Published in final edited form as:

Soc Sci Med. 2017 November ; 193: 1–7. doi:10.1016/j.socscimed.2017.09.045.

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

Rachel Carroll^{1,*}, Andrew B. Lawson², Chandra L. Jackson³, and Shanshan Zhao¹

¹Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, 111 TW Alexander Dr., Research Triangle Park, NC 27709, USA

²Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

³Epidemiology Branch, National Institute of Environmental Health Sciences, 111 TW Alexander Dr., Research Triangle Park, NC 27709, USA

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.

*Corresponding author: Dr. Rachel Carroll, rachel.carroll@nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

breast cancer mortality; spatial frailty; survival analysis; SEER; accelerated failure time model

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 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 & Gamerman, 2006; Henderson et al., 2002; Y. Li & Ryan, 2002; Silva & 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 to 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 sociodemographic 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 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 +/- 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: 52–5,464) 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 & 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 & 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: $\log(t_{ij}) = \lambda_{ij} + \sigma e_{ij}$ where t_{ij} is the survival time, λ_{ij} is the linear predictor of interest, e_{ij} 's are the independent random errors, and σ is a scale parameter. Three different definitions of λ_{ij} were considered, which resulted in three sets of models. The first set of models included only known, individual level covariates as fixed effects for describing the survival outcome. The second set of models attempted to describe an individual's survival with only spatial information via spatial frailty terms. Finally, the third set of models considered both fixed covariates and spatial frailty terms.

All the demographic and clinical covariates considered were selected for their known importance as risk factors to BrCa mortality. However, clinical covariates (i.e. age at diagnosis, cancer grade, ER/PR tumor subtype status, surgery, and radiation therapy) have the potential to lie in the causal pathway as mediators between geographic location and BrCa mortality. As an example, an individual's assigned parish determines where they likely

received treatment and that facility could be more likely to suggest surgery only rather than a combination of surgery and radiation. Further, receipt of treatment could be related to cost and individual socioeconomic status (Ayanian & Guadagnoli, 1996; R. Li et al., 2016), which vary spatially. For this reason, we proposed two options of fixed covariates, a reduced model that only adjusted for demographic covariates race and marital status at diagnosis, notated with “R,” and a full model that adjusted for both demographic and clinical covariates notated with “F.” There were also two options for the spatial frailty term due to different correlation structure: one option assumed a spatially uncorrelated structure (notated with “U”), meaning that there were differences between parishes but neighboring parishes were not necessarily more alike than non-neighbors; the other, notated with “C,” brought in additional correlation between neighboring parishes through a correlated spatial frailty term (Besag & Green, 1993; Eberly & Carlin, 2000). Ultimately, eight fitted models were employed for comparison based on the following specifications, and their parameterization is defined in Supplemental Table 1 of the Supplemental Statistical Methodology along with more details about the statistical methodology.

2.2.2. Assessment of spatial variation—Using the spatial frailty estimates from the model with the best fit, we quantitatively assessed the spatial differences in two ways. In the first, we defined certain parishes as lower income (parish median income below the state median income), river (Red or Mississippi Rivers border or cross through), or lower income river (the intersection of lower income and river). We subsequently calculated percentages of those parishes within the first quartile, less than the mean, greater than the mean, and within the fourth quartile of the spatial frailty term. In the second, two-sample t tests were used to compare distributions of the parish-level variables of interest in the first quartile and fourth quartile of the spatial frailty term. We considered the variables that were significantly different to be potential risk factors for breast cancer and represented by the spatial frailty in our model, in hopes that their identification would encourage future studies to collect these risk factors at the individual level.

2.2.3. Computational details—These analyses were carried out using the R package R2WinBUGS, which calls the Bayesian computing software WinBUGS (Lunn et al., 2013; R Core Team, 2015; Thomas et al., 2014; Thomas et al., 2006). Models were evaluated based on goodness of fit using the deviance information criterion (DIC). For the DIC, a statistically meaningful difference is 3–4 units (Lesaffre & Lawson, 2013), and lower values are more desirable. More information regarding the computation and evaluation techniques is included in the supplemental materials (see the Supplemental Statistical Methodology section).

