한국환경보건학회지, 제37권 제1호(2011)
J. Env. Hlth. Sci., Vol. 37, No. 1, pp 1~11(2011)

초 빙 논 문

공간역학과 환경보건: 공간위치정보 활용에 대한 고찰

한대권*,† · 황승식*****

*텍사스A&M 보건대학 역학과, **텍사스A&M 보건대학 공간역학 프로그램
***인하대 의학전문대학원 사회의학교실
(2011. 2. 4. 접수/2011. 2. 11. 수정/2011. 2. 14. 채택)

Spatial Epidemiology and Environmental Health: On the Use of Spatially Referenced Health and Environment Data

Daikwon Han*,† · Seung-Sik Hwang*****

**Department of Epidemiology and Biostatistics, School of Rural Public Health, Texas A&M Health Science Center*
***Program on GIS and Spatial Statistics, School of Rural Public Health, Texas A&M Health Science Center*
****Department of Social and Preventive Medicine, School of Medicine, Inha University*
(Received February 4, 2011/Revised February 11, 2011/Accepted February 14, 2011)

ABSTRACT

Recent advances in Geographic Information Systems and spatial statistical and analytical methods, along with the availability of spatially referenced health and environmental data, have created unique opportunities to investigate spatial associations between environment exposures and health outcomes at multiple spatial scales and resolutions. However, the increased use of spatial data also faces challenges, one of which is to ensure certainty and accuracy of locational data that meets the needs of a study. This article critically reviews the use of spatially referenced data in epidemiologic studies, focusing on the issue of locational uncertainty generated from the process of geocoding health and environmental data. Primarily, major issues involving the use of spatially referenced data are addressed, including completeness and positional accuracy, potential source of bias and exposure misclassification, and implications for epidemiologic studies. The need for critical assessment and caution in designing and conducting spatial epidemiology studies is briefly discussed.

Keywords: Geographic Information Systems, spatially referenced data, geocoding, locational uncertainty, spatial epidemiology, environmental health

I. 서 론

최근 지리정보시스템(Geographic Information Systems, 이하 GIS로 표기)을 중심으로 빠르게 발전하고 있는 지리공간기술(geospatial technology) - 위성항법장치(global positioning system, 이하 GPS로 표기), 원격탐사기술(remote sensing) 등을 포함하는 - 을 환경보건연구 분야에 활용한 공간분석과 역학연구들이 급속한 성장세를 보여주고 있다. GIS는 우리를 둘러싸는 모든 자연적(physical-natural) 및 인문적(human-cultural) 환경을 포함하는 지리정보를 저장/통합하는 데이터베이스를 구축할 수 있게 하고, 이를 관리/요약하거나 질의하며, 편집된 결과들을 출력할 수 있게 해주는 컴퓨터기반의 정보시스템이다. 환경보건 연구의 특성상, 여러 인문과 자연환경의 자료들을 활용하여 복잡한 환경과 건강의 관계를 살피는 게 주요한 연구과제 중 하나인데, 개인의 건강관련 지표나 행태자료, 그리고 이러한 광범위한 의미의 환경요인들을 공간적으로 표시하고(spatially referenced or georeferenced health and environmental data, 이하 공간위치정보라 함), 이들의 관계/연관성을 기술하고 분석하는 공간역학(spatial epidemiology)이 역학연구분야에서 하나의 새로운 방법론으로 자리 매김을 하고 있다.[1,2] 또한 GIS에 기반을 둔 공간분석 기법

†Corresponding author : Department of Epidemiology & Biostatistics, School of Rural Public Health, Texas A&M Health Science Center
Tel: 1-979-458-0068, Fax: 1-979-458-1877
E-mail: dhan@srph.tamhsc.edu

(GIS-based spatial analysis methods) 개발에 힘입어, 특히 환경역학과 노출측정/평가(environmental epidemiology and exposure assessment) 분야에서 새로운 방법과 도구로서의 역할을 수행하게 되었다. 이러한 추세를 반영하듯 미국에서 최근 10-15년간 이 분야에 많은 환경역학 관련 연구자들이 관심을 두고 연구해 왔는데, 이러한 경향들은 신간 서적 발간,[3-5] 신규 저널의 발간,[6,7] 저널의 특집호 발행,[8-10] 그리고 많은 지리공간 자료와 정보의 온라인 및 웹 포털 공개 증가 등으로 확인할 수 있다.[11]

공간역학의 성장이 가능했던 배경으로 가장 먼저 이용 가능한 지리공간자료(geospatial data)가 늘었고, 또한 이러한 개인의 건강과 환경요인들의 지리공간자료를 처리하여 시각화(visualization)하고 분석 및 모형화(analysis and modeling) 가능한 공간자료 처리/분석 방법론과 도구들의 개발을 들 수 있다. 이는 주어진 자료를 바탕으로, 그 자료에 가장 적합한 방법론을 적용하여 유용한 정보로 변환시켜 주며, 공간현상에 관련된 문제를 설명하고 해결할 수 있으며, 지도나 표와 같이 정리된 형태로 원하는 결과물을 얻기 위한 기술, 수단, 그리고 제반 과정을 포함하는 것이다. 하지만, 다양한 환경보건 연구에의 응용분야와 연구자들에게 이용 가능한 지리공간자료가 많아졌음에도, 공간위치정보의 불확실성(locational uncertainty of georeferenced data)을 만드는 요인들은 많다. 공간위치정보의 불확실성과 위치정확성 오류, 이에 따른 바이어스와 노출 분류오류(exposure misclassification)의 가능성 등을 포함하여, 연구지역/분석단위(spatial scale/resolution), 공간분석 방법론의 선정에 이르기까지, 공간위치정보의 활용에 따른 제반사항들에 대한 세밀한 점검이 요구된다. 따라서 그동안 간과되었던 공간자료의 준비부터 분석에 이르기까지 공간자료의 특성과 처리과정에서 유의해야 할 사항들에 대한 깊이 있는 고찰이 필요한 실정이다.

본 논문에서는 공간역학의 발전을 이끌고 있는 세 가지 학문영역인 지리공간과학(geospatial science), 환경과학(environmental science), 역학(epidemiology) 중 지리공간과학 영역의 문제를 중점으로 하여 역학연구에서 공간위치정보의 올바른 활용과 밀접하게 관련된 불확실성과 오류, 특히 주소매칭(geocoding) 방법을 통한 공간위치정보의 습득/변환하는 과정과 분석에서 발생하는 여러 가지 이슈들을 정리하였다. 미국의 사례를 중심으로 연구동향, 선행연구들의 예를 사용하였고, 이들이 공간분석과 환경역학/노출평가에 주는 함의 점을 고찰 및 토의하는 것을 주안점으로 하였다.

