**HAP720 – PHASE 1**

**NAME – LAKSHMI INDIRA KUPPA**

**CODES**

use HAP720

select \* from dbo.hap720\_claims\_hcpcs

select \* from dbo.hap720\_mpi

select a.\* , b.mpi , b.member\_id into #temptable

from dbo.hap720\_claims\_hcpcs a join dbo.hap720\_mpi b on a.patient\_id = b.patient\_id -----(1507666 rows affected)

select \* from #temptable

select \* from dbo.hap720\_seer

select a.\* , b.diagnosis\_year , b.diagnosis\_age , b.sex , b.race , b.c\_diagnosis

into #final from #temptable a join dbo.hap720\_seer b on a.member\_id = b.member\_id ----(1668007 rows affected)

select \* from #final

SELECT COUNT(\*)

FROM dbo.hap720\_seer s

WHERE NOT EXISTS (

SELECT 1

FROM #temptable t

WHERE s.member\_id = t.member\_id

);

SELECT COUNT(\*)

FROM #temptable t

WHERE NOT EXISTS (

SELECT 1

FROM dbo.hap720\_seer s

WHERE s.member\_id = t.member\_id

);

**RESULTS**

**A computer screen shot of a computer screen

Description automatically generated**

**A computer screen with text and numbers

Description automatically generated**

**A computer screen with text

Description automatically generated**

**A computer screen with a white screen

Description automatically generated**

**SUMMARY**

1. How many patients in the SEER do not have matched records in the Claims?

Number of patients in the SEER without matched records in the Claims are 31526

1. How many patients in the Claims do not have matched records in the SEER?

Number of patients in the Claims do not have matched records in the SEER are 0

1. What does this mean? (i.e., why patients do not exist in SEER?)

Patients in the SEER without matched records in the Claims: This indicate that these patients were diagnosed with cancer or certain conditions tracked by SEER, but their medical treatments related to these conditions might not be captured in the Claims dataset.

Patients in the Claims without matched records in the SEER: These patients have medical claims recorded in the Claims dataset, but there are no corresponding records in the SEER dataset. This could occur if patients received medical treatments or procedures not related to cancer or the specific conditions tracked by SEER. It could also indicate errors in data integration or missing information in either dataset.