ehp ENVIRONMENTAL HEALTH PERSPECTIVES

ehponline.org

Association between Residences in U.S. Northern Latitudes and Rheumatoid Arthritis: A Spatial Analysis of the Nurses' Health Study

Verónica M. Vieira, Jaime E. Hart, Thomas F. Webster, Janice Weinberg, Robin Puett, Francine Laden, Karen H. Costenbader, and Elizabeth W. Karlson

doi: 10.1289/ehp.0901861 (available at http://dx.doi.org/)
Online 25 March 2010

NIEHS
National Institute of Environmental Health Sciences

National Institutes of Health
U.S. Department of Health and Human Services

Association between Residences in U.S. Northern Latitudes and Rheumatoid Arthritis: A Spatial Analysis of the Nurses' Health Study

Verónica M. Vieira[1], Jaime E. Hart[2,3], Thomas F. Webster[1], Janice Weinberg[4], Robin Puett[5,6], Francine Laden[2,3,7], Karen H. Costenbader[8], and Elizabeth W. Karlson[8]

[1]Department of Environmental Health, Boston University School of Public Health, Boston, Massachusetts, USA; [2] Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; [3] Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA; [4]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA; [5]South Carolina Cancer Prevention and Control Program, University of South Carolina, Columbia, South Carolina, USA; [6]Departments of Environmental Health Sciences and Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, USA; [7]Exposure, Epidemiology, and Risk Program, Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, USA; [8]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

Corresponding author and full address of where work was performed:

Verónica Vieira, D.Sc.

Department of Environmental Health

Boston University School of Public Health

715 Albany Street, T4W

Boston, Massachusetts 02118-2526, USA

Tell: 617-638-6479; Fax: 617-638-4857; Email: vmv@bu.edu

Running title: Spatial Analysis of RA

Key words: Geographic information systems (GIS), Generalized additive models, Disease mapping, Prospective cohort study, Rheumatoid Arthritis

Acknowledgments/grant support

This work was supported by grant number 5 P42 ES007381 from the National Institute of Environmental Health (NIEHS), NIH grants R01 AR49880, CA87969, P60 AR047782, K24 AR0524-01, BIRCWH K12 HD051959 (supported by NIMH, NIAID, NICHD, and OD). Its contents are solely the responsibility of the authors and do not necessarily represent the views of NIH.

Competing Interests Declaration: The authors do not have any competing interests.

Abbreviations

AIC: Akaike's Information Criterion

AOR: Adjusted Odds Ratio

COR: Crude Odds Ratio

GAM: Generalized Additive Model

GIS: Geographic Information Systems

NHS: Nurses' Health Study

RA: Rheumatoid Arthritis

UV: Ultra Violet

Outline of manuscript sections

Abstract

Introduction

Materials and Methods

- Study Population
- Spatial Analysis

Results

Discussion

Conclusions

References

Tables

Figure Legends

Figures

Abstract

Background: The etiology of rheumatoid arthritis (RA) remains largely unknown although epidemiologic studies suggest genetic and environmental factors may play a role. Geographic variation in incident RA has been observed at the regional level.

Objective: Spatial analyses are a useful tool for confirming existing exposure hypotheses or generating new ones. To explore further the association between location and RA risk, we analyzed individual level data from U.S. women in the Nurses' Health Study, a nationwide cohort study.

Methods: Participants included 461 incident RA cases and 9,220 controls with geocoded addresses followed from 1988-2002. We examined spatial variation using addresses at baseline in 1988 and at time of case diagnosis/censoring of controls. Generalized additive models were used to predict a continuous risk surface, smoothing on longitude and latitude while adjusting for known risk factors. Permutation tests were conducted to test for the overall importance of location and identify areas of statistically significant risk relative to the whole study area.

Results: A statistically significant area of increased RA risk was identified in the northeast U.S. (p-value=0.034). Risk was generally higher at northern latitudes and increased slightly using nurses' 1988 locations compared to locations at time of diagnosis/censoring. Crude and adjusted models produced similar results.

Conclusions: Spatial analyses suggest women living in higher latitudes may be at greater risk for RA. Further, RA risk may be greater for locations occurring earlier in residential histories. These results illustrate the usefulness of GAM methods in generating hypotheses for future investigation and supporting existing hypotheses.

Background

Rheumatoid arthritis (RA) is a chronic autoimmune disease with unknown etiology, although epidemiologic studies suggest genetic and environmental factors may play a role. Research on other chronic autoimmune diseases including lupus erythematosus, dermatomyositis, polymyositis and vasculitis has shown geographic associations with higher latitudes (Somers et al. 2007; Walsh and Gilchrist 2006; Hengstman et al. 2000; Gatenby et al. 2009). Geographic variation in incident RA has been observed at the regional level according to state of residence (Costenbader et al. 2008). The findings suggested increased risk of RA for women who lived in the midwestern and eastern United States compared to the west, and the association was stronger with residency at age 15 and 30 than at baseline in 1976. A review by Alamanos et al. (2006) also suggests that RA varies geographically among areas of the world, with southern European countries having lower median incidence rates than northern European countries and North America. Ramos-Remus et al. (2007) observed that the mean age of RA onset was much younger among Mexicans compared to Canadians, and results of a study by Anaya et al. (2001) suggest that RA in African populations is rare.