Additionally, we created an R Shiny (Chang et al.) application to display the results of this analysis. In this application, the model with the best fit is used to calculate the survival curves and survival probabilities based on user input for the clinical and demographic covariates adjusted for in the model. This application is available via GitHub from user: carrollrm and repository: LAmortBrCaShiny.

3. Results

Table 1 shows demographic and clinical characteristics by breast cancer mortality status. State-wide, 35,979 (74.1%) women remained alive at the end of follow-up (December 31, 2013), and 6,023 (12.4%) died with breast cancer as the underlying cause of death. The total number of non-breast cancer-specific underlying cause of death was 6,549 (13.5%) where the next highest was heart disease ($n=1,612$; 3.3%) and other cancer deaths accounted for 1,518 (3.1%) of total. The median follow-up time was 54 (range: 0–167) months. When comparing women who died of BrCa to those who survived, all demographic and clinical covariates considered were significantly different per Pearson chi-square tests for the categorical risk factors and a t test for age at diagnosis. Additionally, parish-level representations of these individual level covariates can be found in Supplemental Figure 1; from these, it was apparent that African American race offered largest amount of spatial variation.

Supplemental Table 2 displays DIC estimates for the eight fitted models. Overall, these results suggested that Model 3FU was best fitting, and thus, there was some unexplained spatial variation in the data even after adjusting for important individual-level demographic and clinical risk factors. Model 2U and 2C had high DICs; this suggested that only using spatial frailties was not enough. When comparing 3FU and 3FC models, the goodness of fit measures suggested that the correlated frailty did not improve the fit enough to warrant inclusion; thus, neighboring parishes of diagnosis were not more alike compared to non-neighboring parishes.

Table 2 presents the fixed effect parameter estimates associated with Models 1R, 1F, 3RU, and 3FU. None of the fixed effect parameter estimates changed much when comparing among models with or without spatial frailty terms, and all were well-estimated in terms of narrow credible intervals except the unknown category of radiation therapy. This suggested that spatial information explained additional variability in BrCa mortality, beyond the known individual level risk factors. These estimates are directly related to the logarithm of time, thus, a negative value indicated a decrease in survival time and a positive value indicated an increase in survival time. For example, the average survival time of women who were married at diagnosis was $1.73 (\exp(0.55) = 1.73)$ times higher than single women at diagnosis.

Figure 1 displays the maps of the spatial frailty parameter estimates (u_j) for Models 2U, 3RU, and 3FU. In these maps, the darker shades indicate a longer survival time following BrCa diagnosis. For the southern portion of the map, the estimates were more variable and suggested that several parishes had an increased survival time following BrCa diagnosis that remained unexplained even after adjusting for the covariates in Models 3RU and 3FU. The northwestern parishes along the border exhibited longer survival time while other northern parishes indicated shorter. The results for all models differed in variability but followed the same general distribution (see Supplemental Figure 2). This illustrated that Models 2U and 2C effectively captured the overall spatial pattern of breast cancer mortality and that pattern remained similar after adjusting for demographic and/or clinical covariates (Model 3RU, 3RC, 3FU, and 3FC).

Many of the parishes with shorter survival time appeared to be lower income (parish median income less than the state median income) and along the paths of the Red or Mississippi Rivers (Supplemental Table 3). Explicitly, per the best fitting Model 3FU, 57.1% of the lower income river parishes were among those with decreased survival times ($u_j < 0$) and 42.9% were among the parishes with shortest survival times (in the first quartile of u_j). We also used these estimates to calculate the difference in survival time for the best (Orleans) and worst (Terrebonne) parishes of diagnosis. The average survival time for a woman diagnosed in Orleans Parish was 1.51 times longer than that of a woman diagnosed in Terrebonne Parish. See the Supplemental Results section of the supplemental materials for information about the calculations related to the spatial frailty estimates.

