applicable diagnostic group	Definition
ct	nvarchar(1)	Lymphoproliferative	Whether a CT scan was performed at diagnosis. **Y**: Yes **N**: No **U**: Unknown
bm	nvarchar(1)	Lymphoproliferative	Whether a bone marrow biopsy was performed at diagnosis. **Y**: Yes **N**: No **U**: Unknown
sweats	nvarchar(1)	Lymphoproliferative	Presence of night sweats at diagnosis **Y**: Yes **N**: No **U**: Unknown
b2m	float	Plasma-cell Lymphoproliferative	Level of beta-2 microglobulin at diagnosis **88.8**: Not known **99.9**: Not done
wt_loss	nvarchar(1)	Lymphoproliferative	Presence of weight loss at diagnosis **Y**: Yes **N**: No **U**: Unknown
fever	nvarchar(1)	Lymphoproliferative	Presence of fever at diagnosis. **Y**: Yes **N**: No **U**: Unknown
binet	nvarchar(1)	CLL	Binet stage. **A**, **B**, **C**
stage	nvarchar(4)	Lymphoproliferative, not CLL	Ann Arbor stage,. As recorded in the medical notes, do not use routinely for analyses (use calculated stage in the prognostics table) **I**, **IE**, **II**, **IIE**, **IIS**, **III**, **IIIE**, **IIIS** or **IV** **U**: Unknown
plts	float	CLL Myeloproliferative Myelodysplastic	Platelet count (10^9/L) **8888**: Not known **9999**: Not done
pcv	float	Myeloproliferative Myelodysplastic	Packed cell volume, haematocrit (%) **99.9**: Not done

Variable	Data type	Applicable diagnostic group	Definition
neutrophils	float	Myeloproliferative Myelodysplastic	Absolute neutrophil count (10^9/L) **888**: Not known **999**: Not done
monocytes	float	Myeloproliferative Myelodysplastic	Absolute monocyte count (10^9/L) **888**: Not known **999**: Not done
splenomegaly	nvarchar(1)	Myeloproliferative Myelodysplastic	Presence of splenomegaly **Y**: Yes **N**: No **U**: Unknown
detection_spleen	nvarchar(10)	Myeloproliferative Myelodysplastic	Method of detection for splenomegaly **palpation** **ultrasound** **both** **unknown**
hepatomegaly	nvarchar(1)	Myeloproliferative Myelodysplastic	Presence of hepatomegaly **Y**: Yes **N**: No **U**: Unknown
detection_liver	nvarchar(10)	Myeloproliferative Myelodysplastic	Method of detection for hepatomegaly **palpation** **ultrasound** **both** **unknown**
transfusion	nvarchar(1)	Myelodysplastic	Patient transfusion dependent **Y**: Yes **N**: No **U**: Unknown
rcm	nvarchar(8)	No longer collected	Red cell mass **normal** **raised** **unknown** **not done**
epo	nvarchar(8)	Myeloproliferative	Pre-treatment erythropoietin level **low** **normal** **raised** **not done**

Variable	Data type	Applicable diagnostic group	Definition
cytopenias	float	Myelodysplastic	Number of cytopenias present **0** to **3** **7**: Not applicable **8**: Not known **9**: Not done
karyotype	nvarchar(12)	Myelodysplastic	Karyotypic mortality risk group **good** **intermediate** **poor** **unknown**
cd34	float	Myelodysplastic	Percentage of marrow cd34 (blast cells) **777**: Not applicable **888**: Not known **999**: Not done
bone_disease	nvarchar(18)	Plasma-cell	Type of bone disease present **equivocal** **lytic** **multiple fractures** **multiple types** **osteopenia** **osteoporosis** **plasmacytoma** **single fracture** **none** **unknown**
paraprotein	nvarch(13)	Plasma-cell	Paraprotein type **IgA** **IgA & IgG** **IgD** **IgE** **IgG** **IgG & IgM** **IgM** **non-secretory** **none**
pp_level	float	Plasma-cell	Paraprotein level (g/L) **0:** Unquantifiable **88**: Not done **99**: unknown

