

Residential Mobility and Geospatial Disparities in Colon Cancer Survival

Daniel Wiese¹, Antoinette M. Stroup^{2,3}, Aniruddha Maiti⁴, Gerald Harris², Shannon M. Lynch⁵, Slobodan Vucetic⁴, and Kevin A. Henry^{1,5}



ABSTRACT

Background: Identifying geospatial cancer survival disparities is critical to focus interventions and prioritize efforts with limited resources. Incorporating residential mobility into spatial models may result in different geographic patterns of survival compared with the standard approach using a single location based on the patient's residence at the time of diagnosis.

Methods: Data on 3,949 regional-stage colon cancer cases diagnosed from 2006 to 2011 and followed until December 31, 2016, were obtained from the New Jersey State Cancer Registry. Geographic disparity based on the spatial variance and effect sizes from a Bayesian spatial model using residence at diagnosis was compared with a time-varying spatial model using residential histories [adjusted for sex, gender, substage, race/ethnicity, and census tract (CT) poverty]. Geographic estimates of risk of colon cancer death were mapped.

Results: Most patients (65%) remained at the same residence, 22% changed CT, and 12% moved out of state. The time-varying model produced a wider range of adjusted risk of colon cancer death (0.85–1.20 vs. 0.94–1.11) and resulted in greater geographic disparity statewide after adjustment (25.5% vs. 14.2%) compared with the model with only the residence at diagnosis.

Conclusions: Including residential mobility may allow for more precise estimates of spatial risk of death. Results based on the traditional approach using only residence at diagnosis were not substantially different for regional stage colon cancer in New Jersey.

Impact: Including residential histories opens up new avenues of inquiry to better understand the complex relationships between people and places, and the effect of residential mobility on cancer outcomes.

See related commentary by Williams, p. 2107

Introduction

Identifying cancer survival disparities is critical to focus interventions in specific populations and for clinical and cancer control programs to prioritize efforts with limited resources. Examining whether cancer survival varies geographically can also generate hypotheses about the underlying causes of survival disparities and help identify important demographic and neighborhood factors that play a role. However, it is not known whether incorporating the residential mobility after a cancer diagnosis into spatial models results in different geographic patterns of survival compared with the standard approach of using a single location based on the patient's residence at the time of diagnosis.

Numerous studies have examined associations between cancer survival and neighborhood-based socioeconomic measures, such as poverty. These studies have shown that patient survival is generally worse among patients in poor neighborhoods compared with patients living in more affluent neighborhoods (1–5). Studies specifically examining spatial patterns in survival using geographic methods, such

as spatial scan statistics (4, 6) and spatial models (5, 7, 8), have also noted worse survival clustered in poor communities. To date, studies examining geographic disparities in cancer survival used the patient's residence at the time of diagnosis exclusively to map geographic distributions of patient survival or to define neighborhood characteristics.

Limitations of using a single location have been well documented (9–13), including potentially introducing spatial uncertainty because it assumes the person remained at the same residence and in turn experienced uniform neighborhood-based contexts over time (e.g., unchanging CT-poverty). This may be a potential source of bias in nonspatial and spatial statistical models, calling into question both the accuracy and precision of survival estimates associated with geographic factors. The magnitude and direction of bias largely depends on the volume of residential mobility in the study population, the degree to which mobility results in changes in neighborhood context (e.g., higher or lower poverty rates), and whether the neighborhoods themselves are changing overtime (e.g., gentrification). If a large proportion of cases are moving to neighborhoods with similar characteristics, then bias would be minimal. Failure to account for residential socio-spatial mobility may also impede researchers' ability to account for self-selection into neighborhoods, which may change the overall risk of death in a neighborhood over time (14, 15).

The main reason researchers have not incorporated residential mobility into geographic studies of cancer survival or other outcomes (e.g., stage at diagnosis) is that cancer registries do not collect residential histories. The primary source of location data for cancer patients collected by registries is the residential address at the time of diagnosis. However, recent studies have demonstrated the potential utility of residential histories collected from sources such as electronic medical records (16) and commercial and administrative databases to document residential mobility (16–19). There are numerous examples of applying residential histories in environmental epidemiology for assessing longitudinal exposures (20–24) and risk and geographic

¹Department of Geography and Urban Studies, Temple University, Philadelphia, Pennsylvania. ²New Jersey Department of Health, New Jersey State Cancer Registry, Trenton, New Jersey. ³Rutgers Cancer Institute of New Jersey and Rutgers School of Public Health, Rutgers University, New Brunswick, New Jersey. ⁴Department of Computer and Information Sciences, Temple University, Philadelphia, Pennsylvania. ⁵Division of Cancer Prevention and Control, Fox Chase Cancer Center, Philadelphia, Pennsylvania.

Corresponding Author: Daniel Wiese, Temple University, 1115 West Polett Walk, 308 Gladfelter Hall, Philadelphia, PA 19122. Phone: 215-204-3386; E-mail: daniel.wiese@temple.edu

Cancer Epidemiol Biomarkers Prev 2020;29:2119–25

doi: 10.1158/1055-9965.EPI-20-0772

©2020 American Association for Cancer Research.

clustering of cancers (25–29), but far less use of residential histories to examine disparities in cancer outcomes, such as survival (30, 31). To our knowledge, residential histories have not been used by population-based cancer registries to examine geographic variation in cancer survival.

