Social Science & Medicine 193 (2017) 1–7

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Assessment of spatial variation in breast cancer-specific mortality using Louisiana SEER data

Rachel Carroll [a, *], Andrew B. Lawson [b], Chandra L. Jackson [c], Shanshan Zhao [a]

[a] Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, 111 TW Alexander Dr., Research Triangle Park, NC 27709, USA
[b] Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA
[c] Epidemiology Branch, National Institute of Environmental Health Sciences, 111 TW Alexander Dr., Research Triangle Park, NC 27709, USA

ARTICLE INFO

Article history:
Received 6 June 2017
Received in revised form
15 September 2017
Accepted 26 September 2017
Available online 28 September 2017

Keywords:
Breast cancer mortality
Spatial frailty
Survival analysis
SEER
Accelerated failure time model

ABSTRACT

Background: Previous studies suggest spatial differences in mortality for many types of cancer, including breast cancer. Identifying explanations for these spatial differences results in a better understanding of what leads to longer survival time.
Methods: We used a Bayesian accelerated failure time model with spatial frailty terms to investigate potential spatial differences in breast cancer mortality following breast cancer diagnosis using 2000–2013 Louisiana SEER data.
Results: There are meaningful spatial differences in breast cancer mortality across the parishes of Louisiana, even after adjusting for known demographic and clinical risk factors. For example, the average survival time of a woman diagnosed in Orleans parish was 1.51 times longer than that of a woman diagnosed in Terrebonne parish. Additionally, there is evidence to suggest shorter survival times in lower income parishes along the Red and Mississippi Rivers, as well as parishes with lower socioeconomic status, less access to care and fresh food, worse quality of care, and more workers in certain industries.
Conclusion: The addition of spatial frailties to account for an individual's geographic location is useful when analyzing breast cancer mortality data. Our findings suggest that survival following breast cancer diagnosis could potentially be improved if socioeconomic status differences were addressed, healthcare improved in quality and became more accessible, and certain industrial situations were improved for individuals diagnosed in parishes identified as having shorter average survival times.
Published by Elsevier Ltd.

1. Introduction

Among US women, breast cancer (BrCa) is the most common cancer when excluding non-melanoma cancers of the skin. BrCa accounts for 14.6% of all cancer incidence and presents in approximately 12.4% of women during their lifetime (National Institutes of Health) (http://seer.cancer.gov/statfacts/html/breast.html). There are many known risk factors for BrCa including: older age, white race, older age at first-time birth, and family history (American Cancer Society, 2016) (http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-046381.pdf). Further, improved BrCa survival following diagnosis is associated with several demographic and clinical variables: white race, married at the time of diagnosis, younger age at diagnosis, lower cancer grade, positive estrogen receptor (ER)/progesterone receptor (PR) tumor subtype status, and receiving treatment such as surgery and radiation therapy (Wieder et al., 2016). In addition to these demographic and clinical risk factors, there are others (e.g. genetic information) which are difficult and/or expensive to obtain (American Cancer Society, 2016; Wieder et al., 2016). There is also evidence suggesting that BrCa survival rates differ by geographic location at diagnosis (American Cancer Society, 2016). Many risk factors related to BrCa survival, such as socioeconomic status, race, access to health care, environmental exposures, natural disasters, water quality, and air pollution vary substantially by geographic location (Carroll et al., 2014; Padilla et al., 2016; Zou et al., 2014); so, geography can be used as a surrogate for these unmeasured potential risk factors. However, this information is not widely available nor often used in breast cancer research.

Typically, epidemiological studies employ an assignment or

* Corresponding author.
E-mail address: rachel.carroll@nih.gov (R. Carroll).

https://doi.org/10.1016/j.socscimed.2017.09.045
0277-9536/Published by Elsevier Ltd.

imputation method wherein an individual is assigned a value of an aggregated spatially varying risk factor based on their geographic location, e.g. assigning all individuals within the same county a county-level average measure of air pollution. This method can lead to interesting and useful results, but multiple assignment processes are needed for the risk factors of interest, which must be selected a priori, and correlations between individuals within the same or nearby regions are not regularly considered. Alternatively, we included spatial frailty (or random effect) terms into survival models, which served as latent variables representing combinations of the measured and unmeasured spatially-varying risk factors that are associated with mortality following BrCa diagnosis. This approach allowed us to flexibly explore a wide range of exposures for explaining the extra variation in BrCa-specific mortality through geography (Banerjee et al., 2003; Bastos and Gamerman, 2006; Henderson et al., 2002; Y. Li and Ryan, 2002; Silva and Amaral-Turkman, 2005).

We used data from the state of Louisiana made available by the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute for the years 2000–2013 (Surveillance Epidemiology and End Results (SEER) Program, 2015). These data and region were selected because they offered a large number of patients within the registry as well as a reasonable number of spatial regions to consider with substantial variation in the socio-demographic and environmental risk factors of interest. Further, SEER data is representative of a typical registry that provides individual-level demographic and clinical covariates but no information about socioeconomic status, family history, or environmental exposures, which made these data ideal for examining the uses of spatial frailty terms to represent this unmeasured information.