Variable	Data type	Applicable diagnostic group	Definition
bence_jones	nvarchar(9)	Plasma-cell	Presence and type of urinary free light-chain (Bence-Jones) paraproteins. **none** **kappa** **lambda** **both** **not done** **not known**
serum_flc	nvarchar(9)	Plasma-cell	Presence and type of serum free light-chain (Bence-Jones) paraproteins. **none** **kappa** **lambda** **both** **not done** **not known**
creatinine	nvarchar(4)	Plasma-cell	Serum creatinine (µmol/L) **8888**: Not known **9999**: Not done
Igs	nvarchar(10)	Plasma-cell	Background immunoglobulin level **raised** **normal** **suppressed** **unknown**
skeletal_survey	nvarchar(1)	Plasma-cell	Indicates completion of skeletal survey **Y**: yes **N**: No **U**: Unknown
mri	nvarchar(1)	Plasma-cell	Indicates completion of MRI scan **Y**: yes **N**: No **U**: Unknown
bone_lesions	nvarchar(8)	Plasma-cell	Number of bone lesions present **multiple** **single** **none** **unknown**
serum_kappa	float	Plasma-cell	Serum level of kappa light-chain paraproteins (mg/L) **88888**: Not known **99999**: Not done

Variable	Data type	Applicable diagnostic group	Definition
serum_lambda	float	Plasma-cell	Serum level of lambda light-chain paraproteins (mg/L) **88888**: Not known **99999**: Not done
calcium	float	Plasma-cell	Serum calcium level (mmol/L) **88.8**: Not known **99.9**: Not done

1.3.2 data.DiseaseInvolvement

This table contains information on sites of disease involvement at presentation for all lymphoproliferative disorders and follow-up (relapse) information for a subset of patients. Again, this should be linked to subjects by both EGU_ID and DiagNumber.

Variable	Data type	Definition
EGU_ID	varchar(8) PRIMARY KEY	
DiagNumber	tinyint	Indicates diagnosis to which data relates.
imaging_event_id	int	
dataset	smallint	Chronological ordinal for multiple results
scan_type	nvarchar(50)	Type of scan from which involvement was assessed
timepoint	nvarchar(50)	
date_of_scan	date	
cns	tinyint	Indicates the sites of nodal or extranodal involvement: **0**: Not involved **1**: Involved **NULL:** Not known
git		
gu		
blood		
bone		
liver		
pulmonary_left		
pulmonary_right		
marrow		
orbit_left		
orbit_right		
skin		
salivary_gland_left		
salivary_gland_right		
thyroid		
pericardium		
muscle		
waldeyer		

Variable	Data type	Definition
neck_left		
neck_right		
infraclavicular_left		
infraclavicular_right		
axillary_left		
axillary_right		
arm_left		
arm_right		
thymus		
mediastinal		
hilar_left		
hilar_right		
para_aortic		
mesenteric		
iliac_left		
iliac_right		
inguinal_left		
inguinal_right		
popliteal_left		
popliteal_right		
spleen		
extensive	tinyint	Indicates extensive/systemic involvement **0**: No **1**: Yes **NULL**: Not known
bulky	tinyint	Indicates bulky involvement (largest site > 10cm) **0**: No **1**: Yes **NULL**: Not known

Variable	Data type	Definition
check_scan	tinyint	Indicates that radiology report needs checking by clinician **0**: No **1**: Yes **NULL**: Not applicable
other	int	Indicates involvement outside of the locations listed above. Details of the location of this involvement are held in the 'other notes' field. **0**: No **1**: Yes **NULL**: Not applicable
other_notes	nvarchar(400)	Details of the other sites of extranodal involvement as free text

1.3.3 data.Prognostics

The prognostic table provides all data relating to prognostic indicators and scores. Most of the prognostic scores relate to specific groups or subgroups of haematological malignancies, such as the IPI for lymphomas and the FLIPI specifically for follicular lymphoma. Linkage to subjects should be by EGU_ID and DiagNumber.