To explore further the association between location and RA risk, we analyzed individual-level residential data from U.S. women in the Nurses' Health Study. This prospective cohort study provides information on personal covariates and participant mobility prior to RA onset. Residential histories are particularly useful when exposures of interest are time-dependent. We conducted spatial analyses that considered time measured by calendar year and by year of diagnosis for cases or censoring for controls. Generalized additive models (GAMs), a type of statistical model that combines smoothing with the ability to analyze binary outcome data and adjust for covariates, provide a useful framework for examining point data (Hastie and Tibshirani 1990; Kelsall and Diggle 1998; Webster et al. 2006). Using individual-level information and

location in a generalized additive model, we calculated the crude and adjusted odds ratios for incident RA in the U.S. This method has the advantage of controlling for spatial confounders and allowing for hypothesis testing for the significance of location in the disease maps. The objectives of the present analyses are to examine geographic variation at the individual level and identify potential exposure hypotheses for further investigation.

Methods

Study Population

We investigated the association between residence and incident rheumatoid arthritis (RA) using data from the Nurse's Health Study (NHS), a long-term prospective cohort study of U.S. female nurses. At study inception in 1976, 121,700 nurses, ages 30–55 years, living in 11 states (California, Connecticut, Florida, Massachusetts, Maryland, Michigan, New Jersey, New York, Ohio, Pennsylvania, and Texas) and recruited through the state licensing boards, completed a mailed questionnaire and provided informed consent (Stampfer et al. 2000; Oh et al. 2005; Hart et al. 2009). Every 2 years, follow-up questionnaires are mailed to update information on current residential address, health outcomes and behavioral risk factors. Residential addresses for 103,341 nurses from the Nurses' Health Study questionnaire cycles 1988-2002 were geocoded, contributing 173,624 addresses and 762,511 questionnaire records. Although nurses' addresses were concentrated in the 11 original study states at baseline in 1976, they have moved to all 50 states during follow-up. Medical records of nurses who self-report an RA diagnosis undergo detailed examination for American College of Rheumatology (ACR) diagnostic criteria for RA (Karlson et al. 1995; Karlson et al. 2004). There were 461 cases of confirmed incident RA from 1988 through 2002. Information for participants was censored once they became a case, and their residential history was considered complete at that time. We used incidence density

sampling to randomly select 20 controls per case from among the non-cases for a total of 9,220 controls (Richardson 2004). Selecting 20 controls per case allowed us to analyze the geographic distribution of the population under study, while still keeping numbers reasonable for computation. Women who were non-cases with a geocoded address at the time a case was diagnosed were eligible to be a control. Once selected, information for controls was censored and their residential history was considered complete at that time. As a result, the proportion of participants censored in each year is the same for cases and controls.

Spatial Analysis

We estimated disease odds using generalized additive models, a form of non-parametric or semi-parametric regression with the ability to analyze binary and continuous outcome data while adjusting for covariates (Hastie and Tibshirani 1990; Kelsall and Diggle 1998; Webster et al. 2006). We modeled location, a potential surrogate measure of exposure, using a bivariate smooth (S) of longitude and latitude (x_1) and (x_2)

$$\text{logit}[p(x_1,x_2)] = S(x_1,x_2) + \boldsymbol{\gamma}'\mathbf{z} \qquad [1]$$

where the left-hand side is the log of the disease odds at location (x_1,x_2), $\mathbf{z}$ is a vector of covariates, and $\boldsymbol{\gamma}$ is a vector of parameters. The model is semiparametric because it has the nonparametric smooth but the covariates are modeled parametrically. Without the smooth function, $S(x_1,x_2)$, the model becomes an ordinary logistic regression on the covariates. To examine if timing of residential location impacts RA risk, we conducted analyses using 1988 addresses (the earliest available that were geocoded) and using addresses at time of diagnosis/censoring for all participants.

Spatial confounding occurs when risk factors for a disease are not evenly distributed. A group of core confounders, chosen *a priori* based on the current scientific literature (Costenbader et al. 2006; Uhlig et al. 1999; Liao et al. 2009) or study design, was included in the analyses and

modeled as shown in Table 1 unless otherwise noted: age (in months), non-Caucasian race, age at menarche, parity, total months of lactation, current menopausal status, menopausal hormone use, oral contraceptive use, physical activity, body mass index (BMI, modeled categorically), cigarette pack-years (calculated as the number of packs/day multiplied by number of years of cigarette smoking), current smoking status, and socioeconomic status measured by nurses' educational level, occupation of both parents, marital status, and husband's education (if applicable). These variables were obtained from the questionnaire cycle corresponding to their address used in the analysis (i.e. in the 1988, BMI in 1988 was used).

