

Contribution of socioeconomic and environmental factors to geographic disparities in breast cancer risk in the Nurses' Health Study II

Verónica M. Vieira^a, Trang VoPham^{b,c}, Kimberly A. Bertrand^d, Peter James^e, Natalie DuPré^{b,c,f}, Rulla M. Tamimi^{b,c}, Francine Laden^{b,c,g}, Jaime E. Hart^{c,g}

Background: Evidence of geographic disparities in breast cancer incidence within the United States and spatial analyses can provide insight into the potential contribution of environmental exposures or other geographically varying factors to these disparities. **Methods:** We applied generalized additive models to smooth geocoded residential coordinates while adjusting for covariates. Our analysis included 3,478 breast cancer cases among 24,519 control women from the Nurses' Health Study II. We first examined associations with residential location during adolescence (high school address) or early adulthood (address in 1991). We then assessed the contribution from known individual-level risk factors, measures of socioeconomic status, and occupational and environmental factors that vary spatially and have been linked to breast cancer. Secondary analyses by estrogen receptor and menopausal status were also conducted. **Results:** We identified geographic patterns of breast cancer risk associated with location during adolescence, with increased risk apparent in Michigan, the Northwest, and the New York City area, that shifted to southern New England when addresses during early adulthood were analyzed. Similar results were observed after adjustment for individual- and area-level factors, although spatial associations were no longer statistically significant. **Conclusions:** Breast cancer risk is not spatially uniform across the United States, and incidence patterns varied depending on the timing during life of the residence considered. Geographic disparities persisted even after accounting for established and suspected breast cancer risk factors, suggesting that unmeasured environmental or lifestyle risk factors may explain geographic variation in risk in different parts of the country.

Introduction

Epidemiological investigations focusing on geographic variations of breast cancer risk may help determine social and environmental factors that influence observed disease disparities.

^aProgram in Public Health, Chao Family Comprehensive Cancer Center, Susan and Henry Samueli College of Health Sciences, University of California, Irvine, California; ^bDepartment of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts; ^cChanning Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; ^dSlone Epidemiology Center at Boston University, Boston, Massachusetts; ^eDepartment of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts; ^fDepartment of Epidemiology and Population Health, University of Louisville School of Public Health and Information Sciences, Louisville, Kentucky; ^gExposure, Epidemiology, and Risk Program, Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts.

The results reported herein correspond to specific aims of grant Susan G. Komen for the Cure IIR13264020 to F.L. and R.M.T. This work was also supported by the National Institutes of Health (NIH) National Cancer Institute (NCI) Training Program in Cancer Epidemiology (T32 CA009001) and UM1 CA176726.

*Corresponding Author. Address: Program in Public Health, Chao Family Comprehensive Cancer Center, Susan and Henry Samueli College of Health Sciences, University of California, 653 E. Peltason Dr, AIRB 2084, Irvine, CA 92697. E-mail: vvieira@uci.edu (V.M. Vieira).

Copyright © 2019 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of Environmental Epidemiology. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Environmental Epidemiology (2020) 4:e080

Received: 30 July 2019; Accepted 12 November 2019

Published online 12 December 2019

DOI: 10.1097/EE9.000000000000080

Breast cancer incidence rates among older women in the United States have been reported to be higher in the Northeast than in the South and in urban areas compared with rural areas.^{1–3} Regional variations were observed in the California Teachers Study, with higher incidence rates in San Francisco and coastal Southern California, after accounting for covariates.⁴ In an early prospective analysis of regional variation among postmenopausal women in the Nurses' Health Study (NHS), hazard ratios adjusted for individual-level known breast cancer risk factors were elevated in the Northeast and California (the only state from the West) compared with the South.⁵ The modest elevations in breast cancer risk observed in specific geographic areas suggest that regional variation in breast cancer risk factors (e.g., age, body mass index, reproductive history, hormone

What this study adds

Our analyses disentangle the contribution of spatially varying risk factors, including environmental exposures and sociodemographics, to geographic patterns of breast cancer disparities and identify high risk populations throughout the United States. This is of epidemiologic significance, especially as researchers try to understand the impact of multiple exposure to chemical and non-chemical stressors on breast cancer risk. We observed differences in the spatial distribution of breast cancer risk when we compare associations with location of residence during adolescence and location during early adulthood. Patterns also differed by estrogen receptor and menopausal status, indicating that geographic risk factors vary by type of breast cancer. To our knowledge, our analyses are the first to investigate address-level geographic variation at a national scale while simultaneously controlling for spatial confounding by individual-level risk factors, household and community-level socioeconomic status, and occupational and environmental exposures.

use) do not fully explain the observed patterns. Although provocative, these past studies examined location at a regional level for one time point in the woman's life. The importance of location, a potential proxy for social and environmental factors, is as much a function of exposure timing as it is of the exposures themselves.

The current study seeks to further investigate geographic disparities in breast cancer risk using prospective cohort data from the Nurses' Health Study II (NHSII) by applying generalized additive models (GAMs) to determine associations with geocoded address locations. This spatial approach allows for identification of geographic disease patterns while systematically determining predictors of the pattern. The first objective was to investigate the association between breast cancer risk and location at two time periods, during adolescence and early adulthood. Our second objective was to determine the impact of socioeconomic and environmental factors on the underlying spatial patterns of incident breast cancer risk associated with location in early adulthood, accounting for known individual-level breast cancer risk factors. Identification of geographic patterns and where patterns persist after adjustment for known risk factors may provide additional insight into unexplored risk factors that vary by location.