Two-sample t tests examining differences in parish-level sociodemographic characteristics based on the spatial frailty estimates from Model 3FU are shown in Table 3. Supplemental Figure 4 displays the geographic distributions of these variables. Based on these results, parishes with the highest and lowest spatial frailty estimates differed in access to and quality of care, availability of fresh food, and socioeconomic status categories as well as percent urban population, percent farmland, and percent agriculture, forestry, fishing, or mining workers. Further, the examination of chemical emission variables via two-sample t tests (Supplemental Table 4) suggested that there was evidence of shorter survival times for women diagnosed in parishes with higher ammonia and particulate matter (several types and sizes) emissions related to agriculture and forestry sources. Alternatively, there was evidence of an association between higher emissions (carbon monoxide, nitric oxides, several types and sizes of particular matter, and volatile organic compounds) from a variety of sources and longer survival times. The correlations presented in Supplemental Table 4 suggested that this association was related to areas with better access to healthcare and fresh food, higher socioeconomic status, or some combination of those.

4. Discussion

We found spatial variation in breast cancer mortality among residents in Louisiana, and these differences could not be explained alone by individual-level disease characteristics, such as cancer grade, ER/PR tumor subtype status, surgery, or radiation. Rather, there was a strong indication that these spatial differences were, in part, related to health disparities such as socioeconomic status, access to and quality of care, availability of fresh food, rurality, and working in certain industries (e.g. agriculture and forestry).

The interpretation of the AFT model is attractive in that the effects, fixed and random, have a direct relationship with the logarithm of survival time. The fixed effect parameter estimates produced in these models agreed with what has been documented in the literature in that, on average, there were shorter survival times for women with the following characteristics: older age at diagnosis, African American race, single at the time of diagnosis, high cancer grade, —/— ER/PR tumor subtype status, no surgery performed, and no radiation therapy (American Cancer Society, 2016; Kroenke et al., 2012; Osborne et al., 2005; Parise & Caggiano, 2014). Additionally, relating to the spatial frailty results, an exploration of nationwide BrCa mortality suggested higher rates along the Mississippi and Red River, particularly in those lower income river parishes, (Mokdad et al., 2017), and it has long been

established that BrCa is associated with worse socioeconomic status, access to care, and diet (American Cancer Society, 2016; Breastcancer.org, 2014).

The examination of race in this analysis was interesting. Firstly, the fixed effect estimate suggested that the average survival time of African American women was 46% ($1 - \exp(-0.61) = 0.46$) lower than that of non-African American women. This was likely because African American women typically have a more aggressive form of BrCa (American Cancer Society, 2016). Next, a comparison of the distribution of African American race in the study population (Supplemental Figure 1) to the parish-level percent African American variable (Supplemental Figure 4) from the Area Health Resources Files illustrated similarities between the two in terms of magnitude and distribution. This indicated that the individual-level covariate and the parish-level variable were supplying similar information, where the former had an advantage of being at a finer level. Based on the t test using the frailty from Model 2U (no covariates included, only uncorrelated spatial frailty), percent African American race did appear to be significantly different between low- and high-risk parishes (1st quartile mean (SD): 37.13 (16.5); 4th quartile mean (SD): 25.51 (13.5); p-value: 0.04), as expected. However, after adjusting for individual-level race, African American race no longer indicated that association. This occurred because much of the disparity due to race was already explained by the individual-level fixed effect of race, which left the random effect to represent the unmeasured risk factors of socioeconomic status, access to and quality of healthcare, availability of fresh food, rurality, and working in certain industries.

The main objective of this investigation was to identify spatial differences in this population of Louisiana women diagnosed with BrCa by way of spatial frailty parameters. The secondary explorations of the spatial frailty estimates via percentage calculations and two-sample t test executions allowed us to determine that lower survival times were associated with being diagnosed in lower income parishes along the paths of major rivers as well as socioeconomic status, access to and quality of care, availability of fresh food, rurality, and working in certain industries. We believed that the frailties represented more than these variables identified via the secondary explorations, and the remaining composition is likely some environmental risk factors with weaker associations to BrCa-specific mortality. However, in studies where the environmental influence is stronger, these methods could identify those latent combinations and associations in a way that is more flexible and appropriate for exploratory analyses than the often-employed assignment methods. Finally, other studies have suggested that population-level representations of risk factors could inform on individual-level outcomes (Warnecke et al., 2008), and we believe that the evidence here suggests that frailties also assume that role.