We used a loess smooth which adapts to changes in population density (Hastie and Tibshirani 1990) and determined the optimal amount of data for the smooth, or span size, for each map by minimizing the Akaike's Information Criterion (AIC). A rectangular grid covering the continental U.S. was created using the minimum and maximum longitude and latitude from the study subjects. Because GAMs may exhibit biased behavior at the edges of the data, the study area for spatial model predictions is the continental U.S. excluding regions of low population density along the geographic edges of our study population. Based on sensitivity analyses to determine the impact of sparse data areas on the predicted results, we identified the midwest (Great Plains), northern Maine, and southwest Texas as low population density regions where a participant's nearest neighbor is more than 200 kilometers away along the geographic edges of the study population. We converted from log odds to odds ratios (ORs) using the odds of disease in the whole study area as the reference.

GAMs also provide a framework for hypothesis testing. We first tested the null hypothesis that case status does not depend on the smooth term using a permutation test based on the difference of the deviances of model (1) with and without the smooth term (Webster et al. 2006). We discuss results as "significant" if the associated p-values are less than 0.05, but

acknowledge that some results may be due to chance. If the global deviance test indicated that geographic location is important, we next examined point-wise departures from the null hypothesis using permutation tests to identify areas of the map that exhibit unusually high or low disease odds. We used contour lines to denote areas of significant decreased RA risk (points that ranked in the lower 2.5% of the point-wise permutation distributions indicating low disease odds) or increased RA risk (upper 2.5% of the point-wise permutation distributions indicating high disease odds). We used S-Plus (version 8.0; Insightful Corp. 2007) to perform the generalized additive modeling and a geographic information system (ArcGIS, version 9.3; ESRI, Redlands, CA) to map the results of our analyses. The Institutional Review Boards of Boston University Medical Center and Brigham and Women's Hospital approved the research.

Results

Participants were predominantly white and over the age of 50 (Table 1). Cases were more likely to be former smokers and current users of postmenopausal hormones. Higher proportions of controls were never smokers and breast-fed for at least one year. As expected, more women were postmenopausal at time of diagnosis (for cases) or censoring (for controls) than in 1988. There were also more physically active women and fewer current smokers at time of diagnosis/censoring. Cases were more mobile than controls, with 31.4% (145/461) moving between 1988 and year of diagnosis compared to 26.6% (2,455/9,220) between 1988 and year of censoring for controls. Covariate data were missing for less than 10% of participants with the exception of information on postmenopausal hormone use (17%) and physical activity (23%).

Figure 1 shows the distribution of rheumatoid arthritis cases and controls in the U.S using address at diagnosis/censoring. To preserve confidentiality, the figure was created by randomly

placing residences within a small grid that includes the actual location. Actual locations were used in the analysis.

When geographic variation in RA risk was examined using addresses at diagnosis/censoring, the crude and adjusted analyses (Figures 2a and 2b) predicted similar results. Because of low data density and thus unreliability of the estimates, we did not predict odds of RA for regions shown in white. The association between location and RA was statistically significant for both analyses (crude, global p-values=0.02; adjusted, global p-value=0.034), indicating that odds ratios of RA varied with geographic location. Contour lines denote areas where RA risk relative to the whole study area was significantly increased (red) and decreased (blue) at the 0.05 level. A statistically significant area of increased risk was identified in the upper northeast including Vermont, New Hampshire, and southern Maine. A significant area of decreased risk was located in Pennsylvania. The optimal spans for the crude and adjusted analyses were 0.55 and 0.5 respectively. Crude odds ratios (COR) ranged from 0.76-2.26, only slightly larger than adjusted odds ratios (AOR), which ranged from 0.68-2.17.

Figure 3a shows the results of the adjusted analysis using 1988 residences with the optimal span of 0.55. Again, there were similar spatial patterns of predicted risk between the adjusted and crude analysis (not shown). Crude and adjusted maps predicted comparable ranges in odds ratios relative to the whole study area (COR=0.61-2.39, AOR=0.63-2.37) and both were statistically significant (global p-values of 0.029 and 0.034 respectively). We performed point-wise tests of significance and identified areas of higher risk in the northern areas in the mid-west and northeast denoted by red contour lines. The AIC curve for the adjusted RA model indicated a local minima at span sizes of 0.20 before reaching the global minimum (and optimal span) of 0.55. We repeated the adjusted analysis using a span of 0.20 (Figure 3b). The small span of 0.20 produced a bumpier surface, including an area of highly increased risk along the Ohio River near

West Virginia and northern Kentucky. The model also predicted even higher odds ratios in the northern latitudes, the west, midwest (Great Plains), and northeast. We did not test for statistical significance of location in this model because the optimal span size was not used.

