Counting Haematological Malignancies and Haematological Malignancy Transformations

An NDRS Standard Operating Procedure (SOP)

2023-05-17

Version 2.0
About the NDRS

The National Disease Registration Service (NDRS) is part of NHS England. Its purpose is to collect, collate and analyse data on patients with cancer, congenital anomalies, and rare diseases. It provides robust surveillance to monitor and detect changes in health and disease in the population. NDRS is a vital resource that helps researchers, healthcare professionals and policy makers make decisions about NHS services and the treatments people receive.

The NDRS includes:
- the National Cancer Registration and Analysis Service (NCRAS) and
- the National Congenital Anomaly and Rare Disease Registration Service (NCARDRS)

Healthcare professionals, researchers and policy makers use data to better understand population health and disease. The data is provided by patients and collected by the NHS as part of their care and support. The NDRS uses the data to help:
- understand cancer, rare diseases, and congenital anomalies
- improve diagnosis
- plan NHS services
- improve treatment
- evaluate policy
- improve genetic counselling

National Disease Registration Service
The Leeds Government Hub
7&8 Wellington Place
Leeds
LS1 4AP

For queries relating to this document, please contact:
NDRSerquries@nhs.net
Contents

About the NDRS 1
Contents 2
1. Summary of how to count haematological malignancies 3
2. An aside on terminology 4
3. Defining haematological malignancies 5
4. Transformations 7
Appendix A: Code to count primary haematological malignancies 11
Appendix B: Code to count haematological transformations 13
Appendix C: Code to create the table of haematological transformations - AT_TRANSFORMATION_HAEM table 14
Appendix D: ICD O3 rev 2011 codes to define haematological malignancies – MORPH_HAEM table 30
Appendix E: Codes to define haematological transformations according to the Gavin paper 30
1. Summary of how to count haematological malignancies

In general for most cancers, when counting the incidence of cancer we count the number of rows in AT_TUMOUR. Usually one new diagnosis of cancer is the same as one new primary tumour and is also the same as one row in the table.

This does not currently work for haematological malignancies because of the existence of transformations. ‘What are transformations and why is our data confused around them’ is explained in more detail in Section 4, but broadly some patients have haematological malignancies that are initially diagnosed as one type of malignancy but then develop over time into a nastier / more aggressive malignancy. These subsequent malignancies are not an independent new malignancy, but ‘transformations’ of the earlier malignancy. In our data some rows in AT_TUMOUR are primary haematological malignancies, and some are haematological transformations. If you just count the rows you inflate the count by counting the transformations as well as the true primary malignancies.

To count the number of new primary haematological malignancies (which is the number you usually want for incidence rates and similar) you need to take all the haematological malignancies in AT_TUMOUR, but exclude the rows that are transformations. Code to do this is given in Appendix A. ‘How to define a haematological malignancy’ is explained in more detail in Section 3.

To count the number of haematological transformations you need to use the table, AT_TRANSFORMATION_HAEM. One row in this table is one transformation and it can be counted straightforwardly. Code to do this is given in Appendix B. The code that creates the table, including comments that explain how it works, is given in Appendix C.

We hope to be able to work with the CAS team to take the secondary tumours resulting from a transformation out of AT_TUMOUR, which will make counting primary haematological tumours easier again, but until this is done this SOP should be followed.

This problem is largest for years prior to 2018 (when a change in how we record transformations happened, see section 4 for more details). Following this SOP correctly makes a difference of approximately 3% to the total number of primary haematological malignancies for 2017 (there are 36,003 haematological malignancies recorded from CAS counts, 34,914 excluding transformations). The main differences are for Acute myeloid leukaemia (AML) (2,862 from CAS counts, 2,246 excluding transformations, a reduction of 27%) and Diffuse large B-cell lymphoma (DLBCL) (4,918 from CAS counts, 4,564 excluding transformations, a reduction of 8%) [Numbers run on CAS2210.]

The purpose of this SOP is to provide a method of counting primary haematological malignancies, and haematological malignancy transformations separately for diagnoses.
from 2013. We believe the transformations table AT_TRANSFORMATION_HAEM is working well for identifying transformations from 2013 onwards, and have published 2013-2020 numbers in the Get Data Out publication. If you want to count haematological malignancies for years before the creation of a single NDRS (i.e., pre 2013) we recommend you talk to the haematology lead.

2. An aside on terminology

This SOP talks about ‘haematological malignancies’. The term ‘blood cancers’ is often commonly used by organisations such as Cancer Research UK. The National Statistics talk about ‘new diagnoses of cancer’ or ‘incidence of cancer’ and many of the other NDRS documentation talks about ‘new primary tumours’.

‘Tumour’ v’s ‘Cancer’ v’s ‘Malignancy’

Most cancers are solid tumours. This is not the case for haematological cancers. Cancers like leukaemia are characterised by an abnormally large number of white blood cells, but they do not form a solid lump or growth like a breast tumour or a lung tumour. Therefore referring to ‘tumours’ when talking about haematological malignancies can be misleading.

Some of the conditions included in ‘haematological malignancies’ are behaviour 1 in ICD O3 rev 2011 and may not be generally referred to as ‘cancer’ (e.g., mastocytomas)

In general ‘malignancies’ can be a more accurate term than ‘tumours’ or ‘cancers’ for these conditions.

‘Haematological’ v’s ‘blood’

The two terms can be used fairly interchangeably, but ‘haematological’ is preferred in NDRS work. Some of the conditions do form solid tumours (e.g., lymphomas) – although these are a type of blood cancer, some people use ‘blood cancer’ to mean cancer that does not form a solid tumour, so ‘haematological’ is felt to be clearer.

________________________

3. Defining haematological malignancies

After working with clinicians and coding experts, a standard list of morphology and behaviour codes that define ‘haematological malignancies’ has been produced.

This is a list of 157 pairs of codes (a morphology code and a behaviour code, making a unique pair) coded in ICD O3 rev 2011. These codes can be found in the table ANALYSIS\_SHANHUALIU.GDO\_MORPH\_HAEM@CASREF01 (which we expect to move to GDO.MORPH\_HAEM@CASREF01 soon). They are repeated in Appendix D.

These codes are in agreement with the WHO definitions of haematological malignancies.

To find a cohort of all haematological malignancies, take all the registered malignancies from the CAS, apply the usual counting cancer cases SOP rules for age, gender, geography etc, and inner join to this table on morphology and behaviour so only the haematological ones remain. Do not limit your cohort to only malignancies that are C codes in ICD 10, as some of the haematological malignancies are D codes (and a few are E and L codes!).

This method gives very similar answers to the old method of defining haematological malignancies using ICD 10 codes. This used to define All Blood Cancer as ICD 10 C81-C85, C88, C90-C96, D45-D47.

Comparing these definitions we see that the differences are very small (numbers run for 2019 on CAS2210 as the most recent finalised year not affected by the pandemic).
The new definition remembers to include C86, which only exists as a code in ICD 10 rev 4. It also picks up about 100 cancers that do not get C or D codes in ICD 10, but are haematological malignancies – these are immunoglobulin deposition disease. The new definition excludes 18 poorly coded tumours, which have the very generic morphology code 8000 for ‘neoplasm’ and so we do not truly know what they are, and also excludes 10 Monoclonal gammopathy of undetermined significance (MGUS; ICD O3 rev 2011 9765/1) tumours (which are not registrable conditions and should not be on the cancer registry at all).

The table ANALYSISHANHUALIU.GDO_MORPH_HAEM@CASREF01 also puts these codes into smaller meaningful groups for analysis of different types of haematological malignancies such as ‘Acute myeloid leukaemia (AML)’ and ‘Hodgkin Lymphoma’. A more detailed description of what these groups are can be found in the document ‘Haematological Malignancy Descriptions.docx’. These groupings should be the standard groupings used for any sub-analysis of haematological malignancies.

Once this cohort has been identified by ICD O3 rev 2011 morphology and behaviour codes, you will need to remove transformations to get a clean cohort of primary haematological cases. (See Section 4 and the code in Appendix A).

The code in Appendix A shows how to identify and count all haematological malignancies using this table, and also how to put them into the main groups of sub-types and remove transformations.
4. Transformations

Why is there a SOP on how to count haematological malignancies? The main answer is because of ‘transformations’ and because of a change that happened in registration in 2018.