It is important to consider environmental events that could impact the study. For example, two notable events occurred in or around Louisiana during this study time: Hurricane Katrina (August 23–31, 2005) and the Deepwater Horizon oil spill (April 20, 2010). Both events could have impacted the survival of individuals as well as the distribution of sociodemographic data since these events, particularly Katrina, impacted African Americans and those of lower socioeconomic status more severely. We could not make any adjustments for this or the associated out-migration in the modeling; however, we did fit Model 3FU

separately restricting to those who were either diagnosed before or after Hurricane Katrina, and noticed little change to the fixed effect parameter estimates produced. On the other hand, the uncorrelated random effect did indicate some differences in the distribution of spatial variation between the before and after groups. An option for modeling events such as this could involve introducing a temporal frailty that produces different estimates based on the individual's year of diagnosis, orientation to the event(s) of interest, or both. Incorporating temporal components into survival models is a relatively novel area, particularly with respect to the AFT model; thus, there is potential for future statistical methodological development in this direction.

This study had several limitations. For instance, we simply considered individuals as censored if they experienced non-BrCa related death. One potential solution to this would be to apply multivariate modeling procedures as in the Fine-Gray model (Fine & Gray, 1999) as used by Felix et. al (Felix et al., 2017). However, that approach is only applicable for a proportional hazards model. Some multivariate and competing risks procedures have been developed for the AFT model (Chiou et al., 2014; Cho & Ghosh, 2015), but they are very computationally intensive and, therefore, rarely employed. More importantly, they have not been extended to allow for spatial frailty terms. This is another topic of interest for future statistical methodological development in this area. Other limitations involved the data available. The amount of missing or unknown data, while not massive, needed to be considered, and some explorations did suggest that there were differences between those with complete and unknown information. We attempted to minimize this issue by including unknown categories for many of the covariates so that we would not lose other known information and avoid employing computationally intensive imputation strategies. Finally, the secondary explorations involving t tests, while very useful, could be invalid in certain situations, e.g. imposed correlation in the spatial frailties would violate the independence assumption. Another slightly more complicated method involving a secondary linear mixed model with spatial frailty (Carroll & Zhao, 2017) could circumvent this issue, as well as be further extended to consider the correlations in the risk factors of interest via multivariate modeling.

Despite these limitations, there were many study strengths. First, most disease mapping studies involving breast cancer focus on incidence or mortality without including data on all incident cases followed through a study period. Second, the flexible AFT model presented an ideal platform for including and interpreting the spatial random effects, and established a basis for carrying out the statistical methodology extensions highlighted above. Next, the Louisiana SEER data offered a representative registry data set with a reasonable number of spatial regions ($n=64$) and commonly measured clinical and demographic covariates, making this work generalizable and widely applicable. The spatial frailties provided a way of accounting for some of the limitations of the data by representing unmeasured and unknown risk factors, and the justification and testing involved with explaining these risk factors is quite novel and extremely useful for studies which include frailties of this type. Further, this study provided an example of how to employ these spatial frailty methods rather than the assignment method and the strengths of doing so. Finally, our study illustrated that these methods were applicable to other environmental epidemiology studies for exploring potential environmental risk factors.