Methods

Study population

We investigated the association between location and incident invasive breast cancer risk using data from the NHSII, a long-term prospective cohort study of US female nurses. Participants were geographically distributed throughout the United States, although we restricted the spatial analyses to the contiguous 48 states. At the start of the study in 1989, 116,429 female nurses, 25–42 years old, completed a self-administered questionnaire and were followed up every 2 years to update current residential addresses, health outcomes, and behavioral risk factors. The medical records of participants were reviewed to confirm cases among those who self-reported a diagnosis of invasive breast cancer. We excluded women who developed all other cancers except non-melanoma skin cancer prior to their breast cancer diagnosis. From June 1989 to May 2013, we identified a total of 3,941 women with confirmed incident invasive breast cancer. We used incidence density sampling to randomly select 20 controls per case from among the noncases in the cycle of diagnosis ($n = 78,820$).^{6,7} From this nested case-control study, we then conducted our spatial analyses using addresses at adolescence and early adulthood.

Spatial analyses

We conducted spatial analyses by applying GAMs to determine the association between breast cancer incidence and geocoded addresses^{8,9} and to elucidate the potential role of environmental exposures and other risk factors in any observed geographic disparities.^{6,10–13} GAMs provide an efficient framework for simultaneously including a bivariate smooth of the x and y coordinate of a geocoded residence while adjusting for spatial confounders. We implemented a LOESS (locally weighted scatterplot smooth) with an a priori determined smoothing parameter of 0.2 to capture regional variation in breast cancer risk. The smoothing parameter reflects the proportion of data used to fit the local regression and it is well suited to adapt to variable population density.¹³ We fit a GAM using case status as the binary outcome as implemented in the R package MapGAM.^{10,12} To account for missingness among confounders, we used fully conditional specification implemented by the R package mice to generate five imputed datasets and present the average of the pointwise predictions.¹²

Geographic disparities associated with location during adolescence compared with early adulthood

Our first aim was to evaluate geographic patterns of incident breast cancer risk associated with location for two time period analyses: (1) during adolescence using high school addresses and (2) during early adulthood using addresses in 1991, the first year of cohort follow-up with robust addresses. In 1991, the mean age of participants was 25.7 years. High school addresses were only available for 32% of participants. We linked the participants in the nested case-control study to available high school addresses; the comparison analysis included all available participants with both sets of addresses ($n = 26,323$; 1,342 cases). In addition to the bivariate smooth for location, these spatial analyses included individual-level variables that may have been potential risk factors associated with exposures during adolescence: body mass index (BMI) at age 18, age at menarche, adolescent somatotype (average of somatotypes at ages 10 and 20), alcohol consumption at age 15 and 18, and family history of breast cancer. Events occurring after high school (e.g., screening mammography, postmenopausal hormonal use, parity, age at menopause) cannot affect residential location at high school, indicating that they are not confounders; as such, they were not included in the analyses comparing the contribution of addresses during adolescence and early adulthood to breast cancer risk.¹¹

Assessment of the role of socioeconomic and environmental factors on geographic disparities

Another important aim of the study was to assess the potential contribution of socioeconomic and environmental factors to spatial patterns of breast cancer risk. As these variables were not available for participants during adolescence, we restricted the analyses to use only early adulthood locations (1991 addresses), regardless of whether they were missing high school addresses. We applied a multi-stage spatial modeling approach that compared breast cancer associations across a series of models, each adjusted for additional covariate. We included for our baseline model the variables adjusted for in our analyses comparing adolescence and early adulthood: BMI, age at menarche, adolescent somatotype, alcohol consumption, and family history of breast cancer. This allowed us to assess the impact of our reduced sample size due to limited availability of high school addresses in our analyses comparing adolescence and early adulthood locations described above. Our baseline model associations were compared with models additionally adjusted for (1) socioeconomic status (SES) variables and (2) occupational and environment factors (i.e., fully adjusted model). If the results of the adjusted models revealed differences in the geographic areas of significant breast cancer odds ratios, then it is likely due to the presence of these additional factors in those regions. The SES model (1) included 1990 census tract median home value, median income, population density and whether the participant was married or lived alone as measures of household and community SES in addition to the baseline model covariates.

To assess if geographic disparities in breast cancer incidence were due to certain environmental factors, we then additionally included, in a fully adjusted model, variables related to the participants' environment that have been observed to have suggestive associations with breast cancer, including shift work, outdoor light at night (LAN) and radon exposure.^{14–16} Shift work was categorized as no shift work, <120 months, or 120 months of shift work or more. Cumulative LAN was estimated using 1-km² resolution satellite data to assign nighttime radiance values to participants' residential history.¹⁵ Radon exposure was categorized into quintiles of cumulative average radon based on the Lawrence Berkeley National Laboratory US radon exposure model of county-level indoor radon concentrations.¹⁶ Particulate matter¹⁷ and hazardous air pollutants¹⁸ were not associated with breast cancer incidence in the NHSII and thus

were not considered in the current spatial analyses. Although addresses were for 1991, all variables were continuously updated through 2013.

Lastly, as a sensitivity analysis, we included mammographic screening which has been shown to vary geographically¹⁹ and, in our cohort, is lower in the Southeast. To capture any residual confounding from using only measures of risk factors during adolescence in our baseline model, we also adjusted for potentially spatially varying individual risk factors measured later in life (postmenopausal hormonal use, current BMI, parity and age at first birth).