Generally, for all cancers, cancer registration tries to register the primary tumour. For example if a patient has a lung cancer, and it sadly spreads to their brain, they are not registered with a lung cancer and a brain cancer, because the tumour in their brain is not a primary brain cancer. They are registered with one lung cancer, as that was the primary cancer. Similarly, if a patient has a breast cancer, and it is treated but sadly grows back five years later, the patient is registered with one breast cancer, not two breast cancers. Most relevantly, if a patient was diagnosed with a stage 1 cancer, and despite treatment it kept growing and became a stage 3 cancer, we would not make two registrations, one for the stage 1 cancer and one for the stage 3 cancer, because it would be the same cancer.

This causes problems with haematology, where malignancies can ‘get worse’ and get new names, but are still broadly the same malignancy. The most common example of this is Myelodysplastic syndrome (MDS) transforming to Acute myeloid leukaemia (AML). These are both diseases where there is a problem with immature blood cells, ‘blasts’, in the bone marrow. AML is generally diagnosed when there is involvement of more than 20% of the blood and/or bone marrow by leukaemic myeloblasts. Not all MDS transforms to AML, and not all AML is preceded by a diagnosis of MDS, but sometimes MDS can ‘get worse’ and it can transform to AML. In this case it makes more sense to think of this as the earlier MDS diagnosis transforming / progressing than it does to think of the patient as having two independent unrelated malignancies.

Because of this we changed the way the registry records haematological transformations in 2018. For diagnosis years before 2018, if a patient had MDS and AML, we would make two registrations for two malignancies, as though they were separate unrelated malignancies.
In 2018, following European Network of Cancer Registry guidelines and an improved understanding of transformations, we stopped registering these cases as two separate malignancies. This means that for patients with a haematological malignancy transformation (e.g., MDS followed by AML) if two patients had an identical pair of malignancies, the number of tumours reported in AT_TUMOUR made a sudden change in 2018:

<table>
<thead>
<tr>
<th>First tumour</th>
<th>Second tumour</th>
<th>Number of rows in AT_TUMOUR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before 2018</td>
<td>Before 2018</td>
<td>2 rows</td>
</tr>
<tr>
<td>Before 2018</td>
<td>2018 onwards</td>
<td>1 row</td>
</tr>
<tr>
<td>2018 onwards</td>
<td>2018 onwards</td>
<td>1 row</td>
</tr>
</tbody>
</table>

You can see the effect of this if you do a naïve count of the number of cases of AML by year in the AT_TUMOUR table, there is a large drop between 2017 and 2018.

The transformation is still registered, but it is not registered as a new tumour – it is registered as a ‘transformation event’ on the primary tumour. This event is recorded in the SPRINGMVC3.TRANSFORMATION table.

---

To correct for this, we have attempted to identify all transformations in a unified table for analysts to use: AT_TRANSFORMATION_HAEM. We have done this in two ways:

- Any transformation event registered as a transformation event is included in AT_TRANSFORMATION_HAEM.
- For any pair of malignancies where the second malignancy appears to be a transformation from the first malignancy (e.g., AML following MDS), the second malignancy is included in the transformation table. The definition of which pairs of tumours count as a transformation is taken from the Gavin et al paper ‘Towards optimal clinical and epidemiological registration of haematological malignancies: Guidelines for recording progressions, transformations and multiple diagnoses’.

The second part of this requires knowing what both registered malignancies are. We have a definition of haematological transformations in both ICD O2 and ICD O3 rev 2011. This allows us to find pairs of malignancies that impute a transformation when the earlier malignancy has been registered any time from 1995 onwards. However, prior to 1995 our data is only coded in ICD 9. We do not have a look up for valid first malignancies (precursor conditions) in ICD 9, and so our table does not capture these transformations. This has two main effects on the table:

- Data for the years very near to the coding change will be very incomplete, and the majority of transformations will be missing. This is a major driver of why we do not recommend analysis of transformations before 2013.
- Transformations which have taken a very long time will be missing from our imputed transformations in later years. For example, if a malignancy first diagnosed in 1994 transforms in 2017, it is expected that this will be missing from the transformation table and miscounted as a new primary tumour. For 2018 onwards where transformations have been manually registered, whether or not transformations where the precursor condition was registered pre-1995 are recorded as a transformation will depend on how registration officers have recorded them, but good practice would now be to add these to the original primary tumour.

When building the AT_TRANSFORMATION_HAEM table, we discovered that for some patients the same transformation appeared to be recorded more than once. For example if there was a patient with MDS registered in 2014, and then AML registered in 2016 and AML registered again in 2017. Because of this we have added a flag to identify duplicates to the transformation table – the DEDUP flag. Transformations we believe are genuine are flagged ‘1’ and duplicates are flagged ‘0’. This mirrors the DEDUP flag in the main AT_TUMOUR table.

We could not remove the duplicate transformations from this table entirely, as this table is used to remove transformations from the primary malignancies count in the code in Appendix A. If the duplicate transformations were removed from this table, the code to
count primary malignancies would not take them out, and so they would reappear inaccurately as primary malignancies.

The majority of transformations from 2018 are identified using the transformation event table. Because registering transformations is new and difficult and a manual process, currently there appear to be some cases in 2018 onwards where two tumours have been registered, but only one tumour should have been registered according to the rules. We are sending these to the QA team to review and fix, but we are also continuing to remove them as transformations in the code for the time being, as after discussions with Brian Rous this was felt to make the most accurate time series. If this was not the case, we would only need to apply the code to remove transformations for the years 2013-2017, but currently we are applying it for all years 2013 onwards.

If the transformations in the transformation table are removed from the count of AML (i.e., following the code in Appendix A) the timeseries is much steadier.

![Incidence of AML by year](image)

Full details of how the transformation table was built and the assumptions made can be found in the code in Appendix C.
Appendix A: Code to count primary haematological malignancies

Please note that at the time this SOP was written, some tables were still in personal tablespaces on the CAS. We will aim to update this SOP when these tables are moved, but if this code does not run, this would be the first thing I would check. We expect the following moves:

- The definition of haematological malignancies
  ANALYSISISHANHUALIU.GDO_MORPH_HAEM@CASREF01 to move from Hanhua’s space to the GDO space and become GDO.MORPH_HAEM@CASREF01
- The transformations table
  ANALYSISISPOLLYJEFFREY.AT_TRANSFORMATION_HAEM@CAS2210 to move from Polly’s space to the general analysis space
  ANALYSISNCR.AT_TRANSFORMATION_HAEM@CAS2210 or AV2020.AT_TRANSFORMATION_HAEM@CASREF01

Code

This is sample code to count the number of primary haematological malignancies in 2017 based on CAS2210 / av2020.

The main features are:

- an ‘inner join’ to define the haematological malignancies by the MORPH_HAEM table
- a ‘not exists’ to take out the transformations defined by the AT_TRANSFORMATION_HAEM table.

```sql
SELECT count(*)
FROM av2020.at_tumour_england@casref01 avt
INNER JOIN analysishanhuialiu.gdo_morph_haem@casref01 hl ON avt.morph_icdo3rev2011=hl.morph_icdo3rev2011 AND avt.behaviour_icdo3rev2011=hl.behaviour --using the haem lookup
WHERE avt.diagnosisyear >= 2013 -- restrict to diagnoses of 2013 onwards
AND not exists (
    SELECT ht.transformed_tumourid
    FROM analysispollyjeffrey.at_transformation_haem@cas2210 ht
    WHERE avt.tumourid=ht.transformed_tumourid
)```
SOP – Counting Haematological Malignancies

--below are CAS_SOP_Counting_cancer_cases.pdf rules
and avt.ctry_code = 'E' -- England residents using country code
and avt.statusofregistration = 'F' -- Finalised cases
and avt.dedup_flag=1 -- Excluding duplicates
and avt.age between 0 and 200 -- Sensible age: age between 0 and 200
and avt.sex in (1,2) -- Known sex
and diagnosisyear = 2017 -- restrict to diagnoses of 2017
;

The code can be tweaked for different years/different snapshots.

If you want to count more granular groups than ‘all haematological malignancies’ the ‘SPLIT’ columns can be used in the MORPH_HAEM table.

select
nvl(split_4, nvl (split_3, nvl (split_2, split_1))) as haem_type,
count(*)
from av2020.at_tumour_england@casref01 avt
inner join analysishanhuai.lu.gdo_morph_haem@casref01 hl on
avt.morph_icdo3rev2011=hl.morph_icdo3rev2011 and
avt.behaviour_icdo3rev2011=hl.behaviour -- using the haem lookup
where
avt.diagnosisyear >= 2013 -- restrict to diagnoses of 2013 onwards
-- excluding transformation events
and not exists (select
ht.transformed_tumourid
from analysispollyjeffrey.at_transformation_haem@cas2210 ht
where
avt.tumourid=ht.transformed_tumourid)

--below are CAS_SOP_Counting_cancer_cases.pdf rules
and avt.ctry_code = 'E' -- England residents using country code
and avt.statusofregistration = 'F' -- Finalised cases
and avt.dedup_flag=1 -- Excluding duplicates
and avt.age between 0 and 200 -- Sensible age: age between 0 and 200
and avt.sex in (1,2) -- Known sex
and diagnosisyear = 2017 -- restrict to diagnoses of 2017
group by nvl(split_4, nvl (split_3, nvl (split_2, split_1)))
order by count(*) desc
;
Appendix B: Code to count haematological transformations

Please note that at the time this SOP was written, some tables were still in personal tablespaces on the CAS. We will aim to update this SOP when these tables are moved, but if this code does not run, this would be the first thing I would check. We expect the transformations table ANALYSISPOLLYJEFFREY.AT_TRANSFORMATION_HAEM@CAS2210 to move from Polly’s space to the general analysis space ANALYSISNCR.AT_TRANSFORMATION_HAEM@CAS2210 or AV2020.AT_TRANSFORMATION_HAEM@CASREF01

Also note that the haematological transformations table was designed to produce a cohort of English transformations. This means that currently this code has no restriction on geography (and the transformation table has no geography field). If we were to extend the transformations table to Welsh cases we would need to review how to identify these in the table.

Code

This is sample code to count the number of haematological transformations in 2017 based on CAS2210 / av2020.