To address the uncertain geographic context problem as it relates to place-based disparities research (32) and evaluate how uncertainty in a patient's residential location after diagnosis might impact survival estimates and geographic patterns, we incorporated residential histories into Bayesian spatial models to identify whether the survival of patients diagnosed with regional stage colon cancer in New Jersey varied geographically after adjusting for key individual-level risk factors and neighborhood poverty level. We compared differences in geographic variation in colon cancer survival between models using residence at the time of diagnosis alone to models that incorporated residential mobility, accounting for whether the patients moved out of state after diagnosis.

Materials and Methods

Study population

Colon cancer cases were obtained from the New Jersey State Cancer Registry (NJSCR), which is a population-based cancer registry established in October 1978 to monitor cancer among the more than 8.9 million residents of New Jersey (33). The study population includes all New Jersey residents 18 years and older diagnosed between January 1, 2006, and December 31, 2011, with histologically confirmed, first primary regional stage colon cancer as defined according to the International Classification of Diseases for Oncology, Third Edition (ICD-O3 C180-C189, C260; excluding histology codes 9050-9055, 9140, 9590-9992; ref. 34; $N = 4,041$). The Rutgers University institutional review board (IRB) approved all study activities.

Regional stage was defined according to Surveillance, Epidemiology, and End Results (SEER) derived summary stage 2000 and includes regional, direct extension only, regional lymph nodes only, and regional, direct extension and regional lymph nodes. This definition includes the American Joint Committee on Cancer (AJCC) 6th Edition (35) groupings IIA, IIB, IIIA, IIIB, and IIIC. The study sample was limited to regional stage colon cancers to control the influence that stage will have on the survival estimates. Doing so allows for ease of interpretation of the geographic disparities that may otherwise be attributable to spatial variation in stage.

Individual-level factors included age at diagnosis, gender (male, female), and race/ethnicity [non-Hispanic (NH) White, NH Black, NH Asian/Pacific Islander (API), NH other, and Hispanic (any race)]. Cases with any stage at diagnosis other than regional stage or had no survival time (i.e., only ascertained through death certificates or autopsy) were excluded from analysis (36). Vital status, including date of death and cause of death (if deceased) or date of last contact (if alive) is updated routinely by the NJSCR through linkages with state and national sources including death data from the New Jersey Department of Health Office of Vital Statistics and Registry and the National Death Index, hospital discharge files, Centers for Medicare and Medicaid Services, Social Security Administration Services for Epidemiologic Researchers, and motor vehicle registration files. Cases were followed until their deaths or until December 31, 2016. Deaths attributed to colon cancer were coded based on ICD-10 code C18 (34).

Residential history data from LexisNexis

Residential histories were acquired through a data linkage with a commercial database developed by LexisNexis, Inc. LexisNexis

(<https://www.lexisnexis.com/en-us/products/public-records.page>) has developed a specific resource for researchers to obtain residential histories for adults aged 18 and older (16, 18, 19, 37). Up to 20 of the most recent addresses between 1946 and 2018 with documented start and stop dates were returned for each case. All residential addresses were geocoded to the 2010 census tract (CT) boundaries using the North American Association of Central Cancer Registries (NAACCR) AGGIE Geocoder (38).

Of the 4,041 regional stage colon cancer cases, 98 (2.4%) had no residential information available. For the remaining 3,949 (97.6%), we applied a preprocessing technique described previously (31).

Neighborhood socioeconomic status data

Similar to other studies (4, 39), we used the percentage of the CT population 18 years and older living below the federal poverty level as a measure of neighborhood deprivation (CT-poverty). CT-poverty was obtained from publicly available U.S. Census and American Community Survey (ACS) data. U.S. Census 2010 and the ACS 5-year average data 2006–2010, 2007–2011, 2008–2012, 2009–2013, and 2010–2014 were used for residencies between 2006 and 2010. ACS 2011–2015, 2012–2016, and 2013–2017 were used for residencies between 2011 and 2016. For each case, every residential record received a corresponding CT-poverty value based on the earliest date of the residential appearance in the data set. In situations when cases remained at a single CT over multiple years, we assigned the annual CT-poverty values to capture changes within the neighborhood (e.g., changes due to gentrification). We developed two different poverty measurements. The first measurement used the standard CT-poverty from the location at the time of diagnosis (DX CT-poverty). For the second, all corresponding CTs were included; therefore, each patient could have multiple New Jersey CTs during the follow-up period (time-varying CT-poverty).

Statistical analysis

Patient survival times were calculated in months as the difference between the date of diagnosis and the date of last contact or death. Patients were censored at the date of death if they died from causes other than colon cancer, the date the patient was lost to follow-up, or at the end of the follow-up period (December 31, 2016), whichever occurred first. In the time-varying model, cases were also censored at the time they moved out of the State of New Jersey; and, assigned start and end dates for every CT location, and calculated the survival time spent at each CT. In other words, start and end dates set the follow-up intervals and corresponding covariate values (e.g., CT-poverty), as well as the vital status at the end of every interval (1 = dead, 0 = alive).

Bayesian geospatial models were applied to survival time as an extension of conventional Cox regression survival models described by Kneib and Fahrmeir (40). This model includes the geographic location of the patient's CT, which is used as the spatial function and estimates the spatial effect. The spatial function provides a way to estimate the geographic variation in the risk of death after controlling for individual- and neighborhood-level covariates (41). It measures the instantaneous event rate or the probability that an individual would experience an event (e.g., death from regional stage colon cancer) after diagnosis. The spatial function is based on stationary Gaussian random fields (bivariate penalized splines). P-spline smoothed spatial effects were incorporated into the model based on an adjacency matrix of geographic neighbors by CTs (weights based on rook's case; refs. 42–44). The risk of death is the exponentiated smoothed posterior mean for the CT based on those living in that CT. For the standard DX

CT-poverty model, the spatial effect was estimated using the CT from the time of the diagnosis. To implement the time-varying CT-poverty model, we redesigned the model as suggested by Belitz and colleagues (43) for time-dependent covariates, while including only New Jersey CTs.