As our spatial confounders for the early adulthood analyses included group-level census tract SES measures, analyzing all address data may induce spatial clustering due to geographically linked factors. To account for this, we sampled from our eligible cases and non-cases in our nested case-control study ($n = 82,761$; 3,941 cases) one participant per census tract using the `sampcont` function in the `MapGAM` package in R.¹⁰ After sampling and excluding participants with missing data for geocoded 1991 residential coordinates (0.2%), the final analyses for the early adulthood addresses included 3,478 cases and 21,041 non-cases. GAMs were then fit to the independent individual-level data using inverse probability weighting.^{20,21}

In addition to our primary spatial analyses that examined risk of all incident invasive breast cancer cases, we performed secondary analyses of estrogen receptor (ER)-positive ($n = 2,084$) and ER-negative cases ($n = 509$), each with 20,994 non-cases. We also stratified by menopausal status at breast cancer diagnosis. Among premenopausal women, there were 1,786 cases and 12,062 non-cases; among postmenopausal women, there were 885 cases and 9,738 non-cases. All secondary analyses used addresses during early adulthood and were fully adjusted.

Mapping of breast cancer odds ratios

We predicted continuous breast cancer odds ratios across the United States, excluding regions of extremely low population density along the geographic edges of our study population.²¹ Odds ratios were calculated using the odds of breast cancer incidence in the whole study area as the reference, holding confounders constant at the reference level for categorical covariates and the median value for continuous covariates.¹² The `MapGAM` package also allows for calculation of confidence intervals for the point estimates of the risk map. Geographic areas where the confidence interval excludes one are indicated on the map with black contour lines. The absence of these contours indicates that the odds ratios are not statistically significant. All maps were mapped using the same odds ratio scale ranging from 0.5 to 1.6 for better comparison across the analyses. We present odds ratio ranges for the average across the five imputed datasets.

Analyses were conducted using R Package 3.4.0 (R Foundation for Statistical Computing, Vienna, Austria) and maps were created using the `colormap` function in the `MapGAM` package.^{20,21} The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital, Harvard T.H. Chan School of Public Health, the University of California, Irvine, and those of participating registries as required. Return of the questionnaires implied informed consent, and all participants (or next-of-kin) provided written approval to obtain medical records.

Results

The distribution of selected characteristics among study participants are provided in the Table. Cases of breast cancer had a higher proportion of family history of breast cancer and personal history of BBD compared with non-cases. Consequently, they were also more likely to have a screening mammography compared with non-cases. Cases were also of higher SES, as

Table. Select characteristics of NHSII participants ($n = 24,519$) included in spatial analyses of early adulthood (1991) addresses.

Characteristic	Cases ($n = 3,478$)	Non-cases ($n = 21,041$)
Age at diagnosis or index year (years) (mean \pm SD)	49.0 \pm 6.9	47.5 \pm 7.6
White (%)	95.7	93.9
Family history of breast cancer (%)	18.5	10.5
Age at menarche (%)		
<12 years	25.7	24.8
12 years	30.2	30.0
13 years	28.0	27.2
14+ years	15.7	17.8
Missing	0.4	0.4
Parity and age at first birth (%)		
Nulliparous	18.0	16.1
1–2 children and <25 years	13.9	14.3
1–2 children and 25–29 years	19.8	18.0
1–2 children and \geq 30 years	15.4	13.1
3–4 children and <25 years	9.3	10.3
3–4 children and 25–29 years	10.0	9.0
3–4 children and \geq 30 years	3.2	2.3
5+ children and <25 years	0.7	1.0
5+ children and \geq 25 years	0.6	0.7
Missing	9.3	15.3
Menopausal status and hormone use (%)		
Premenopausal	56.1	58.3
Never users	6.6	6.0
Past users	13.2	14.3
Current users	13.1	9.5
Missing	11.0	11.9
Body mass index (BMI) (kg/m ²) at age 18 (mean \pm SD)	20.7 \pm 2.7	21.3 \pm 3.2
Missing (%)	1.1	1.4
Current BMI (%)		
<18.5 kg/m ²	1.3	1.6
18.5–19.9 kg/m ²	5.6	5.5
20–21.9 kg/m ²	19.4	19.6
22–24.9 kg/m ²	21.5	20.7
25–29.9 kg/m ²	26.9	26.8
30–34.9 kg/m ²	13.5	14.0
35+ kg/m ²	8.8	11.6
Missing	3.0	0.3
Personal history of biopsy-confirmed benign breast disease (BBD) (%)	29.6	18.4
Missing	0.02	0.0
Alcohol consumption at age 15–17 years (g/day) (mean \pm SD)	0.8 \pm 3.2	1.1 \pm 3.6
Missing (%)	0.8	1.1
Alcohol consumption at age 18–22 years (g/day) (mean \pm SD)	4.7 \pm 7.6	4.9 \pm 8.1
Missing (%)	0.6	0.7
Screening mammography (%)		
No mammogram since last questionnaire	12.3	16.8
Screening mammogram	68.0	55.6
Missing	19.8	27.6
Currently married (%)	61.8	55.1
Lives alone (%)	8.4	7.4
Census tract population density (per km ²) (mean \pm SD)	3,350 \pm 7,829	3,718 \pm 9,811
Census tract level median income (\$) (mean \pm SD)	65,963 \pm 24,960	62,387 \pm 24,771
Census tract median home value (\$1,000) (mean \pm SD)	172.6 \pm 125.8	166.0 \pm 128.6
Missing (%)	0.06	0.07
Cumulative average LAN (nW/m ² /sr) (mean \pm SD)	28.3 \pm 25.1	28.9 \pm 26.6
Missing (%)	0.03	0.05
Radon quintiles of exposure (%)		
First	21.0	24.0
Second	20.4	21.9
Third	19.8	18.4
Fourth	18.6	18.0
Fifth	20.1	17.6
Missing	0.03	0.1
Shift work		
No shift work	31.1	30.6
120 months of shift work	61.8	62.6
120+ months of shift work	7.2	6.7
Missing (%)	0.0	0.1

indicated by their marital status and higher census tract median income and home values. Data for most covariates included in the analyses were missing for <2% of participants (BMI at age 18, age at menarche, adolescent somatotype, alcohol consumption, family history of breast cancer, census tract median home value and median income, shift work, cumulative LAN, radon exposure). However, postmenopausal hormonal use, screening mammography, parity, and age at first birth were missing for ~10%–30% of participants.