```sql
select count(*)
from analysispollyjeffrey.at_transformation_haem@cas2210
where transformation_year = 2017 -- restrict to transformations of 2017
and dedup_flag=1 -- Excluding duplicates
;
```

The main features are

- A dedup flag that removes duplicate transformations
- The implicit knowledge that all transformations are for English residents

The code can be tweaked for different years/different snapshots.

If you want to count more granular groups than ‘all haematological transformations’ the ‘TRANSFORMATION_SPLIT’ columns can be used. Transformation split 1 is what the haematological malignancy has transformed into (secondary malignancy), transformation split 2 is what the haematological malignancy transformed from (primary/precursor malignancy).
select transformation_split_1, transformation_split_2, count(*)
from analysispollyjeffrey.at_transformation_haem@cas2210
where
transformation_year = 2017 -- restrict to transformations of 2017
and dedup_flag=1 -- Excluding duplicates
group by transformation_split_1, transformation_split_2
order by transformation_split_1, transformation_split_2
;

Appendix C: Code to create the table of haematological transformations - AT_TRANSFORMATION_HAEM table

Generally we recommend you just use the table of transformations and do not try and rebuild it from scratch unless you really need to. It was developed by Sally Vernon, Andrew Bacon, Charlotte Eversfield and Hanhua Liu, so if there are things in it you want to understand and don’t, please feel free to ask them. The code is included here for transparency and in case someone needs to understand it in more detail in the future.

The code currently lives in the analysis svn in Analytical code repository\Haematology\AT_TRANSFORMATION_HAEM although we expect it to move into the standard CAS build soon.

There are comments in the code which should explain what it is doing and the assumptions it is making.

/*-----------------------------------------------------------------------------------------------
at_transformation_haem--------------------------------------------------------------------------------
Creating a standard haematological transformations table for analysis
By imputing transformations from 1995 onwards using the Gavin paper and counting transformations registered in the transformation table by NDRS
--------------------------------------------------------------------------------
Development:
Tables were developed by Andrew Bacon, with clinical and pathological input from Mark Bishton and Brian Rous. QA was performed by Sally Vernon.

This code is designed to produce a transformation table on any standard CAS snapshot. Table was developed on CAS2109 and CAS2210. Producing this table on end of year snapshots is important for correcting reporting incidence of haematological malignancies.

---
Overview
---
Motivation for this table

To date there is no uniform definition of the transformation of haematological malignancies (Okosun et al., 2018). However, the basic principle of a transformation is that a tumour cell, subject to a complex combination of genetic mutations, can undergo a process of clonal evolution that profoundly changes the architecture of the cell and results in a transformation to a different tumour type (Gavin et al., 2015).
The key question is whether this transformed tumour is a new primary tumour or a natural progression of the earlier tumour. If it is a new tumour, then two tumour registrations would be made. If it is a progression of the earlier tumour, then only one tumour registration would be made.

Prior to 2018 two tumour registrations were always made. From 2018 onwards, best practice is to make one tumour registration and one transformation registration:

Changes to COSD rules and recording transformation events:
Following consultation with Laboratory Information Management System Suppliers and the Chair of the Royal College of Pathologists (Working Group on Cancer Services), alterations were made to the structure and contents of the COSD dataset. Changes included the recording of transformation events, to conform to the recommendations established by the European Network of Cancer Registration, published in February 2014.

Previously, the national cancer registration and analysis service recorded a transformation event as a new primary malignancy. As part of COSD v9.0, due for release in April 2020, with full conformance expected by the 1st of July; the option now exists to record a transformation event, in addition to recording a relapse or new tumour. A transformation event can be recorded if the haematological malignancy transforms into a new morphological entity (different diagnostic group) three-months after the initial registration.

This change in practice causes a noticeable reduction in some primary tumours, particularly AML (20% reduction in registered AML between 2017 and 2018)
In order to provide reliable haematological time series cases that would now be registered as transformations must be identified so they can be removed from the primary tumour time series

This code produces a table of transformations and likely transformations. Transformations can be counted annually and should be removed from the primary tumour cohort.

---
Top level guide to the code
---

Transformations are identified in one of two ways:
1) They are imputed from a pair of primary tumours
2) They are extracted from the ENCORE Transformations table

To do (1), we need to agree which pairs of primary tumours suggest a transformation

For example, MDS followed by AML

We want to follow the transformation rules specified in the National Cancer Registration and Analysis Service (NCRAS) Haematology guidance as much as possible to do this

These rules come from the Gavin et al., (2015) paper. We have coded the Gavin et al. paper up into a table:

```
analysisandrewbacon.gdo_transformation_haem@casref01
```

If a single patient has a pair of tumours defined as a transformation by the Gavin paper. The first one is assumed to be the primary tumour and the second is included in the transformation table.

The Gavin paper defines transformations in ICD O3. ICD O3 codes were only used consistently from 2013 onwards. Given that there is no defined time to transformation (could be a month, could be 20+ years) we need to extend our current ICD-0-3 methodology to ICD-0-2.

---
Overview of tables used / created
---

Table 1:
```
analysishanhuailiu.gdo_morph_haem@casref01
```

Table defining which cases count as a haematological cancer, based on icd03rev2011. Any case with morphology and behaviour in this table is classified as a haem malignancy. The table also puts all haematological cases into subtypes according to cell lineage.
Table 2: analysisandrewbacon.gdo_transformation_haem@casref01

Table defining when a registered tumour is counted as a precursor condition / transformation if a patient has multiple tumours registered before 2018 based on Gavin et al. paper.

Table 3: analysisandrewbacon.at_transformation_haem

The output of this script, with one row per transformation

```sql
DROP TABLE at_transformation_haem PURGE;

CREATE TABLE at_transformation_haem
(
  transformationid VARCHAR2(30) NOT NULL,
  patientid VARCHAR2(10) NOT NULL,
  precursor_tumourid VARCHAR2(10) NOT NULL,
  precursor_diagnosis_date DATE,
  precursor_year NUMBER(4),
  precursor_morphology VARCHAR2(4),
  precursor_behaviour VARCHAR2(1),
  precursor_description VARCHAR2(256),
  precursor_subtype VARCHAR2(256),
  precursor_broad_subtype VARCHAR2(256),
  precursor_lineage VARCHAR2(70),
  transformed_tumourid VARCHAR2(10) NOT NULL,
  transformation_date DATE,
  transformation_year NUMBER(4),
  transformation_morphology VARCHAR2(4),
  transformation_behaviour VARCHAR2(1),
  transformation_description VARCHAR2(256),
  transformation_subtype VARCHAR2(256),
  transformation_broad_subtype VARCHAR2(256),
  transformation_lineage VARCHAR2(70),
  trans_source VARCHAR2(70),
  gavin_flag VARCHAR2(70),
  dedup_flag VARCHAR2(70),
  transformation_split_1 VARCHAR2(256),
  transformation_split_2 VARCHAR2(256)
) NOLOGGING COMPRESS BASIC;
```

Inserting transformations using the Gavin paper.

Section 1a: Finding precursor conditions. Any valid ICD-O-3 precursor code, of which there are 44 at the moment.