Discussion

Results of the spatial analysis are consistent with an earlier regional study (Costenbader et al. 2008) that found increased risk of RA for those women who lived in the midwest and northeast United States, compared to west of the Rocky Mountain range, and the association was stronger with residency at age 15 and 30 than at baseline in 1976. Costenbader et al. also observed elevated risk in the mid-Atlantic compared to west of the Rocky Mountain range, which the current spatial analysis did not observe. Although both studies used the NHS dataset, possible reasons for the difference in results include study population (the earlier study included women diagnosed with RA beginning in 1976 compared to 1988 in the current study), reference group (west of the Rocky Mountain range compared to the entire study area), and geographic scale (regional versus individual level analyses). The time periods of the addresses were different as well. We examined risk of RA using addresses in 1988 (when the mean age for the current study population was 54) and addresses at diagnosis/censoring, two time points that were not considered in earlier regional study. Although the Nurses' Health Study began in 1976, geocoded addresses are only available beginning in 1988, limiting our ability to perform extensive space-time analyses (Vieira et al. 2008).

While spatial patterns were similar for addresses in 1988 and at time of diagnosis/censoring (Figures 2b and 3a), slightly higher odds ratios were observed for the 1988 analysis suggesting that long term exposure may be more important than recent exposure. We observed even higher odds ratios when we restricted the 1988 analysis to women who were diagnosed/censored at least

8 years later (1996 or later - data not shown). Although this restricted analysis was limited by small case numbers (n=227), it supports our hypothesis that earlier rather than recent exposure may be more important. Regardless of timing, a statistically significant area in the upper northeast including Vermont, New Hampshire, and southern Maine was identified as having consistently elevated RA risk relative to the whole study area, and an additional analysis (Figure 3b) predicted increased odds ratios for the more northern latitudes of the U.S. A geographic association with northern latitudes has also been observed for multiple sclerosis and Crohn's disease, other autoimmune diseases that may be mediated by reduced vitamin D from decreased solar exposure and the immune effects of vitamin D deficiency (Hernan 1999; Munger 2006; Armitage et al. 2004; McLeod et al. 1994;Patel et al. 2007; Arnson et al. 2007; Ponsonby et al. 2005; Sioka et al. 2009). The studies of vitamin D dietary intake and incident RA come to contradictory conclusions. Merlino et al. (2004) found a strong protective effect of high vitamin D intake in diminishing incident RA, while a study by Costenbader et al (2008) revealed no association between intake and incident RA. However neither study assessed vitamin D from solar exposure.

Geographic variation may also be due to other environmental exposures or residual spatial confounding. Spatial confounding occurs when risk factors for a disease are not evenly distributed. For example, a cluster of lung cancer may be due to an increased density of smokers. Crude and adjusted analyses produced similar geographic patterns of RA risk, and missing covariate data was not a concern in our analyses. Although we adjusted for individual level socioeconomic status, some authors argue for the inclusion of group-level contextual variables (e.g., Krieger et al. 2002). By linking residential location to census data, one could test the importance of these variables relative to individual-level covariates. We are currently working on methods involving generalized additive mixed models to incorporate a smooth of

location into a multi-level model adjusted for individual- and community-level risk factors. Our findings also may be due to geographic differences in the location of rheumatology specialists or in diagnosing practices.

These spatial analyses have some potential limitations. GAMs may exhibit biased behavior at the edges of the data, although our work with synthetic data suggested little to no bias when a loess smooth is used (Webster et al. 2005). To reduce the likelihood of bias from edge effects, we did not predict odds ratios in regions of low data density, which restricted the extent of northern latitudes available for our analysis. We used the Akaike Information Criterion to choose an "optimal" span, but when we used a smaller span of 0.20 in our analyses, we were able to discern greater spatial variation that may be of importance. While there is some benefit to having a non-adhoc method for span selection, analyses should not be limited to just one span. In the current analyses, we identified areas with significantly increased or decreased risk using pointwise hypothesis tests only if global tests were statistically significant, but performing multiple testing at each location may result in an increase in the type I error rate. Also, many epidemiologists prefer confidence intervals when evaluating the precision of point estimates in addition to p-values (Rothman 1998). It should be possible to compute variance bands (also known as confidence bands) for our maps, but displaying three surfaces of odds ratios makes it difficult to visually interpret points where the bands do not include one (Hastie and Tibshirani 1990).

Prospective cohort studies are one of the standard epidemiologic tools for investigating associations between disease and exposure. By combining such data with advanced statistical techniques, we were able to address many criticisms of spatial studies. Self-reported cases were confirmed by examining medical records, and controls selected from among the non-cases provided an estimate of the underlying, non-uniform population from which the cases arose.

