-- icnarc\_\_01\_\_format\_dates

-- This script creates 'icnarc\_01b\_dates\_formatted', a copy of the source table where the dates are correctly formatted

-- Note that tables are created with the naming convention

-- icnarc\_XXy\_description, where

-- XX is the number, e.g. '01' relating to the script number

-- y is a letter, e.g. 'a' relating to the order of table creation within the script

-- Tables are dropped when no longer needed

-- Steps:

-- 1) Change any null values to actual NULLs

-- 2) Cast new dates

-- DATES: There are 21 columns containing the word date (see post-script) and no other columns that could be considered a date,

-- following visual inspection of first 200 results of each column, and use of DISTINCT() when unsure, e.g. DOB\_estimated (contains yes/no/null).

-- 14 date columns are NOT empty (empty is either 00/00/00 or null; see post-script)

-- TIMES: There are 7 columns containing the word time (see post-script) and 1 more which could be considered a time (excluding 'Timeliness\_of\_admission\_to\_your\_unit')

-- This script focuses on dates.

-- Not all the of the dates are just dates, some (n = 4) are date times (with 00:00:00.000 times).

-- I figured out which was which by looking at the distinct values in each date column.

-- Maybe this could be done programmatically by calculating the character length of the values.

-- 1) Change any instance of '00/00/00' or '00/00/00 00:00:00.000' to null

-- and exclude the 7 columns that I identified manually as having no meaningful values (see post-script)

CREATE TABLE `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_01a\_dates\_nulled`

AS

SELECT \*

-- exclude meaningless values (see post-script)

EXCEPT(

Booked\_Unit\_Admission\_Date,

Outreach\_1st\_visit\_date,

Date\_unit\_left,

Date\_Extubated,

Bed\_ready\_date,

Date\_brainstem\_death,

Date\_of\_death,

Event\_date ----- excluded near the end of the cleaning process because it turned out no-one knew what it was

)

-- replace empty dates with nulls

REPLACE(

NULLIF(Date\_of\_birth, '00/00/00 00:00:00.000') AS Date\_of\_birth,

NULLIF(Actual\_date\_of\_delivery, '00/00/00') AS Actual\_date\_of\_delivery,

NULLIF(Expected\_date\_of\_delivery, '00/00/00') AS Expected\_date\_of\_delivery,

NULLIF(Date\_1st\_managed\_by\_unit\_team, '00/00/00') AS Date\_1st\_managed\_by\_unit\_team,

NULLIF(Date\_of\_last\_critical\_care, '00/00/00') AS Date\_of\_last\_critical\_care,

NULLIF(Hospital\_admission\_date, '00/00/00 00:00:00.000') AS Hospital\_admission\_date,

NULLIF(Original\_hosp\_admission\_date, '00/00/00') AS Original\_hosp\_admission\_date,

NULLIF(Original\_unit\_admission\_date, '00/00/00') AS Original\_unit\_admission\_date,

NULLIF(Outreach\_last\_visit\_date, '00/00/00') AS Outreach\_last\_visit\_date,

NULLIF(Unit\_admission\_date, '00/00/00 00:00:00.000') AS Unit\_admission\_date,

-----NULLIF(Event\_date, '00/00/00 00:00:00.000') AS Event\_date,

NULLIF(Date\_body\_removed, '00/00/00') AS Date\_body\_removed,

NULLIF(Date\_of\_first\_critical\_care, '00/00/00') AS Date\_of\_first\_critical\_care,

NULLIF(Date\_ult\_unit\_discharge, '00/00/00') AS Date\_ult\_unit\_discharge

)

FROM `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.tbl\_ICNARC\_V2\_nulled`;

-- Delete the tbl\_ICNARC\_V2\_nulled table

drop table `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.tbl\_ICNARC\_V2\_nulled`;

-- 2) Format dates (four are datetime format so require casting, others can be parsed)

CREATE TABLE `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_01b\_dates\_formatted`

AS

SELECT \*

REPLACE(

DATE(CAST(Date\_of\_birth AS DATETIME)) AS Date\_of\_birth,

PARSE\_DATE('%d/%m/%Y', Actual\_date\_of\_delivery) AS Actual\_date\_of\_delivery,

PARSE\_DATE('%d/%m/%Y', Expected\_date\_of\_delivery) AS Expected\_date\_of\_delivery,

PARSE\_DATE('%d/%m/%Y', Date\_1st\_managed\_by\_unit\_team) AS Date\_1st\_managed\_by\_unit\_team,

PARSE\_DATE('%d/%m/%Y', Date\_of\_last\_critical\_care) AS Date\_of\_last\_critical\_care,

DATE(CAST(Hospital\_admission\_date AS DATETIME)) AS Hospital\_admission\_date,

PARSE\_DATE('%d/%m/%Y', Original\_hosp\_admission\_date) AS Original\_hosp\_admission\_date,

PARSE\_DATE('%d/%m/%Y', Original\_unit\_admission\_date) AS Original\_unit\_admission\_date,

PARSE\_DATE('%d/%m/%Y', Outreach\_last\_visit\_date) AS Outreach\_last\_visit\_date,

DATE(CAST(Unit\_admission\_date AS DATETIME)) AS Unit\_admission\_date,

-----DATE(CAST(Event\_date AS DATETIME)) AS Event\_date,

PARSE\_DATE('%d/%m/%Y', Date\_body\_removed) AS Date\_body\_removed,

PARSE\_DATE('%d/%m/%Y', Date\_of\_first\_critical\_care) AS Date\_of\_first\_critical\_care,

PARSE\_DATE('%d/%m/%Y', Date\_ult\_unit\_discharge) AS Date\_ult\_unit\_discharge

)

FROM `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_01a\_dates\_nulled`;

-- Hand check distinct values to make sure it worked on each column

-- SELECT DISTINCT Date\_ult\_unit\_discharge,

-- FROM `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_dates\_formatted`

-- ORDER BY Date\_ult\_unit\_discharge;

-- Delete the icnarc\_01a\_dates\_nulled table

drop table `yhcr-prd-phm-bia-core.CY\_MYSPACE\_EmW.icnarc\_01a\_dates\_nulled`;

-- Now run 'icnarc\_\_02\_\_format\_times'

---- POST-SCRIPT

--- DATE COLUMNS (n = 21)

---- Usable date columns (n = 14)

--Date\_of\_birth (Date time)

--Actual\_date\_of\_delivery (Date)

--Expected\_date\_of\_delivery (Date)

--Date\_1st\_managed\_by\_unit\_team (Date)

--Date\_of\_last\_critical\_care (Date)

--Hospital\_admission\_date (Date time)

--Original\_hosp\_admission\_date (Date)

--Original\_unit\_admission\_date (Date)

--Outreach\_last\_visit\_date (Date)

--Unit\_admission\_date (Date time)

-----Event\_date (Date time) [removed]

--Date\_body\_removed (Date)

--Date\_of\_first\_critical\_care (Date)

--Date\_ult\_unit\_discharge (Date)

---- Unusable date columns (n = 7)

--Booked\_Unit\_Admission\_Date (two distinct values: '2006-06-12' for 1 record, null for the rest)

--Outreach\_1st\_visit\_date (the only distinct value is ’00/00/00’)

--Date\_unit\_left (the only distinct value is ’00/00/00’)

--Date\_Extubated (the only distinct value is ’00/00/00’)

--Bed\_ready\_date (the only distinct value is ’00/00/00’)

--Date\_brainstem\_death (the only distinct value is null)

--Date\_of\_death (the only distinct value is null)

--Event\_date -- added near the end of the cleaning process because it turned out no-one knew what it was