```sql
INSERT /*+ append */ INTO at_transformation_haem
SELECT transformationid,
  patientid,
  precursor_tumourid,
  precursor_diagnosis_date,
  precursor_year,
  precursor_morphology,
  precursor_behaviour,
  precursor_description,
  precursor_subtype,
  precursor_broad_subtype,
  precursor_lineage,
  transformed_tumourid,
  transformation_date,
  transformation_year,
  transformation_morphology,
  transformation_behaviour,
  transformation_description,
  transformation_subtype,
  transformation_broad_subtype,
  transformation_lineage,
  trans_source,
  gavin_flag,
  dedup_flag,
  transformation_split_1,
  transformation_split_2
FROM analysisandrewbacon.gdo_transformation_haem@casref01
```

Guidance:

Select all of the verified ICD-O-2 codes for precursor conditions. There are 34 of these codes at the moment (o2 is less granular than o3).
FROM analysisandrewbacon.gdo_transformation_haem@casref01

-- Guidance:
-- Select all of the verified ICD-O-3 codes for precursor conditions.

-- Notes:
-- Unique pairs of ICD-O-3 codes that define a transformation event. Both the
-- precursor conditions and transformations are recorded in ICD-O-3. To date
-- there are 771 possibilities. This is basically the heart of how we impute
-- transformations, everything else is simply about aligning classification
-- systems.

, gdo_transformations_o3_o3 AS
(SELECT DISTINCT precursor_morph_o3
, precursor_behaviour_o3
, precursor_description_o3
, transformation_morph_o3
, transformation_behaviour_o3
, transformation_description_o3
FROM analysisandrewbacon.gdo_transformation_haem@casref01
)

-- Guidance:
-- The section of code below, identifies unique pairs of codes that define a
-- transformation across multiple classifications (ICD-O-2; ICD-O-3). There are
-- 522 possibilities.

-- Notes:
-- There are two ICD-O-3 transformations where their ICD-O-2 mapping has been
-- deliberately excluded, as the O-3 code is very specific and the O-2 is very
-- vague. These are:
-- ICD-O-3 9817/3 B lymphoblastic leukaemia/lymphoma with t(5;14)(q31;q32); IL3-IGH is o2 9821/3
-- Acute lymphoblastic leukaemia NOS
-- ICD-O-3 9898/3 Myeloid leukemia associated with Down Syndrome in o2 is sometimes 9960/1
-- Chronic myeloproliferative disease

-- It is also the forward mapping from O-2 to O-3
-- There is also an o2 transformation where the mapping to o3 depends on the
-- precursor condition! o2 9685/3 Malignant lymphoma, lymphoblastic - the
-- precursor condition can show it is a B cell or not. So o3 can be 9728/3
-- Precursor B-cell lymphoblastic lymphoma or 9727/3 Precursor cell
-- lymphoblastic lymphoma, NOS. This is why we can't do a transformation lookup
-- without the precursor conditions.

, gdo_transformations_o3_o2 AS
(SELECT DISTINCT precursor_morph_o3
, precursor_behaviour_o3
, precursor_description_o3
, transformation_morph_o2
, transformation_behaviour_o2
, transformation_description_o2
, transformation_morph_o3
, transformation_behaviour_o3
, transformation_description_o3
FROM analysisandrewbacon.gdo_transformation_haem@casref01
WHERE transform_o2_o3_mapping_flag = 1
)

-- Notes:
-- gdo_transformation_haem includes a flag (precursor_o2_o3_mapping_flag = 1)
-- that allows us to map O-2 definitions to O-3. These mappings were devised
-- using the mega mapping table. There are 11 precursor conditions in ICD-O-2
-- that have been mapped to a preferred ICD-O-3 code.

, precursor_mapping AS
(SELECT DISTINCT precursor_morph_o2
, precursor_behaviour_o2
, precursor_description_o2
, precursor_morph_o3
, precursor_behaviour_o3
, precursor_description_o3
FROM analysisandrewbacon.gdo_transformation_haem@casref01
WHERE precursor_o2_o3_mapping_flag = 1
)

-- Guidance:
-- ICD-O-2 was used by the registries between 1995-2012. Tumours registered from
SOP – Counting Haematological Malignancies

-- 2013 are captured using ICD-O-3.
-- Notes:
-- ICD-O-3 definitions for transformations have been added, despite being
-- outside of the GDO date range because we need to correctly map
-- transformations over the years - a transformation in 2015 could have a

-- The following table includes all patients diagnosed with a precursor
-- condition between 1995-2012, as defined in gdo_transformation_haem. It also
-- includes January 2013 to allow one month of follow-up for transformation
-- recorded in 2012. This is not needed for GDO but is useful in general.

-- The subtyping and cell lineage of each precursor condition is listed in
-- analyseishanhualliu.gdo_morph_haem. Here our output will consist of one row per
-- tumour / precursor condition. There were 247,153 rows when run on CAS2109 and
-- 247,421 when run on CAS2210. Which is approximately 12,000 precursor
-- conditions a year. However, it is worth stating that there are around 9,000
-- cases in 1995 compared to 19,000 in 2012. There are a number of factors that
-- will have contributed to the dramatic rise in transformations, but mainly
-- it is because of:
-- 1. Introduction of electronic records
-- 2. Merger of 8 regional registries, to one unified registry, with unified
-- standards for data collection.

, o2_precursor_conditions AS
(SELECT /*+ use_hash (t tr)*/
  t.patientid AS precursor_patientid,
  t.tumourid AS precursor_tumourid,
  t.diagnosisdatebest AS precursor_diagnosis_date,
  t.diagnosisyear AS precursor_year,
  p.precursor_morph_o3 AS precursor_morphology,
  p.precursor_behaviour_o3 AS precursor_behaviour,
  p.precursor_description_o3 AS precursor_description,
  hlp.split_3 AS precursor_subtype,
  hlp.split_2 AS precursor_broad_subtype,
  hlp.split_1 AS precursor_lineage
FROM analysisncr.at_tumour_england t
INNER JOIN gdo_precursors_o2 tr ON t.morph_icd10_o2 = tr.precursor_morph_o2
AND t.behaviour_icd10_o2 = tr.precursor_behaviour_o2
LEFT OUTER JOIN precursor_mapping p ON p.precursor_morph_o2 = tr.precursor_morph_o2
AND p.precursor_behaviour_o2 = tr.precursor_behaviour_o2
LEFT OUTER JOIN analysishanhualliu.gdo_morph_haem@casref01 hlp ON p.precursor_morph_o3 = hlp.morph_icdo3rev2011
WHERE dedup_flag = 1
-- Mandatory restrictions
AND age BETWEEN 0 AND 200
AND statusofregistration = 'F'
-- Finalised Registrations
AND diagnosisdatebest BETWEEN TO_DATE('1995-01-01', 'YYYY-MM-DD') AND TO_DATE('2013-01-31', 'YYYY-MM-DD')
-- ICD-O3 implemented in 2013. Allowing for 31 days in Jan for patients
with synchronous presentation
)

-- Guidance:
-- The table below includes descriptions of all precursor conditions described
-- in the Cancer Analysis System (CAS) from 2013 onwards, as defined by Gavin et
-- al. This table lists each precursor condition on a single row. There were
-- 154,092 rows when run on CAS2109 and 173,115 when the table is run on
-- CAS2210. Which is approximately 20,000 precursor conditions a year. Counts of
-- precursor conditions remain pretty steady over the years, with exception to
-- 2020 which is likely to have been impacted by COVID-19.

, o3_precursor_conditions AS
(SELECT /*+ use_hash (t tr)*/
  t.patientid AS precursor_patientid,
  t.tumourid AS precursor_tumourid,
  t.diagnosisdatebest AS precursor_diagnosis_date,
  t.diagnosisyear AS precursor_year,
  tr.precursor_morph_o3 AS precursor_morphology,
  tr.precursor_behaviour_o3 AS precursor_behaviour,
  tr.precursor_description_o3 AS precursor_description,
  hlp.split_3 AS precursor_subtype,
  hlp.split_2 AS precursor_broad_subtype
FROM analysisncr.at_tumour_england t
INNER JOIN gdo_precursors_o2 tr ON t.morph_icd10_o2 = tr.precursor_morph_o2
AND t.behaviour_icd10_o2 = tr.precursor_behaviour_o2
LEFT OUTER JOIN precursor_mapping p ON p.precursor_morph_o2 = tr.precursor_morph_o2
AND p.precursor_behaviour_o2 = tr.precursor_behaviour_o2
LEFT OUTER JOIN analysishanhualliu.gdo_morph_haem@casref01 hlp ON p.precursor_morph_o3 = hlp.morph_icdo3rev2011
WHERE dedup_flag = 1
-- Mandatory restrictions
AND age BETWEEN 0 AND 200
AND statusofregistration = 'F'
-- Finalised Registrations
AND diagnosisdatebest BETWEEN TO_DATE('1995-01-01', 'YYYY-MM-DD') AND TO_DATE('2013-01-31', 'YYYY-MM-DD')
-- ICD-O3 implemented in 2013. Allowing for 31 days in Jan for patients
with synchronous presentation
)
SOP – Counting Haematological Malignancies

-- Most recent annual snapshot was chosen to help capture any registry creep
INNER JOIN gdo_precursors_o3 tr ON t.morph_icdo3rev2011 = tr.precursor_morph_o3
-- Match on morphology (ICD-O-3)
AND t.behaviour_icdo3rev2011 = tr.precursor_behaviour_o3
-- Match on behaviour for those cases where the behaviour distinguishes the type of disease
LEFT OUTER JOIN analysishanhualiu.gdo_morph_haem@casref01 hlp ON tr.precursor_morph_o3 = hlp.morph_icdo3rev2011
-- AND tr.precursor_behaviour_o3 = hlp.behaviour
WHERE dedup_flag = 1
AND age BETWEEN 0 AND 200
AND statusofregistration = 'F'