Because the data are from a prospective cohort, selection bias is not a concern in this study. However, results for this spatial analysis of a female nurse cohort may not be generalizable to other populations. Point-based data were used, avoiding aggregation within administrative boundaries. We were able to control for a large number of covariates, which can only be done to a limited degree using other cluster analysis tools like the scan statistic (Kulldorff 1998). Residential history information allowed us to take calendar year into account, potentially quite important for diseases with environmental risk factors. Although spatial analyses are useful for generating new hypotheses or supporting existing hypotheses, areas of increased and decreased odds ratios should be considered exploratory. Further analysis that examines the relationship between vitamin D exposure and RA is warranted to explore these results.

Conclusions

Using generalized additive models and GIS, we generated maps of rheumatoid arthritis (RA) risk relative to the whole study area. When available, prospective cohort studies provide extensive data on potential risk factors and residential histories that address many methodological criticisms of cluster studies. We identified a significant area of increased odds ratios in the northeast, and additional analyses suggest that women living in more northern latitudes may be at greater risk for RA relative to the whole study area. Similar geographic associations have been observed with other chronic autoimmune diseases including multiple sclerosis. The results of the current analysis illustrate the application of GAMs and GIS to visualize geographic variation in RA risk, adjust for known confounders, and test for the statistical significance of location. Our method is particularly useful in generating hypotheses for further investigation and supporting existing hypotheses, especially when residential histories are available.

References

Alamanos Y, Voulgari PV, Drosos AA. 2006. Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review. Semin Arthritis Rheum 36(3):182-8.

Anaya JM, Correa PA, Mantilla RD, Jimenez F, Kuffner T, McNicholl JM. 2001. Rheumatoid arthritis in African Colombians from Quibdo. Semin Arthritis Rheum 31(3):191-8.

Armitage EL, Aldhous MC, Anderson N, Drummond HE, Riemersma RA, Ghosh S, et al. 2004. Incidence of juvenile-onset Crohn's disease in Scotland: association with northern latitude and affluence. Gastroenterology 127:1051-1057.

Arnson Y, Amital H, Shoenfeld Y. 2007. Vitamin D and autoimmunity: new aetiological and therapeutic considerations. Ann Rheum Dis 66(9):1137-42.

Costenbader KH, Chang SC, Laden F, Puett R, Karlson EW. 2008. Geographic variation in rheumatoid arthritis incidence among women in the United States. Arch Intern Med 168(15):1664–1670.

Costenbader KH, Feskanich D, Holmes M, Karlson EW, Benito-Garcia E. 2008. Vitamin D intake and risks of systemic lupus erythematosus and rheumatoid arthritis in women. Ann Rheum Dis 67(4):530-5.

Costenbader KH, Feskanich D, Mandl LA, Karlson EW. 2006. Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. Am J Med 119(6):503 e501–509.

Gatenby PA, Lucas RM, Engelsen O, Ponsonby AL, Clements M. 2009. Antineutrophi cytoplasmic antibody-associated vasculitides: Could geographic patterns be explained by ambient ultraviolet radiation? Arthritis Rheum 61(10):1417-1424.

Hart JE, Laden F, Puett RC, Costenbader KH, Karlson EW. 2009. Exposure to traffic pollution and increased risk of rheumatoid arthritis. Environ Health Perspect 117(7):1065–1069.

Hastie T, Tibshirani R. 1990. Generalized Additive Models. London: Chapman and Hall.

Hengstman GJ, van Venrooij WJ, Vencovsky J, Moutsopoulos HM, van Engelen BG. 2000. The relative prevalence of dermatomyositis and polymyositis in Europe exhibits a latitudinal gradient. Ann Rheum Dis 59(2):141-2.

Hernan MA, Olek MJ, Ascherio A. 1999. Geographic variation of MS incidence in two prospective studies of US women. Neurology 53(8):1711-8.

Kamen DL, Cooper GS, Bouali H, Shaftman SR, Hollis BW, Gilkeson GS. 2006. Vitamin D deficiency in systemic lupus erythematosus. Autoimmun Rev 5(2):114-7.

Karlson EW, Mandl LA, Hankinson SE, Grodstein F. 2004. Do breast-feeding and other reproductive factors influence future risk of rheumatoid arthritis? Results from the Nurses' Health Study. Arthritis Rheum 50(11):3458-3467.

Karlson EW, Sanchez-Guerrero J, Wright EA, Lew RA, Daltroy LH, Katz JN, et al. 1995. A connective tissue disease screening questionnaire for population studies. Ann Epidemiol 5(4):297-302.

Kelsall J, Diggle P. 1998. Spatial variation in risk of disease: a nonparametric binary regression approach. J Roy Stat Soc C-App 47:559-573.

Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. 2002. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level matter? Am J Epidemiol 156:471-482.

Kulldorff M. 1997. A spatial scan statistic. Commun Statist Theory and Methods 26: 1481-1496.

Liao KP, Alfredsson L, Karlson EW. 2009. Environmental influences on risk for rheumatoid arthritis [review]. Curr Opin Rheumatol 21:279-83. PMC Journal –In Process.