II. 고찰 및 토의

1. 공간위치정보와 공간분석방법론의 사용

이용 가능한 지리공간자료의 증가와 공간분석 방법론들의 발전에 힘입어 공간역학 방법론을 활용한 환경보건 연구들은 다양한 분야에서 수행되고 있다. GIS는 고유의 분석과 자료처리 기능들에 기반, 공간-비공간자료(locational-attribute data)들의 저장/연계, 그리고 상호관계(spatial correlations)를 분석할 수 있고, 이를 통하여 시공간 상 질병의 변화를 추적/감시할 수 있는 감시 체계(surveillance systems), 탐색적 공간분석 방법론들을 활용한 인구집단 간 또는 지역단위 간의 건강 위험요인들의 격차 분석(health disparity), 공간모형화(spatial modeling)를 통해 이들의 복잡한 관계를 설명할 수 있는 가설의 생성, 그리고 설정된 가설을 역학연구를 통해 밝혀낼 수 있는 기반을 제공한다.[5,12] 일반적으로 공간분석 방법론은 연구의 목적과 방법에 따라 탐색적(exploratory) 또는 확증적(confirmatory)으로, 차원에 따라 시간과 공간으로, 자료의 형태에 따라 점(point) 또는 면(area)으로 분류되며, 각각의 범주에 사용될 수 있는 다양한 형태의 공간분석 방법론들이 개발되었다.[13-15] GIS를 이용한 질병 지도화와 군집 분석 기법(disease mapping and clustering analysis methods)은 탐색적 공간분석 기법(exploratory spatial analysis methods)을 이용하여 질병의 공간적 유형을 통해 노출-건강의 인과관계를 파악하기 위한 가설을 창출하는 방법으로 많이 사용되었다.[5,16] 일례로 환경보건 연구에 가장 광범위한 분야에서 활용되었던 탐색적 공간분석 방법의 하나인 공간 스캔 통계량(spatial scan statistic)은 공간, 시간, 그리고 시공간 군집(spatial & spatio-temporal clustering) 방법론을 포함하여, 지역과 개인단위 각각의 자료 형태에 따른 공간통계모델이며, 지역 단위 자료를 위한 포아송(Poisson) 모형, 개인 단위를 위한 베르누이(Bernoulli) 모형을 제공한다.[17] 개인단위 위치정보에 기반을 둔 점 자료와 이를 지역단위(예를 들어, 읍면동이나 시군구) 수준의 집합된 자료(aggregated areal data)를 이용하는 소규모 지역 수준의 공간분석(small area analysis)은 일반적으로 자료의 이용 가능성이나 연구의 목적에 따라 결정되며, 연구의 설계와 분석에서 고려해야 할 사항들 또한 달라진다. 다른 예로 초기환경 요인들(early-life factors)에의 노출과 유방암의 관계를 살펴보기 위한 연구의 일환으로 환자-대조군들의 군집 유형을 탐색적으로 살펴본 연구는 점유형 분석(point pattern analysis) 중 하나인 k-function 방법을 사용, 환자-대조군(case-control)의 차이를 이용한 군

집 유형과 환경요인들의 영향을 설명하고자 하였다.[18] 이렇듯 개인의 위치에 기반을 둔 공간분석을 하는 데 체계적이고 정확한 공간위치정보의 파악은 필수 선행 요건이며, 주소매칭이라는 과정을 통해 위치정보를 습득/지도화하고, 그 분포를 파악하는 가장 적절한 공간 통계방법론을 사용하게 된다.

공간분석뿐만 아니라 노출측정/평가에서도 공간위치정보를 활용할 때 고려할 사항들은 많다. 공간위치정보의 습득과 활용에 관한 여러 가지 문제들을 논의할 내용으로 넘어가기 전에 한 가지 덧붙일 것은, 환경역학 연구에서 활용되고 있는 노출측정/평가 방법론들(exposure assignment and assessment methods)에 대한 것이다. 전반적인 노출측정/평가 방법의 고찰은 본 논문의 범위를 벗어나는 것이므로, 여기서는 간략하게 언급한다. 다양한 종류의 공간분석 방법들에 기반을 둔 노출측정/평가 방법론은 주로 대기오염역학(air pollution epidemiology) 분야를 중심으로 개발/활용되고 있으며, 근접성 분석방법(proximity analysis)과 같은 단순한 형태의 위치/거리기반 방법과 노출 여부를 규정하기 위해 오염원으로부터의 거리에 따라 완충지대(buffering)를 이용한 노출지역(exposure zone) 선정방법들, 그리고 지리통계방법론(geostatistical methods)에 근거한 여러 종류의 내삽법(interpolation-type)이나 회귀분석방법에 기반을 둔 방법론들(land-use regression modeling), 오염요인들의 특성과 과정, 그리고 생물학적 요인들을 포함한 확산 경로(dispersion pathways), 생물학적 지표(biomarkers), 누적노출(cumulative exposure)방법론에 이르기까지 보다 정교한 노출측정/평가 방법론들이 환경보건 연구에 활용되고 있다.[19-25] 거주지 기반 노출측정/평가방법은 대리 측정 지표(surrogate measure)로의 한계는 있으나, 기존의 연구에서 여러 환경과학과 노출 연구 방법론들과 함께 활용됐으며, 특히 최근 관심이 증대되고 있는 생애주기 역학(life course epidemiology) 접근법을 이용한 만성병 연구와, 이와 밀접하게 연관된 초기환경 요인들의 영향에 대한 연구들의 경우 - 현재의 생체지표(biomarker)를 활용할 수 없는 과거의 노출과 건강의 관계를 살피는 연구들에서 대리 측정지표로 활용되고 있다.[26-28]

본 논문에서는 공간위치정보 활용의 중요성을 고찰하기 위해 연구대상자들의 거주지나 기타 다른 형태의 활동기반 위치정보를 활용하여 노출을 측정/평가하는 연구들의 예를 주로 사용한다. 따라서 이러한 연구에서는 노출의 측정을 위한 환자-대조군들 거주지나 다른 활동기반장소의 정확한 위치정보, 연구의 관심 오염물질(예 - 전기시설물과 같은 전자기장(electromagnetic fields, 이하 EMF로 표기)의 발생원) 위치정보 모두가 중요하다. 환자-대조군 연구에서 수집한 주소정보가 있다면 이들의 거주지 정보에 기반한 공간위치정보를 습득하고, 거주지로부터 환경오염원인 전기시설물까지의 거리를 계산하며, 이를 통해 노출 여부를 규정하거나, 노출과 질병의 관계를 환자-대조군 간 비교분석 하게 된다. 이때 대리 측정지표로 사용되는 환자-대조군의 주거지(질병 발생 시 또는 사망 시)와 오염원들에 대한 신뢰할 수 있는 주소 위치정보의 습득, 그리고 불확실성을 최소화하고 정확성을 확보하며 생겨날 수 있는 노출 분류오류를 최소화할 수 있는 공간위치정보로의 변환을 위한 방법론 등이 주요한 사전 작업이 된다.

2. 공간위치정보의 습득: 주소매칭(geocoding)

환자군과 대조군의 암 발생률을 지도화하거나 비만과 환경요인들의 공간적 관계를 고찰하고자 하는 환경보건 연구에서 주소매칭은 공간분석 수행을 위한 첫 번째 작업이다. 이는 공간위치정보를 습득/변환하는 과정에 사용되는 방법으로, 일반적인 과정에 대한 상세정보, 그리고 암 연구에서 주소매칭이 갖는 중요성을 요약한 논문들이 최근에 출판되었다.[29,30] 다음에서는 주소매칭의 과정을 간략히 요약한다. 연구자들은 먼저 질병자료, 이를테면 암 환자의 주소 정보와 환경오염원이나 관련된 위험 요인(risk factors)의 위치정보를 표 형식으로 얻게 된다. 이들의 공간적 연관성(spatial association)을 고찰하기 위한 연구에서 공간분석 방법론을 적용하려면 먼저 환자군의 주거지와 유해 시설물들의 자료를 공간위치정보로 변환하여 지도 상에 표시하여야 한다. 따라서 주소매칭 작업에는 표 형식의 주소 정보와 연구지역의 기본 지도(reference map)가 필요하며, 이들의 주소정보를 지리 좌표(geographic coordinates)로 변환하는 작업이 필요하다. 환자-대조군 연구에서 설문을 통해 얻은 환자군과 대조군의 주소 정보를 지리 좌표(경도와 위도 형식)로 변환하고, 정확한 위치를 지도 상에 표시하는 작업을 하게 된다.