Estimation of regression models is based on Markov chain Monte Carlo simulation techniques, corresponding to full Bayesian inference, and obtained by specifying prior distributions for all unknown parameters. For each model, 10,000 iterations were run, with the first 2,000 samples used as a burn-in. Every 20th sample from the remaining 8,000 samples was saved and used to construct the posterior distribution for each of the parameter estimates in the model. The 95% confidence intervals (CI) were calculated on the basis of the posterior distribution of the 1,000 samples to identify significant risk of death and the CTs with significantly higher or lower estimates than the state average (<1 lower risk, >1 higher risk). All models were implemented with R using BayesX (45), BayesXsrc (42), and R2BayesX (46, 47) packages. The exponentiated spatial effects of each CT from the models were mapped to visualize the risk of death from regional stage colon cancer. Then, we calculated a difference map, highlighting differences between the risk of death estimates using CT-poverty from diagnosis and time-varying values. In addition, we compared models using geographic disparity (GD) percentage, a method originally proposed by Chien and colleagues (7), to assess the geographic variance in regional stage colon cancer survival after accounting for DX CT-poverty and time-varying CT-poverty. It is calculated as the square root of the spatial variance, where higher GD suggests wider geographic variability unexplained by independent variables, and assumes larger geographic disparities in the study area. We also compared model's fit using deviance information criterion (DIC). DIC is defined as the sum of the posterior expected deviance and the effective number of parameters. A better model fit is assumed with a lower value of DIC (48).

Results

Study population

Table 1 summarizes the characteristics of the study population ($n = 3,949$). The average age at diagnosis was 66. There were fewer males (47.6%) than females (52.4%). Around three quarters (73.6%) of the study population were NH-White, 12.4% NH-Black, 8.1% Hispanic origin (any race), 3.6% NH-API, and 2.4% Other race. Approximately 27.5% of all the patients died from the colon cancer by the end of follow-up, with a median survival of 66 months (range 1–139).

During the follow-up period, 65.5% remained at their diagnosis CT, 22.4% changed CTs within New Jersey, and 12.1% left New Jersey during the study period. Among those who moved, 18.5% only moved once, 12.3% moved twice, and 3.6% moved 3 or more times after cancer diagnosis. The average time spent at the CT at diagnosis was 7.5 years. At the time of cancer diagnosis, nearly half (46%) lived in CTs with <5% poverty and 25.9% lived in CTs with poverty rates 10% or higher.

Geographic clustering and spatial effects

Table 2 summarizes the statewide range of risk of death and geographic disparities percentage for each of the spatial models (Model a using DX CT-poverty and Model b using time-varying CT-poverty). The DX CT-poverty model produced fewer CT-specific risk of death estimates ($N = 1,604$) compared with the time-varying CT-poverty model with residential histories ($N = 1,742$). The time-varying model includes all the CT locations where the cases lived, resulting in more CTs where risk of death from regional stage colon cancer can be estimated. The adjusted risk estimates in Model a ranged from

Table 1. Study population characteristics.

	Overall ($n = 3,949$)
Age	
Mean (SD)	65.8 (13.3)
Median (min-max)	68.0 (21.0–85.0)
Gender	
Male	1,878 (47.6%)
Female	2,071 (52.4%)
Race/ethnicity	
NH-White	2,902 (73.5%)
NH-Black	488 (12.4%)
Hispanic (any race)	325 (8.2%)
NH-API	141 (3.6%)
Other	93 (2.4%)
Regional stage subcategory	
Regional, direct extension only	1,339 (33.9%)
Regional, lymph nodes only	1,268 (32.1%)
Regional, both	1,342 (34.0%)
Vital status	
Censored	2,862 (72.5%)
Colon cancer death	1,087 (27.5%)
Survival time (months)	
Mean (SD)	62.3 (38.0)
Median (min-max)	66.0 (1.00–139)
CT changes (type of “moves”)	
CT at date of diagnosis only	2,587 (65.5%)
Change in residential CT within NJ	885 (22.4%)
Change in residential CT outside NJ	477 (12.1%)
CT changes (number of “moves”)	
Single residency	2,587 (65.5%)
1 Change in residential CT	732 (18.5%)
2 Changes in residential CT	486 (12.3%)
3+ Changes in residential CT	144 (3.6%)
CT-poverty at diagnosis	
<5% (low)	1,816 (46.0%)
5%–<10%	1,110 (28.1%)
10%–<20%	688 (17.4%)
≥20% (high)	335 (8.5%)

Abbreviations: API, Asian/Pacific Islander; CT, census tract; NH, non-Hispanic; NJ, New Jersey; SD, standard deviation.

0.94 to 1.11—a 0.17 point range from low to high based on survival data for 1,604 CTs. Whereas, the adjusted risk estimates in Model b ranged from 0.83 to 1.2, which is a 0.37 point range from low to high based on survival data for 1,742 CTs. The GD percentage was 14.2% and 25.5% in Model a and Model b, respectively. The DIC indicated a better fit in Model a (14,269) compared with Model b (16,489). Fixed effects of CT-poverty were similar between the models and not significant at $P < 0.05$ level.

