-- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* This Routine Calculates Likelihood Ratios \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

USE HAP725  -- The name of your database is likely to be different

/\*-- Create new tables from DxAge\_x tables and use CAST to change data types

/\* These commands are made into comments so that they do not accidentally re-run

DROP TABLE #temp

SELECT \* INTO #temp FROM [dbo].[DxAge\_1] -- 4233546 rows

INSERT INTO #temp SELECT \* FROM [dbo].[DxAge\_2] -- 5223128 rows

INSERT INTO #temp SELECT \* FROM [dbo].[DxAge\_3] -- 4179754 rows

INSERT INTO #temp SELECT \* FROM [dbo].[DxAge\_4] -- 3807014 rows

DROP TABLE dbo.final

SELECT CAST([id] as int) as id

      , [icd9]

      , CASE AgeAtDx

              WHEN 'NULL' THEN null

              ELSE CAST(AgeAtDx as float) END as AgeAtDx

      , CASE AgeAtFirstDM

              WHEN 'NULL' THEN null

              ELSE CAST(AgeAtFirstDM as float) END as [AgeAtFirstDM]

      , CASE AgeAtDeath

              WHEN 'NULL' THEN null

              ELSE CAST(AgeAtDeath as float) END as [AgeAtDeath]

INTO dbo.final

FROM #temp

\*/

SELECT Count(\*) FROM dbo.final --(17,443,442 rows)

-- Identify zombies

DROP TABLE #Z

SELECT DISTINCT Id

INTO #Z

FROM dbo.final

WHERE AgeAtDeath<AgeAtDx -- Death before Dx

GROUP BY ID

SELECT TOP 5 \* FROM #Z ORDER BY id DESC



-- 168 unique patients with wrong date of death

-- Exclude zombies from final table

DROP TABLE #data

SELECT a.\*

INTO #data

FROM dbo.final a left join #Z b ON a.id=b.id

WHERE b.id is null

SELECT TOP 3 \* FROM #data order by ID

-- (17,432,694 row(s) affected)



-- Remove patients with more than 365 diagnosis in a year and diagnosis with age being wrong

DROP TABLE #Data2

SELECT DISTINCT ID

INTO #Data2

FROM #Data

GROUP BY ID, Cast(AgeAtDx as Int)

HAVING Count(Icd9) >365

SELECT TOP 10 \* FROM #Data2

-- (56 row(s) affected)



DROP TABLE #Data3

SELECT a.\*

INTO #Data3

FROM #Data a left join #Data2 b on a.id=b.id

WHERE b.id is null and AgeAtDx is not null AND AgeAtDx >0

-- removing also problems with age at diagnosis

SELECT TOP 3 \* FROM #Data3 WHERE AGeAtDx>0

-- 17,432,694 is reduced to 17,379,713 reduced to 17,379,218



-- Select training and validation set

SELECT \*

INTO dbo.training

FROM #Data3

WHERE Rand(ID) <=.8

SELECT TOP 5 \* FROM dbo.training WHERE ID=467828

-- (13,760,073 row(s) affected) ---- (13759944 rows affected)



 -- Find unique IDs in training set

DROP TABLE #trainID

SELECT DISTINCT ID

INTO #trainID

FROM dbo.training

--  (657,885 row(s) affected) --- (657883 rows affected)

-- Create Validation set

SELECT a.\*

INTO dbo.vSet

FROM #Data3 a left join #trainID b ON a.id=b.id

WHERE b.id is null

-- (3619145 row(s) affected)

 -- Calculate # dead and # alive in training set

DROP TABLE #cnt1

select ID, CASE WHEN Max(ageatdeath)>0 THEN 1 ELSE 0 END AS Dead

       , CASE WHEN Max(ageatdeath)>0 THEN 0 ELSE 1 END AS Alive

       , CASE WHEN Max(AgeAtDeath) IS NULL THEN 1 ELSE 0 END AS Alive2

INTO #cnt1

FROM dbo.training

GROUP BY ID

SELECT TOP 3 \* FROM #cnt1



-- (657885 row(s) affected)

DROP TABLE #cnt2

SELECT SUM(Alive) AS PtsAlive, Sum(Dead) AS PtsDead

INTO #Cnt2

FROM #cnt1

SELECT \* FROM #Cnt2

 /\* Unique patients alive or Dead\*/



-- \*\*\*\*\*\*\*\* Calculate Likelihood Ratio \*\*\*\*\*\*\*\*\*

-- Select patients who died 6 month after diagnosis

DROP TABLE #DeadwDx

SELECT ICD9, count(distinct ID) as PtsDead6

INTO #DeadwDx

FROM dbo.training

WHERE AgeatDeath-AgeatDx<=.5 -- This is 6 months in age measured in years

GROUP BY ICD9

SELECT TOP 5 \* FROM #DeadwDx



-- (6400 row(s) affected)---6297 rows

-- Select diagnosis where patient did not die or did not die within 6 months

DROP TABLE #AlivewDx

SELECT ICD9, count(distinct ID) as PtsAlive6

INTO #AlivewDx

FROM dbo.training

WHERE AgeatDeath-AgeatDx>.5 or AgeAtDeath is null -- Not dead in 6 months or not dead

GROUP BY ICD9

SELECT TOP 5 \* FROM #AlivewDx ORDER BY ICD9



--(10439 row(s) affected)

-- Combine the tables for dead and alive patients

Drop Table #Dx

SELECT CASE a.Icd9 WHEN null THEN b.icd9 ELSE a.icd9 END as icd9

, PtsDead6

, PtsAlive6

INTO #Dx

FROM #alivewDx a FULL OUTER JOIN #DeadwDx b

       ON a.icd9=b.icd9 --Full join keeps record even if not in either table

SELECT TOP 5 \* FROM #Dx



-- (10480 row(s) affected) 🡺 10480 rows affected

-- Calculate Likelihood Ratios

-- Set LR to maximum when all in DX are dead

-- Set LR to minimum when all in Dx are alive

SELECT Icd9

, PtsDead6

, PtsAlive6

, PtsDead

, PtsAlive

, CASE

       WHEN PtsAlive6 is null THEN PtsDead6+1

       WHEN PtsAlive6=0 THEN PtsDead6+1

       WHEN PtsDead6 is null THEN 1/(PtsAlive6 +1)

       WHEN PtsDead6= 0 THEN 1/(PtsAlive6 +1)

       ELSE

       (cast(PtsDead6 as float)/Cast(PtsDead as float))/(Cast(PtsAlive6 as Float)/Cast(PtsAlive As Float)) END AS LR

-- % of Dx among dead divided by % of Dx among alive patients

INTO dbo.LR

FROM #Dx cross join #Cnt2

SELECT top 10 \* FROM dbo.LR ORDER BY LR desc

--(10480 row(s) affected)