The geographic distributions of adjusted breast cancer odds ratios at the participants' addresses during adolescence (Figure 1A) and early adulthood (Figure 1B) show significant variation. After adjusting for family history of breast cancer, BMI at age 18, age at menarche, alcohol consumption, and adolescent somatotype, odds of incident breast cancer at their location during adolescence were increased in the Northwest, Michigan and Iowa, and the greater New York City area relative to the average odds in the study area. Decreased odds were observed in northern New York and across West Virginia (Figure 1A; OR range: 0.1–2.0). The early adulthood analysis revealed similar patterns of breast cancer risk, although increased risk had shifted from the New York City area to southern New England. Risk in the Northwest remained elevated but was no longer significant whereas risk in Iowa was significantly increased in the early adulthood analysis (Figure 1B; OR range: 0.3–2.0).

Figure 2 shows the contribution of SES and environmental factors to geographic patterns of breast cancer odds ratios associated with the participants' location during early adulthood. Similar to results shown in Figure 1B (which used a subset of participants of the NHSII cohort with both addresses), odds of breast cancer were generally lower in the Southeast and higher in Iowa, Ohio, and southern New England relative to the average odds in the study area. Geographic disparities in risk remained even after adjustment for SES and environmental factors, although odds ratios in these areas were no longer statistically significant except for southern New England. Compared with the baseline results (adjusted for the BMI, age at menarche,

adolescent somatotype, alcohol consumption, and family history of breast cancer, Figure 2A; OR range: 0.6–1.4), inclusion of household and community-level SES variables slightly attenuated the ORs (range: 0.7–1.3), and the statistically significant area of increased risk in Ohio was no longer significant (Figure 2B). Additional adjustment for occupational and environmental factors (shift work, cumulative LAN, and radon exposure) did not appreciably change the magnitude of the association between breast cancer and location during early adulthood (Figure 2C; OR range: 0.6–1.3). Furthermore, mammographic screening and individual known risk factors included in the sensitivity analyses did not explain the geographic disparities, but areas of increased risk in the Southeast and southern New England were no longer statistically significant after inclusion (Figure 2D; OR range: 0.6–1.3).

When we examined cases stratified by ER status, we observed similar patterns of reduced ORs in the Southeast with ER-positive cases but for a much larger region (Figure 3A; OR range: 0.4–1.8). Geographic distributions of ORs were very different for ER-negative cases (Figure 3B; OR range: 0.3–2.7), with high ORs in the Southwest, but the number of cases was much smaller and results less stable. Premenopausal breast cancer risk was also decreased in the Southeast and increased in Ohio and the southern New York area (Figure 4A; OR range: 0.5–1.6). Postmenopausal breast cancer risk was decreased in Florida (Figure 4B; OR range: 0.1–2.0). Confidence intervals for the odds ratios associated with these spatial patterns did not exclude one when averaged across the five imputed analyses.

Discussion

Our analyses suggest that women have different breast cancer risks depending on where in the United States they live at the time of adolescence or early adulthood. Women living in the New York City area during adolescence had increased ORs, while those living in southern New England during early adulthood (average age was in their mid-twenties) were at increased