공간위치정보로의 변환과 습득은 다음과 같은 여러 단계를 거치게 된다. 연구 목적에 맞는 잘 정비된 최신 연구지역 기본 지도의 선정과 획득, 주소 자료의 표준화 및 정리, 주소정보 추가 및 갱신(update), 코드화/좌표부여, 자료매칭/정합, 매칭된 공간위치정보의 정확성 평가를 위한 민감도 분석(sensitivity analysis)을 수행하게 된다. 이외에도 자료의 손실을 최소화하고 가능한 한 많은 자료를 변환시키고(completeness, 이하 완결성이라 함) 동시에 공간위치정보의 정확성(accuracy)을 확보하기 위한 여러 보조수단과 방법들이 사용된다. 이러

한 과정에서 공간위치정보의 불확실성과 오류(완결성과 정확성)에 영향을 주는 많은 요인이 있게 된다. 먼저 사용되는 자료의 질과 관련된 요인들이 중요하다. 앞서 간략히 언급한 것처럼 사용되는 두 가지 정보원인 - 기본 지도와 연구대상자나 유해시설물들의 주소정보 -의 질이 중요하다. 사용되는 연구지역의 기본 지도에서 기인한 오류가 있다. 미국은 연방 통계청에서 제공하는 Topologically Integrated & Geographic Encoding and Referencing(TIGER)가 이러한 수치 기본도의 바탕을 이루며, 여기서 생성된 도로망 지도를 기본으로 사용한 매칭방법(street geocoding)이 가장 일반적으로 사용되어 왔다.[29] 도로 방식의 주소체계에 기반을 둔 매칭법으로 도로망 지도의 도로 구역(segments)을 근거로 내삽법을 사용하여 도로에서 상대적 위치에 따라 집/건물들의 위치정보를 표현하게 된다. 이때 도로명이 명시되지 않은 경우, 존재하지 않은 도로나 도로 구역, 새로운 도로의 누락 등과 같은 기본 지도가 포함하는 지리정보의 완결성에 관련된 오류가 있게 되며, 또한 이러한 기본도들에 내재되어 있는 위치정보의 정확성 오류(positional error) 역시 하나의 요인이 된다.

마찬가지로 주소 정보의 오류에 따른 잘못된 위치 정보의 생성 역시 자료의 질에 영향을 주는 중요한 이유로, 이는 역학 연구에서 정보 바이어스의 일종인 보고 바이어스(reporting bias)와 회상 바이어스(recall bias)가 개입하여 발생하게 된다. 설문을 통해 얻은 환자군과 대조군의 주소가 완전치 않거나 또는 잘못된 경우, 표준화되지 않은 주소정보의 사용(비공식적인 도로명 또는 약자 사용 등) 등 조사과정에서 다양한 오류가 생길 수 있다. 회상 바이어스는 특히 생애주기 접근 방법을 적용한 역학 연구에서 많이 발생한다. 앞서 언급한 사례인 뉴욕주 유방암 환자-대조군 연구에서는 환경 요인과 유방암의 관계를 살펴보고자 환자군과-대조군 모두의 평생 거주력(lifetime residential history) 정보를 수집하였다. 훈련된 면접요원들이 엄격하게 설문조사를 시행했지만 사람들이 태어나서부터 지금까지 살아온 모든 주소를 완벽하게 기억하지 못하여 오류 발생은 불가피하였다.[31] 필요한 수준의 완결성과 정확성을 확보하기 위해 앞서 언급한 것처럼 표준화된 주소매칭 방법론을 개발/활용하였고, 역사적으로 개인들의 주소를 기재해 놓은 재산지도(property map)와 목록(directory)의 사용, 기업에서 만들어 놓은 최신 연구지역 지도, 필요에 따라선 GPS를 이용한 현장측정 등과 같은 다양한 보조수단을 적극적으로 활용하였다.

마지막으로 매칭작업 과정에서 사용된 방법론과 매칭 알고리즘/기준치 설정에 따른 오류들이 발생할 수 있다. 연구목적에 따라서는 통상적인 GIS 프로그램에서 제공하는 주소매칭 기능이나 기업에서 제공되는 서비스를 이용할 수도 있다. GIS 프로그램에서 제공하는 기능에도 매칭 알고리즘의 선정과 많은 기타 선택사항(options)들을 사용할 수 있다. 일례로 GIS 프로그램에서 제공하는 도로지도를 기본도로 사용하는 방식의 경우 알고리즘은, 도로 이름의 철자 민감도(spelling sensitivity), 매칭 후보자들의 최저점수(minimun candidate score), 매칭의 최저점수(minmum match score) 등 매칭의 기본 값(default value)을 설정하게 되며, 기본 값을 변화시킴에 따라 매칭률은 큰 차이를 보여주게 될 것이다. 사전에 프로그래밍이 된 기본 값을 활용하여 한꺼번에 매칭하거나(batch match), 여기서 매칭이 안 되었을 경우, 반복적 과정(interactive matching process)을 통해 가장 적절한 후보를 찾게 된다. 이러한 작업은 상당한 수준의 인력과 시간을 요구하며(labor-intensive process), 작업하는 사람에 따른 변이(human error)를 보일 수 있다는 점 등의 문제점을 안고 있다.

3. 공간분석에의 함의: 지리적 선택 바이어스 (geographic bias)

그렇다면 왜 주소매칭이 환경보건연구에서 중요한 문제인가? 주소매칭의 완결성은 곧 역학연구의 타당도(validity)에 영향을 주는 선택 바이어스와 직결되기 때문이다. 환경역학 연구에서 완결성을 높이고자 매칭 기준을 완화하면, 결과적으로 매칭률을 더 높일 수 있는 반면, 정확성은 보장할 수 없다. 잘못된 위치에 환자군을 위치시켜 노출측정에 큰 오류를 발생시킬 수도 있기 때문이다. 텍사스 지역의 대기 오염과 기형아 출산을 주제로 한 환자-대조군 연구는 선택 바이어스(selection bias)의 영향을 살펴보았다. 산모의 주거지 주소를 이용하여 GIS 프로그램 주소매칭기능의 자동 매칭(batch match)된 환자-대조군들만 포함하였을 때와, 기본값을 완화하고 매칭된 결과를 모두 포함하였을 때를 비교해보니, 대기 오염과 기형아 출산의 관계 추정값(point estimate)의 변화는 크게 보이지 않았으나, 통계적 검정력(statistical power)이 증가하였음을 보고하였다.[32] 또한, 매칭된 것과 되지 않은 연구대상자들이 교란변수들, 특히 라틴계 인종(Latino ethnicity)이 차이가 있음을 발견하였고, 주소매칭의 결과가 선택 바이어스로 연결될 수 있음을 실증적으로 보고하였다.

