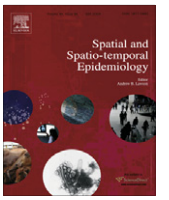
Spatial and Spatio-temporal Epidemiology 3 (2012) 163–171

Contents lists available at SciVerse ScienceDirect

Spatial and Spatio-temporal Epidemiology

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



Spatial–temporal analysis of non-Hodgkin lymphoma risk using multiple residential locations

David C. Wheeler [a,*], Lance A. Waller [b], Wendy Cozen [c], Mary H. Ward [d]

[a] Department of Biostatistics, School of Medicine, Virginia Commonwealth University, One Capitol Square, 7th Floor, Room 733, 830 East Main Street Richmond, VA 23298, USA
[b] Department of Biostatistics, Rollins School of Public Health, Emory University, 1518 Clifton Rd., N.E., 3rd Floor, Atlanta, GA 30322, USA
[c] Department of Preventive Medicine and Pathology, and Norris Comprehensive Cancer Center, USC Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA
[d] Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health (NIH), Department of Health and Human Services (DHHS), Executive Plaza South, Room 8006, Bethesda, MD 20892, USA

ARTICLE INFO

Article history:
Available online 21 April 2012

Keywords:
Cancer
Generalized additive model
Spatial risk
Latency
Exposure

ABSTRACT

Exploring spatial–temporal patterns of disease incidence and mortality can identify areas of significantly elevated or decreased risk, providing potential etiologic clues. Several methodological issues arise in spatial–temporal analysis of cancer, including population mobility, disease latency, and confounding, but applying modern statistical methods to case-control studies with residential histories can address these issues. As an example, we present a spatial–temporal analysis of non-Hodgkin lymphoma (NHL) risk using data from Los Angeles County, one of four centers in a population-based case-control study. Using residential histories, we fitted generalized additive models (GAMs) adjusted for known risk factors to model spatially the probability that an individual had NHL and identify areas of significantly elevated NHL risk. In previous analyses using models with single lag times, the lag time of 20 years yielded the most significant decrease in model deviance. To better assess cumulative effects of unmeasured environmental exposures over space and time, we considered models that allowed for multiple residences per subject through spatial smoothing functions of residential location at different times. We found that the model with the best goodness-of-fit included components for residential change and residential duration, although the model that included residential duration was not meaningfully better than the model that included only residential change. The estimated cumulative spatial risk surface from the model with residential change amplified the risk surface in some areas compared with the surface based on the model with a single component for the most significant time lag.

1. Introduction

Many cancers have risk factors that are distributed unevenly in the environment. Examples include bladder cancer and arsenic (Silverman et al., 2006), lung cancer and radon (Spitz et al., 2006), and leukemia and benzene (Linet et al., 2006). It is reasonable, therefore, to expect spatial pattern in cancer risk, which may be explained by the uneven distribution of risk factors that may be known or unknown. When risk factors are unknown, studying spatial–temporal patterns in risk may reveal clues about disease etiology. There are numerous examples in the literature of spatial analyses to study patterns in cancers, including childhood leukemia (Alexander, 1993; Bithell and Vincent, 2000; Wheeler, 2007) and bladder cancer (Jacquez et al., 2006).

In spatial analyses of cancer risk, the residence at diagnosis is typically used as a surrogate for unknown environmental exposures, defined broadly to include

* Corresponding author. Tel.: +1 804 828 9827; fax: +1 804 828 8900.
E-mail address: dcwheels@gmail.com (D.C. Wheeler).

1877-5845/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.sste.2012.04.009

lifestyle factors as well as pollutants. However, due to the long latency, or lag time, between exposure to a relevant risk factor and diagnosis of cancer, and due to residential mobility, it is reasonable to believe that residential locations many years before cancer diagnosis are potentially more relevant for cancer risk. Researchers in geography (Bentham, 1988; Han et al., 2004; Sabel et al., 2009) and public health (Jacquez et al., 2005; Paulu et al., 2002; Vieira et al., 2005) have recognized the importance of population mobility when studying disease patterns. Ignoring migration when studying health outcomes with long latencies can lead to exposure misclassification, biased risk estimates, and diminished study power (Tong, 2000).