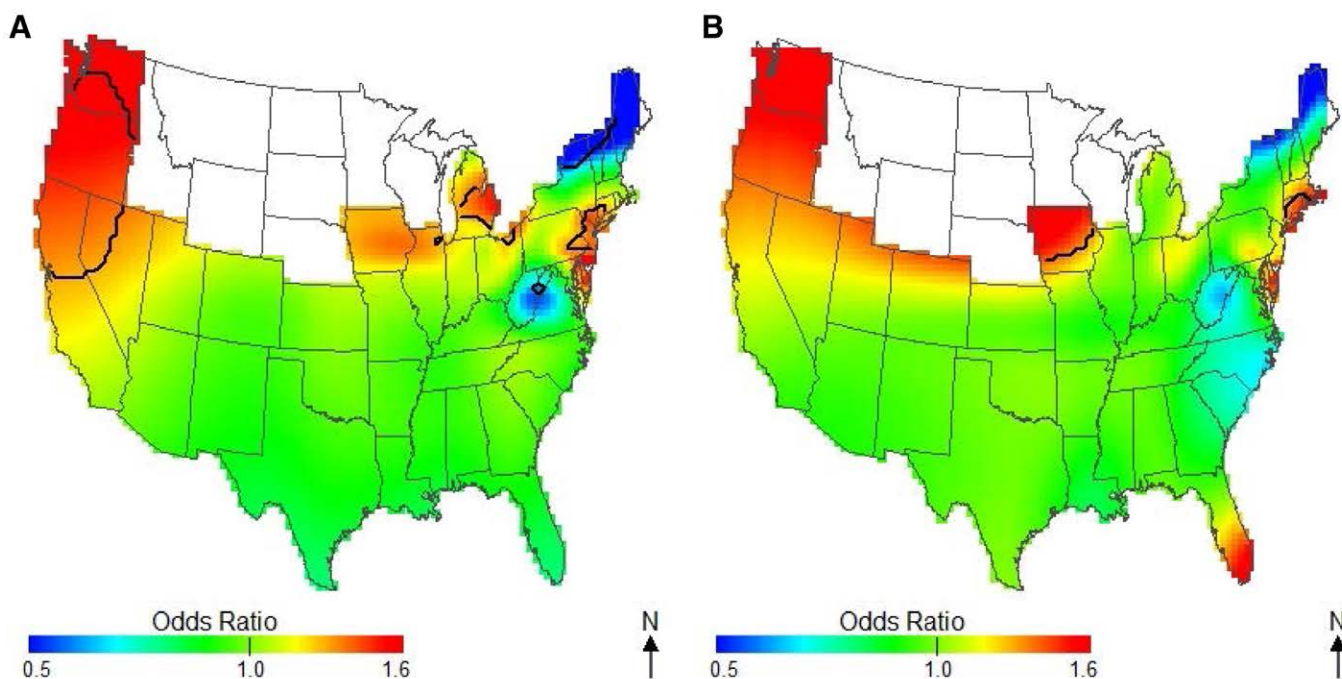


Figure 1. Geographic distribution of adjusted incident breast cancer odds ratios associated with participants' location during adolescence (A) and early adulthood (B) for subset of NHSII participants with both addresses ($n = 26,323$). The address during high school and at study baseline was used to examine location in the adolescence and early adulthood analyses, respectively. Both analyses were adjusted for family history of breast cancer, alcohol consumption at age 15 and 18, BMI at age 18, age at menarche, and adolescent somatotype. Black contour lines indicate areas where the upper and lower confidence bands exclude one.

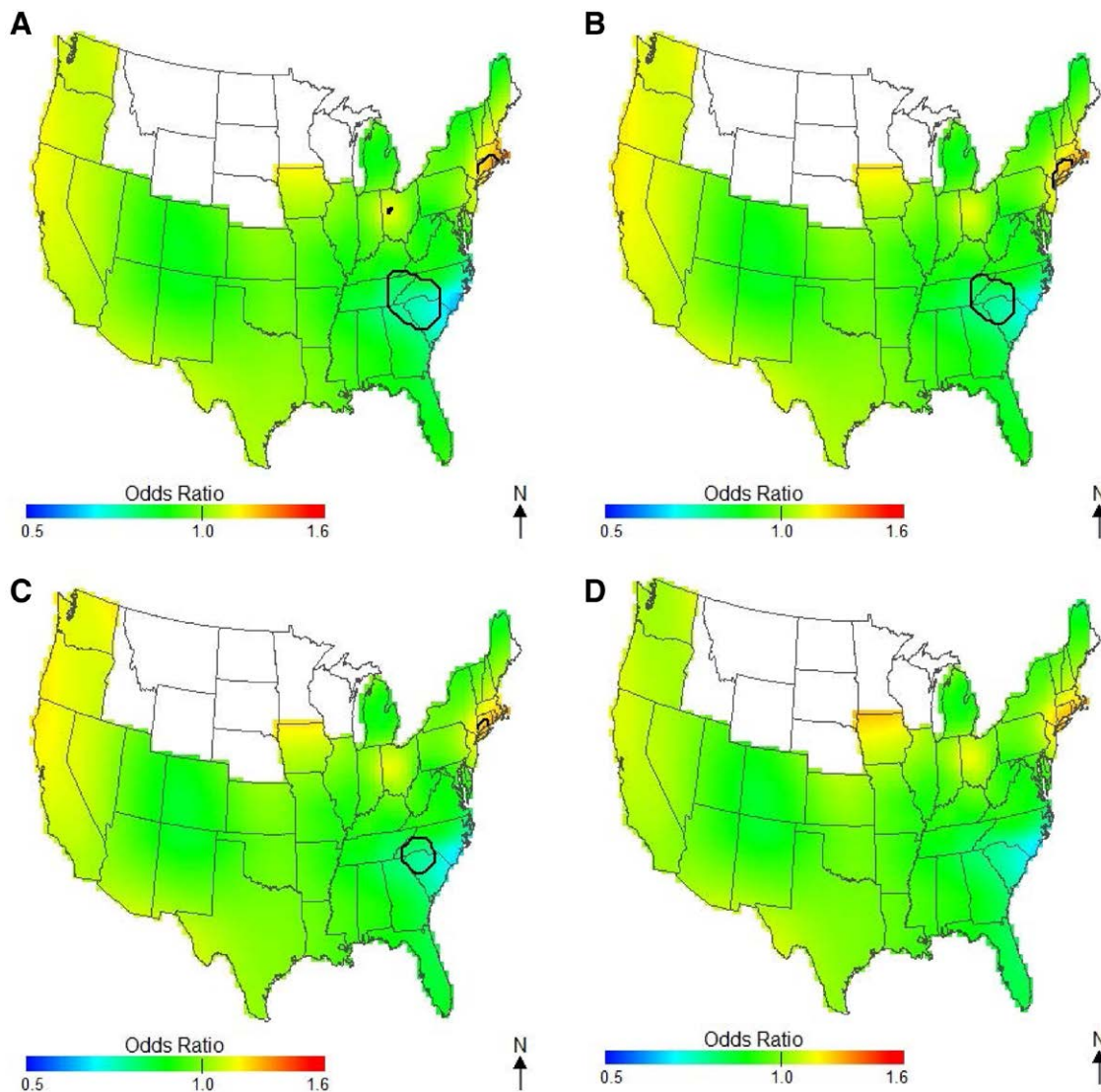


Figure 2. Contribution of covariates to geographic distribution of incident breast cancer odds ratios associated with NHSII participants' location during early adulthood. Analyses of 1991 addresses included adjustment for baseline risk factors previously controlled for: family history of breast cancer, alcohol consumption at age 15 and 18, BMI at age 18, age at menarche, and adolescent somatotype (A), followed by further adjustment for SES variables (B), environmental exposures (C), and additional adulthood individual level risk factors including screening mammography, postmenopausal hormonal use, parity, and age at menopause (D). Black contour lines indicate areas where the upper and lower confidence bands exclude one.

