地理學論叢 제44호 (2004. 9)

우리나라 소아암의 공간적 분포에 관한 연구
-군집성을 중심으로-

이 미 혜*

A Study on Distribution of Childhood Cancer in Korea

Lee, Mi-Hye

요약 : 본 논문은 우리나라의 소아암의 공간적 분포를 군집성 여부를 중심으로 하여 살펴보는 것을 목적으로 한다. 이를 위해 1993년부터 1999년까지 종합병원급에서 진단된 소아암을 대상으로, 그 전국적인 분포를 1995년 기준의 읍면동 단위로 살펴보았다. 자료를 분석함에 있어 인구요소에 따른 비례적인 발병을 제외하고, 암 자료가 가진 적은 양의 발병수를 고려하기 위해, 확률지도와 베이즈 추정법을 이용하여 우리나라의 소아암의 분포도를 작성하였다. 또한 소아암의 군집 여부를 알아보기 위해 베삭과 뉴웰의 방법론과 턴불의 방법론, 국지적 모란 지수를 통해, 우리나라의 소아암의 군집성 여부를 검증하고 그 위치를 통계적으로 탐색하였다. 이에 따라 우리나라의 소아암은 전반적으로 경상북도와 강원도 지역에서 높게 나타났으며, 전라북도 지역은 상대적으로 낮게 나타났다. 군집의 위치는 부산시 해운대구 일대와 경기도 고양시 일대 지역을 중심으로 전국 각지에서 소아암 군집이 관찰되었다.

주요어 : 보건지리학, 공간통계, GIS, 군집탐색, 소아암, 암 분포

Abstract : This study aimed to explore the distribution of childhood cancer in Korea and examine whether there are clusters in Korea. Data is diagnosis on childhood cancer-leukemia, brain tumor, lymphomas-at general hospital from 1993-1999, whole country. Analyses are applied through probability mapping, Bayes estimation, Besag and Newell' s method, Turnbull' s method and loca morna' s I. Results are there are high incidence rates in Kyungsang buk-do and Kangwondo and there are clusters in study area. The clustered areas are discovered.

Key words : medical geography, cluster anlysis, epidemiology, spatial epidemiology, GIS, childhood cancer, cancer distribution

* 서울대학교 대학원 지리학과 석사과정 졸업

1. 서론

1) 연구배경과 목적

선진국은 물론 우리나라에서도 암은 소아사망의 주요 원인이 되고 있다(신상만 외, 1988). 따라서 소아암에 대한 질병 모니터링이 요구되지만 이에 대한 준비는 미비한 실정이다. 질병을 가진 인구는 지리적으로 불균등하게 분포하기 때문에 인구의 공간적인 분포는 역학에서 중요하게 다루어져야 할 요소이다. "질병, 사망, 출산 등 여러 가지 역학적 현상의 지리적 분포는 원인에 관한 가설을 세운다거나, 지역사회진단 또는 관리사업을 계획하는데 매우 중요한 역할을 한다(김정순, 1998)". 따라서 본 연구의 목적은 국내 소아암 발생의 공간적인 분포를 지도화하며, 그 공간적 군집여부를 알아보는 데 있다.

이를 위해, 본 연구에서는 먼저, 소아암의 분포를 살펴보기 위해서 소아암 자료 자체에 대한 기술적인 요약을 실시하고, 이러한 소아암의 분포를 발생률과 표준화이환비(Standardized Morbidity Ratio), 상대위험도(Relative Risk) 등의 역학적 지수와 관련된 단계구분도를 작성하여, 소아암의 분포를 살펴본다. 그러나 기존 역학에서 사용된 방법들은 공간적인 이질성과 지리적 군집의 가능성을 찾기에 부족하다. 또한 질병의 발생률은 그 인구구성 뿐만 아니라 모집단의 크기 자체에도 영향을 받는다. 무엇보다, 작은 단위지역에서의 질병 관찰값(Small Numbers Problem)은 값이 적은 경우(질병이 드문 경우), 값의 편차가 매우 심하기 때문에, 값의 신뢰도가 낮다. 따라서 이러한 인구요소를 제외하고, 소아암 자체만의 분포를 안정적으로 관찰할 수 있는 방법에 대해 탐색해 본다.

마지막으로 1993년부터 1999년까지 발생한 우리나라의 소아암에 대한 공간적인 임의성(군집성)을 검정한다. 즉, 소아암의 발생이 공간적으로 군집되어 있는지 여부와 군집을 이루고 있다면, 그 군집의 위치가 어디인지를 탐색하는 작업을 수행한다. 이를 위해 전역적 군집(Global Clustering)과 지역적 군집(Local Clustering) 검정을 실시한다.