5. Conclusion

The results from this study suggest that survival following breast cancer diagnosis was heterogeneous across the parishes of Louisiana, and there was evidence of shorter survival times for women diagnosed in several of the lower income parishes along the Red and Mississippi Rivers. Additionally, our results indicated that this heterogeneity was clearly distinguishable from known clinical and demographic risk factors. With respect to interpretation, our spatial frailty results suggested that shorter survival following BrCa diagnosis was associated with lower income parishes along the main rivers as well as parishes with worse socioeconomic status, worse access to and quality of healthcare, less availability of fresh food, more rural population, more workers in certain industries, and more chemical emissions related to those same industries. Ultimately, our findings indicate that the use of spatial frailties accompanied by secondary assessments leads to improved, interesting, and useful results for indicating risk factors that should be considered for individual level collection in future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- American Cancer Society. Breast Cancer Facts & Figures 2015–2016. Atlanta: American Cancer Society, Inc; 2016.
- Ayanian JZ, Guadagnoli E. Variations in breast cancer treatment by patient and provider characteristics. *Breast Cancer Res Treat.* 1996; 40:65–74. [PubMed: 8888153]
- Banerjee S, Wall MM, Carlin BP. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics.* 2003; 4:123–142. [PubMed: 12925334]
- Bastos LS, Gamerman D. Dynamic survival models with spatial frailty. *Lifetime Data Anal.* 2006; 12:441–460. [PubMed: 17031498]
- Besag J, Green PJ. Spatial statistics and Bayesian computation. *J Roy Stat Soc B.* 1993; 55:25–37.
- Healthy eating after diagnosis improves survival. Ardmore, PA: 2014. Breastcancer.org
- Bureau of Health Workforce. Area Health Resource Files (AHRF). Rockville, MD: US Department of Health and Human Services, Health Resources and Services Administration; 2015.
- Carroll R, Lawson AB, Voronca D, Rotejanaprasert C, Vena JE, Aelion CM, et al. Spatial environmental modeling of autoantibody outcomes among an African American population. *Int J Environ Res Public Health.* 2014; 11:2764–2779. [PubMed: 24608900]
- Carroll R, Zhao S. Spatial heterogeneity: Gaining relevance from the random. 2017 Submitted.
- Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. shiny: Web application framework for R.
- Chiou SH, Kang S, Kim J, Yan J. Marginal semiparametric multivariate accelerated failure time model with generalized estimating equations. *Lifetime Data Anal.* 2014; 20:599–618. [PubMed: 24549607]
- Cho Y, Ghosh D. Weighted estimation of the accelerated failure time model in the presence of dependent censoring. *PLoS One.* 2015; 10:e0124381. [PubMed: 25909753]
- Christensen R, Johnson W. Modelling accelerated failure time with a Dirichlet process. *Biometrika.* 1988; 75:693–704.
- Eberly LE, Carlin BP. Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Stat Med.* 2000; 19:2279–2294. [PubMed: 10960853]

