ALASKA CANCER REGISTRY
https://health.alaska.gov/dph/VitalStats/Pages/cancer/registry.aspx

NPCR 2024 SUCCESS STORY

Enhancing Cancer Data: Linking COVID-19 Cases to Improve Public Health Insights

AUTHORS: Rosa M. Avila and Julie Cleaton

SUMMARY

The Alaska Cancer Registry (ACR) launched a project to enhance data quality in response to new COVID-19-related fields added for cancer cases diagnosed in 2020 and 2021. ACR noted several reporting errors throughout the data collection process. To address this, ACR linked COVID-19 data from the Department of Public Health (DPH) section of epidemiology with their cancer registry data. The project demonstrated how integrating public health data can improve data quality to identify public health issues and refine registry data for better outcomes.

CHALLENGES

The new COVID-19 data fields lacked sufficient documentation and edits for internal consistency checks, leading to reporting errors.

- Linking large datasets required sophisticated deduplication and probabilistic matching methods, with manual review of matches for accuracy.
- The project could not evaluate the impact of COVID-19 infections on cancer treatment due to data constraints.
- CDC reported that health care disruptions during the COVID-19 pandemic led to delays in diagnosis.[1] Therefore, researchers analyzing this data will need to consider selection bias (cases diagnosed in 2020 and 2021 may not represent the true population) and temporal bias (some diagnosed cases might have shifted to later time periods).

SOLUTION

The ACR team used R quantitative software for this data analysis. COVID-19 case data (n=213,573) were first cleaned by standardizing the formats of each data field. Multiple infections were deduplicated to a dataset of 202,730 unique individuals using probabilistic matching and manual review. The analysts then linked COVID-19 cases with the cancer registry data (n=7,108), using similar probabilistic matching techniques, which identified 1,287 linked cases, of which 914 were cancer diagnoses preceding COVID-19 infections. This approach revealed substantial improvements in COVID-19 data completeness and accuracy in our cancer database.

RESULTS

- This strategy improved data quality.
- COVID-19 case data were cleaned and linked to cancer cases diagnosed in 2020 and 2021 using methods previously outlined. The number of positive COVID-19 test results increased from 223 to 914—a 310% increase. The percentage of cancer patients tested with a date associated increased from 83.0% to 100% completeness.
- The COVID-19 data elements collected by the central cancer registries are likely to be incomplete if only medical records were used to collect that data in cancer abstracts. If researchers plan to use these data to evaluate the effect of COVID-19 on people with cancer, they could consider linking cancer registry data with state COVID-19 data.

CONCLUDING REMARKS

- Next steps include evaluating the effect of COVID-19 on people with cancer by linking this dataset with vaccination data.
- The COVID-19 data are unique in that testing was widely available to the public, free of charge, and reported to DPH in 2020 and 2021. This study would be difficult to reproduce for other diagnosis years now that home testing is available and test results are not reported.

1 Centers for Disease Control and Prevention. Highlights from 2021 Cancer Incidence with Comparisons to Previous Years. Centers for Disease Control and Prevention, U.S. Department of Health and Human Services; 2024.

NATIONAL PROGRAM of CANCER REGISTRIES
NPCR
CANCER DATA DRIVING ACTION

U.S. Department of Health and Human Services
Centers for Disease Control and Prevention