AND diagnosisdatebest >= TO_DATE('2013-01-01', 'YYYY-MM-DD')
-- ICD03 implemented in 2013
-- AND ctry_code = 'E'
-- No restriction on England. If we know they had a precursor condition we don't care if it was in England or some other country as long as the transformation was recorded in England.

-- Section 1b: Finding transformations

-- Three tables to do this:

-- O2-O2 for when both precursor condition and transformation are 1995-2012
-- O2-O3 for when precursor is 1995-2012 and transformation is 2013 onwards
-- O3-O3 for when both precursor condition and transformation are 2013 onwards

-- Guidance:
-- The following table is used to identify all transformation events recorded in ICD-O-2, where both the precursor condition and the transformation are recorded between 01/01/1995 and 31/12/2012 as defined by gdo_transformation_haem.
-- Diagnosis dates are recorded according to either the date the diagnosis was agreed at MDT; or, the day the diagnosis was signed off by the senior pathologist. We therefore routinely allow a 31 day window to account for flexibility in reporting. This sentiment has been adopted here too, whereby a precursor condition can be captured up to 31 days after the transformation.
-- The general rules is that a precursor condition should be before a transformation, but cannot be more than a month after and still be counted.
-- Multiple O3 lookups are avoided by using the precursor_o2_o3_mapping_flag = 1
-- One row is one unique pair of precursor tumourid and transformed tumourid.
-- There were 4,689 rows when run on CAS2109 (4,691 on CAS2210). This is about 300 transformations a year. Counts in 1995 are low because there is no data from the years before for precursor conditions. Counts rise to 561 a year in 2012.
-- The subtyping and lineage of the transformed neoplasm are determined using analysishanhualiu.gdo_morph_haem.

, o2_o2_transformations AS
(SELECT /*+ use_hash (pc t) use_hash (t tr)*/
 pc.patientid
 , pc.precursor_tumourid
 , pc.precursor_diagnosis_date
 , pc.precursor_year
 , pc.precursor_morphology
 , pc.precursor_behaviour
 , pc.precursor_description
 , pc.precursor_subtype
 , pc.precursor_broad_subtype
 , pc.precursor_lineage
 , t.tumourid AS transformed_tumourid
 , t.diagnosisdatebest AS transformation_date
 , t.diagnosisyear AS transformation_year
 , tr.transformation_morph_o3 AS transformation_morphology
 , tr.transformation_behaviour_o3 AS transformation_behaviour
 , tr.transformation_description_o3 AS transformation_description
 , hlt.split_3 AS transformation_subtype
 , hlt.split_2 AS transformation_broad_subtype
 , hlt.split_1 AS transformation_lineage
FROM o2_precursor_conditions pc
FROM analysisncr.at_tumour_england t
-- precursor_lineage AS
FROM analysishanhualiu.gdo_morph_haem@casref01
LEFT OUTER JOIN analysishanhualiu.gdo_morph_haem@casref01 hlp
ON tr.precursor_morph_o3 = hlp.morph_icdo3rev2011
AND tr.precursor_behaviour_o3 = hlp.behaviour
INNER JOIN analysisncr.at_tumour_england t ON pc.patientid = t.patientid  -- Matching on patient ID to ensure we pull all related tumours
INNER JOIN gdo_transformations_o3_o2 tr ON pc.precursor_morphology = tr.precursor_morph_o3
    AND pc.precursor_behaviour = tr.precursor_behaviour_o3  -- Joining converted o2 precursors on to o3 definitions of disease
    AND t.morph_icd10_o2 = tr.transformation_morph_o2
    AND t.behaviour_icd10_o2 = tr.transformation_behaviour_o2  -- Extra level of linkage needed to minimise cross matching
LEFT OUTER JOIN analysishanhualiu.gdo_morph_haem@casref01 hlt ON tr.transformation_morph_o3 = hlt.morph_icdo3rev2011
    AND tr.transformation_behaviour_o3 = hlt.behaviour
WHERE t.diagnosisdatebest - pc.precursor_diagnosis_date >= -31
    AND pc.precursor_tumourid != t.tumourid  -- Registration would capture some transformations AS new primaries, so lets make sure the tumourids dont match
    AND t.ctry_code = 'E'  -- NDRS
Mandatory restrictions:
    AND t.dedup_flag = 1
    AND t.statusofregistration = 'F'
    AND t.age BETWEEN 0 AND 200
    AND t.diagnosisyear BETWEEN 1995 AND 2012  -- ICD-0-2
wAS the principle coding system between 1995 and 2012
)

-- Notes:  
The following table includes all tumour-pairs that could be a transformation,  
-- where the precursor is recorded between 1995-2012 and the transformation  
-- reported from 2013 onwards, as defined by gdo_transformation_haem  
-- O3-precursors and O3-transformations. Please note that a precursor condition  
-- should be before a transformation, but we have allowed a 31 day window either  
-- side of the precursor diagnosis date to allow for differences in reporting.  

-- Subtyping and lineage are determined by analysishanhualiu.gdo_morph_haem.  
-- Multiple O3 lookups are avoided by using the precursor_o2_o3_mapping_flag = 1  
-- One row is one unique pair of precursor_tumourid and transformed_tumourid.  
-- There were 2,418 rows when run on CAS2109 (2,410 on CAS2210)  
-- This is about 400 transformations a year for 2013-2018 (662 in 2013, falling  
-- annually to 350 in 2017, as 2012 gets further away. There should be none 2018  
-- onwards, but we have 92,26,12 on CAS2210 which is an improvement on 92,41,13  
-- on CAS2109).

, o2_o3_transformations AS
(SELECT /*+ use_hash (pc t) use_hash (tr t)*/
    pc.patientid, pc.precursor_tumourid, pc.precursor_diagnosis_date, pc.precursor_year, pc.precursor_morphology, pc.precursor_behaviour, pc.precursor_description, pc.precursor_subtype, pc.precursor_broad_subtype, pc.precursor_lineage, t.tumourid AS transformed_tumourid, t.diagnosisdatebest AS transformation_date, t.diagnosisyear AS transformation_year, tr.transformation_morph_o3 AS transformation_morphology, tr.transformation_behaviour_o3 AS transformation_behaviour, tr.transformation_description_o3 AS transformation_description, hlt.split_3 AS transformation_subtype, hlt.split_2 AS transformation_broad_subtype, hlt.split_1 AS transformation_lineage FROM o2_precursor_conditions pc
INNER JOIN analysisncr.at_tumour_england t ON pc.patientid = t.patientid  -- Match on patient ID to get all related tumours
INNER JOIN gdo_transformations_o3_o3 tr ON pc.precursor_morphology = tr.precursor_morph_o3
    AND pc.precursor_behaviour = tr.precursor_behaviour_o3  -- Precursors were converted from o2, so matching on o3
    AND t.morph_icd3rev2011 = tr.transformation_morph_o3  -- Transformed in ICD-0-3
    AND t.behaviour_icd3rev2011 = tr.transformation_behaviour_o3  -- Extra level of linkage needed to minimise cross matching
LEFT OUTER JOIN analysishanhualiu.gdo_morph_haem@casref01 hlt ON tr.transformation_morph_o3 = hlt.morph_icd3rev2011
    AND tr.transformation_behaviour_o3 = hlt.behaviour
WHERE t.diagnosisdatebest - pc.precursor_diagnosis_date >= -31

20
SOP – Counting Haematological Malignancies

AND pc.precursor_tumourid != t.tumourid -- Lets make sure the tumourids don't match
AND t.ctry_code = 'E' -- Mandatory restrictions
AND t.dedup_flag = 1
AND t.statusofregistration = 'F'
AND t.age BETWEEN 0 AND 200
AND t.diagnosisyear >= 2013 -- Changes to ENCR guidelines means that transformations are recorded in a separate table from 2018
AND pc.precursor_year <= 2012 -- Take out the January 2013 tumours or would get same transformation in o2o3 and o3o3

-- The table below includes all tumour-pairs that could be a transformation, where the precursor and the transformation were diagnosed from 2013 onwards, as defined by gdo_transformation_haem O3-precursors and O3-transformations.

-- One row is one unique pair of precursor_tumourid and transformed_tumourid
-- There were 2,728 rows when run on CAS2109 (2,660 on CAS2210). This is about 500 transformations a year, although there are very few in 2013 (only 165, -- this because there isn't a long follow-up period). Counts rise a lot year on -- year to 755 in 2017 (also as expected). Then in the 2018-2020 when imputing -- transformations should no longer be required, we have 68/168/44 CAS2109 and -- 38/75/61 CAS2210. We hope that as registration staff become more familiar with the recording of transformations the need for imputation won't be -- necessary in later years.

, o3_o3_transformations AS
(SELECT /*+ use_hash (pc t) use_hash (t tr)*/
pc.patientid
, pc.precursor_tumourid
, pc.precursor_diagnosis_date
, pc.precursor_year
, pc.precursor_morphology
, pc.precursor_behaviour
, pc.precursor_descr
, pc.precursor_subtype
, pc.precursor_broad_subtype
, pc.precursor_lineage
, t.tumourid AS transformed_tumourid
, t.diagnosisdatebest AS transformation_date
, t.diagnosisyear AS transformation_year
, tr.transformation_morph_o3 AS transformation_morphology
, tr.transformation_behaviour_o3 AS transformation_behaviour
, tr.transformation_description_o3 AS transformation_description
, hlt.split_3 AS transformation_subtype