The adjusted risk of death from colon cancer from each model was mapped (**Fig. 1**) to compare geographic distribution of adjusted risk estimates from Model a (**Fig. 1A**) and Model b (**Fig. 1B**). For both models, there were no statistically significant areas of higher or lower risk of colon cancer death. However, risk estimates for the southern and northwestern regions as well as the northeastern metropolitan area that borders New York City clearly become more pronounced when residential mobility is included in the model. For example, in the southern region, the risk of colon cancer death ranged from 1.02 to 1.05 in the DX CT-poverty model and increased to 1.08–1.2 in the time-varying CT-poverty model. In the northeastern region, the area with the highest risk expands to a larger geographic area around the Newark metropolitan area. Similarly, in the northwestern region, areas with the

Table 2. Model comparison.

Models	Number of CTs	Range of adjusted risks of death	GD	DIC	CT-poverty fixed effect (95% CI)
Model a (DX CT-poverty) ^a	1,604	0.94-1.11	14.2%	14,269	1.003 (0.996-1.011)
Model b (time-varying CT-poverty) ^b	1,742	0.83-1.2	25.5%	16,489	1.003 (0.994-1.012)

Abbreviations: CI, confidence interval; CT, census tract; DIC, deviance information criterion; GD, geographic disparities.

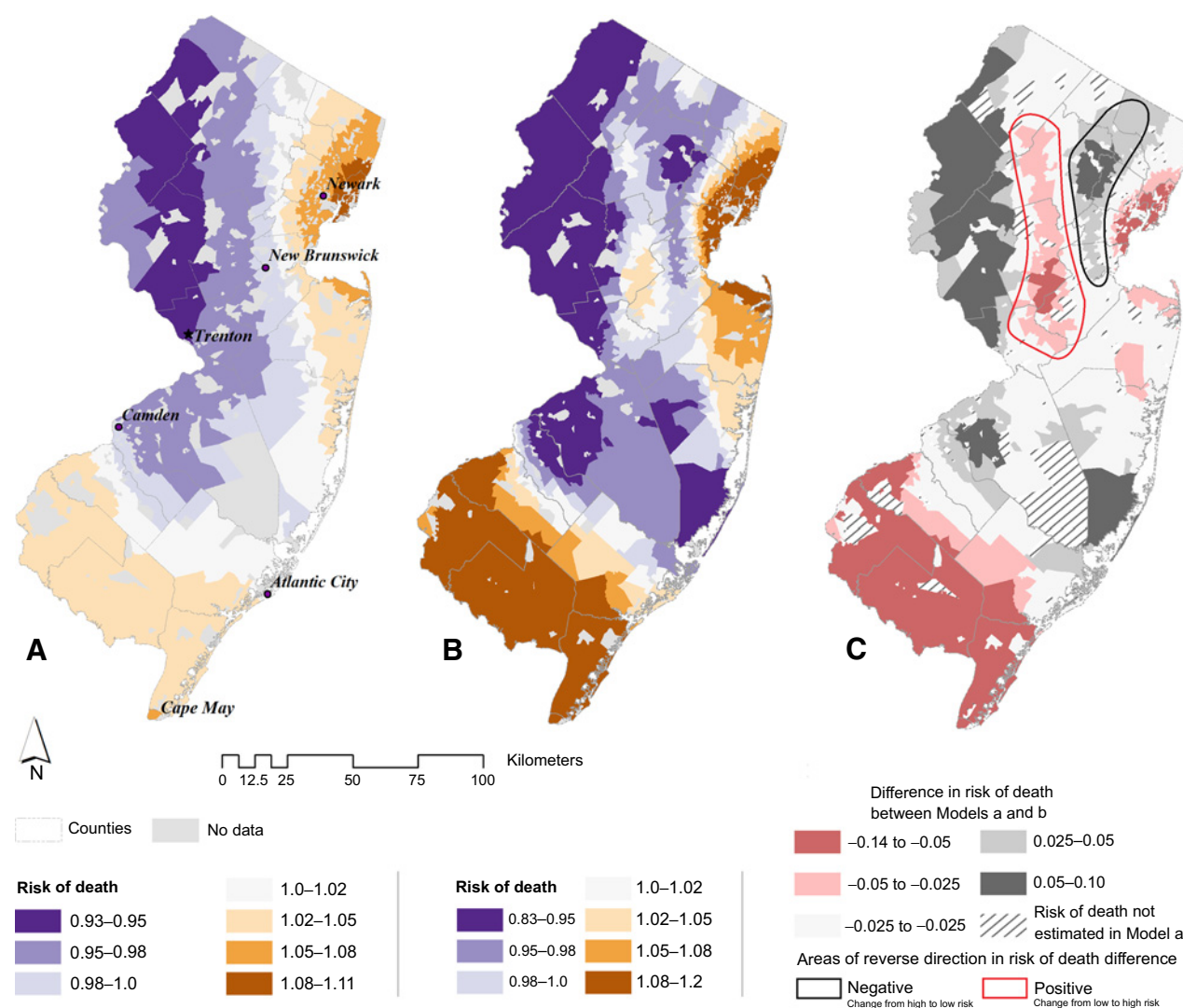
^aAdjusted for sex, age, regional substage at diagnosis, race/ethnicity, and CT-poverty at CT of diagnosis.

^bAdjusted for sex, age, regional substage at diagnosis, race/ethnicity, change in CT (mover status), and time-varying CT-poverty using residential histories.

lowest risk expand geographically. The range of risks in the lowest and upper most categories, however, widen from 0.93–0.95 (Fig. 1A) to 0.83–0.95 (Fig. 1B) and 1.08–1.11 (Fig. 1A) to 1.08–1.2 (Fig. 1B).

