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,5,6,7 Francine Laden,2,3,8 Karen H. Costenbader,9 and Elizabeth W. Karlson9

1Department of Environmental Health, Boston University School of Public Health, Boston, Massachusetts, USA; 2Channing Laboratory, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts, USA; 3Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA; 4Department of Bio-statistics, Boston University School of Public Health, Boston, Massachusetts, USA; 5South Carolina Cancer Prevention and Control Program, University of South Carolina, Columbia, South Carolina, USA; 6Department of Environmental Health Sciences, and 7Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, USA; 8Exposure, Epidemiology, and Risk Program, Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, USA; 9Division of Rheumatology, Immunology, and Allergy, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts, USA

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 further explore 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; participants were followed from 1988 to 2002. We examined spatial variation using addresses at baseline in 1988 and at the time of case diagnosis or the censoring of controls. Generalized additive models (GAMs) were used to predict a continuous risk surface by smoothing on longitude and latitude while adjusting for known risk factors. Permutation tests were conducted to evaluate the overall importance of location and to identify, within the entire study area, those locations of statistically significant risk.

RESULTS: A statistically significant area of increased RA risk was identified in the northeast United States (p-value = 0.034). Risk was generally higher at northern latitudes, and it increased slightly when we used the nurses’ 1988 locations compared with those at the time of diagnosis or 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 that occur earlier in residential histories. These results illustrate the usefulness of GAM methods in generating hypotheses for future investigation and supporting existing hypotheses.

KEY WORDS: disease mapping, generalized additive models, geographic information systems (GIS), prospective cohort study, rheumatoid arthritis. Environ Health Perspect 118:957–961 (2010). doi:10.1289/ehp.0901861 [Online 25 March 2010]
cycles were geocoded, which contributed 173,624 addresses and 762,511 questionnaire records. Although nurses’ addresses were concentrated in the 11 original study states at baseline in 1976, by the follow-up period in 2002, the women were located in all 50 states.

The medical records of nurses were reviewed to identify those who self-reported a diagnosis of RA. In this study, we used the diagnostic criteria of the American College of Rheumatology for RA (Karlson et al. 1995, 2004). For 1988 through 2002, we identified a total of 461 women with confirmed incident RA. When the participants reported having RA, their information was censored, 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 noncases 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 and to keep the numbers reasonable for computation. Women who were noncases 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. Thus, the proportion of participants censored in each year was the same for cases and controls.

Table 1. Selected characteristics of cases and controls in 1988 and at the time of diagnosis or censoring.

| Characteristics                        | Cases (n = 461) | Controls (n = 9,220) |
|----------------------------------------|----------------|---------------------|
| Mean age (years)                        | 54.8 ± 6.1     | 54.4 ± 6.0          |
| BMI (kg/m²)                            | 23.5 ± 2.0     | 22.5 ± 2.4          |
| Mean age at menarche (years)           | 12.4 ± 2.4     | 12.4 ± 2.4          |
| Pack-years of smoking (mean)           | 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 (%)         | 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 used: 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        |

Dx, diagnosis or censoring.

*Age was originally modeled in months but converted to years for this paper. *Among ever-smokers. *Among postmenopausal women.
Figure 1 shows the distribution of RA cases and controls in the United States using address at diagnosis or censoring. To preserve confidentiality, the figure was created by randomly placing residences within a small grid that included the actual location. Actual locations were used in the analysis.

When geographic variation in RA risk was examined using addresses at diagnosis or censoring, the crude and adjusted analyses (Figure 2A and 2B, respectively) predicted similar results. Because of low data density and thus unreliable 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 ORs 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 ORs (CORs) ranged from 0.76 to 2.26, only slightly larger than adjusted ORs (AORs), which ranged from 0.68 to 2.17.

Figure 3A shows the results of the adjusted analysis using 1988 residences with the optimal span of 0.55. Again, similar spatial patterns of predicted risk were found between the adjusted and crude analysis (not shown). Crude and adjusted maps predicted comparable ranges in ORs 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 pointwise tests of significance and identified areas of higher risk in the northern areas in the midwest 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 surface with more spatial variation in risk, including an area of high ORs along the Ohio River near West Virginia and northern Kentucky. The model also predicted even higher ORs in the northern latitudes, the west, the midwest (Great Plains), and the 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 conducted by Costenbader et al. (2008a) that found increased risk of RA for those women who lived in the midwest and northeast United States, compared with west of the Rocky Mountain range, and the association was stronger with residency at age 15 and 30 years than at baseline in 1976. They also observed elevated risk in the mid-Atlantic region compared with the area west of the Rocky Mountain range, which the current spatial analysis did not observe. Although both studies used the NHS data set, possible reasons for the difference in results include study population (the earlier study included women diagnosed with RA beginning in 1976 compared with 1988 in the current study), reference group (west of the Rocky Mountain range compared with the entire study area),...
Vieira et al.

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 from 1988 (when the mean age for the current study population was 54 years old) and those at diagnosis or censoring. These two time points were not considered in the earlier regional study (Costenbader et al. 2008a). Although the NHS began in 1976, addresses were only geocoded beginning in 1988, which limited our ability to perform extensive space-time analyses (Vieira et al. 2008).