McLeod JG, Hammond SR, Hallpike JF. 1994. Epidemiology of multiple sclerosis in Australia: with NSW and SA survey results. Med J Aust 160:117-122.

Merlino LA, Curtis J, Mikuls TR, Cerhan JR, Criswell LA, Saag KG. 2004. Vitamin D intake is inversely associated with rheumatoid arthritis: results from the Iowa Women's Health Study. Arthritis Rheum 50(1):72-7.

Munger KL, Levin LI, Hollis BW, Howard NS, Ascherio A. 2006. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. Jama 296(23):2832-8.

Oh K, Hu FB, Manson JE, Stampfer MJ, Willett WC. 2005. Dietary fat intake and risk of coronary heart disease in women: 20 years of follow-up of the nurses' health study. Am J Epidemiol 161(7):672-9.

Patel S, Farragher T, Berry J, Bunn D, Silman A, Symmons D. 2007. Association between serum vitamin D metabolite levels and disease activity in patients with early inflammatory polyarthritis. Arthritis Rheum 56(7):2143-9.

Ponsonby AL, Lucas RM, van der Mei IA. 2005. UVR, vitamin D and three autoimmune diseases—multiple schlerosis, type 1 diabetes, rheumatoid arthritis. Photochem Photobiol 81(6):1267:75.

Ramos-Remus C, Sierra-Jimenez G, Skeith K, Aceves-Avila FJ, Russell AS, Offer R, et al. 2007. Latitude gradient influences the age of onset in rheumatoid arthritis patients. Clin Rheumatol 26:1725-28.

Richardson DB. 2004. An incidence density sampling program for nested case-control analyses. Occup Environ Med 61(12):e59.

Rothman KJ, Greenland S. 1998. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott-Raven.

Sioka C, Kyritsis AP, Fotopoulos A. 2009. Multiple sclerosis, osteoporosis, and vitamin D. J Neurol Sci 287(1-2):1-6..

Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC. 2000. Primary prevention of coronary heart disease in women through diet and lifestyle. N Eng J Med 343(1):16-22.

Somers EC, Thomas SL, Smeeth L, Schoonen WM, Hall AJ.2007. Incidence of systemic lupus erythematosus in the United Kingdom, 1990-1999. Arthritis Rheum 57(4):612-8.

Uhlig T, Hagen KB, Kvien TK. 1999. Current tobacco smoking, formal education, and the risk of rheumatoid arthritis. J Rheumatol 26(1):47–54.

Vieira V, Webster T, Weinberg J, Aschengrau A. 2008. Spatial-temporal analysis of breast cancer in upper Cape Cod, Massachusetts. International Journal of Health Geographics 7(1):46.

Walsh SJ, Gilchrist A. 2005. Geographical clustering of mortality from systemic lupus erythematosus in the United States: contributions of poverty, Hispanic ethnicity and solar radiation. Lupus 15(10):662-70.

Webster T, Vieira V, Weinberg J, Aschengrau A. 2006. A method for mapping population-based case control studies using generalized additive models. Int J Health Geogr 5(1):26.

Table 1. Selected characteristics of cases and controls in 1988 and at time of diagnosis/censoring (Dx)

	Cases (n=461)		Controls (n=9,220)	
Address No. (%)	1988	Dx	1988	Dx
Age [years (mean)]	54.8	61.4	54.4	61.0
BMI	23.5	26.0	22.5	24.5
Age (years) at menarche (mean)	12.4	12.4	12.4	12.4
Pack-years of smoking (mean)[a]	26.3	27.8	23.0	24.6
Caucasian race (%)	94.1	94.1	93.8	93.8
Smoking status (%)				
Current	18.4	13.2	19.1	13.8
Former	40.8	48.6	34.2	40.0
Never	37.7	37.7	43.8	43.6
Parity/lactation (%)				
Nulliparous	6.7	6.7	7.0	7.0
Parous, never breast-fed	34.1	34.1	29.5	29.5
Parous, breast-fed 1–11 months	36.9	36.9	35.0	35.0
Parous, breast-fed ≥ 12 months	12.8	12.8	15.8	15.8
Menopausal status (%)				
Premenopausal	20.8	6.3	25.8	10.7
Postmenopausal	75.3	93.1	68.7	85.9
Unknown status	3.9	0.6	5.5	3.4
Postmenopausal hormone use (%)[b]				
Never used	50.7	33.0	52.4	33.7
Past use	13.4	19.7	11.7	17.9
Current use	26.2	40.6	19.3	31.8
Oral contraceptive use (%)				
Never used	48.6	48.6	50.2	50.2
Ever use	48.6	48.6	45.0	45.0
Physical activity (metabolic equivalent hours/week, %)				
< 3	17.6	20.8	17.0	17.2
3 to < 9	23.6	20.0	19.7	18.1
9 to < 18	16.9	20.2	15.6	16.0
18 to < 27	10.2	10.8	8.7	9.9
≥ 27	15.2	19.3	12.9	15.9
Father's occupation (%)				
Professional/manager	23.6	23.6	25.4	25.4
Other job	76.4	76.4	74.6	74.6
Mother's occupation (%)				
Housewife	67.5	67.5	64.2	64.2
Other job	32.5	32.5	35.8	35.8
Education (%)				
Nurse	84.2	84.2	72.4	72.4
Other	15.8	15.8	27.6	27.6
Marital status (%)				