Figure 1C summarizes the difference in the risk of death estimates between the adjusted DX CT-poverty model (1a) and the adjusted time-varying CT-poverty model (1b). The negative values indicate

areas where the risk of death estimates from Model b were higher than the risk estimates from Model a (DX CT-poverty < time-varying CT-poverty). Whereas, positive values indicate areas where the risk of death estimates from Model b were lower than the risk estimates from Model a (DX CT-poverty > time-varying CT-poverty). Overall, the range of differences in risk of death estimates were small, ranging from



-0.14 and +0.10. There were two areas in North Jersey where the estimates in Model a and Model b change direction. In the area around parts of Somerset County, noted in the pink solid boundary, the risk of colon cancer death changes from low (<1.0) in Model a to high (>1.0) in Model b. Whereas, the area around Morris County and along the border of Essex County, in the solid gray boundary, the risk of colon cancer death changes from high (>1.0) in Model a to low (<1.0) in Model b.

Discussion

To our knowledge, this is the first geospatial cancer survival study that combined residential history data with cancer surveillance data from a population-based cancer registry. We used residential history data to characterize changes in neighborhood poverty level postdiagnosis and applied it to geospatial survival models. Cancer disparities research using population-based cancer registry data traditionally use residential location at the time of diagnosis to assign neighborhood measures of socioeconomic or sociodemographic status and to conduct geographic analysis. The limitation of this approach is that it assumes patients remain at the same location during the follow-up period after diagnosis. This approach contributes to the uncertain geographic context problem, which arises from the lack of knowledge about the appropriate spatial and temporal configuration for assessing the influence of the environment on health outcomes (49).

In this study, we used spatial modeling techniques to examine geographic variation in regional stage colon cancer survival, and focused on whether the results based on the conventional “residence-at-diagnosis” approach to estimate area poverty would differ from results based on a time-varying model that incorporated changing poverty estimates from residential histories. After adjusting for age, gender, race/ethnicity, regional sub-stage, CT change, and CT-poverty, the time-varying geospatial model produced notable differences in the geographic patterns in risk of death as evidence by the increase in the effect sizes of high (South Jersey) and low (Northwest Jersey) risk and the expansion of affected geographic areas (Northeast Jersey). When we examined the spatial variance of each approach, we found that the standard model using poverty level at diagnosis alone explained more of the geographic disparity (i.e., lower spatial variance) than the time-varying model with changing CT-poverty levels; suggesting that standard approaches using residence at diagnosis explained a greater proportion of the geographic disparities in the risk of death from colon cancer in New Jersey. This finding could be a result of the length of the follow-up time, censoring of cases that moved out of the state during the follow-up period, and removing cases from CTs as they move in and out of the area over time. Although we followed cases for 11 years, this may not be long enough for residential mobility to have an impact on survival time. It is likely that the use of a single CT-poverty measure captured at the time of diagnosis is sufficient in this population. In the standard time-to-event models, all cases would have contributed survival time in the same CT until the end of follow-up, death, or lost to follow-up. Whereas, in the time-varying model, survival time for each case will vary by the number of residential locations from diagnosis to the end of follow-up and the length of time in each location. Decreasing both the number of observations and follow-up time within CTs may add to larger variance (i.e., geographic disparity) and loss in precision (i.e., changing effect sizes).

Our finding that, in some areas in New Jersey, the risk estimates of colon cancer death changed in direction from low risk to high risk or from high risk to low risk after incorporating residential history data

requires further research to assess the accuracy of estimates in these geospatial models. However, despite this finding, the regional patterns and effect sizes were not entirely dissimilar between the two models. This is due to the relative stability of individuals postdiagnosis (65.5% remained in the same CT). There was also minimal socio-spatial mobility as cases often moved to CTs with similar levels of poverty (68% of the cases that changed CT during the follow-up period had an absolute change in CT-poverty of less than 10 percentage points). Indeed, researchers using either model would have come to the same general conclusion, but future work using simulated and real data should examine the impact of differential socio-spatial mobility on spatial and nonspatial models.

Our study had several strengths. We used high quality, population-based cancer surveillance data from the NJSCR, which conducts follow-up for all cases diagnosed in the State of New Jersey. With an average follow-up rate of 97%, our study population is less prone to bias due to loss to follow-up. We were also able to obtain residential histories on 97.6% of our study population, limiting selection bias. With residential history data, we were able to adjust our models for out migration of patients after diagnosis. Previous studies using population-based cancer registry data do not exclude cases moving out of the study area (6, 50), which may lead to an underestimation of the true risk of death among New Jersey residents. In this analysis, about 12% of the cases moved out of state before the end of follow-up. The majority of these cases had higher 5-year survival rates compared with those that remained in New Jersey (82% vs. 72%); and, were from less impoverished areas (average poverty 7.6% vs. 8.1%). Future research should look more carefully at potential biases in statewide survival estimates from not accounting for residential mobility. Our study also produced consistent results with previous studies as regards individual-level factors such as age, gender, race/ethnicity associated with survival outcomes (4, 51–54) as well the fixed-effect of area-based poverty as cases living in the high-poverty neighborhoods had a higher risk of death from colon (7, 55) and other cancers (8, 51, 56–58).