, hlt.split_2 AS transformation_broad_subtype
, hlt.split_1 AS transformation_lineage
FROM o3_precursor_conditions pc
INNER JOIN analysisncr.at_tumour_england t ON pc.patientid = t.patientid -- Match on patient to get all related tumours
INNER JOIN gdo_transformation_haem_o3_o3 tr ON pc.precursor_morphology = tr.precursor_morph_o3
AND pc.precursor_behaviour = tr.precursor_behaviour_o3
AND t.morph_icdo3rev2011 = tr.transformation_morph_o3
AND t.behaviour_icdo3rev2011 = tr.transformation_behaviour_o3
-- Extra level of linkage needed to ensure there is no cross matching
LEFT OUTER JOIN analysishanhualiu.gdo_morph_haem@casref01 hlt ON tr.transformation_morph_o3 = hlt.morph_icdo3rev2011
AND tr.transformation_behaviour_o3 = hlt.behaviour
WHERE t.diagnosisdatebest - pc.precursor_diagnosis_date >= -31
AND pc.precursor_tumourid != t.tumourid -- Registration would class transformations as new primaries, so lets make sure the tumourids dont match
AND t.ctry_code = 'E' -- Mandatory restrictions
AND t.dedup_flag = 1
AND t.statusofregistration = 'F'
AND t.age BETWEEN 0 AND 200
AND t.diagnosisyear >= 2013
)

-- European Network of Cancer Registries: Recommendations for Registration of Haematological Malignancies.
1. Multiple data sources such as blood, bone marrow, flow cytometry, molecular and cytogenetic tests from haematology and designated molecular laboratories in addition to histopathology, cytology, clinical records and death notifications should be used to register haematological malignancies (HM).

2. When additional information is received on the same HM patient the option exists to record as a transformation, same tumour or new tumour.
   a. Allocate as Transformation if first HM transforms into a new morphological entity (different diagnostic group) after a three month window of first registration:
      - Only the first tumour’s morphology and date of diagnosis to be considered as incident for analysis and reporting
      - The transformed tumour must not be counted as a new tumour and therefore not included in the incidence statistics.
   b. Allocate as same tumour with more specific /revised morphology,
      - if within same diagnostic group, date of diagnosis remains unchanged.
      - if a transformation occurs within three months after the incidence date, the morphology code of the transformed malignancy should replace that of the first tumour and be recorded as the first primary and not a transformation. Date of diagnosis remains unchanged as that of the first tumour.
   c. Allocate as a new tumour registration with new incident date when
      - HM with malignant behaviour (code /3) occurs after a previous haematological disease with uncertain behaviour (code /1) or
      - The change is not a transformation or a revised diagnosis of an existing tumour; or,
      - Clinical opinion regarding a new tumour is available and the detail of that decision is recorded.

3. Regular survival analysis methods do not necessarily apply in the case of patients with HM where transformations have occurred as the patient has to be alive until the diagnosis of multiple tumour or transformation occurs. The information of these changes may be used as time-dependent covariates. There are special methods for such multiple tumour analyses.

4. The ENCR recommendations for coding of incidence date should be followed.
5. Basis of diagnosis should follow the ENCR recommendations.
6. Record all dates and diagnoses of transformations in the registry.

---
This table has one row for each transformation in the transformations table.
-- It finds the details of the precursor condition from the tumour table and
-- puts this in to definitions of ICD03_rev2011 as much as possible. It finds
-- the details of the transformation from the transformations table and assumes
-- they are in ICD03_rev2011 (there are a few hacks to clean this).

-- There are 4,552 transformations on CAS2210 (2,973 on CAS2109). It is about
-- 1,031 in 2018, 1,223 in 2019, and 1,437 in 2020.

, reg_transformations_raw AS
  (SELECT t.patientid
   , t.tumourid          AS precursor_tumourid
   , t.diagnosisdatebest AS precursor_diagnosis_date
   , t.diagnosisyear     AS precursor_year
   FROM tumour t
   )

    , CASE WHEN morph_icdo3rev2011 IS NOT NULL THEN morph_icdo3rev2011
      WHEN p.precursor_morph_o2 IS NOT NULL THEN p.precursor_morph_o2
      WHEN morph_icd10_o2 IN ('9699','9689','9590','9702') AND behaviour_icd10_o2 = '3' THEN morph_icd10_o2
      WHEN morph_icd10_o2 IS NOT NULL THEN morph_icd10_o2
      WHEN coding_system_desc = 'ICD-0-2' THEN morph_coded
    AS precursor_morphology

    , CASE WHEN behaviour_icdo3rev2011 IS NOT NULL THEN behaviour_icdo3rev2011
      WHEN p.precursor_morph_o2 IS NOT NULL THEN p.precursor_morph_o3
      WHEN morph_icd10_o2 IN ('9699','9689','9590','9702') AND behaviour_icd10_o2 = '3' THEN behaviour_icd10_o2
      WHEN morph_icd10_o2 IS NOT NULL THEN morph_icd10_o2
      WHEN coding_system_desc = 'ICD-0-2' THEN morph_coded
    AS precursor_diagnosis
WHEN behaviour_icd10_o2 IS NOT NULL THEN behaviour_icd10_o2
WHEN coding_system_desc = 'ICD-0-2' THEN behaviour_coded
END AS precursor_behaviour

, CASE WHEN morph_icd10_2011 IS NOT NULL THEN 'ICD-0-3 (2011)'
WHEN p.precursor_morph_o2 IS NOT NULL THEN 'ICD-0-3 (2011)'
WHEN morph_icd10_2011 IN ('9699','9689','9590','9702') AND behaviour_icd10_o2 = '3' THEN 'ICD-0-3 (2011)'
WHEN morph_icd10_2011 IS NOT NULL THEN 'ICD-10/O-2'
WHEN coding_system_desc = 'ICD-0-2' THEN 'ICD-10/O-2'
END AS precursor_coding

transformationid                        AS transformed_tumourid
, transformation_date                     AS transformation_date
, EXTRACT(YEAR FROM(transformation_date)) AS transformation_year

FROM springmvc3.transformation tr
LEFT OUTER JOIN springmvc3.event e ON tr.transformationid = e.eventid
LEFT OUTER JOIN springmvc3.tumourevents te ON e.eventid = te.eventid
LEFT OUTER JOIN analysisncr.at_tumour_england t ON te.tumourid = t.tumourid
LEFT OUTER JOIN tumour t1 ON te.tumourid = t1.tumourid
LEFT OUTER JOIN precursor_mapping p ON p.precursor_morph_o2 = nvl(t.morph_icd10_o2, t.morph_coded)

AND p.precursor_behaviour_o2 = nvl(t.behaviour_icd10_o2, t.behaviour_coded)
WHERE t.statusofregistration = 'F'
AND t.ctry_code = 'E'
AND t.diagnosisyear >= 1995

Some patients are registered with multiple transformations
- For example they have a essential thrombocytemia which transforms
- to MDS which transforms to AML
- Talking with clinicians, we want the MDS to be the precursor to the AML
- Not the ET.
- So we need to take patients with more than one transformation
- and make earlier transformations into precursor conditions of later transformations
- There are 335 transformations on CASREF01 which are created by this method

, reg_transformations_multiples
as

SELECT
tr1.patientid as patientid
, tr2.transformed_tumourid as precursor_tumourid
, tr2.transformation_date as precursor_diagnosis_date
, tr2.transformation_year as precursor_year
FROM reg_transformations_raw tr1
left outer join reg_transformations_raw tr2
-- transformation is for the same patient
on tr1.patientid = tr2.patientid
-- transformation happened earlier than this transformation
and tr2.trans_rank < tr1.trans_rank
-- an earlier transformation actually exists
where tr2.patientid is not null
-- transformations on the same day are not precursors to each other
and tr1.transformation_date != tr2.transformation_date
)

-- We then need to add in the extra multiple transformations to the main
-- table of registration transformations:

, reg_transformations_raw_all
as
(
select
patientid
, precursor_tumourid
, precursor_diagnosis_date
, precursor_year
, precursor_morphology
, precursor_behaviour
, precursor_coding
, transformed_tumourid
, transformation_date
, transformation_year
, transformation_morphology
, transformation_behaviour
from reg_transformations_raw
UNION
select
patientid
, precursor_tumourid
, precursor_diagnosis_date
, precursor_year
, precursor_morphology
, precursor_behaviour
, precursor_coding
, transformed_tumourid
, transformation_date
, transformation_year
, transformation_morphology
, transformation_behaviour
from reg_transformations_multiples
)

-- Guidance:
-- Get rid of about 247 transformations that are not haem malignancis (brain
-- etc) by joining to the haem table. This takes us down to 4,305
-- transformations. Also take out 59 where the transformed morphology is the
-- same as the precursor morphology, which takes us down to 4,246
-- transformations.

-- Check they are all in TCD03_REV2011 at this point - they are on CAS2210.
-- 4,185 are transformations recognised in the Gavin paper (on CAS2210), 61 are
-- weird ones (on CAS2210).
SOP – Counting Haematological Malignancies