risk. Geographic disparities observed in Iowa, Michigan, Ohio, and the Southeast were no longer statistically significant after fully adjusting for potential risk factors. For example, the area of significant increased ORs in Ohio was no longer apparent after median income was adjusted for. Patterns also differed by ER and menopausal status, indicating that geographic risk factors vary by type of breast cancer, although numbers of ER-negative and post-menopausal cases were small. Thus, we are cautious in interpreting differences among these types as they are likely underpowered and more susceptible to edge effects.^{8–10}

To our knowledge, our analyses are the first to investigate address-level geographic variation at a national scale while simultaneously controlling for spatial confounding by individual-level risk factors, household and community-level SES, and occupational and environmental exposures (shift work, cumulative LAN, radon exposure). These environmental exposures were selected because of recent studies that suggested associations with breast cancer risk in the NHSII.^{14–16} Other recent studies of particulate matter¹⁷ and hazardous air pollutants¹⁸ did not support an association with breast cancer incidence in the NHSII and were not considered in the current spatial

analyses. Our results showing elevated breast cancer risk in the New York and southern New England areas are consistent with results from studies of other cohorts that identified increased breast cancer risk in the New York City and Long Island areas of New York,²² in parts of Connecticut²³ and throughout Cape Cod County in Massachusetts.^{24,25} In Cape Cod County, drinking water contamination has been implicated as a potential risk factor,^{26,27} suggesting that the geographic disparities identified in our study may be related to local environmental exposures.

Although patterns of breast cancer risk varied in their significance, the magnitude of breast cancer risk appeared to be independent of several established individual risk factors as well as socioeconomic and environmental predictors, indicating that these factors are not driving the disparities we observed and that other geographically distributed factors may be associated with breast cancer ORs among women in the NHSII. While we cannot rule out inadequate measurement of the included variables in our cohort, the persistence of geographic disparities after controlling for confounders has also been seen in other spatial studies.^{4,5} The combination of individual geographic and risk factor data from a prospective cohort study with a point-level

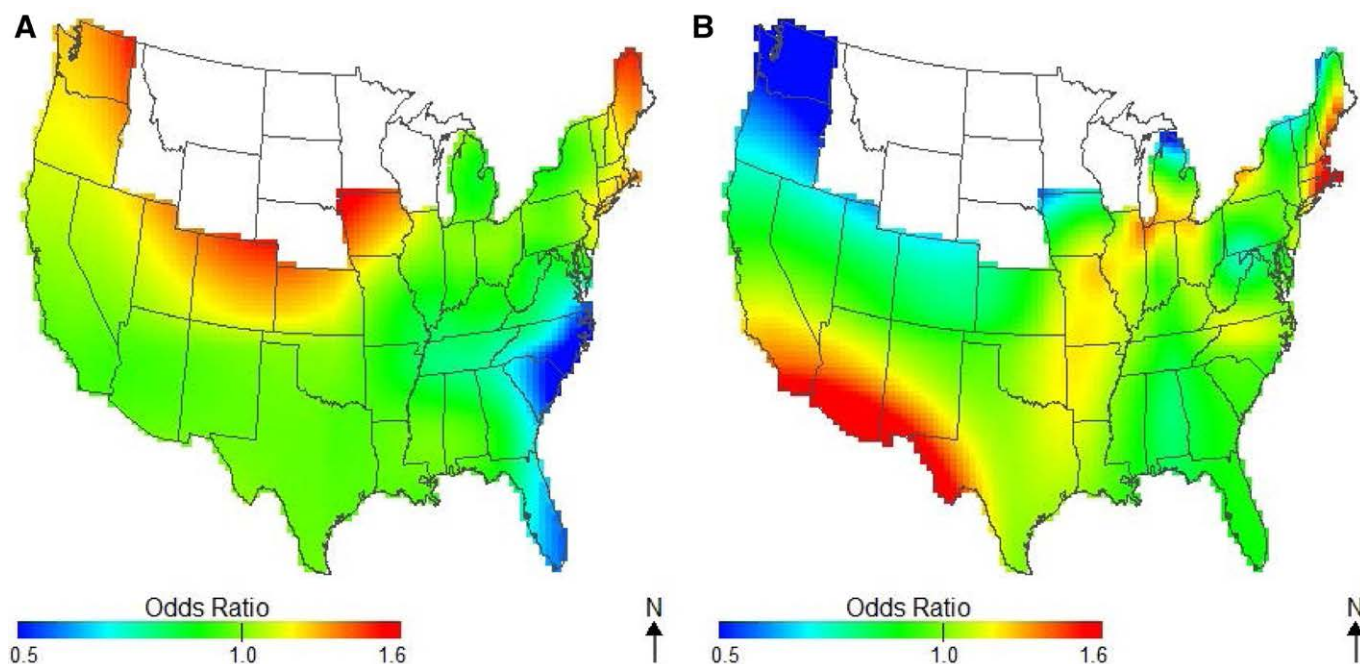


Figure 3. Differences by ER status in the geographic distribution of incident breast cancer odds ratios associated with NHSII participants' location during early adulthood. Geographic distribution of ER-positive (A) and ER-negative (B) adjusted incident breast cancer odds ratios associated with NHSII participants' location during early adulthood. Analyses of 1991 addresses included adjustment for SES variables, environmental exposures, and individual level risk factors.