2)연구 대상 및 연구 방법

본 연구의 대상 지역은 전국이며, 사용되는 소아암 자료는 진단일을 기준으로 한 암발생 자료이다. 발생의 대상 시점은 1993년부터 1999년까지로, 해당 기간 동안의 의료보험청구자료 중 대학병원의 의무기록자료와 비교하여 소아암으로 진단받은 환자들을 대상으로 하였다. 이상의 과정과 자료 가공 단계를 거쳐 읍면동 단위까지의 주소가 확인된 자료 3,333건을 대상으로 분석하였다. 환자군의 모집단으로는 환자 발생 기간의 중간 시점인 1995년 인구주택총조사의 0세부터 14세까지의 인구를 대상으로 하였다.

본 논문에서 사용된 소프트웨어는 자료 가공과 이의 시각적인 제시를 위해 ArcView 3.1과 ArGIS 8.3을 사용하였으며, 자료의 통계적인 분석을 위해서는 S-plus 6.2와 ClusterSeer2를 사용하였다.

2. 공간역학

공간역학은 보건지리(medical geography)의 한 분야로 질병의 공간적인 변이를 기술하고 그 원인을 이해하는 데 목적을 가진다. 공간역학의 분야는 주제면에서 질병지도화, 지리적 상관성연구, 점원

이나 선형원과 관련된 위험평가, 질병 군집탐색으로 나뉜다(Elliott 등, 2000). 한편, Kulldorff(1998)는 다음과 같은 기준으로 공간역학의 연구주제를 세분하였다(〈표-1〉).

〈표-1〉 공간역학의 제분야

시각화분야	공간분석분야
질병지도화를 위한 기술적인 방법 지도 평활화 기법	공간 회귀 공간-역학 모형화 소지역 추정 공간적 임의성 검정

Kulldorff, 1998

시각화 분야에서 질병의 지도화는 자료에 대한 시각화이자 분포 패턴을 알려주는 첫시작점이다. 이를 위해 확률 지도나 경험적 베이지언 평활화 지도를 생성할 수 있다.

공간분석분야에서 공간적인 임의성 검정은 연구지역을 통틀어 상대위험도는 일정하며 각각의 질병은 독립적으로 일어난다는 귀무가설을 검증하는 것이다.

이러한 귀무가설을 검증하는 방법론들은 크게 포아송 확률모형을 이용하거나 베르누이 확률 모형에 기초하고 있다. 어떤 방법론들은 상황에 따라서 두 모형 중 하나를 선택해 분석을 실시할 수 있다. 포아송 모형에서는 각각의 지역의 질병수는 포아송 분포를 따르는 귀무가설에 따라 생성되며, 각각의 지역의 질병수의 초과는 인구 규모의 비율이나 연인원으로 주어진다. 베르누이 모형에서는 이진변수로 환자/대조군이 나타내지며 일정한 수의 개인이 고려된다. 베르누이 모형에 기반한 방법론들은 모든 개인의 위치가 필요하고 주어진 위치에서 환자의 임의적인 분포를 검증한다.

임의성 검정은 공간적인 군집성 검정과 동일한 의미를 가진다. '군집'은 일반적으로 작은 지역 내에 특정 질병의 식별할 수 있는 집합체라고 정의되어진다(Sankoh and Becher, 2002).

즉, 질병 수가 초과발생한다고 하는 기준은 어디까지인가와 그러한 초과량은 과연 우연적인 결과인 것인가, 아니면 유의미한 결과인가를 통계적으로 검증하는 것이다. 군집성 검정은 공간, 시간, 시공간 단위로 이루어 질 수 있으며, 이 중 공간만을 위한 군집성 검정의 연구유형은 다음과 같이 구분할 수 있다(Kulldorff, 1998).

- 질병 군집(disease clustering)이 존재하는지 여부에 대한 탐색
- 군집 탐색
- 특정 점이나 선 주위에 상대위험도가 증가하는지에 대한 탐색

첫 번째 유형의 문제를 다루는 방법은 전역적 군집화(Global Clustering)이라 불리며, 연구지역을 통틀어 전체적으로 질병이 모이는 과정이 존재하는지 검증하는 것이다. 이것은 군집의 위치와는 상관없이 질병 근처에 다른 질병의 위험도가 증가할 때 발생한다.

따라서 전역적 분석은 군집의 위치에 관련한 추론은 이루어지지 않는다. 군집의 위치에 대한 검정을 위해서는 지역적 군집(Local Clustering) 탐색을 실시한다.

