2020 NPCR NEW JERSEY CANCER STATE REGISTRY SUCCESS STORY

STORY TOPIC/FOCI: Expanding data sources, improve data quality, and utilizing improved data for cancer prevention and control planning.

STORY CATEGORY: Public Health Impact

STORY TITLE: A Not-So-New Partner in Cancer Surveillance? Leveraging Lexis Nexis to Enhance Demographic Information for the Benefit of Public Health

STORY AUTHOR(S):
Antoinette M. Stroup, PhD; Lisa E. Paddock, PhD, MPH; Mireille Lemieux, M.Sc., CTR; Gerald Harris, PhD

SUMMARY

The New Jersey State Cancer Registry (NJSCR) has successfully developed a partnership with Lexis Nexis, Inc. (LN) to enhance demographic data for the benefit of public health. We leverage LN consumer data from public records (e.g., real estate/tax assessor records, mortgage, motor vehicle registrations, driver's license records, voter registration, etc.) to improve the accuracy and completeness of variables such as address at diagnosis and race. As a special study, we also leveraged LN consumer data to evaluate the utility of residential history data in assessing cancer risk and survival.

CHALLENGE

Cancer registries have historically relied on reporting facilities such as hospitals and physician offices to provide patient demographic information, including address at diagnosis and race. However, these data are not reported consistently and are prone to errors; likely due to increases in pathology-only and physician-only reporting and providers electing not to document race and ethnicity of patients. This has significant downstream impacts on data quality and statistics. Errors in address at diagnosis, for example, may lead to our inability to verify the State of diagnosis, reduce the statewide completeness estimate, and prevent the case from being geocoded. Lack of geocoded data impacts our ability to generate reliable geographic cancer statistics. Race is also a critical variable because New Jersey is one of the more racially diverse states in the United States and is home to the largest proportion of foreign-born residents (1). Race-specific cancer statistics such as incidence are prone to bias and may lead to an underestimation of rates if there are significant numbers of cases missing race values.

Although NJSCR has been able to limit the proportion of missing race to less than 3% and continues to meet the National Program of Cancer Registry's (NPCR) Registry of Distinction Standard, and North American Association of Central Cancer Registry (NAACCR) Gold certification, meeting these goals requires a significant amount of time and effort to obtain the information from hospitals, physicians, and other reporting sources.

Epidemiological studies that use population-based cancer registry data are also limited by the lack of retrospective residential histories to evaluate risk for cancers with long latency periods (2-4). And, with cancer survivors living longer, residential histories *post* diagnosis may contribute to our understanding of survivorship, quality of life, access to subsequent treatment for recurrence, and, of course, long term survival outcomes.

Sources

1. U.S. Census Bureau QuickFacts (2019). U.S. Census Bureau. https://www.census.gov/quickfacts/fact/table/NJ,US/PST045219 (Accessed on October 29, 2020).
2. Boscoe FP. The use of residential history in environmental health studies. In: Maantay JA, McLafferty S, eds. *Geospatial Analysis of Environmental Health*. Dordrecht, the Netherlands: Springer Science+Business Media; 2011:93–110.
3. Jacquez GM, Kaufmann A, Meliker J, et al. Global, local and focused geographic clustering for case-control data with residential histories. *Environ Health*. 2005;4(1):4.
4. Hurley, S., Hertz, A., Nelson, D. O., Layefsky, M., Von Behren, J., Bernstein, L., . . . Reynolds, P. (2017). Tracing a path to the past: exploring the use of commercial credit reporting data to construct residential histories for epidemiologic studies of environmental exposures. *American Journal of Epidemiology, 185*(3), 238-246.
5. Wiese, D., Stroup, A. M., Maiti, A., Harris, G., Lynch, S. M., Vucetic, S., & Henry, K. A. (2020). Socioeconomic Disparities in Colon Cancer Survival: Revisiting Neighborhood Poverty using Residential Histories. *Epidemiology*. doi:10.1097/ede.0000000000001216

SOLUTION

NJSCR initially set out to improve address data with the sole purpose of correcting and verifying address at diagnosis to ensure that the case was counted as a New Jersey incident case and to improve geocoding quality. To do so, we secured 7-10 individual subscription licenses to access the LN Accurint database online. Staff were trained to conduct individual lookups in order to edit address information in the registry database management system (SEER*DMS). This lookup process has since expanded to correcting and verifying other key demographic data including date of birth, social security number, name, and phone number.