Variable	Date type	Diagnostic group applicable to (where relevant)	Notes/explanation
EGU_ID	nvarchar(10) PRIMARY KEY		
DiagNumber	tinyint		Indicates diagnosis to which data relates.
BSymptoms	int	Lymphoproliferative	Presence of at least one B-symptom (fever, weight loss, sweats) **0**: No/unknown **1**: Yes **NULL**: Non-calculable

Variable	Date type	Diagnostic group applicable to (where relevant)	Notes/explanation
Karyotype	nvarchar(12)	Myelodysplastic	Corrected version of field found in data.PresentationClinical, supplemented with further data from cytogenetics. MDS Karyotype risk group **good** **intermediate** **poor** **unknown** AML Karyotype risk group (deprecated, but remnant entries remain) **favourable** **intermediate** **adverse** **unknown**
CD34	float	Myelodysplastic	Percentage of marrow cd34 (blast cells) **777**: Not applicable **888**: Not known **999**: Not done
CTScan	int	Lymphoproliferative	CT scan performed (codified) **0**: No **1**: Yes **9**: Unknown
PETScan	int	Lymphoproliferative	PET scan performed **0**: No **1**: Yes
Binet	nvarchar(1)	CLL	Binet CLL risk category **A**, **B**, **C** **NULL**: Non-calculable
Rai	nvarchar(3)	CLL	Rai CLL risk score **0**, **I**, **II**, **III**, **IV** **NULL**: Non-calculable
RaiRiskGroup	nvarchar(20)	CLL	Rai risk category **Low** **Intermediate** **High**

Variable	Date type	Diagnostic group applicable to (where relevant)	Notes/explanation
Stage	nvarchar(5)	Lymphoproliferative	Calculated Ann Arbor stage of disease at presentation taking highest of Stage derived from CT and PET **I**, **II**, **III**, **IV** **NULL**: Non-calculable
IPI	varchar(50)	NHL (aggressive)	The IPI risk category based on the IPI_Score **Low** **Low/Intermediate** **Intermediate/High** **High** **NULL**: Non-calculable
IPI_Score	int		IPI prognostic score, using both CT and PET data, combined. **0** to **5** **NULL**: Non-calculable
FLIPI_Score	int	Follicular lymphoma	Follicular lymphoma IPI prognostic score, only given for those with complete data **0** to **5** **NULL**: Non-calculable
FLIPI	varchar(50)		FLIPI risk category based on FLIPI_Score **Low** **Intermediate** **High** **NULL**: Non-calculable
HIPI_Score	int	Hodgkin lymphoma	HIPI prognostic score, only given for those with complete data **0** to **7** **NULL**: Non-calculable
HIPI_Cat_2	varchar(50)		HIPI category based on HIPI_Score **Low (0** to **2)** **Moderate/High (3** to **7)** **NULL**: Non-calculable

Variable	Date type	Diagnostic group applicable to (where relevant)	Notes/explanation
HIPI_Cat_4	varchar(50)		HIPI category based on HIPI_Score , where marrow involvement does not contribute. **Low (0** to **1)** **Low-intermediate (2)** **High-intermediate (3)** **High (4** to **7)** **NULL**: Non-calculable
MDS_IPSS_Score	float	Myelodysplastic	MDS IPSS prognostic score, only given for those with complete data **0**, **0.5**, **1**, **1.5**, **2**, **2.5**, **3** **NULL**: Non-calculable
MDS_IPSS_Risk	nvarchar(14)		MDS IPSS risk category based on MDS_IPSS_Score **Low** **Intermediate-1** **Intermediate-2** **High** **NULL**: Non-calculable
Myeloma_ISS_Stage	nvarchar(3)	Myeloma	International staging system for myeloma **I**, **II**, **III** **NULL**: Non-calculable

1.3.4 data.Presentation_Molecular

Presentation data specifically relating to molecular measurements, linkage to subjects by EGU_ID and DiagNumber.

Variable	Data type	Definition
EGU_ID	nvarchar(10) PRIMARY KEY	
Diagnosis	nvarchar(510)	
DiagNumber	tinyint	Chronological ordinal of the diagnosis
flt3	varchar(10)	flt3 internal tandem duplication status
t_flt3_days	int	Number of day after diagnosis that flt3 test was performed
npm1	varchar(10)	nmp1 mutation status
t_npm1_days	int	Number of day after diagnosis that test was performed