- Felix AS, Bower JK, Pfeiffer RM, Raman SV, Cohn DE, Sherman ME. High cardiovascular disease mortality after endometrial cancer diagnosis: Results from the Surveillance, Epidemiology, and End Results (SEER) Database. *Int J Cancer*. 2017; 140:555–564. [PubMed: 27741565]
- Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999; 94:496.
- Henderson R, Shimakura S, Gorst D. Modeling spatial variation in Leukemia survival data. *J Am Stat Assoc*. 2002; 97:965–972.
- Kroenke CH, Michael Y, Tindle H, Gage E, Chlebowski R, Garcia L, et al. Social networks, social support and burden in relationships, and mortality after breast cancer diagnosis. *Breast Cancer Res Treat*. 2012; 133:375–385. [PubMed: 22331479]
- Lesaffre, E., Lawson, AB. *Bayesian Biostatistics*. West Sussex, U.K: Wiley; 2013.
- Li R, Daniel R, Rachet B. How much do tumor stage and treatment explain socioeconomic inequalities in breast cancer survival? Applying causal mediation analysis to population-based data. *Eur J Epidemiol*. 2016; 31:603–611. [PubMed: 27165500]
- Li Y, Ryan L. Modeling spatial survival data using semiparametric frailty models. *Biometrics*. 2002; 58:287–297. [PubMed: 12071401]
- Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D. *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: CRC Press; 2013.
- Mokdad AH, Dwyer-Lindgren L, Fitzmaurice C, Stubbs RW, Bertozzi-Villa A, Morozoff C, et al. Trends and Patterns of Disparities in Cancer Mortality Among US Counties, 1980–2014. *J Am Med Assoc*. 2017; 317:388–406.
- National Institutes of Health. SEER stat fact sheets: Female breast cancer. Rockville, MD: National Institutes of Health;
- Onicescu G, Lawson A, Zhang J, Gebregziabher M, Wallace K, Eberth JM. Bayesian accelerated failure time model for space-time dependency in a geographically augmented survival model. *Stat Methods Med Res*. 2015
- Orbe J, Ferreira E, Nunez-Anton V. Comparing proportional hazards and accelerated failure time models for survival analysis. *Stat Med*. 2002; 21:3493–3510. [PubMed: 12407686]
- Osborne C, Ostir GV, Du X, Peek MK, Goodwin JS. The influence of marital status on the stage at diagnosis, treatment, and survival of older women with breast cancer. *Breast Cancer Res Treat*. 2005; 93:41–47. [PubMed: 16184457]
- Padilla CM, Kihal-Talantikit W, Perez S, Deguen S. Use of geographic indicators of healthcare, environment and socioeconomic factors to characterize environmental health disparities. *Environ Health*. 2016; 15:79. [PubMed: 27449640]
- Parise CA, Caggiano V. Breast Cancer Survival Defined by the ER/PR/HER2 Subtypes and a Surrogate Classification according to Tumor Grade and Immunohistochemical Biomarkers. *J Cancer Epidemiol*. 2014; 2014:469251. [PubMed: 24955090]
- R Core Team. *R Foundation for Statistical Computing*. Vienna, Austria: 2015. *R: A language and environment for statistical computing*.
- Silva GL, Amaral-Turkman MA. Bayesian analysis of an additive survival model with frailty. *Commun Stat A-Theor*. 2005; 33:2517–2533.
- N.C. Institute. , editor. *Surveillance Epidemiology and End Results (SEER) Program. SEER*Stat Database: Incidence - SEER 9 Regs Research Data, Nov 2015 Sub (1973–2013)*. Bethesda, MD: 2015.
- Thomas A, Best N, Lunn D, Arnold R, Spiegelhalter D. *GeoBUGS user manual*. 2014
- Thomas A, O'hara B, Ligges U, Sturtz S. Making BUGS Open. *R News*. 2006; 6:12–17.
- U.S. Environmental Protection Agency. *Pollutant Emissions Summary Files for Earlier NEIs*. Washington, D.C: USEPA; 2005.
- University of Wisconsin Population Health Institute. *County Health Rankings & Roadmaps*. 2017
- Warnecke RB, Oh A, Breen N, Gehlert S, Paskett E, Tucker KL, et al. Approaching health disparities from a population perspective: the National Institutes of Health Centers for Population Health and Health Disparities. *Am J Public Health*. 2008; 98:1608–1615. [PubMed: 18633099]

- Wieder R, Shafiq B, Adam N. African American Race is an Independent Risk Factor in Survival from Initially Diagnosed Localized Breast Cancer. *J Cancer*. 2016; 7:1587–1598. [PubMed: 27698895]
- Zhang J, Lawson AB. Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. *J Appl Stat*. 2011; 38:591–603. [PubMed: 21475617]
- Zou B, Peng F, Wan N, Mamady K, Wilson GJ. Spatial cluster detection of air pollution exposure inequities across the United States. *PLoS One*. 2014; 9:e91917. [PubMed: 24647354]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Research Highlights

- Breast cancer mortality and many of its risk factors varied spatially
- Spatial variation was estimated using frailties in survival analysis methods
- Assessment of the spatial variation led to important findings
- Breast cancer mortality was related to environmental factors and health disparities
- Other interesting details related to this assessment were discussed

Figure 1.

Estimated Unmeasured Spatial Variation (Uncorrelated Spatial Frailty) for A) Model 2U (no covariates included); B) Model 3RU (reduced set of covariates included: African American race and age at diagnosis); and C) Model 3FU (full set of covariates included: all individual-level risk factors).^{a, b}

^a Parishes with positive, higher estimates (darker shading) indicate longer survival time that is unexplained by the individual-level clinical and demographic risk factors. Conversely, parishes with negative, lower estimates (lighter shading) indicate shorter unexplained survival time.

