National Cancer Database (NCDB)

Access approval process: Must be from a CoC accredited program and requires a agreement to a data use policy as well as a letter of support from a CoC accredited program.

Need for specialized software: Data exists in PUF’s (Participant User Files) and are HIPPA compliant data files that are available to licensed users. For the publicly accessible data an online data analysis tool is available for users to draw inference from.

Granularity of publicly accessible data: Aside from a small subset of data that is publicly available, the majority of the data exists in a walled environment. The publicly accessible includes benchmark reports and via the tool users can select from 16 explanatory variables regarding patient status along with hospital cohort by state.

Institutional goals & governing bodies: CoC oversees ACDB and is a multidisciplinary consortium of professional organizations with the goal of striving to improve cancer care through setting standards, prevention, research, education, and the monitoring of comprehensive cancer care.

Data analysis of selected cancer variables:

The online data exploration tool allows the user to select cancer site (breast) and case type (All diagnosis types) and allows the user to select up to three analysis variables along with the year. As a baseline the cancer selected across databases was breast cancer.

The tool is extremely intuitive, albeit a bit limited in scope (no ability to model relationships).

After variables are selected, the user can view the generated plot along with a data table. The online tool has multiple plotting options and can generate most of the commonly known chart types (bar, pie, scatter, area) the tool also has further options to customize the chart (titles, colors, fonts, and gridlines). This element of the tool features a high level of customization for an online dynamic chart generation tool.

In this example, we look at trends in breast cancer diagnosis since screening guidelines changes were instituted (government changed recommended screening intervals in 2009)

Table

Description automatically generated

Based on the plot above, we can see that either diagnoses are up due to increased screening or some other factor. The point being this information is publicly available and has a high degree of accuracy.

Surveillance, Epidemiology, and End Results

Access approval process: Institutional users can access after completing a registration form, a provided signing official and then acquiring a SEER\*stat username.

Need for specialized software: SEER\*Stat statistical software. Alternatively, users can browse the public data via the web browser.

Granularity of publicly accessible data: SEER cancer incidence and survival data originate from population-based cancer registries covering approximately 34.6% of the population in the United States. Data is publicly accessible, however the number of applicants for research data are limited. Public users may use the online tool to examine the data.

Institutional goals & Governing bodies: SEER is an authoritative source for cancer statistics in the United States. SEER provides information on cancer statistics in an effort to reduce the cancer burden among the U.S population. SEER is supported by the Surveillance Research Program in NCI’s Division of Cancer Control and Population Sciences (DCCPS).

Data analysis of selected cancer variables:

The SEER\*Explorer tool allows the user to select cancer site (breast) and a statistic to explore (SEER Incidence) and allows the user to select multiple explanatory variables along with the year, gender, state at diagnosis, rate type, and a precision option. As a baseline the cancer selected across databases was breast cancer.

The tool is extremely intuitive, user friendly and has robust data export options (export data, export plots, or share the web generated chart). The tool has many more customization options compared to the NCDB and features many statistical tools.

After variables are selected, the user can view the generated plot along with a data table. The tool determines the chart type based on the “Choose a Statistic to Explore” option chosen. This is somewhat more limited than NCDB due to the inability to customize the chart. However, this element of the tool features a high level of customization for an online dynamic chart generation tool.

In this example, we look at trends in breast cancer diagnosis since screening guidelines changes were instituted (government changed recommended screening intervals in 2009). The chart presents similar information in a slightly different method.

Chart, bar chart

Description automatically generated