SEER Stat- Cancer Incidence and mortality file

Important sheets:

“Incidence”

“Adeno\_SCC\_All\_SEER”

\*\* **We can delete the other sheets, if they aren’t applicable to our analysis**

**“Incidence” Sheet**

Column 1: (contains: All: 8010-8671, 8940-8941)- types of cancer

In this column, I separated out ‘all cancer’ and ‘SCC’

Column 2: age (can change these groupings)

Column 3: Rate- Cervical cancer incidence rates per 100,000 women (we can also pull data for males and other types of cancer)

Column 4: Count- number of cancer cases (cumulative) reported to SEER cancer registries

Column 5: Pop- US standardized population

Other important information:

* First year columns give all data from 1973-2012 combined and then rest of columns separate out into individual years
* From 1973- 1992, only 9 registries reported cancer incidence to SEER. From 1992- 2000, 12 registries reported cancer incidence to SEER. From 2000-2012, 18 registries reported. We can decide if we want to only use 9 registries from 1973-2012 or if we want to include all additional registries as they come in (of course, more registries = bigger sample size)

**“Adeno\_SCC\_All\_SEER” Sheet**

* same as above except that cancers are separated out into adenocarcinoma, SCC, and all cancers

NHANES

* Not possible to download PDFs of the codebooks for the surveys, but all variable descriptions can be found here: <http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm>
  + Look for corresponding year and data set (demography and lab data)
* It’s possible to download any of these data sets, so if we find another data set with variables we want, we can download and merge it to what we have

NHIS

* Codebooks with variable descriptions all uploaded to repository
* More information about the surveys can be found here: <http://www.cdc.gov/nchs/nhis/about_nhis.htm>