^b These values were quantitatively compared by using the exact parish-level estimates. For example, the estimates associated with Orleans and Terrebonne Parishes from Model 3FU were 0.214 and -0.197, respectively. From this, we calculated that the average survival time of women in Orleans parish was 1.508 ($\exp(0.214 - (-0.197)) = 1.508$) times longer than that of women in Terrebonne parish.

Table 1
Summary of Demographic and Clinical Characteristics by Breast Cancer Survival Status

	BrCa survivors (N=42,528)				BrCa mortality (N=6,023)				P value ^d
	N	Col %	Mean (SD)		N	Col %	Mean (SD)		
Age at Diagnosis			62.2 (15.2)				61.4 (13.5)		0.0001
Race									<1e-15
African American	11405	26.8			2519	41.8			
Other	31123	73.2			3504	58.2			
Marital Status ^b									<1e-15
Single	6134	14.4			1262	21.0			
Currently Married	22530	53.0			2490	41.3			
Previously Married	12036	28.3			2019	33.5			
Unknown	1828	4.3			252	4.2			
Breast Cancer Grade									<1e-15
Low	22858	53.7			1916	31.8			
High	12726	30.0			2775	46.1			
Unknown	6944	16.3			1332	22.1			
ER/PR									<1e-15
+/+	22580	53.1			1894	31.4			
+/-	4446	10.5			712	11.8			
-/+	495	1.2			127	2.1			
-/-	6663	15.7			1671	27.7			
Unknown ^c	8344	19.6			1619	26.9			
Surgery									<1e-15
No	1791	4.2			1621	26.9			
Yes	40474	95.2			4341	72.1			
Unknown	263	0.6			61	1.0			
Radiation Therapy									<1e-15
No	22905	52.3			3867	64.2			
Yes	19397	47.2			2131	35.4			

BrCa survivors (N=42,528)			BrCa mortality (N=6,023)			P value ^a
N	Col %	Mean (SD)	N	Col %	Mean (SD)	
226	0.5		25	0.4		
Unknown						

^a P values were obtained from a two-sided two-sample t test for age and two-sided Pearson chi-square tests for categorical variables

^b Three individuals that stated they had an unmarried or domestic partner were included in the currently married level of marital status and those that stated they were separated at diagnosis were included in previously married.

^c Those with unknown values for one or both tests were included in the unknown category of ER/PR.

Table 2

Parameter Estimates for the Demographic and Clinical Risk Factors Related to Breast Cancer Survival.

	Model 1R ^a		Model 1F ^a		Model 3RU ^a		Model 3FU ^a	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
African American								
No (Referent)	0.00		0.00		0.00		0.00	
Yes	-0.88 *	-0.98, -0.78	-0.57 *	-0.66, -0.48	-0.90 *	-1.00, -0.80	-0.61 *	-0.70, -0.51
Marital status								
Single (Referent)	0.00		0.00		0.00		0.00	
Currently Married	0.74 *	0.60, 0.87	0.54 *	0.42, 0.66	0.76 *	0.63, 0.88	0.55 *	0.43, 0.68
Previously Married	0.11	-0.03, 0.24	0.21 *	0.07, 0.35	0.13	-0.002, 0.26	0.23 *	0.09, 0.35
Unknown	0.08	-0.16, 0.34	0.50 *	0.25, 0.75	0.12	-0.14, 0.38	0.53 *	0.29, 0.78
Age at Diagnosis ^b			-0.24 *	-0.29, -0.20			-0.25 *	-0.29, -0.20
Breast Cancer Grade								
Low (Referent)	0.00		0.00		0.00		0.00	
High			-0.93 *	-1.04, -0.82			-0.94 *	-1.04, -0.84
Unknown			-0.25 *	-0.38, -0.13			-0.25 *	-0.37, -0.12
ER/PR								
+/+ (Referent)	0.00		0.00		0.00		0.00	
+/-			-1.05 *	-1.33, -0.74			-1.05 *	-1.34, -0.75
-/+			-0.64 *	-0.79, -0.50			-0.64 *	-0.78, -0.50
-/-			-0.98 *	-1.11, -0.87			-0.98 *	-1.09, -0.86
Unknown ^c			-0.28 *	-0.40, -0.17			-0.28 *	-0.39, -0.16
Surgery								
No (Referent)	0.00		0.00		0.00		0.00	
Yes			3.26 *	3.14, 3.39			3.27 *	3.14, 3.40
Unknown			2.10 *	1.60, 2.61			2.09 *	1.60, 2.58
Radiation								
No (Referent)	0.00		0.00				0.00	