Our study was limited to New Jersey residents diagnosed with regional stage colon cancer from 2006 to 2011, and may not be generalizable to other state populations with different geographic, socioeconomic, and demographic profiles, nor to local or distant stage colon cancer or other cancer sites. Results may also be nongeneralizable to patient populations that were diagnosed before or after our study cohort due to changes in both treatment regimens for colon cancer and area-level socioeconomic status. We also relied solely on residential histories collected from LexisNexis and did not include self-reported information that may be used to validate and/or augment LexisNexis data. However, recent studies found good concordance (82%–92%) between LexisNexis addresses and addresses collected from study participants (16, 17). For our study, data concordance between the LexisNexis locations and the cases' CTs at time of diagnosis reported by the NJSCR was around 83%, and it increased to 93% when comparisons were limited to a 6-month window before and after the diagnosis date. Although these potential data errors could have contributed to nondifferential misclassification, the extent of the bias in either direction would likely be minimal due to the low proportion of cases affected.

The results are also limited by the available follow-up time (maximum 11 years). Longer follow-up could increase the number of observed deaths as well as increase the number of potential moves over the follow-up period. If more people moved and we were looking at long-term survival (e.g., 20–25 years), then residential histories may have a larger impact on predicting survival and explaining the spatial variance. The study also did not include

individual-level measures of poverty or income, which previous studies have shown to play a significant role in determining colon cancer survival outcomes (5, 59, 60)

Residential locations and CT-poverty prior to the diagnosis could have also impacted survival outcomes. We had access to residences prior to diagnosis, but in this study, we were limited by the specific spatial model employed, which required a start at the New Jersey diagnosis location. We are presently exploring alternative spatial models and approaches to assess the role of residential mobility prior to diagnosis on colon cancer survival outcomes.

Finally, the neighborhood geographic measure of poverty was based on CTs. A change in geographic unit (CT to block group) may result in different conclusions than those that are based on CTs (i.e., modifiable areal unit problem; refs. 32, 61, 62).

In conclusion, including residential histories opens up new and important avenues of inquiry to better understand the complex relationships between people and places, and to evaluate the effects of residential mobility on cancer outcomes. Our findings suggest that residential mobility after regional stage colon cancer diagnosis has the potential to change geographic patterns of the risk of colon cancer death, increasing the spatial variance and ultimately leading to greater unexplained geographic disparities. Including residential mobility may also allow for more precise spatial risk of death estimates. However, in this study, which was limited in both geographic scope and cancer type, the results based on the traditional approach using only address at diagnosis were not substantially different from the approach using residential mobility data. Both approaches provide

relevant results for understanding cancer survival disparities and prioritizing interventions.

Disclosure of Potential Conflicts of Interest

G. Harris reports grants from National Science Foundation during the conduct of the study. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

D. Wiese: Conceptualization, formal analysis, validation, investigation, visualization, methodology, writing—original draft, writing—review and editing. **A.M. Stroup:** Data curation, funding acquisition, validation, project administration, writing—review and editing. **A. Maiti:** Validation, methodology, writing—review and editing. **G. Harris:** Data curation, validation, writing—review and editing. **S.H. Lynch:** Validation, writing—review and editing. **S. Vucetic:** Funding acquisition, validation, writing—review and editing. **K.A. Henry:** Conceptualization, data curation, supervision, funding acquisition, validation, project administration, writing—review and editing.

Acknowledgments

This work was supported by the National Science Foundation (1560888). NJSCR data were collected using funding from NCI and the Surveillance, Epidemiology, and End Results (SEER) Program (HHSN2612013000211), the CDC's National Program of Cancer Registries (NPCR; 5U58DP003931-02/NU58DP006279-02-00), as well as the State of New Jersey and the Rutgers Cancer Institute of New Jersey.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 19, 2020; revised June 24, 2020; accepted July 29, 2020; published first August 5, 2020.