선택 바이어스가 공간역학 연구에 주는 오류를 지리적 선택 바이어스(geographic bias)라 부르고, 이는 주소매칭의 과정에서 생겨날 수 있는 중요한 오류이다.[33] 공간분석에 기반을 둔 환경역학 연구에서는 주소매칭

이 연구분석 대상자 선정에 하나의 기준이 될 수 있다. 매칭이 되지 않을 때 연구대상자들을 제외하여 문제가 발생할 수 있는데, 결과적으로 공간패턴에 영향을 주게 되고, 제외된 연구대상자들이 랜덤(random)하게 분포하는 것이 아니라 공간적으로 군집되어 있는 경우 심각한 지리적 선택 바이어스가 생긴다. 이와 관련하여 공간분석방법, 특히 공간군집분석(spatial clustering analysis)의 검정력 역시 낮춘다는 여러 실증 연구들이 보고되었다.[34,35]

완결성이 정확성을 의미하지는 않는다. 상세한 주소 정보가 이용 가능하지 않을 경우나 특수한 형태의 주소자료 경우(예를 들면 미국의 농촌지역은 우편사서함인 PO Box를 주소로 보고하는 사람들이 많다) 완결성을 높이기 위한 대안으로 센서스에 기반을 둔 지역단위의 중심점(centroid)을 이용하여 대략적인 공간위치정보를 습득하기도 한다.[36,37] 이 방법은 주소매칭의 완결성을 증가시켜 주나, 정확성을 감소시킨다. 요구되는 정확성은 환경역학 연구의 목적과 자료의 형태에 따라 다르겠고, 특히 오염물질의 특성이나 관련 사항들이 고려되어야 한다. 농약(pesticide) 노출과 건강의 관련성에 대한 환경역학 연구에서는 노출을 분류/평가하기 위해 환자군과 대조군의 거주지 주변에 대략 500미터의 완충 지역을 설정 노출지역(exposure zone)으로 분류하였고, 오염물질의 특성, 그리고 농약 살포시 바람에 의한 표류(drift) 현상을 고려한 노출 여부의 한계치(threshold)를 선정하여 분석하였다.[26] 공간위치정보의 사용과 관련된 불확실성을 최소화하고 위치정확성을 확보할 수 있는 주소매칭 방법론의 사용과 이를 이용한 노출측정 방법의 사용이 노출과 건강의 관계를 고찰하는 환경보건 연구에서 중요한 문제이며, 이들의 관계를 보여주는 통계적 검정력은 연관성의 강도 뿐 아니라 공간위치정보의 완결성과 정확성에도 영향을 받게 되기 때문이다.[39]

4. 공간위치정보의 정확성과 환경오염물질: 노출측정 분류오류(exposure misclassification)

공간위치정보의 정확성은 공간위치정보에 기반을 둔 노출분류/측정 방법의 사용과 관련해서도 중요한 함의를 갖는다. 노출측정 분류오류로 인한 연구의 검정력을 낮출 수 있고, 위험도 추정치를 편향시킬 가능성이 있기 때문이다. 최근 공간위치정보의 불확실성과 정확성 오류에 대한 많은 연구가 발표되어 높은 관심을 반영하고 있다.[37,38,40-44] 위치정확성(positional accuracy)은 주소매칭 방법의 사용에 따른 공간위치정보의 오차(uncertainty and error of locational data)를 의미하며, GPS나 원격탐사를 통해 얻은 실측정값을 기준값(gold standard)으로 사용하여 비교/분석한다. 환경역학 연구들의 일환으로 수집된 환자-대조군들의 자기 보고식(self-reported) 주소 정보를 기반으로 GIS프로그램이나 기업에서 제공되는 서비스를 이용하여 습득된 공간위치정보가 실측정값과 차이의 정도를 거리/방향등을 통해 비교하며, 이러한 위치정보의 정확성 오차가 노출측정 분류 오류에 주는 영향을 고찰한 연구들이다. Table 1에 공간위치정보의 정확성과 노출분류 오류를 평가한 연구들을 정리하였으며, 다음에서는 일부 주요 연구들을 간략히 소개한다. 200명의 유방암 환자-대조군 표본을 이용하여 위치 정확성을 평가한 연구는 GPS 측정치와 GIS 프로그램을 이용하여 얻은 주소 간 평균 38 m 미만의 차이가 있었고, 일반적으로 농촌지역 주소의 위치 정확성이 낮았으며(52 m), 환자-대조군 간 비차별분류오류(non-differential misclassification)의 결과가 있음을 보고하였다.[40] 농약 노출과 비호지킨 림프종(non-hodgkin lymphoma) 환자-대조군 자료를 이용한 연구도 이와 비슷한 결과를 보고하였다.[38] 특히 농촌지역에서 위치정확성 자료의 문제점과 함께, 환경오염물질에의 노출 여부를 분류할 때 발생할 수 있는 노출측정 분류오류를 주의할 것을 제안하였다. 한편 학생들의 노출측정/평가를 위해 학교의 공간위치정보를 사용하여, 특히 대기오염 노출측정 연구에서 중요한 도로에서의 거리에 기반을 둔 노출측정 분류오류를 분석한 연구에서는, 높은 수준의 위치정확성 오류(평균 151-178 m의 차이)와 근거리(<500 m)에서의 노출분류 오류가능성을 보고하였다. 따라서 더 높은 수준의 정확성을 확보할 수 있는 지리정보 원(data source)과 공간위치정보 습득을 위한 방법론의 사용을 제안하였다.[44,45]

모든 분야의 환경보건 연구에서 높은 수준의 위치 정확성을 요구하지는 않는다. 연구의 목적과 범위, 관심대상인 환경오염물질의 특성에 따라 요구되는 공간위치정보의 정확성이 달라진다. 앞에서 간략히 언급하였듯이, 공간역학에서 광범위하게 사용되는 있는 소규모 지역수준의 공간분석(small area analysis) 연구의 경우, 대규모 역학 연구(large epidemiologic study)에서 환경과 건강의 관련성에 대한 가설을 만들어 내고, 궁극적으로 보건정책(health policy) 수립에 객관적인 자료를 제공하려는 연구 목적상 - 요구되는 공간위치정보의 정확성은 높지 않을 수 있다. 개인들의 사회경제적 요인들이 환경오염물질의 노출과 건강에 주는 영향을 연구하는 환경불평등 연구(environmental justice)가 대표적인 사례이다. 이러한 연구에서는 넓은 연구지역과 인구집단을 대상으로, 종종 인구 센서스에 기반을 둔 자료

를 활용하여, 환경오염물질에 노출된 인구집단의 파악과 주요변수 간의 연관성을 살펴보는 데 주목적이 있다.[46,47] 또한, 환경보건 정책의 개발을 위한 연구들, 예를 들면 도시-농촌지역 간 의료시설에의 접근성(spatial accessibility) 차이, 비만과 환경연구에서 소규모 지역단위 간 녹지공간 접근성의 차이 등, 연구가설과 지역단위 수준에서 보건정책의 문제점들을 파악하는 데 주목적이 있는 연구들의 경우도 마찬가지다. 환경역학 연구들 역시 환경오염물질의 특성에 따라 상대적으로 요구되는 공간위치정보의 정확성이 낮은 수준에서 가능한 연구들이 있다. 핵발전소에서 배출되는 방사선 가스의 노출과 백혈병 발생의 관련성을 살펴본 연구들은 대략 24-37 km의 거리를 이용하여 노출 여부를 분류하여 연구하였다.[48,49] 또한, 기형아 출산과 유해화학물질의 관련성을 살펴본 많은 환경역학 연구들에서도 환자-대조군의 노출 여부 기준으로 대략 1.5-3 km 범위의 거리를 사용하였다.[50-53]