When analyzing cancer risk in epidemiologic studies which include data on residential histories, one may use historic residential locations for individuals to study spatial risk over time. Residential histories can be informative for both the location and the timing of potential environmental exposures associated with residential locations, which is especially useful when the average latency for a cancer with suspected environmental causes is unknown. In such an analysis, one can adjust for known or suspected risk factors that are typically collected in epidemiologic studies and then examine the unexplained risk for spatial–temporal patterns (Kelsall and Diggle, 1998). Researchers have conducted this type of research using generalized additive models (GAMs) and residential histories for several cancers (Vieira et al., 2005; Webster et al., 2006; Wheeler et al., 2011). Researchers have either used all the residential locations available for each subject to estimate one risk surface without modeling lag times (Vieira et al., 2005; Webster et al., 2006) or have estimated a spatial risk surface for one time period or lag time in a model for several time periods of interest, yielding a different risk surface for each model (Vieira et al., 2008; Wheeler et al., 2011). Using residential histories, it should be possible to model spatial risk surfaces at several different times together in one model. Such an approach can be thought of as providing an estimate of unmeasured life-course environmental exposures, which is in concordance with the increasing popular vision in epidemiology of the "exposome" that seeks to characterize the totality of environmental exposures for disease risk (Rappaport and Smith, 2010; Wild, 2005). The rationale is that relevant cumulative environmental exposures could occur over multiple residential locations across time and models should attempt to encompass such life-course environmental exposures.

The research presented in this paper extends earlier work (Wheeler et al., 2011) to include multiple residential locations per subject in one model of the spatial variation of non-Hodgkin lymphoma risk in a population-based case-control study with residential histories. The objective of this research was to consider different approaches to modeling spatial variation in cancer risk using multiple residential locations per subject and assess the impact of allowing for a more cumulative measure of spatial risk. We used an analysis approach based on generalized additive models with several spatial smoothing functions to model residual spatial variation in cancer risk at multiple exposure times jointly after adjusting for known risk factors. We analyzed the cumulative spatial risk of NHL in one of the four study centers, Los Angeles, of the case-control study.

2. Methods

2.1. Study population

The National Cancer Institute (NCI)-Surveillance, Epidemiology and End Results (SEER) NHL study is a case-control study of 1321 cases aged 20 to 74 years that were diagnosed between July 1, 1998 and June 30, 2000 in four SEER cancer registries, including Detroit, Iowa, Seattle, and Los Angeles County. The study has been described previously (Chatterjee et al., 2004; Morton et al., 2008; Wheeler et al., 2011). Briefly, population controls (1057) were selected from residents of the SEER areas using random digit dialing (<65 years of age) or Medicare eligibility files (65 and over) and were frequency matched to cases by age (within 5-year groups), sex, race, and SEER area. Among eligible subjects contacted for an interview, 76% of cases and 52% of controls participated in the study. Cases and controls with a history of NHL or known HIV infection were not included in the study. The goal of the NCI-SEER NHL study was to investigate potential environmental and genetic risk factors for NHL.

Computer-assisted personal interviews were conducted during a visit to each subject's home to obtain lifetime residential and occupational histories, medical history, and other information including date of birth, gender, race, education, and pest treatments including home treatment for termites before 1988 (a surrogate for the insecticide chlordane). Written informed consent was obtained during the home visit and human subjects review boards approved the study at the NCI and at all participating institutions. Historic addresses were collected in a residential history section of an interviewer-administered questionnaire. Participants were mailed a residential calendar in advance of the interview and were requested to provide the complete address of every home in which they lived from birth to the current year, listing the years they moved in and out (De Roos et al., 2010). Interviewers reviewed the residential calendar with respondents and probed to obtain missing address information. Residential addresses were matched to geographic address databases to yield geographic coordinates that were used in this analysis.

Our previous analysis of spatial risk in NHL in the four study centers found that the lag time with the most significant association with risk overall was 20 years before diagnosis. Among the four study areas, this lag period was most statistically significant for the Los Angeles study center (Wheeler et al., 2011). Therefore, we have focused our analysis on this lag time in the Los Angeles study area (Fig. 1). In previous work, we limited the analysis to subjects residing in one of the study areas for at least 20 years prior to study enrollment to maximize the power to detect local variation in spatial risk within each center at different lag times. Retaining this selection criterion for consistency, we analyzed 190 cases and 161 controls in the Los Angeles study area.