2. Materials and methods

2.1. Data description

2.1.1. Individual-level data

The SEER program has publicly available cancer data from 17 registries across the US. The specific registry that provides data for Louisiana currently includes individuals diagnosed with BrCa between January 1, 2000 and December 31, 2013. To produce the appropriate survival outcome, we combined the documented survival time in months together with a BrCa mortality underlying cause of death indicator. Thus, an individual is considered censored either at their last follow-up time, the end of the study period (December 31, 2013), or time of death due to other causes. The final sample size was reduced from 49,176 to 48,551 based on the inclusion criteria of female with known survival time (231 women with unknown survival time and 394 males were removed). In this analysis, unknown values for covariates were treated as a separate category rather than excluding women with missing data. The amount of unknown data ranged from 0.5% (radiation) to 20.5% (ER/PR). The unknown in ER/PR was the largest (cancer grade and marital status were next highest at 17.0% and 4.3%, respectively), and this was largely due to missing in PR status. However, it was important to include this information so that we could maximize the number of individuals contributing within each parish for the spatial component. Finally, different attempts at organizing the categories of ER/PR to minimize the amount in the unknown category resulted in nearly identical estimates with slightly more complicated interpretations (e.g. including ER positive/PR unknown in the $\pm$ category).

The SEER data also provides the Federal Information Processing Standard county code for each registrant. With this information, we assigned an individual to one of the 64 Louisiana parishes. On average, there were 759 (range: 52-5464) registrants per parish. The parish with the smallest number of deaths had 8 while the parish with the largest number of deaths had 684. The mean age at diagnosis was 62 years and parish-specific means ranged from 59 to 65.

The demographic and clinical covariates considered in this analysis were selected for their known associations with BrCa mortality (American Cancer Society, 2016; National Institutes of Health; Wieder et al., 2016) and their availability in the SEER database. Specifically, the covariates included were: race (African American vs. other), marital status at diagnosis (single, currently married, previously married, unknown), age at diagnosis, cancer grade (low, high, unknown), ER/PR tumor subtype status (+/+, +/−, −/+, −/−, unknown), BrCa surgery (no, yes, unknown), and radiation therapy (no, yes, unknown). Chemotherapy treatment was not included in the analysis because it was not uniformly available in the database.

2.1.2. Parish-level data

To investigate potential features of the environment associated with spatial differences in BrCa survival, we acquired the following parish-level variables from the Area Health Resources Files (Bureau of Health Workforce, 2015) from various years across the study time: percent of persons 25 years or older with four or more years of college education (2005-09); total number of hospitals (public or private, 2006); number of hospitals per square mile; total number of hospitals with BrCa screening and mammography machines (2006); number of BrCa screening hospitals with mammography machines per square mile; number of hospitals with an approved American Cancer Society program (2011); number of hospitals with an approved American Cancer Society program per square mile (2011); number of hospitals with Medicare certification (2003); number of hospitals with Medicare certification per square mile (2003); median household income (2006); percent Medicaid eligible persons; percent urban population; percent farmland (2002); percent of persons living in poverty (2010); percent African American population (2010); and percent persons working in agriculture, forestry, fishing, hunting, or mining (2006-09). We also examined 150 parish-level modeled estimates of 2005 chemical emissions variables available from the Environmental Protection Agency (U.S. Environmental Protection Agency, 2005); these included carbon monoxide, ammonia, nitroxide, particulate matter, sulfur dioxide, and volatile organic compound emissions from specific sources such as agriculture and forestry, coal, residential, commercial industrial, etc. The final parish-level variables we examined were current (2017) total number of grocery stores (includes: Winn-Dixie, Piggly Wiggly, Ford's, Market Basket, Brookshire's, Cannata's, Matherne's Market, Ramey's Marketplace, Rouses, Robert Fresh Market, and Mac's Fresh Market), total number of grocery stores per square mile, and a ranking of hospital quality based on health factors (2010) (University of Wisconsin Population Health Institute, 2017).

2.2. Statistical analysis

2.2.1. Statistical model

We used the accelerated failure time (AFT) model (Christensen and Johnson, 1988), which allows for a direct relationship of the logarithm of survival time with both the fixed and random effects (Onicescu et al., 2015; Orbe et al., 2002; Zhang and Lawson, 2011). This capability along with the models' general flexibility in terms of assumptions has led to the AFT model's increase in popularity and, for our purposes, an ideal interpretation of spatial frailty estimates. The AFT model for an individual i diagnosed in parish j can be written as: $\log(t_{ij}) = \lambda_{ij} + \sigma\varepsilon_{ij}$ where t_{ij} is the survival time, λ_{ij} is the