반면, 상대적으로 짧은 거리상에서 큰 폭의 공간 변이를 보여주는 환경오염물질들이 있다. 가장 좋은 예 중 하나로 소아암과 EMF의 노출 간 연관성에 대한 연구들이다. 이들은 EMF의 출처가 되는 시설물인 송전선이나 배전 시설로부터의 거리인 대략 50-150 m를 노출 여부 기준치로 사용하였는데, 100 m의 거리는 대략 0.2 uT의 전자기장 노출에 해당하는 것으로 보고되었다.[54-56] 교통관련 대기오염의 건강 영향에 대한 연구들 역시 높은 정확성을 요구하게 된다. 차량으로부터 배출되는 대기오염 물질들이 소아암, 천식 등의 질환과의 연관성을 살펴본 연구에서, 주요 도로로부터 환자-대조군들 거주지까지의 거리나, 거주지를 중심으로 도로밀도와 교통량을 일정거리 이내에서 계산하는 방법 등, 거리를 기반으로 하는 노출측정 방법들은 대략 50-200 m를 기준치로 사용하였다.[24,25,57-59] 이는 교통관련 대기오염물질의 농도(concentration)에 따른 것으로, 주요 대기오염 물질인 이산화질소는 거리의 증가에 따라 기하급수적으로 감소하는 공간적 경사(spatial gradient)를 보이기 때문이다.[60,61]

5. 공간위치정보의 올바른 활용을 위한 제언

공간분석에 기반을 둔 환경보건연구에는 올바른 주소매칭 방법의 선정과 사용에 따른 노출측정/평가, 그리고 노출-건강관계에 미치는 영향에 대한 가능성을 항상 유의해야 한다. 따라서 연구의 설계에서부터 공간위치정보의 습득, 주소매칭의 완결성/정확성, 그리고 매칭된 자료를 이용한 공간분석에서 바이어스와 오류들까지 여러 단계에 걸쳐 공간위치정보가 연구의 타당도와 신뢰도에 줄 수 있는 영향을 면밀히 검토해야 한다. 먼저 공간역학 연구의 설계 단계에서는, 1) 연구목적과 환경오염물질의 특성을 고려한 공간위치정보의 질과 주소매칭 방법론의 선정, 2) 최대한 완결성을 높일 수 있는 표준화된 방법론의 개발과 사용, 기본도의 선정, 인터넷 비교분석, 상업화된 데이터베이스의 사용 등 적절한 보조 수단들의 이용, 3) 그리고 연구목적에 맞는 연구지역의 공간분석 단위(spatial scale/resolution) 선정 후 요구되는 공간위치정보의 정확성을 평가한다.

주소매칭과 분석 단계에서도 다음과 같은 여러 가지 사항들에 대한 세심한 주의가 필요하다. 1) 거주지 기반 근접성 분석과 같이 공간위치정보를 노출측정/평가에 활용한 연구는 항상 노출측정 분류오류의 가능성을 염두에 두고, 이를 평가/보고해야 하며, 2) 여러 가지 바이어스, 특히 지리 선택 바이어스, 노출 분류 오류에 대한 주의가 요구된다. 사용된 기본도의 오류로 인한 지리 선택 바이어스는 더욱 세심한 관찰이 요구된다. 노출과 건강의 연관성에 대한 통계검정력과 편향된 추정 값의 문제뿐 아니라, 매칭되지 않아 제외된 대상자들의 공간분포가 체계적인 형태를 보여주는 경우(예를 들면 군집해 있는 경우) 매우 심각한 오류의 소지로 발전할 수 있다. 따라서 항상 연구에 사용된 공간위치정보의 출처와 질뿐 아니라, 주소매칭의 상세한 과정, 그리고 매칭이 되지 않은 자료들이 공간분석 결과에 어떠한 영향을 줄 수 있는지를 점검하고, 보고해야 한다. 유방암과 환경 요인의 관계에 대한 공간역학 연구에서는 주소매칭의 상세한 과정 - 완결성을 높이기 위한 방법론의 개발, 매칭의 성공률, 그리고 매칭이 되지 않은 자료들(주소정보가 전혀 없는 경우, 그리고 정보는 있으나 연구지역에서 벗어난 지역의 주소정보를 포함)의 정보, 그리고 매칭되지 않은 환자군과 대조군을 연구에서 제외함으로써 생길 수 있는 바이어스에 대한 고찰을 논문에 포함시켰다.[18] 백혈병과 석유화학물질 노출의 관계에 대한 연구에서도 주소매칭의 상세 과정과 바이어스에 대한 고찰을 보고하였고,[62] 학계는 이제 이러한 절차가 공간위치정보를 활용한 환경역학 연구에서 필요하다는데 인식을 같이하고 있다.

주소매칭의 완결성과 정확성을 둘 다 확보하는 방법의 사용이 이상적이겠으나, 여러 가지 고려사항들이 있겠다. 먼저 완결성을 위한 오류를 유의하면서, 정확성을 확보할 수 있는 연구의 설계가 중요하겠다. 특히 다양한 도시규모의 환경보건 연구를 수행할 시, 이들이 주는 영향을 고려한 연구방법론의 선정과 분석이 요구된다. 앞서 언급한 바와 같이 일반적으로 농촌지역 위치정보, 그리고 주소매칭은 도시지역과 비교하면 상대

적으로 여러 가지 이유로 정확성에 문제가 있을 수 있다.[38,40,41] 예를 들면 부정확한 주소정보의 사용, 사용된 기본 지도가 자주 갱신되지 않는 경우가 있다. 또한, 주소매칭의 오류가 노출측정 분류오류에 미치는 영향은 해당 노출요인의 공간분포/변이(geographic variation)에 좌우된다. 따라서 노출측정 분류오류를 최소화하려면 연구설계 단계에서 요구되는 연구지역 단위 및 연구지역의 특성(오염물질 또는 인구집단의 공간적 분포)에 대한 사전 작업이 필요하다.[2] 사용되는 자료의 질은 연구목적에 맞는 공간위치정보의 올바른 활용을 위한 주소매칭 작업의 성공 여부(완결성)를 좌우한다. 이 과정에서 사용되는 기본도와 연구대상자 두 가지 모두의 공간위치정보의 완결성과 정확성이 중요하며, 앞에서 언급하였듯이 표준화된 방법론의 개발, 기타 보조작업의 이용 등은 완결성을 올리는 데 이바지하게 된다. 물론 거주지 기반 노출측정 연구에서의 사용과 같은 원리가 다른 활동기반이 되는 다른 장소들 - 직장, 학교 등 - 그리고 오염물질 원 등의 위치정보(앞서 언급한 EMF시설의 예)에 적용되며, 공간위치정보 완결성과 정확성은 이러한 건강과 환경 두 가지 모두의 요인들에 적용된다. 따라서 연구에 사용되는 모든 위치정보를 가능한 동일한 출처의 기본도를 이용하여 주소매칭을 하는 것도 오차를 줄이는 하나의 방법이 되겠다.