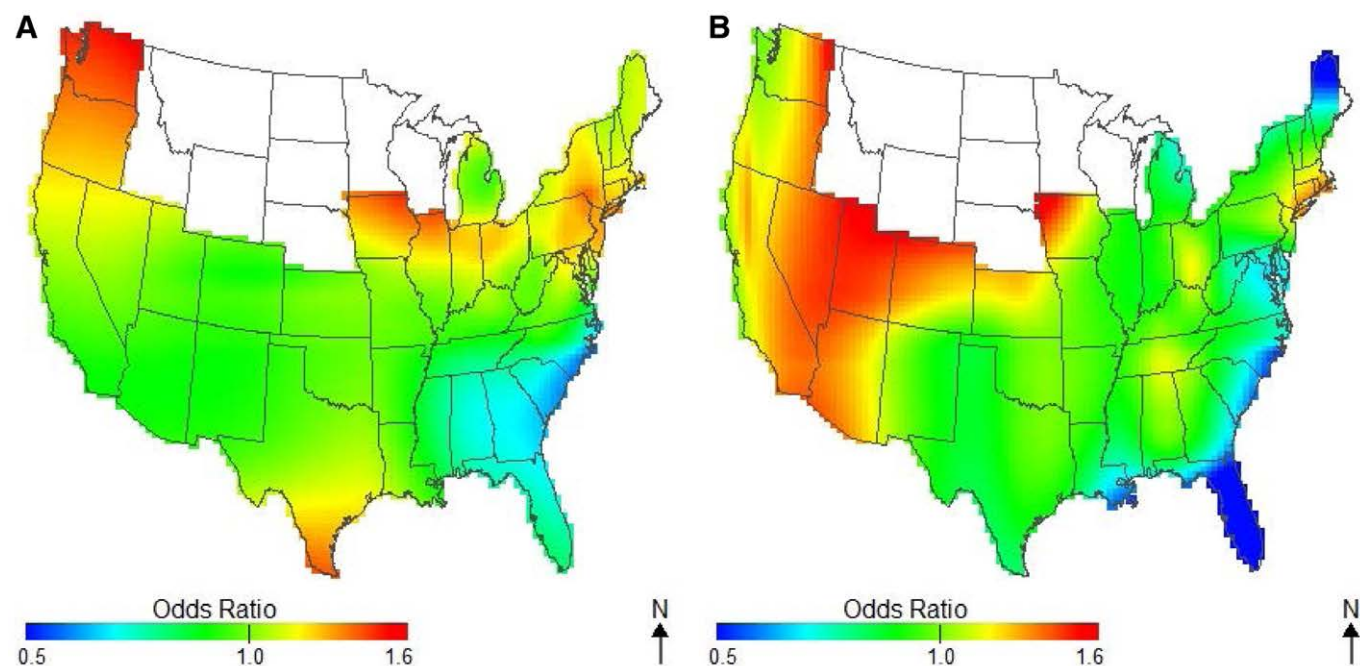


Figure 4. Differences by menopausal status in the geographic distribution of incident breast cancer odds ratios associated with NHSII participants' location during early adulthood. Geographic distribution of premenopausal (A) and postmenopausal (B) adjusted incident breast cancer odds ratios associated with NHSII participants' location during early adulthood. Analyses of 1991 addresses included adjustment for SES variables, environmental exposures, and individual level risk factors.

spatial modeling approach allowed us to examine geographic disparities in breast cancer risk that would not be feasible when using aggregated residential data.

Despite the methodological strengths, our current geographic analyses have some potential limitations. Although address data were available for women throughout the United

States, data were sparse in some regions (e.g., the upper Midwest). We excluded these areas from our predictions and are cautious in interpreting data from low-population density regions. We did not control for race as our study population is 95% white; this limits generalizability of our results to other populations. Geographic patterns of breast cancer risk may

differ for nonwhite racial groups. In addition, patterns may be more pronounced for groups with limited access to screening compared with nurses. Although we included mammographic screening because of its strong spatial component,¹⁹ nurses are likely to seek out screening more than other women.²⁸ Our analyses are also limited by missing high school addresses for many of the participants and the lack of sociodemographic and environmental data for that adolescent time period. Our assessment of the contribution of these factors to breast cancer risk was restricted to location of the woman during early adulthood using their address in 1991. Similarly, we do not have full residential histories for participants prior to enrollment at the study inception, restricting our ability to conduct an extensive spatiotemporal analysis as has been done in other studies in Cape Cod, MA²⁵ and Denmark.²⁹ Nonetheless, using the MapGAM package in R, we generated maps of breast cancer risk in the United States and identified areas with statistically significant breast cancer odds ratios. Patterns may be explained by spatial variation in inadequately measured or unmeasured environmental exposures. Further research is needed to understand geographic disparities in breast cancer and possible spatial risk factors.

Conclusions

Breast cancer risk is not spatially uniform across the United States and varied depending on timing of residence. Geographic disparities persisted even after accounting for established and suspected breast cancer predictors, suggesting that unmeasured environmental or lifestyle risk factors are distributed unevenly in different parts of the country.

Conflicts of interest statement

The authors declare that they have no conflicts of interest with regard to the content of this report.

Acknowledgements

We would like to thank the participants and staff of the Nurses' Health Study II for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, and WY. The authors assume full responsibility for analyses and interpretation of these data.