References

- Warner ET, Gomez SL. Impact of neighborhood racial composition and metropolitan residential segregation on disparities in breast cancer stage at diagnosis and survival between black and white women in California. *J Community Health* 2010;35:398–408.
- Feinglass J, Ryzdzewski N, Yang A. The socioeconomic gradient in all-cause mortality for women with breast cancer: findings from the 1998 to 2006 National Cancer Data Base with follow-up through 2011. *Ann Epidemiol* 2015;25:549–55.
- Gomez SL, Shariff-Marco S, DeRouen M, Keegan THM, Yen IH, Mujahid M, et al. The impact of neighborhood social and built environment factors across the cancer continuum: current research, methodological considerations, and future directions. *Cancer* 2015;121:2314–30.
- Henry KA, Niu X, Boscoe FP. Geographic disparities in colorectal cancer survival. *Int J Health Geogr* 2009;8:48.
- Lian M, Schootman M, Doubeni CA, Park Y, Major JM, Stone RA, et al. Geographic variation in colorectal cancer survival and the role of small-area socioeconomic deprivation: a multilevel survival analysis of the NIH-AARP Diet and Health Study cohort. *Am J Epidemiol* 2011;174:828–38.
- Niu X, Roche LM, Pawlish KS, Henry KA. Cancer survival disparities by health insurance status. *Cancer Med* 2013;2:403–11.
- Chien LC, Schootman M, Pruitt SL. The modifying effect of patient location on stage-specific survival following colorectal cancer using geosurvival models. *Cancer Causes Control* 2013;24:473–84.
- Wiese D, Stroup AM, Crosbie A, Lynch SM, Henry KA. The impact of neighborhood economic and racial inequalities on the spatial variation of breast cancer survival in New Jersey. *Cancer Epidemiol Biomarkers Prev* 2019;28:1958–67.
- Boscoe FP, Ward MH, Reynolds P. Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer. *Int J Health Geogr* 2004;3:28.
- Schootman M, Gomez SL, Henry KA, Paskett ED, Ellison GL, Oh A, et al. Geospatial approaches to cancer control and population sciences. *Cancer Epidemiol Biomarkers Prev* 2017;26:472–5.
- Tong S. Migration bias in ecologic studies. *Eur J Epidemiol* 2000;16:365–9.
- Jacquez GM, Kaufmann A, Meliker J, Goovaerts P, AvRuskin G, Nriagu J. Global, local and focused geographic clustering for case-control data with residential histories. *Environ Health* 2005;4:4.
- Wheeler DC, Ward MH, Waller LA. Spatial-temporal analysis of cancer risk in epidemiologic studies with residential histories. *Ann Assoc Am Geogr* 2012;102:1049–57.
- Bergström L, Van Ham M. Understanding neighbourhood effects: Selection bias and residential mobility. In: Van Ham M, Manley D, Bailey N, Simpson L, Maclennan D, editors. *Neighbourhood effects research: new perspectives*. New York: Springer; 2010. p. 79–99.
- Clark WA, Van Ham M, Coulter R. Spatial mobility and social outcomes. *J Hous Built Environ* 2014;29:699–727.
- Hurley S, Hertz A, Nelson DO, Layefsky M, Von Behren J, Bernstein L, et al. Tracing a path to the past: exploring the use of commercial credit reporting data to construct residential histories for epidemiologic studies of environmental exposures. *Am J Epidemiol* 2017;185:238–46.
- Jacquez GM, Slotnick MJ, Meliker JR, AvRuskin G, Copeland G, Nriagu J. Accuracy of commercially available residential histories for epidemiologic studies. *Am J Epidemiol* 2011;173:236–43.
- Wheeler DC, Wang A. Assessment of residential history generation using a public-record database. *Int J Environ Res Public Health* 2015;12:11670–82.
- Stinchcomb D, Roeser A. NCI/SEER residential history project technical report. Rockville (MD): Westat, Inc.; 2016. Contract No.: HHSN261201500371P. Sponsored by the National Cancer Institute. Available from: https://www.westat.com/sites/default/files/NCISAS/NCI_Res_Hist_Proj_Tech_Rpt_v2sec.pdf.
- Brokamp C, LeMasters GK, Ryan PH. Residential mobility impacts exposure assessment and community socioeconomic characteristics in longitudinal epidemiology studies. *J Expo Sci Environ Epidemiol* 2016;26:428.
- Gallagher LG, Webster TF, Aschengrau A, Vieira VM. Using residential history and groundwater modeling to examine drinking water exposure and breast cancer. *Environ Health Perspect* 2010;118:749–55.
- Hughes AE, Pruitt SL. The utility of EMR address histories for assessing neighborhood exposures. *Ann Epidemiol* 2017;27:20–6.