반복성(repeatability)의 확보 역시 고려되어야 할 하나의 주요한 작업이다. 연구에 사용된 공간위치정보의 습득과 변환 방법론을 다른 방법을 사용한 결과들을 비교/분석하여 검증하는 절차가 완결성, 정확성을 평가하는 데 유용한 방법이 되겠다.[37,44] 최근 높은 수준의 정확성을 요구하는 환경역학 연구에서 사용될 수 있는 다른 방법들과의 비교와 사용이 제안되고 있는데, 항공사진과 같은 실측정값의 기준이 되는 방법들의 이용 가능성, 지적도의 중심점을 이용한 방법(address point), 응급 911 데이터베이스의 활용 등이 연구되었다.[30,39,41,63] 예를 들면 전자 지적도(digital parcel map)를 이용한 공간위치정보는 높은 수준의 정확성을 확보할 수 있는 대안으로 제시되기도 하며, 사전작업의 요구에 따른 비용과 시간문제, 연구지역에 따라 다른 전자지적도의 구축 여부에 따른 이용 가능성 등은 선결되어야 할 문제들이다. 따라서 점차 많아지는 대규모 역학연구에서 공간정보자료의 준비와 처리에 드는 비용과 실현가능성(feasibility) 역시 고려해야 할 사항이다.[64] 환경보건 연구에서 질병과 환경오염 물질에 따라 요구되는 위치자료의 정확성, 향상된 노출평가를 가능케 해주는 방법론 설정은 공간환경 역학연구를 위한 첫걸음이며, 이는 환경보건 현장에서도 중요한 문제가 되고 있다. 미국은 Healthy People 2010의 여러 목적중 하나로 연방/주정부 보건관련 부서들의 보다 효율적인 건강과 환경관련 지리공간자료의 사용을 위한 체계적인 주소매칭 방법의 개발, 그리고 이를 통한 공간위치정보와 GIS사용의 진작을 골자로 하는 주소매칭과 관련된 제안(geocoding initiative)을 시행하였다. 우리나라의 경우도 체계적이고 객관적인 주소 정보의 정리, 이에 수반되는 주소매칭 방법의 개발과 사용, 이를 통한 양질의 공간위치정보 습득과 활용을 위한 표준화된 방법론의 개발이 요구되고 있다. 최근 새로운 도로방식 주소체계의 도입으로 그 필요성이 어느 때보다도 절실하게 대두되고 있으며, 본 논문에서 제시된 미국의 사례와 선행연구 결과를 토대로 여러 방법들의 장/단점을 고려한 표준화된 공간위치정보의 활용을 위한 방법의 개발, 그리고 이를 통한 학계와 현장에서의 향상된 공간정보 활용의 계기가 되기를 기대해 본다.

III. 요약 및 결론

GIS를 중심으로 한 지리공간기술의 발전에 힘입어 공간위치정보 이용이 증대되었고, 더불어 공간분석 방법론의 발전이 공간역학의 성장을 이끌어 왔다. 사용되는 공간자료의 질, 공간자료의 특성과 처리과정에 대한 이해와 관련된 함의 등은 적합한 분석 방법론의 사용과 함께 가장 먼저 해결되어야 할 중요한 문제들이다. 본 논문에서는 이러한 활용을 위해 주요 논점들을 정리하였다. 특히 지리공간과학의 근본적인 문제를 중심으로, 지리공간자료의 위치정확성/오류와 불확실성에 관련된 문제들을 소개, 고찰하였다. 역학에 대한 이해없이 GIS를 이용하여 건강과 환경의 관계에 대한 지도화나 단순 공간관계 분석에 그치는 점에 주의해야 하는 바와 같이, 공간자료의 속성과 사용된 방법론, 특히 공간위치정보를 습득하고 활용하는 데 있어 생길 수 있는 오류들에 대한 이해가 없는 역학연구 활용 역시 유의해야 한다. 현대 사회에서 지도는 강력한 매체이다. 지도를 통해 나타내고자 하는 복잡한 정보를 훨씬 효과적으로 전달할 수 있다. 하지만 지도화와 시각화가 GIS와 동일어는 아니다. 마찬가지로 주소매칭은 아주 실용적인 도구이다. 10여 년 전쯤 GIS와 공간분석의 개념과 방법론들이 알려지기 시작했을 때, 필자가 관련 학회에 가서 받았던 가장 많은 질문은 주소매칭에 대한 것이었다. 많은 사람이 주소매칭이 GIS라는 생각을 하고 있고, 특히 주소매칭의 성공률에만 관심이 있었다. 하지만, 지도화가 GIS가 아니듯이, 주소매칭 역시 GIS가 아니다. 주소매칭은 그 과정이 아주 단순할 수도 있

표 1. 공간위치정보의 정확성과 노출분류 오류

참고문헌	연구대상/환경요인	기준값	연구방법[1]	주요결과[2]	기타사항
38	비호지킨 림포마 234명 표본; 농약노출	GPS	GIS프로그램/기업매칭값과 비교	평균 62 m 차이; 도시 (56)/농촌(88)의 차이; 기업매칭값 더 큰 차이	낮은 수준의 노출분류 오류, 근거리 (<100) 제외
40	유방암 200명 표본; 대기오염 등의 환경요인	GPS	GIS 프로그램 값과 비교	평균 38m 차이; 도시 (32)/농촌(52)의 차이	비노출 분류오류; 과거 환경요인에의 노출평가
41	3000명 표본; 환경요인 명시되지 않음	항공사진	GIS프로그램/지적도 이용 값과 비교	도시(38)-교외(78)-농촌(201) 차이; 지적도 이용값 더 작은 차이	인구밀도에 따른 정확성 평가
42	기형아 출산 200명 표본; 대기오염등의 환경요인	GPS	두 종류의 GIS프로그램값과 비교	평균 48-69 m 차이	낮은 수준의 노출분류 오류 가능성 (민감도/특이도)
43	261 천식환자군 표본; 환경요인 명시되지 않음	항공사진	GIS프로그램/기업매칭값 들과 비교	평균 26-31 m의 차이; 편향성 보고	센서스 단위[3]에 할당하는 데 있어서의 오류 가능성 평가
44	126개 학교 표본; 교통관련 대기오염원	지적도기반 실측정값	GIS프로그램/기업매칭값과 비교	평균 151-178 m의 차이; 5% 특이점을 제외 후 196-306 차이	반복성 평가; 근거리 (<500) 에서 높은 수준의 노출분류 오류 가능성
37	142 표본; 환경요인 명시되지 않음	실측정값 (10개)	기업매칭값간, 실측정값/ 통계청 지도와 비교	기업매칭값간 비교: 비용, 시간, 정확성 상당한 차이	반복성 평가; 센서스 단위에 할당하는 데 있어서의 오류 가능성 평가

[1]GIS프로그램/기업매칭값은 TIGER나 이에 기반한 상업화된 도로망 지도를 기본도로 사용한 주소매칭 방법을 사용하여 얻은 값들이며, 실측정값과의 거리/방향 측정을 통해 오차를 비교/분석하였다; [2]보고된 평균은 모두 중간값(median)이며 단위는 미터이다; [3]센서스 단위로는 Census tract, census block group등이 사용되었다.

는 과정이지만, 이와 관련된 공간적/역학적 함의는 아주 크다. 단순한 프로그램에 의존한 주소매칭이나 그 공간위치정보의 특성에 대한 이해없는 활용은 경계하여야 할 것이다.