References

1. Takahashi MH, Thomas GA, Williams ED. Evidence for mutual interdependence of epithelium and stromal lymphoid cells in a subset of papillary carcinomas. *Br J Cancer*. 1995;72:813–817.
2. Graham S, Zielezny M, Marshall J, et al. Diet in the epidemiology of postmenopausal breast cancer in the New York State Cohort. *Am J Epidemiol*. 1992;136:1327–1337.
3. Bako G, Dewar R, Hanson J, Hill G. Population density as an indicator of urban-rural differences in cancer incidence, Alberta, Canada, 1969–73. *Can J Public Health*. 1984;75:152–156.
4. Reynolds P, Hurley S, Goldberg DE, et al. Regional variations in breast cancer among California teachers. *Epidemiology*. 2004;15:746–754.
5. Laden F, Spiegelman D, Neas LM, et al. Geographic variation in breast cancer incidence rates in a cohort of U.S. women. *J Natl Cancer Inst*. 1997;89:1373–1378.
6. Vieira VM, Hart JE, Webster TF, et al. Association between residences in U.S. northern latitudes and rheumatoid arthritis: a spatial analysis of the Nurses' Health Study. *Environ Health Perspect*. 2010;118:957–961.

7. Richardson DB. An incidence density sampling program for nested case-control analyses. *Occup Environ Med*. 2004;61:e59.
8. Hastie T, Tibshirani R. *Generalized Additive Models*. London: Chapman and Hall; 1990.
9. Kelsall J, Diggle P. Spatial variation in risk of disease: a nonparametric binary regression approach. *J Roy Stat Soc C-App Statist*. 1998;47:559–573.
10. Hoffman K, Weisskopf MG, Roberts AL, et al. Geographic patterns of autism spectrum disorder among children of participants in Nurses' Health Study II. *Am J Epidemiol*. 2017;186:834–842.
11. Hoffman K, Webster TF, Weinberg JM, et al. Spatial analysis of learning and developmental disorders in upper Cape Cod, Massachusetts using generalized additive models. *Int J Health Geogr*. 2010;9:7.
12. Vieira VM, Fabian MP, Webster TF, Levy JJ, Korrick SA. Spatial variability in ADHD-related behaviors among children born to mothers residing near the new bedford harbor superfund site. *Am J Epidemiol*. 2017;185:924–932.
13. Webster T, Vieira V, Weinberg J, Aschengrau A. Method for mapping population-based case-control studies: an application using generalized additive models. *Int J Health Geogr*. 2006;5:26.
14. James P, Bertrand KA, Hart JE, Schernhammer ES, Tamimi RM, Laden F. Outdoor light at night and breast cancer incidence in the Nurses' Health Study II. *Environ Health Perspect*. 2017;125:087010.
15. Wegrzyn LR, Tamimi RM, Rosner BA, et al. Rotating night-shift work and the risk of breast cancer in the Nurses' Health Studies. *Am J Epidemiol*. 2017;186:532–540.
16. VoPham T, DuPré N, Tamimi RM, et al. Environmental radon exposure and breast cancer risk in the Nurses' Health Study II. *Environ Health*. 2017;16:97.
17. Hart JE, Bertrand KA, DuPré N, et al. Long-term particulate matter exposures during adulthood and risk of breast cancer incidence in the Nurses' Health Study II Prospective Cohort. *Cancer Epidemiol Biomarkers Prev*. 2016;25:1274–1276.
18. Hart JE, Bertrand KA, DuPré N, et al. Exposure to hazardous air pollutants and risk of incident breast cancer in the nurses' health study II. *Environ Health*. 2018;17:28.
19. Chandak A, Nayar P, Lin G. Rural-Urban disparities in access to breast cancer screening: a spatial clustering analysis. *J Rural Health*. 2019;35:229–235.
20. Bai L, Bartell SM, Bliss RL, Vieira VM. MapGAM: Mapping Smoothed Odds Ratios from Individual-Level Data, R package version 1.0. 2016. <https://CRAN.R-project.org/package=MapGAM>.
21. Bai L, Gillen DL, Bartell SM, Vieira VM. Mapping smoothed spatial effect estimates from individual-level data: MapGAM. *R journal*. 2019 (in press).
22. Kulldorff M, Feuer EJ, Miller BA, Freedman LS. Breast cancer clusters in the northeast United States: a geographic analysis. *Am J Epidemiol*. 1997;146:161–170.
23. Rybnikova N, Stevens RG, Gregorio DI, Samociuk H, Portnov BA. Kernel density analysis reveals a halo pattern of breast cancer incidence in Connecticut. *Spat Spatiotemporal Epidemiol*. 2018;26:143–151.
24. Vieira V, Webster T, Weinberg J, Aschengrau A, Ozonoff D. Spatial analysis of lung, colorectal, and breast cancer on Cape Cod: an application of generalized additive models to case-control data. *Environ Health*. 2005;4:11.
25. Vieira VM, Webster TF, Weinberg JM, Aschengrau A. Spatial-temporal analysis of breast cancer in upper Cape Cod, Massachusetts. *Int J Health Geogr*. 2008;7:46.
26. 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;118:749–755.
27. Gallagher LG, Vieira VM, Ozonoff D, Webster TF, Aschengrau A. Risk of breast cancer following exposure to tetrachloroethylene-contaminated drinking water in Cape Cod, Massachusetts: reanalysis of a case-control study using a modified exposure assessment. *Environ Health*. 2011;10:47.
28. Tsai RJ, Luckhaupt SE, Sweeney MH, Calvert GM. Shift work and cancer screening: do females who work alternative shifts undergo recommended cancer screening? *Am J Ind Med*. 2014;57:265–275.
29. Nordborg RB, Meliker JR, Ersbøll AK, Jacques 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.