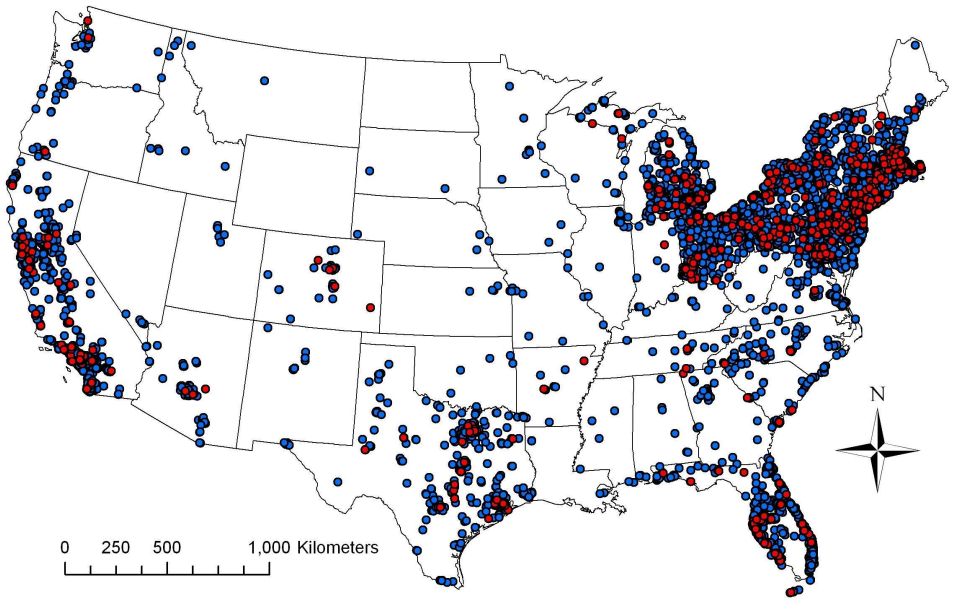
Married	71.6	71.6	63.2	63.2
Other	28.4	28.4	36.8	36.8
Husband's education (%)				
Missing or not applicable	22.8	22.8	34.6	34.6
< High school	6.1	6.1	3.9	3.9
High school	31.7	31.7	25.7	25.7
> High school	39.5	39.5	35.7	35.7

[a]Among ever-smokers.
[b]Among postmenopausal women.

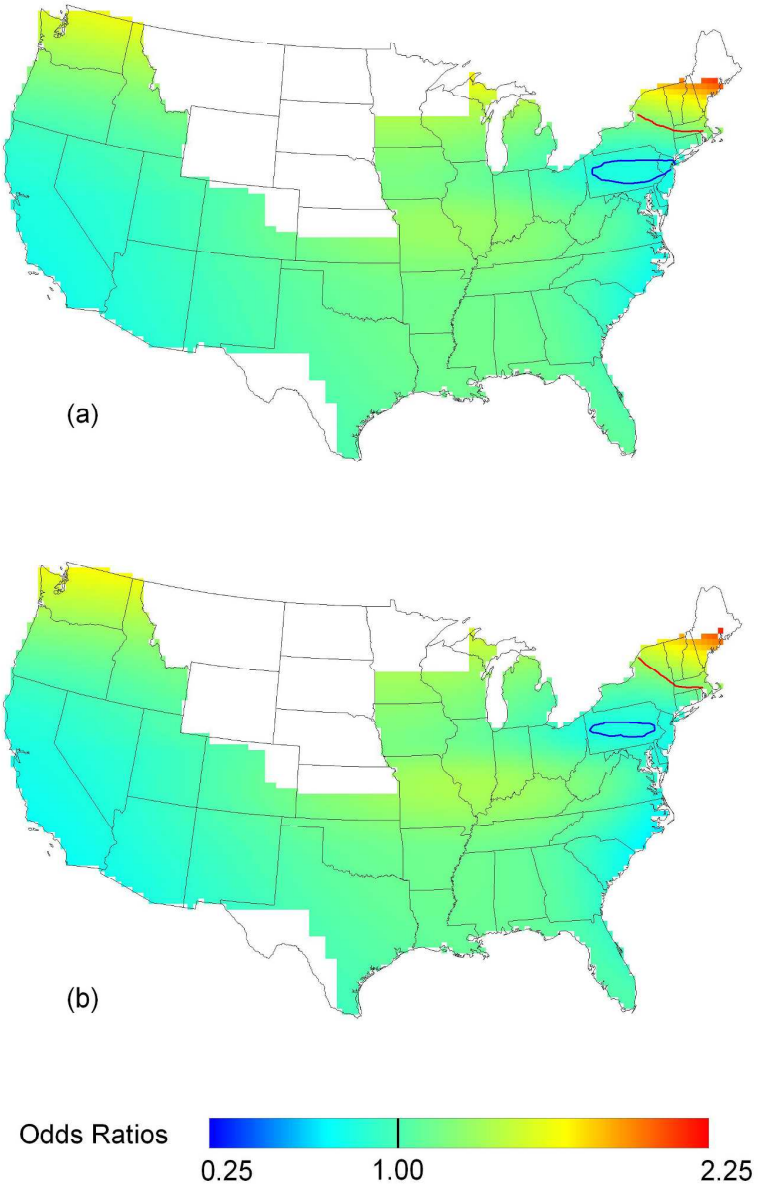
Figure 1. Distribution of cases and controls for rheumatoid arthritis. Each point represents the residences for cases (red) at diagnosis and controls (blue) at time of censoring. Locations have been geographically altered to preserve confidentiality.

