SEER Project - PIcc meeting

Methods

Study design

* **Retrospective Analysis of existing public available data (highly curated) 🡪 to find a baseline of how cancer is diagnosed**
* SEER database 🡪 considered one of the largest databases and to be representative of the population of US
  + (Additional description of SEER database)
  + Real world data on how cancer is diagnosed in the US today
* Pre-competitive = trying to get trends outlined, not portray a specific technology. Provide a “ground truth”/ “gold standard”
  + While there are other modalities for diagnosis, light microscopy represents a most used approach. We want to demonstrate this point with data.
  + Need to provide a **baseline** so we can assess the impact of new technologies
* SEER data to analyze NAACCR variables
  + SEER data can be linearly traced (direct link from state registry)
  + NAACR has higher population representation (NAACCR covers the whole population, while SEER covers 30-40%)
  + Consider: both SEER and NAACCR and compare

Data & Variables

* ICD-O
  + Separated into two pools of data, solid tumors and heme tumors
    - Supplemental table (as dictionary)
* Source/tumor site features
  + i.e. Age, sex, ancestral info., location (geographic trends), additional socioeconomic factors
* NAACCR #490 – codes 1-4 microscope vs codes 5-9 non-microscope
  + This is super essential because this is how we came to the conclusion in the data. These codes differentiate the light microscopy diagnoses from the non-light microscopy diagnoses.

Analytical Plan

* Primary endpoint = **fraction of cases with microscopic diagnosis**
* Define secondary endpoints =
  + Fraction by solid; fraction by heme
  + By code (ICD)
  + By site (anatomic)
  + Other population metric (socioeconomic status, state, age, etc.)
    - Additional population metrics are heterogenous across time scales and across data

Statistical

* Two separate groups complete data pull
  + Alex/Emma
  + Ahmad
* By tumor area not ICD code
* Additional contributors?? Our approach is checked for reproducibility
  + 95% concordance acceptable

Next steps:

* Perform reproducibility pulls