공간역학의 발전은 지리공간과학, 환경과학, 역학이라는 세 가지 학문영역에서 제공하는 이론과 방법론의 토대 위에, 균형잡힌 접근성과 동반 상승효과를 통해 궁극적인 목표인 환경과 건강의 관계를 밝혀내는 데 도움을 줄 수 있는 향상된 공간역학 방법론의 개발이 가능하다.[2] 최근 급속한 속도로 발전하고 활용되고 있는 공간역학 분야는 하나의 논문으로 정리되기 어려울 정도다. 따라서 본 논문의 목적은 광범위한 분야에 대한 피상적인 소개보다는 실용적이면서도 근본적인 이슈 중 하나인 공간위치정보의 활용에 대한 심층적인 고찰을 하고자 하였다. 본 논문에서 자세히 다루지 못한 많은 응용분야와 방법론적 고찰은 차후의 과제로 남겨둔다. 대기오염역학 연구분야를 중심으로 급속히 발전하는 GIS-기반 공간분석 방법론을 이용한 노출측정/평가 방법론들의 체계적인 정리가 필요하고, 소규모 지역수준 공간 분석과 관련된 문제들 - 탐색적 공간분석 방법, 공간단위의 선정에 관련된 문제들, 공간자료의 특성, 공간적 자기상관(spatial autocorrelation)들의 문제 - 이 환경보건 연구, 특히 넓은 영역의 환경 요인 간의 복잡한 관계를 분석하는 데 활용될 수 있는 방법론으로서의 의미와 시사점들 역시 정리되어야 한다. 공간역학 방법론들의 올바른 활용을 위해서는 공간자료의 특성, 공간분석 방법들의 용도를 이해하는 일이 반드시 선행되어야 한다.

참고문헌

1. Elliott, P. and Wartenberg, D. : Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives*, **112**(9), 998-1006, 2004.
2. Nuckols, J. R., Ward, M. H. and Jarup, L. : Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environmental Health Perspectives*, **112**(9), 1007-1015, 2004.
3. Cromley, E. K. and McLafferty, S. L. : GIS and Public Health. New York, Guilford, 2002.
4. Khan, O. A. : Geographic Information Systems and Health Applications. Hershey, PA: Idea Group Publishing, 2003.
5. Elliott, P., Wakefield, J. C., Best, N. G. and Briggs, D. J. : Spatial Epidemiology: Methods and Applications. Oxford: Oxford University Press, 2000.
6. International Journal of Health Geographics. http://www.ij-healthgeographics.com
7. Spatial and Saptio-temporal Epidemiology. http://www.sciencedirect.com/science/journal/18775845
8. Krieger, N. : Place, space, and health: GIS and epidemiology. *Epidemiology*, **14**, 384-385, 2003.
9. Richards, T., Pickle, L. and Rushton, G. : Prostate cancer and Geographic Information Systems (GIS). *American Journal of Preventive Medicine*, **30**, 1-118, 2006.
10. Ricketts, T. : Symposium: geographic information systems (GIS). *Annual Review of Public Health*, **24**, 1-82, 2003.
11. Croner, C. M. : Public health, GIS, and the internet. *Annual Review of Public Health*, **24**, 57-82, 2003.
12. Lawson, A., Biggeri, A., Bohning, D., Lesaffre, E., Viel, J. F. and Bertollini, R. : Disease mapping and risk assessment for public health. New York, Wiley, 1999.
13. Moore, D. A. and Carpenter, T. E. : Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiologic Reviews*, **21**(2), 143-161, 1999.
14. Bailey, T. C. and Gatrell, A. C. : Interactive Spatial Data Analysis. London: Longman, 1995.
15. Cressie, N. : Statistics for Spatial Data. New York, Wiley, 1993.
16. Rushton, G. : Public health, GIS, and spatial analytic tools. *Annual Review of Public Health*, **24**, 43-56, 2003.
17. Kulldorff, M. and Nagarwalla, N. : Spatial disease clusters: detection and inference. *Statistics in Medicine,* **14**, 799-810, 1995.
18. Han, D., Rogerson, P. A., Nie, J., Bonner, M. R., Vena, J. E., Muti, P., Trevisan, M., Edge, S. and Freudenheim, J. L. : Geographic clustering of residence in early life and subsequent risk of breast cancer (United States). *Cancer Causes and Control*, **15**, 921-929, 2004.
19. Beyea, J. : Geographic exposure modeling: a valuable extension of geographic information systems for use in environmental epidemiology. *Environmental Health Perspectives,* **107**(Suppl 1), 181-190, 1999.
20. Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J. and Giovis, C. : A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology,* **15**(2), 185-204, 2005.
21. Briggs, D. : The role of GIS: coping with space (and time) in air pollution exposure assessment. *Journal of Toxicology and Environmental Health A*, **68**(13-14), 1243-1261, 2005.
22. Ryan, P. H. and LeMasters, G. K. : A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhalation Toxicology,* **19** (Suppl 1), 127-33, 2007.
23. Huang, Y. L. and Batterman, S. : Residence location as a measure of environmental exposure: a review of air pollution epidemiology studies. *Journal of Expo-*

