Data Use Agreement for SEER Radiation Therapy and Chemotherapy Information

November 2018 Data Submission

The population-based Surveillance, Epidemiology, and End Results (SEER) registries collect information on radiation therapy (RT) and chemotherapy given as part of the first course of treatment. RT data are classified by the type of RT received or "no/unknown – no evidence of radiation was found in the medical records examined". Chemotherapy data are categorized as either "yes – patient had chemotherapy" or "no/unknown – no evidence of chemotherapy was found in the medical records examined". These data are available upon request after acknowledging the limitations associated with analyses of the data. Two main limitations affect recommended analyses using the SEER RT and chemotherapy data: 1) the completeness of the variables; and 2) the biases associated with unmeasured reasons for receiving or not receiving RT/chemotherapy. Below we further describe the issues and analyses that could be problematic.

Completeness of the Variables

One recent publication comparing SEER data with SEER-Medicare data reported that overall sensitivity was 80% for SEER RT data and 68% for SEER chemotherapy data. Sensitivity varied by cancer site, stage, and patient characteristics. The overall positive predictive value was high (>85%) for all treatments and cancer sites except chemotherapy for prostate cancer. This analysis used a fairly broad definition for chemotherapy use based on Medicare claims, and further sensitivity analysis is ongoing. [1]

Although sensitivity was moderate, specificity was high, meaning that if RT or chemotherapy was captured in SEER, it was most likely received by the patient. But if it was not captured in SEER, then we do not know whether it was not received by the patient or whether it was missed by the registry. As treatment is increasingly received outside of the hospital setting, there is a diminishing likelihood that it is captured completely. Because we cannot accurately distinguish between "no treatment" and "unknown if patients received treatment," the variables that are released upon request are classified as "yes" or "no/unknown".

Examples of analyses that would NOT be supported by the RT/chemotherapy data, due to the incompleteness of the variable, include:

- Estimates of population prevalence of treatment or patterns of care in the population without appropriate comment on the limitations of the data (e.g., clearly labeling both treatment categories as "yes" and "no/unknown" wherever they appear).

- Estimates of compliance with guidelines

- Comparison of treatment levels in different groups, e.g., investigating health disparities, without adequately stated limitations

- Comparison of outcomes by treatment received

Since we have high confidence that an individual received RT/chemotherapy if the variable is listed as "yes", analyses such as identifying a cohort of patients who received treatment in order to identify risk of adverse events, including risk of second cancers, would be supported by the data.

Biases Associated with Who Receives Treatment

Unlike clinical trials, many factors involved in determining the course of treatment will not be captured in the registry data. Such factors include: patient preferences, physician recommendations, comorbidities, and proximity to treatment providers. Because the data collected do not include these and other factors that are related to why a patient did or did not receive RT/chemotherapy, we do not recommend comparing outcomes conditioned on treatment or comparative effectiveness research using the SEER data without careful consideration of possible biases and appropriate adjustments, potentially using data beyond standard SEER data (e.g. SEER-Medicare linked data). For example, survival differences observed for patients who did vs. did not receive chemotherapy cannot be attributed to the efficacy or effectiveness of treatment without controlling for the factors that determined treatment receipt. Similarly, observed differences cannot be generalized to describe the benefit an individual would expect to receive from chemotherapy treatment.

Reference:

1. Noone AM, Lund JL, Mariotto A, Cronin K, McNeel T, Deapen D, Warren JL. Comparison of SEER Treatment Data with Medicare Claims. Med Care 2014 Mar 15. [Epub ahead of print]

I have read and understand the limitations of the SEER RT and chemotherapy data described above and will include a description of relevant limitations in any analyses published using the SEER data. I acknowledge that the SEER Program has advised me that there are substantive concerns about using these data to address certain research questions as described above. I understand that any findings from such analyses may be inaccurate or misleading.

I will send to NCI any publication that uses SEER RT or chemotherapy variables available through a custom data request. NCI will use this information to track the use of RT and chemotherapy variables collected in SEER.