```
, precursor_diagnosis_date
, precursor_year
, precursor_morphology
, precursor_behaviour
, precursor_coding
, h2.morph_icdo3rev2011_desc AS precursor_description
, h2.split_3 AS precursor_subtype
, h2.split_2 AS precursor_broad_subtype
, h2.split_1 AS precursor_lineage
, transformed_tumourid
, transformation_date
, transformation_year
, transformation_morphology
, transformation_behaviour
, h.morph_icdo3rev2011_desc AS transformation_description
, h.split_3 AS transformation_subtype
, h.split_2 AS transformation_broad_subtype
, h.split_1 AS transformation_lineage
, CASE WHEN tr.precursor_morph_o3 IS NOT NULL THEN 1
ELSE 0
END AS gavin_flag
FROM reg_transformations_raw_all r

-- Only haem transformations
LEFT OUTER JOIN analysishanhualiu.gdo_morph_haem@casref01 h ON r.transformation_morphology = h.morph_icdo3rev2011
  AND r.transformation_behaviour = h.behaviour
-- Talked to Brian R, and the transformation where the precursor is not haem,
-- but the transformation is haem is valid and should be counted as a haem
-- transformation. Only haem precursors (only takes out one extra tumour over
-- the above condition).
LEFT OUTER JOIN analysishanhualiu.gdo_morph_haem@casref01 h2 ON r.precursor_morphology = h2.morph_icdo3rev2011
  AND r.precursor_behaviour = h2.behaviour
LEFT OUTER JOIN gdo_transformations_o3_o3 tr ON r.precursor_morphology = tr.precursor_morph_o3
  AND r.precursor_behaviour = tr.precursor_behaviour_o3
  AND r.transformation_morphology = tr.transformation_morph_o3
  AND r.transformation_behaviour = tr.transformation_behaviour_o3
WHERE h.morph_icdo3rev2011 IS NOT NULL
AND (precursor_morphology||precursor_behaviour != transformation_morphology||transformation_behaviour)
)