Individual lookups, however, are not sustainable if there are hundreds or thousands of records that are missing data for key demographic variables such as race (Table 1). NJSCR began to address this issue in 2020 by partnering with LN to conduct a batch linkage to acquire race information from Driver's License and Voter Registration records. After establishing a linkage agreement and setting up a secure, encrypted file transfer protocol between the registry and LN, we conducted a pilot test of N=400 cases diagnosed in 2018 with known race. We used the pilot data to assess linkage patterns and evaluate the quality of the race codes returned by LN. We then provided a larger dataset of 10,000 records with unknown race (RACE 1 = 99) spanning diagnosis year 2013-2020.

To leverage population-based data from cancer registries and enhance residential history data, we conducted a batch linkage with LN to acquire residential history data for adult (18+) Non-Hodgkin Lymphoma (NHL) and colon cancers diagnosed from 2006-2014.

RESULTS

Unknown Race – Of the 400 patients with known race that we sent to LN as a test file, 103 (25.75%) matched to at least 1 driver's license (DL) record, 297 (74.25%) matched to at least 1 voter registration (VR) record, and 318 (79.5%) matched to either a DL or VR record. Unfortunately, of the 318 matched records, LN was only able to return 10 records with race from a DL or VR record, a return of only 3.1%. However, 8 of 10 records (80%) were concordant with NJSCR data as White (N=7) and Black (N=1). The remaining 2 records were classified as Other by LN (vs. White in SEER*DMS) and Hispanic (vs. White in SEER*DMS). The larger dataset of 10,000 records were sent to LN for linkage and is still pending.

Residential Histories – We submitted 17,067 colon cancer and 12,121 NHL cases for our project. LN returned up to 20 addresses for each of 11,490 NHL (98.5%) and 16,750 colon (98.1%) cancer cases. Few (n=58) had no geographic information available from NJSCR or LN. We used these data to conduct a seminal study using residential history data to evaluate socioeconomic disparities in regional stage colon cancer (RSCC) survival, which was recently published in *Epidemiology* (5). We used census track poverty (<5%, 5-10%, 10-20%, > 20%) as a measure of area-based SES and conducted multivariable Cox Proportional Hazards time-varying models to estimate the hazards of RSCC death. Of the 4,041 RSCC cases, 92 (2%) had no LN residential information, 65% remained in the same census tract from diagnosis to date of last contact or date of death (non-movers), 19% changed census tracts once, and 16% changed census tracts two or more times. Additionally, 42% of cases that lived in the highest poverty areas moved, compared to only 32% in the lowest poverty areas. We also found that Non-Hispanic Blacks (44%), Asian/Pacific Islanders (43%), and Hispanics (38%) were more likely to move than Non-Hispanic Whites (32%). After adjusting for age, sex, regional stage sub-category, race/ethnicity, and mover status, our time-varying models showed that RSCC cases in the highest poverty areas had a 30% greater risk of colon cancer death than those living in the lowest poverty areas (HR=1.30 95% CI 1.04-1.64). When the analysis was limited to movers (i.e., changed census tracts), the hazard of colon cancer death increased to 1.62 (95% 1.07-2.44).

SUSTAINING SUCCESS

Utilization of LN data is ongoing and remains an integral source to verify important patient data. We will build on this effort in the coming weeks by evaluating the quality of the linkages and the race information returned on 10,000 records; use updated race and address information to create more reliable geographic- and race-specific cancer statistics; and, share our results with New Jersey's Task Force on Cancer Control and Prevention to develop public health priorities. We also plan to import our NHL and colon cancer residential histories into SEER*DMS and publish our findings for NHL subtypes.

REGISTRY CONTACT INFORMATION

New Jersey State Cancer Registry
609-633-0500
https://www.state.nj.us/health/ces/reporting-entities/njscr/