	Model 1R ^a		Model 1F ^a		Model 3RU ^a		Model 3FU ^a	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Yes			0.18 [*]	0.09,0.27			0.18 [*]	0.09,0.26
Unknown			-0.41	-1.06,0.29			-0.37	-1.03,0.32

^aThese models relate to: 1R – reduced covariate set (African American race and marital status); 1F – full covariate set (African American race, marital status, age at diagnosis, BrCa grade, ER/PR status, surgery, and radiation; 3RU – reduced covariate set, uncorrelated frailty; 3FU – full covariate set, uncorrelated frailty.

^bAge at diagnosis was standardized for analysis (mean 0, standard deviation of 1).

^c5898 of the 6244 women with unknown ER/PR tumor subtype status were missing both. Sensitivity analysis indicated no difference when merging those with only one status unknown into different categories or excluding them from the analysis.

^{*}Indicates statistical significance

Table 3

Differences in Parish-level Sociodemographic Characteristics for the First and Fourth Quartiles of Parishes Based on the Spatial Frailty Estimates from Model 3FU.

Risk Factors	1st Quartile Mean (SD)	4th Quartile Mean (SD)	<i>P</i> value^a
Access to Care			
Total	3.25 (3.5)	6.25 (6.2)	0.11
per square mile	0.005 (0.005)	0.02 (0.02)	0.04
Total BrCa ^b	0.88 (1.1)	1.56 (1.3)	0.12
BrCa per square mile ^b	0.001 (0.001)	0.004 (0.004)	0.01
Quality of Care ^c			
ACS program	0.44 (0.9)	1.06 (1.5)	0.16
ACS program per square mile	0.0005 (0.001)	0.003 (0.005)	0.07
Medicare certification	1.56 (1.3)	3.00 (2.6)	0.06
Medicare cert. per square mile	0.002 (0.002)	0.001 (0.01)	0.07
Health factors	31.31 (17.5)	22.19 (19.8)	0.18
Availability of Fresh Food			
Number of Groceries	2.63 (2.7)	5.50 (5.7)	0.08
Groceries per square mile	0.003 (0.004)	0.01 (0.02)	0.04
Socioeconomic Status			
Median Income ^d	33.46 (6.8)	39.92 (8.2)	0.02
% High Education ^e	14.88 (5.8)	19.39 (7.5)	0.07
% Medicaid Eligible	29.19 (7.5)	25.64 (7.9)	0.20
% Poverty ^f	22.15 (6.9)	17.47 (5.0)	0.04
% Agriculture etc Work ^g	8.27 (4.7)	4.08 (3.2)	0.01
% African American	31.22 (14.5)	30.70 (16.8)	0.93
% Urban Population	51.89 (23.6)	71.09 (30.2)	0.05
% Farmland	33.81 (22.7)	20.04 (10.8)	0.04

^aThe *P* values are from two-sided two-sample *t* tests for identifying a difference in means.

^bBrCa hospitals are a subset of total hospitals which perform BrCa screenings and have mammography machines.

^cHospital quality risk factors based on: number of hospitals with ACS cancer programs, number of hospitals with Medicare certifications, and a hospital quality ranking based on health factors.

^dMedian household income in thousands of dollars

^ePercent of persons 25 years or older with four or more years of college education

^fPercent of persons living in poverty

^gThis includes workers in: agriculture, forestry, fishing, hunting, and mining industries