23. Hystad P, Demers PA, Johnson KC, Carpiano RM, Brauer M. Long-term residential exposure to air pollution and lung cancer risk. *Epidemiology* 2013;24:762–72.
24. Meliker JR, Slotnick MJ, AvRuskin GA, Schottenfeld D, Jacquez GM, Wilson ML, et al. Lifetime exposure to arsenic in drinking water and bladder cancer: a population-based case-control study in Michigan, USA. *Cancer Cause Control* 2010;21:745–57.
25. Jacquez GM, Meliker JR, Avruskin GA, Goovaerts P, Kaufmann A, Wilson ML, et al. Case-control geographic clustering for residential histories accounting for risk factors and covariates. *Int J Health Geogr* 2006;5:32.
26. Meliker JR, Jacquez GM. Space-time clustering of case-control data with residential histories: insights into empirical induction periods, age-specific susceptibility, and calendar year-specific effects. *Stoch Environ Res Risk Assess* 2007;21:625–34.
27. Sloan CD, Nordsborg RB, Jacquez GM, Raaschou-Nielsen O, Meliker JR. Space-time analysis of testicular cancer clusters using residential histories: a case-control study in Denmark. *PLoS One* 2015;10:e0120285.
28. Nordsborg RB, Sloan CD, Shahid H, Jacquez GM, De Roos AJ, Cerhan JR, et al. Investigation of spatio-temporal cancer clusters using residential histories in a case-control study of non-Hodgkin lymphoma in the United States. *Environ Health* 2015;14:48.
29. Nordsborg RB, Meliker JR, Ersboll AK, Jacquez GM, Poulsen AH, Raaschou-Nielsen O. Space-time clusters of breast cancer using residential histories: a Danish case-control study. *BMC Cancer* 2014;14:255.
30. Zhang D, Matthews CE, Powell-Wiley TM, Xiao Q. Ten-year change in neighborhood socioeconomic status and colorectal cancer. *Cancer* 2019;125:610–7.
31. Wiese D, Stroup A.M., Maiti A., Harris G., Lynch S.M., Vucetic S., et al. Socioeconomic disparities in colon cancer survival: revisiting neighborhood poverty using residential histories. *Epidemiology* 2020;31:728–35.
32. Sahar L, Foster SL, Sherman RL, Henry KA, Goldberg DW, Stinchcomb DG, et al. GScience and cancer: state of the art and trends for cancer surveillance and epidemiology. *Cancer* 2019;125:2544–60.
33. New Jersey State Cancer Registry. 2018 NJ State Cancer Registry. Trenton (NJ): State of New Jersey, Department of Health. Available from: <https://www.state.nj.us/health/ces/reporting-entities/njsr/>.
34. Percy C, Holten Vv, Muir CS, World Health Organization. International classification of diseases for oncology. Geneva (Switzerland): WHO; 1990.
35. Greene FL, Balch CM, Fleming ID, Fritz A, Haller DG, Morrow M, et al. AJCC cancer staging handbook: TNM classification of malignant tumors. New York: Springer Science & Business Media; 2002.
36. American Cancer Society. Cancer facts & figures 2019. Atlanta (GA): American Cancer Society; 2019.
37. Jacquez GM, Slotnick MJ, Meliker JR, AvRuskin G, Copeland G, Nriagu J. Accuracy of commercially available residential histories for epidemiologic studies. *Am J Epidemiol* 2010;173:236–43.
38. Texas A&M University. NAACCR Geocoder Data Dictionary. NAACCR; 2016.
39. Boscoe FP. Towards the use of a census tract poverty indicator variable in cancer surveillance. *J Registry Manag* 2010;37:148–51.
40. Kneib T, Fahrmeir L. A mixed model approach for geoadditive hazard regression. *Scand J Stat* 2007;34:207–28.
41. Brezger A, Lang S. Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics & Data Analysis* 2006;50:967–91.
42. Adler D, Kneib T, Lang S, Umlauf N, Zeileis A. BayesXsrc: R package distribution of the BayesX C++ sources. R package version 2012;2:1.
43. Belitz C, Brezger A, Klein N, Kneib T, Lang S, Umlauf N. BayesX: methodology manual; 2015. Available from: <https://doi.org/10.1198/1061860043010>.
44. Lang S, Brezger A. Bayesian P-splines. *J Comput Graph Statist* 2004;13:183–212.
45. Umlauf N, Adler D, Kneib T, Lang S, Zeileis A. Structured additive regression models: an R interface to BayesX. Working Papers in Economics and Statistics; 2012. Available from: <https://cran.r-project.org/web/packages/R2BayesX/R2BayesX.pdf>.
46. Umlauf N, Kneib T, Heinzl F, Lang S, Zeileis A. R2BayesX: estimate structured additive regression models with BayesX. R package version 01–2 2013. Available from: <https://rdrr.io/cran/R2BayesX/man/R2BayesX-package.html>.
47. Umlauf N, Kneib T, Lang S, Zeileis A. Package 'R2BayesX'. 2012. Available from: <https://cran.r-project.org/web/packages/R2BayesX/index.html>.
48. Jin X, Carlin BP, Banerjee S. Generalized hierarchical multivariate CAR models for areal data. *Biometrics* 2005;61:950–61.
49. Kwan MP. The uncertain geographic context problem. *Ann Assoc Am Geograph* 2012;102:958–68.
50. Allemani C, Harewood R, Johnson CJ, Carreira H, Spika D, Bonaventure A, et al. Population-based cancer survival in the United States: data, quality control, and statistical methods. *Cancer* 2017;123:4982–93.
51. Niu X, Pawlish KS, Roche LM. Cancer survival disparities by race/ethnicity and socioeconomic status in New Jersey. *J Health Care Poor Underserved* 2010;21:144–60.
52. White A, Vernon SW, Franzini L, Du XL. Racial disparities in colorectal cancer survival: to what extent are racial disparities explained by differences in treatment, tumor characteristics, or hospital characteristics? *Cancer* 2010;116:4622–31.
53. White A, Joseph D, Rim SH, Johnson CJ, Coleman MP, Allemani C. Colon cancer survival in the United States by race and stage (2001–2009): findings from the CONCORD-2 study. *Cancer* 2017;123:5014–36.
54. Du XL, Fang S, Vernon SW, El-Serag H, Shih YT, Davila J, et al. Racial disparities and socioeconomic status in association with survival in a large population-based cohort of elderly patients with colon cancer. *Cancer* 2007;110:660–9.
55. Henry KA, Sherman RL, McDonald K, Johnson CJ, Lin G, Stroup AM, et al. Associations of census-tract poverty with subsite-specific colorectal cancer incidence rates and stage of disease at diagnosis in the United States. *J Cancer Epidemiol* 2014;2014:823484.
56. Shariff-Marco S, Yang J, John EM, Sangaramoorthy M, Hertz A, Koo J, et al. Impact of neighborhood and individual socioeconomic status on survival after breast cancer varies by race/ethnicity: the Neighborhood and Breast Cancer study. *Cancer Epidemiol Biomarkers Prev* 2014;23:793–811.
57. Shariff-Marco S, Gomez SL, Sangaramoorthy M, Yang J, Koo J, Hertz A, et al. Impact of neighborhoods and body size on survival after breast cancer diagnosis. *Health Place* 2015;36:162–72.
58. Schootman M, Jeffe DB, Lian M, Gillanders WE, Aft R. The role of poverty rate and racial distribution in the geographic clustering of breast cancer survival among older women: a geographic and multilevel analysis. *Am J Epidemiol* 2009;169:554–61.
59. Aarts MJ, Lemmens VEPP, Louwman MWJ, Kunst AE, Coebergh JWW. Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. *Eur J Cancer* 2010;46:2681–95.
60. Robinson JRM. Effects of individual and neighborhood socioeconomic status on outcomes following a colorectal cancer diagnosis. 2019. Available from: <https://digital.lib.washington.edu/researchworks/handle/1773/45188>.
61. Pawitan G, Steel DG. Exploring the MAUP from a spatial perspective. Centre for Statistical and Survey Methodology. University of Wollongong. Working Paper 2009;20:28p. Available from: <https://ro.uow.edu.au/cssmwp/40>.
62. Buzzelli M. Modifiable areal unit problem. International Encyclopedia of Human Geography 2020:169–73. Available from: <https://europepmc.org/article/pmc/pmc7151983#free-full-text>.