sure Analysis and Environmental Epidemiology, **10**(1), 66-85, 2000.
24. Reynolds, P., Von Behren, J., Gunier, R. B., Goldberg, D. E. and Hertz, A. : Residential exposure to traffic in California and childhood cancer. *Epidemiology*, **15**, 6-12, 2004.
25. Von Behren, J., Reynolds, P., Gunier, R. B., Rull, R. P., Hertz, A., Urayama, K. Y., Kronish, D. and Buffler, P. A. : Residential traffic density and childhood leukemia risk. *Cancer Epidemiology, Biomarkers and Prevention,* **17**(9), 2298-2301, 2008.
26. Ward, M. H., Nuckols, J. R., Weigel, S. J., Maxwell, S. K., Cantor, K. P. and Miller, R. S. : Identifying populations potentially exposed to agricultural pesticides using remote sensing and a geographic information system. *Environmental Health Perspectives,* **108**, 5-12, 2000.
27. Brody, J. G., Aschengrau, A., McKelvey, W., Rudel, R. A., Swartz, C. H. and Kennedy, T. : Breast cancer risk and historical exposure to pesticides from wide-area applications assessed with GIS. *Environmental Health Perspectives,* **112**(8), 889-897, 2004.
28. Bonner, M. R., Han, D., Nie, J., Rogerson, P. A., Vena, J. E., Muti, P., Trevisan, M., Edge, S. and Freudenheim, J. L. : Breast cancer risk and exposure in early life to polycyclic aromatic hydrocarbons using total suspended particulates as a proxy measure. *Cancer Epidemiology, Biomarkers and Prevention,* **14**, 53-60, 2005.
29. Drummond, W. J. : Address matching: GIS technology for mapping human activity patterns. *Journal of American Planning Association,* **61**, 240-251, 1995.
30. Rushton, G., Armstrong, M. P., Gittler, J., Greene, B. R., Pavlik, C. E., West, M. M. and Zimmerman, D. L. : Geocoding in cancer research: a review. *American Journal of Preventive Medicine,* **30**(Suppl 2), S16-24, 2006.
31. Han, D., Rogerson, P. A., Bonner, M. R., Nie, J., Vena, J. E., Muti, P., Trevisan, M. and Freudenheim, J. L. : Assessing spatio-temporal variability of risk surfaces using residential history data in a case control study of breast cancer. *International Journal of Health Geographics,* **4**, 9, 2005.
32. Gilboa, S. M., Mendola, P., Olshan, A. F., Harness, C., Loomis, D., Langlois, P. H., Savitz, D. A. and Herring, A. H. : Comparison of residential geocoding methods in population-based study of air quality and birth defects. *Environmental Research*, **101**(2), 256-262, 2006.
33. Oliver, M. N., Matthews, K. A., Siadaty, M., Hauck, F. R. and Pickle, L. W. : Geographic bias related to geocoding in epidemiologic studies. *International Journal of Health Geographics*, **4**, 29, 2005.
34. Zimmerman, D. L. : Statistical methods for incompletely and incorrectly geocoded cancer data. In *Geocoding Health Data: The Use of Geographic Codes in Cancer Prevention and Control, Research and Practice*, Rushton, G., Armstrong, M. P., Gittler, J., Greene, B. R., Pavlik, C. E., West, M. M. and Zimmerman, D. L. (eds). CRC Press: Boca Raton, FL, 2008.
35. Waller, L. A. : Statistical power and design of focused clustering studies. *Statistics in Medicine*, **15**, 765-782, 1996.
36. Hurley, S. E., Saunders, T. M., Nivas, R., Hertz, A. and Reynolds, P. : Post office box addresses: a challenge for geographic information system-based studies. *Epidemiology,* **14**(4), 386-391, 2003.
37. Krieger, N., Waterman, P., Lemieux, K., Zierler, S. and Hogan, J. : On the wrong side of the tracts? evaluating the accuracy of geocoding in public health research. *American Journal of Public Health*, **91**(7), 1114-1116, 2001.
38. Ward, M. H., Nuckols, J. R., Giglierano, J., Bonner, M. R., Wolter, C., Airola, M., Mix, W., Colt, J. S. and Hartge, P. : Positional accuracy of two methods of geocoding. *Epidemiology,* **16**(4), 542-547, 2005.
39. Mazumdar, S., Rushton, G., Smith, B. J., Zimmerman, D. L. and Donham, K. J. : Geocoding accuracy and the recovery of relationships between environmental exposures and health. *International Journal of Health Geographics*, **7**, 13, 2008.
40. Bonner, M. R., Han, D., Nie, J., Rogerson, P., Vena, J. E. and Freudenheim, J. L. : Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology,* **14**(4), 408-412, 2003.
41. Cayo, M. R. and Talbot, T. O. : Positional error in automated geocoding of residential addresses. *International Journal of Health Geographics*, **2**, 10, 2003.
42. Zhan, F. B., Brender, J. D., De Lima, I., Suarez, L. and Langlois, P. H. : Match rate and positional accuracy of two geocoding methods for epidemiologic research. *Annals of Epidemiology,* **16**(11), 842-849, 2006.
43. Schootman, M., Sterling, D. A., Struthers, J., Yan, Y., Laboube, T., Emo, B. and Higgs, G. : Positional accuracy and geographic bias of four methods of geocoding in epidemiologic research. *Annals of Epidemiology,* **17**(6), 464-470, 2007.
44. Zandbergen, P. A. and Green, J. W. : Error and bias in determining exposure potential of children at school locations using proximity-based GIS techniques. *Environmental Health Perspectives*, **115**(9), 1363-1370, 2007.
45. Zandbergen, P. A. : Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads. *BMC Public Health,* **7**, 37, 2007.
46. Maantay, J. : Mapping environmental injustices: pitfalls and potential of geographic information systems in assessing environmental health and equity. *Environmental Health Perspectives*, **110**, 161-171, 2002.
47. Briggs, D., Abellan, J. J. and Fecht, D. : Environmental inequity in England: small area associations between socio-economic status and environmental pollution. *Social Sciences and Medicine,* **67**(10),

1612-1629, 2008.
48. Morris, M. S. and Knorr, R. S. : Adult leukemia and proximity-based surrogates for exposure to Pilgrim plant's nuclear emissions. *Archives of Environmental Health*, **51**(4), 266-274, 1996.
49. McLaughlin, J. R., Clarke, E. A., Nishri, E. D. and Anderson, T. W. : Childhood leukemia in the vicinity of Canadian nuclear facilities. *Cancer Causes and Control*, **4**(1), 51-58, 1993.
50. Suarez, L., Brender, J. D., Langlois, P. H., Zhan, F. B. and Moody, K. : Maternal exposures to hazardous waste sites and industrial facilities and risk of neural tube defects in offspring. *Annals of Epidemiology,* **17**(10), 772-777, 2007.
51. Croen, L. A., Shaw, G. M., Sanbonmatsu, L., Selvin, S. and Buffler, P. A. : Maternal residential proximity to hazardous waste sites and risk for selected congenital malformations. *Epidemiology*, **8**(4), 347-354, 1997.
52. Elliott, P., Briggs, D., Morris, S., de Hoogh, C., Hurt, C., Jensen, T. K., Maitland, I., Richardson, S., Wakefield, J. and Jarup, L. : Risk of adverse birth outcomes in populations living near landfill sites. *BMJ*, **323**, 363-368, 2001.
53. Dolk, H., Vrijheid, M., Armstrong, B., Abramsky, L., Bianchi, F., Garne, E., Nelen, V., Robert, E., Scott, J. E., Stone, D. and Tenconi, R. : Risk of congenital anomalies near hazardous-waste landfill sites in Europe: the EUROHAZCON study. *Lancet*, **352**, 423-427, 1998.
54. Tynes, T. and Haldorsen, T. : Electromagnetic fields and cancer in children residing near Norwegian high-voltage power lines. *American Journal of Epidemiology,* **145**, 219-226, 1997.
55. Feychting, M. and Ahlbom, A. : Magnetic fields and cancer in children residing near Swedish high voltage power lines. *American Journal of Epidemiology*, **138**, 467-481, 1993.
56. Wartenberg, D., Greenberg, M. and Lathrop, R. : Identification and characterization of populations living near high-voltage transmission lines: a pilot study. *Environmental Health Perspectives,* **101**, 626-632, 1993.
57. English, P., Neutra, R., Scalf, R., Sullivan, M., Waller, L. and Zhu, L. : Examining associations between childhood asthma and traffic flow using a geographic information system. *Environmental Health Perspectives*, **107**(9), 761-767, 1999.
58. Wilhelm, M. and Ritz, B. : Residential proximity to traffic and adverse birth outcomes in Los Angeles county, California, 1994-1996. *Environmental Health Perspectives,* **111**(2), 207-216, 2003.
59. Reynolds, P., Elkin, E., Scalf, R., Von Behren, J. and Neutra, R. R. : A case-control pilot study of traffic exposures and early childhood leukemia using a geographic information system. *Bioelectromagnetics*, **Suppl 5**, S58-68, 2001.
60. Gilbert, N. L., Woodhouse, S., Stieb, D. M. and Brook, J. R. : Ambient nitrogen dioxide and distance from a major highway. *The Science of the Total Environment*, **312**, 43-46, 2003.
61. Briggs, D. J., de Hoogh, C., Gulliver, J., Wills. J., Elliott, P., Kingham, S. and Smallbone, K. : A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. *The Science of the Total Environment,* **253**, 151-167, 2000.
62. Yu, C. L., Wang, S. F., Pan, P. C., Wu, M. T., Ho, C. K., Smith, T. J., Li, Y., Pothier, L. and Christiani, D. C. : Kaohsiung Leukemia Research Group. : Residential exposure to petrochemicals and the risk of leukemia: using geographic information system tools to estimate individual-level residential exposure. *American Journal of Epidemiology*, **164**(3), 200-207, 2006.
63. Dearwent, S. M., Jacobs, R. R. and Halbert, J. B. : Locational uncertainty in georeferencing public health datasets. *Journal of Exposure Analysis and Environmental Epidemiology*, **11**(4), 329-334, 2001.
64. McElroy, J. A., Remington, P. L., Trentham-Dietz, A., Robert, S. A. and Newcomb, P. A. : Geocoding addresses from a large population-based study: lessons learned. *Epidemiology*, **14**(4), 399-407, 2003.