--------------------------------------------------------------------------------
-- Combining the transformations, with a record of where they came from
--------------------------------------------------------------------------------
-- I make this 14,007 transformations on CAS2210 (12632 on CAS2109)

, at_transformation_base AS
(SELECT patientid
, precursor_tumourid
, precursor_diagnosis_date
, precursor_year
, precursor_morphology
, precursor_behaviour
, precursor_description
, precursor_subtype
, precursor_broad_subtype
, precursor_lineage
, transformed_tumourid
, transformation_date
, transformation_year
, transformation_morphology
, transformation_behaviour
, transformation_description
, transformation_subtype
, transformation_broad_subtype
, transformation_lineage
, 'Imputed o2-o2' as trans_source
, 1 as gavin_flag
FROM o2_o2_transformations
UNION

SELECT patientid
, precursor_tumourid
, precursor_diagnosis_date
```

SOP – Counting Haematological Malignancies

, precursor_year
, precursor_morphology
, precursor_behaviour
  , precursor_description
, precursor_subtype
, precursor_broad_subtype
, precursor_lineage
, transformed_tumourid
, transformation_date
, transformation_year
, transformation_morphology
, transformation_behaviour
, transformation_description
, transformation_subtype
, transformation_broad_subtype
, transformation_lineage
  , 'Imputed o2-o3' as trans_source
, 1 as gavin_flag
FROM o2_o3_transformations
UNION
SELECT patientid
, precursor_tumourid
, precursor_diagnosis_date
, precursor_year
, precursor_morphology
, precursor_behaviour
  , precursor_description
, precursor_subtype
, precursor_broad_subtype
, precursor_lineage
, transformed_tumourid
, transformation_date
, transformation_year
, transformation_morphology
, transformation_behaviour
, transformation_description
, transformation_subtype
, transformation_broad_subtype
, transformation_lineage
  , 'Imputed o3-o3' as trans_source
, 1 as gavin_flag
FROM o3_o3_transformations
UNION
SELECT patientid
, precursor_tumourid
, precursor_diagnosis_date
, precursor_year
, precursor_morphology
, precursor_behaviour
  , precursor_description
, precursor_subtype
, precursor_broad_subtype
, precursor_lineage
, transformed_tumourid
, transformation_date
, transformation_year
, transformation_morphology
, transformation_behaviour
, transformation_description
, transformation_subtype
, transformation_broad_subtype
, transformation_lineage
  , 'Reg_transform' as trans_source
, gavin_flag
FROM reg_transformations_haem
}

-- Exactly the same table as at_transformation_base, but with the
-- transformations in GDO tree groups.
, at_transformation_mapped AS
(SELECT patientid
, precursor_tumourid
, precursor_diagnosis_date
, precursor_year
-- Warning this code is fragile to changes in names of groups in master tables

CASE WHEN transformation_broad_subtype LIKE '%Hodgkin%' THEN 'Diffuse large B-cell lymphoma (DLBCL), Hodgkin, and other high grade'
WHEN transformation_subtype LIKE '%DLBCL%' THEN 'Diffuse large B-cell lymphoma (DLBCL), Hodgkin, and other high grade'
WHEN transformation_broad_subtype LIKE '%AML%' THEN 'Acute myeloid leukaemia (AML)'
ELSE 'Other transformations'
END AS transformation_split_1

CASE WHEN transformation_broad_subtype LIKE '%AML%' AND precursor_broad_subtype = 'Myelodysplastic syndromes (MDS)' THEN 'Transformation from myelodysplastic syndromes (MDS)'
WHEN transformation_broad_subtype LIKE '%AML%' AND precursor_broad_subtype = 'Myeloproliferative neoplasms (MPN)' THEN 'Transformation from myeloproliferative neoplasms (MPN)'
WHEN transformation_broad_subtype LIKE '%AML%' THEN 'Transformation from other myeloid neoplasms'
WHEN transformation_broad_subtype LIKE '%Hodgkin%' and precursor_subtype LIKE '%CLL%' THEN 'Richter transformation to diffuse large B-cell lymphoma (DLBCL) or Hodgkin lymphoma from chronic lymphocytic leukaemia (CLL) or small lymphocytic lymphoma (SLL)'
WHEN transformation_subtype LIKE '%DLBCL%' AND precursor_subtype LIKE '%CLL%' THEN 'Richter transformation to diffuse large B-cell lymphoma (DLBCL) or Hodgkin lymphoma from chronic lymphocytic leukaemia (CLL) or small lymphocytic lymphoma (SLL)'
WHEN transformation_subtype LIKE '%follicular%' THEN 'Transformation to diffuse large B-cell lymphoma (DLBCL) from Follicular lymphoma'
WHEN transformation_subtype LIKE '%DLBCL%' AND precursor_subtype LIKE '%marginal%' THEN 'Transformation to diffuse large B-cell lymphoma (DLBCL) from Marginal lymphoma'
WHEN transformation_subtype LIKE '%DLBCL%' THEN 'Transformation to diffuse large B-cell lymphoma (DLBCL) from other precursors'
WHEN transformation_broad_subtype LIKE '%Hodgkin%' THEN 'Transformation to diffuse large B-cell lymphoma (DLBCL) or Hodgkin lymphoma from other precursors'
ELSE NULL
END AS transformation_split_2

from at_transformation_base

-- Need to clean duplicates. Start by working out which patients have -- duplicates, by grouping by patient ID and measuring some useful stuff.

SELECT patientid AS row_count
COUNT(*) AS prec_count
COUNT(DISTINCT transformed_tumourid) AS tran_count
MAX(transformation_split_1) AS max_t1
MIN(transformation_split_1) AS min_t1
MAX(transformation_split_2) AS max_t2
MIN(transformation_split_2) AS min_t2
MAX(transformation_morphology||transformation_behaviour) AS max_tm
MIN(transformation_morphology||transformation_behaviour) AS min_tm
MAX(transformation_date) AS max_td
MIN(transformation_date) AS min_td
MAX(precursor_diagnosis_date) AS max_pd
MIN(precursor_diagnosis_date) AS min_pd
MAX(precursor_morphology||precursor_behaviour) AS max_pm
MIN(precursor_morphology||precursor_behaviour) AS min_pm
MAX(precursor_subtype) AS max_ps
SOP – Counting Haematological Malignancies

, MIN(precursor_subtype) AS min_ps
FROM at_transformation_mapped
GROUP BY patientid
)
, at_tran_rank AS
(SELECT t.*
, row_count
, prec_count
, tran_count
, max_t1
, min_t1
, max_t2
, min_t2
, max_tm
, min_tm
, max_td
, min_td
, max_pm
, min_pm
, max_ps
, min_ps
, max_td - min_td AS datediff_t
, max_pd - min_pd AS datediff_p
, RANK() OVER (PARTITION BY t.patientid ORDER BY t.transformation_date, t.transformed_tumourid) AS trans_rank
, RANK() OVER (PARTITION BY t.patientid, t.transformation_split_1 ORDER BY t.transformation_date, t.transformed_tumourid) AS trans_rank_2
, RANK() OVER (PARTITION BY t.patientid ORDER BY t.precursor_diagnosis_date DESC, t.precursor_tumourid ) AS prec_rank
, RANK() OVER (PARTITION BY t.patientid, t.transformation_split_1, t.transformed_tumourid ORDER BY t.precursor_diagnosis_date desc, t.precursor_tumourid) AS prec_rank_2
FROM at_transformation_mapped t
LEFT OUTER JOIN at_tran_patient f ON f.patientid = t.patientid
)
, at_transformation AS
(SELECT t.*
-- When there’s only one transformation per patient, just take it
CASE WHEN row_count = 1 THEN 1
-- Two transformations in different groups, take both
WHEN row_count = 2 AND tran_count = 2 AND max_t1 != min_t1 THEN 1
-- When there are two transformations and they are the same precursor with two
-- transformation endpoints, and one of them is 99303 (myeloid sarcoma) and one
-- of them is 98613 (AML), take the AML if it is within 92 days.
WHEN row_count = 2 AND tran_count = 2 AND prec_count = 1 AND max_tm='99303' AND min_tm = '98613' AND datediff <= 92 AND transformation_morphology = '9930' THEN 0
WHEN row_count = 2 AND tran_count = 2 AND prec_count = 1 AND max_tm='99303' AND min_tm = '98613' AND datediff <= 92 AND transformation_morphology = '9861' THEN 1
-- When there are two transformations in the same group but one is earlier than
-- the other take the earlier one.
WHEN row_count = 2 AND tran_count = 2 AND datediff_t > 0 AND max_t1 = min_t1 AND trans_rank = 1 THEN 1
WHEN row_count = 2 AND tran_count = 2 AND datediff_t > 0 AND max_t1 = min_t1 AND trans_rank = 2 THEN 0
-- When there are two transformations, and they are the same precursor with two
-- transformation end points and they are on the same day, and one is 9840 (AML
-- M6 type) and one is 9861 (AML NOS) take the M6 type.
WHEN row_count = 2 AND tran_count = 2 AND prec_count = 1 AND max_tm='98403' AND min_tm = '98613' AND datediff <= 0 AND transformation_morphology = '9861' THEN 0
WHEN row_count = 2 AND tran_count = 2 AND prec_count = 1 AND max_tm='98403' AND min_tm = '98613' AND datediff <= 0 AND transformation_morphology = '9840' THEN 1
-- When there are two transformations, and they are the same precursor with two
-- transformation end points and they are on the same day with different
-- histology take the higher histology.
WHEN row_count = 2 AND tran_count = 2 AND prec_count = 1 AND datediff_p = 0 AND max_t1 = min_t1 AND max_tm!= min_tm AND transformation_morphology||transformation_behaviour = max_tm THEN 1
### SOP – Counting Haematological Malignancies

When there are two transformations, and they are the same precursor with two transformation end points and they are on the same day with the same histology use tumourID as a tie-break, take the lowest tumour ID.

When there is one transformation but two possible precursor conditions and the precursor conditions are in the same precursor_subtype (i.e., are basically the same) take the precursor condition that is earliest (i.e., happened first).

When there is one transformation but two possible precursor conditions but they're both on the same day with the same morphology code, take the one with the lower tumourID as a tie-break.

When there is a person with more than two transformations. Divide the transformations into fundamentally different types (using the three top level GDO groups). Take the earliest transformation in each group. Take the latest precursor condition that was allowed for that transformation.

Tidy up the final table. This makes 14,067 transformations on CAS2210, of which 13,694 are not duplicates.
Appendix D: ICD O3 rev 2011 codes to define haematological malignancies – MORPH_HAEM table

This table is the master list of which ICD O3 rev 2011 codes of morphology and behaviour combinations should be counted as haematological malignancies. It is used by the code in Appendix A to define ‘all haematological malignancies’. It also has columns for which sub-groups they should be assigned to. It is a copy of the ANALYSISISHANHUALIUGDO_MORPH_HAEM@CASREF01 table.

It can be found in the word document version of this SOP, which is available from the haematology lead, or on-line at https://www.cancerdata.nhs.uk/getdataout inside the grouping document for haematological cancers.

Appendix E: Codes to define haematological transformations according to the Gavin paper

This table is the master list of which pairs of malignancies should be counted as haematological malignancy transformations. It is used by the code in Appendix C to define ‘haematological malignancy transformations’ and create the transformation table. It follows the Gavin et al paper ‘Towards optimal clinical and epidemiological registration of haematological malignancies: Guidelines for recording progressions, transformations and multiple diagnoses’. It is a copy of the ANALYSISISANDREWBACON.GDO_TRANSFORMATION_HAEM@CASREF01 table.
It can be found in the word document version of this SOP, which is available from the haematology lead, or on-line at https://www.cancerdata.nhs.uk/getdataout inside the grouping document for haematological cancers.