Spatial patterns were similar for addresses in 1988 and at the time of diagnosis or censoring (Figures 2B and 3A), although slightly higher ORs were observed for the 1988 analysis. This finding suggests that long-term exposure may be more important than recent exposure. We observed even higher ORs when we restricted the 1988 analysis to women who were diagnosed or 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 that included 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 ORs for the more northern latitudes of the United States. A geographic association with northern latitudes has also been observed for multiple sclerosis and Crohn’s disease. These autoimmune diseases may be mediated by a reduction in vitamin D through decreased solar exposure and the immune effects of vitamin D deficiency (Armitage et al. 2004; Aronson et al. 2007; Hernán et al. 1999; Kamen et al. 2006; McLeod et al. 1994; Munger et al. 2006; Patel et al. 2007; Ponsonby et al. 2005; Sioka et al. 2009). The studies of dietary intake of vitamin D and incident RA have come to contradictory conclusions. Merlino et al. (2004) found a strong protective effect of high vitamin D intake in diminishing incident RA, whereas a study by Costenbader et al. (2008b) 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 were 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 multilevel 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. 2006). To reduce the likelihood of bias from edge effects, we did not predict ORs in regions of low data density, which restricted the extent of northern latitudes available for our analysis. We used the AIC 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. Although there is some benefit to having a non-adjacent 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. In addition, many epidemiologists prefer confidence intervals when evaluating the precision of point estimates in addition to p-values (Rothman and Greenland 1998). It should be possible to compute variance bands (also known as confidence bands) for our maps, but displaying three surfaces of ORs 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 make it difficult 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, Voulgaris 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):182–188.

Anaya JM, Correa PA, Martilla RD, Jimenez F, Kuffner T, McNicholl JM. 2001. Rheumatoid arthritis in African Colombians from Quibdo. Semin Arthritis Rheum 31(1):191–198.

Armitage EL, Alhoods 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 altitude. Gastroenterology 127(4):1051–1057.

Arnson Y, Amital H, Shoenfeld Y. 2007. Vitamin D and autoimmunity: new aetiological and therapeutic considerations. Ann Rheum Dis 66(9):1317–1142.

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

Costenbader KH, Keskinar D, Holmes M, Karlson EW, Benito-Garcia E. 2008b. Vitamin D intake and risks of systemic lupus erythematosus and rheumatoid arthritis in women. Ann Rheum Dis 67(4):350–355.

Costenbader KH, Keskinar 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.e1–e9.

Gatenby PA, Lucas RM, Engelsen O, Ponsonby AL, Clements M. 2004. Incidence of juvenile 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 Stat 47(1):71–93.

Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson RT. 2002. Geocoding and monitoring of US socio-economic inequalities in mortality and cancer incidence: comparison of census and ACS geographic level matter? Am J Epidemiol 156(5):471–482.

Kullendorf M. 1997. A spatial scan statistic. Commun Stat Theory Methods 26(8):1481–1498.

Liao KP, Altellodson L, Karlson EW. 2009. Environmental influences on risk for rheumatoid arthritis. Curr Opin Rheumatol 21:279–283.

McLeod JG, Hammond SR, Hallpike JF. 1994. Epidemiology of multiple sclerosis in Australia. With NSW and SA survey results. Med J Aust 160(8):319–325.

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):72–77.

Mungel KR, Levin LI, Hollis BW, Howard NS, Ascherio A. 2006. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. JAMA 296(23):2832–2838.

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):872–679.

Patel S, Farragher T, Berry J, Bush NR, Silman A, Symmons D. 2007. Association between serum vitamin D metabolites and disease activity in patients with early inflammatory polyarthritis. Arthritis Rheum 56(7):2142–2149.

Ponsonby AL, Lucas RM, van der Mei IA. 2005. UV, vitamin D and three autoimmune diseases—multiple sclerosis, type 1 diabetes, rheumatoid arthritis. Photochem Photobiol 81(6):1265–1275.

Ramos-Reurca E, Sierra-Jimenez G, Skeith K, Acves-Aviia FJ, Russell AS, Offer R, et al. 2007. Latitude gradient influences the age of onset in rheumatoid arthritis patients. Clin Rheumatol 26(12):1725–1728.

Richardson DB. 2004. An incidence density sampling program for nested case-control analyses. Occup Environ Med 61(12):e9; doi:10.1136/okee.2004.014472 [Online 24 June 2004].

Rieman KJ, Greenland S. 1998. Modern Epidemiology. 2nd ed. Philadelphia: Lipcott-Raven.

Sicska C, Kyntys AF, Fotopoulos A. 2009. Multiple sclerosis, osteoporosis, and vitamin D. J Neurol Sci 281(1–2):21–6.

Somers EC, Thomas SL, Smeth L, Schoonen WM, Hall AJ. 2007. Incidence of systemic lupus erythematosus in the United Kingdom, 1990–1995. Arthritis Rheum 57(6):1204–1218.

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

Uhlig T, Hagen KB, Kien TK. 1999. Current tobacco smoking, formal education, and the risk of rheumatoid arthritis. Rheumatology 38(1):47–54.

Vieira V, Weber T, Weinberg J, Aschengrau A. 2008. Spatial-temporal analysis of breast cancer in upper Cape Cod, Massachusetts. Int J Health Geogr 7(1):46; doi:10.1186/1476-072X-7-46 [Online 13 August 2008].

Wakely SJ, Girgis A. 2006. Geographical clustering of mortality from systemic lupus erythematosus in the United States: contributions of poverty, Hispanic ethnicity and solar radiation. Lupus 15(10):962–970.

Webster T, Vieira V, Weinberg J, Aschengrau A. 2006. Method for mapping population-based case-control studies: an application using generalized additive models. Int J Health Geogr 5(1):26; doi:10.1186/1476-072X-5-26 [Online 9 June 2006].