Figure 2. Results for Addresses at Time of Diagnosis/Censoring. Odds ratios are relative to the whole study area. a) Crude, optimal span of 0.55 (global p-value=0.02). b) Adjusted, optimal span of 0.50 (global p-value=0.034). Contour lines denote areas of significantly increased (red) and decreased (blue) risk at the 0.05 level. Geographic patterns are similar for crude and adjusted analyses.

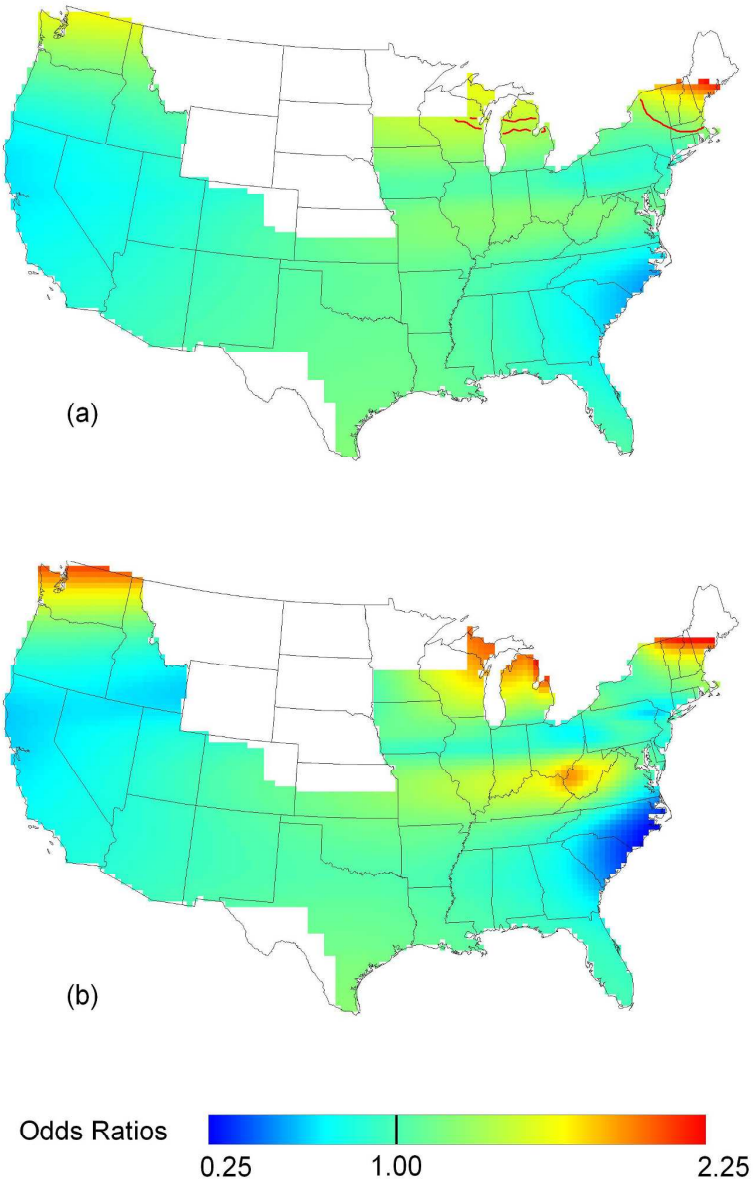
Figure 3. Results for Addresses in 1988. Odds ratios are relative to the whole study area. a) Adjusted, optimal span of 0.55 (global p-value=0.034). Contour lines denote areas of significantly increased (red) and decreased (blue) risk at the 0.05 level. b) Adjusted, span of 0.20. Small span size results in more spatial variation in risk.



202x129mm (300 x 300 DPI)



663x1038mm (150 x 150 DPI)



657x1042mm (150 x 150 DPI)