마지막 유형의 문제를 다루는 방법은 보통 초점화 검정(focused test)이라 불리며, 이는 어떤 특정 지점과 관련하여 그 지점 주변에 지역적인 군집이 존재하는지에 대한 가설을 다룬다.

이상의 방법론들에 대해 Kulldorff는 이들 방법론

들이 주변 효과(Edege Effect)[2] 에 대해 안정적인 방법론이라고 소개하였다.

1) 전역적 군집 검정

전역적 군집 검정을 위한 주요 방법론으로는 Moran's I(1950), Ripley's K Function(1980), Besagn and Newell(1991), Whittemore(1987), k-NN(Cuzick & Edwards, 1990), Tango, Mantel & Bailer(1970), Esseen's Test 등이 있다.

Moran's I와 리플리의 K 함수는 개별적인 발생율을 다루는 함수지만, 지역적으로 합산된 자료 역시 다룬다.

Besag과 Newell(1991)이 고안한 전역적 군집 방법은 포아송 분포를 기초로 하고 임의적이지 않은 질병의 군집을 조사하기 위한 것이다. 이를 위해 각각의 지역에 차례대로 원형의 권역(window)을 중심에 놓고 질병수가 사용자가 정의하는 질병 임계점인 k에 다다를 때까지 권역을 확장시킨다.

k가 임계치에 다다르면, 이 때 권역 안에 들어간 인구수를 이러한 질병 빈도하에 평균적으로 예상되는 인구수와 비교한다. Besag과 Newell의 귀무가설은 "연구지역 내 질병의 수는 포아송분포를 따른다"는 것이며, 연구가설은 "몇몇 지역에서 질병 수가 초과한다"는 것이다.

따라서, 귀무가설 아래에서 권역 내의 인구수는 질병의 수와 비례하여야 하며, 그렇지 않다면 귀무가설은 기각된다.

Besag과 Newell의 방법론에서 검정 통계량은 I과 r이다. I은 지역적 수준의 군집이며 r은 단순히 유의한 I의 수로 I를 통해 전역적 군집 여부를 판단한다. 검정 통계량 I는 아래의 식을 통해 도출된다.

$$P(L \leq l) = 1 - \sum_{x=0}^{k-1} \frac{e^{\lambda} \lambda^{x}}{x!}$$

즉, l이 귀무가설 아래서의 값 L에 이르거나 L보다 클 확률이다. 즉, L보다 l이 작을 확률을 1에서 빼서 계산된다. λ는 전체 지역의 평균 질병수이고, e는 지수 함수이다. 얻어진 통계량은 몬테-칼로 임의화법(Monte Carlo Randomization)으로 자료를 임의화하여 얻어진 참조 분포와 비교하여 유의성을 검정한다.

2) 지역적 군집 검정

지역적 군집성을 검정하는 방법론으로는 Local Moran's I, Turnbull의 방법론, 공간적 탐지(Spatial scan; Kulldorff 1997) 등이 있다.

Turnbull의 방법은 질병의 군집을 위해 연구 지역 내의 인구를 합산한다. 원형의 권역(window)이 각각의 지역에 차례대로 놓아지고, 그 안의 인구가 사용자가 지정한 한도(R)까지 도달하도록, 주변 지역을 포섭해가며 확대된다. 이러한 원형 창은 중첩될 수 있기 때문에, 창 안의 수는 독립적이지 않다. Turnbull의 방법의 귀무가설은 "일정한 인구 영역 안의 질병수는 공통적인 비율을 가진 포아송 분포를 따르지만, 영역이 중첩될 때는 통계적으로 독립적이지 않다." 이다. 이에 대한 연구가설은 "일정한 인구 영역 안의 질병수는 공통적인 비율을 가진 포아송 분포로 예측되는 질병수를 초과한다." 이다.

2) 주변효과(Edge Effect)는 연구지역의 경계지역, 즉 지도 상의 경계지역에 해당하는 값들은 연구지역 밖에 존재할 수 있는 다른 관찰값들에 안정적이지 않은 상태를 말한다. 이는 나아가 질병군집탐색에 있어서 연구지역내의 관찰값들만 고려함으로써, 군집의 형성을 왜곡할 수 있다.

Turnbull의 방법의 통계량 MR은 인구 크기 R을 가진 창 내에서 관찰된 최대질병수이다. 고정된 인구 크기를 가진 원형 권역들은 각각의 위치(주로 해당 지역의 중심점)를 방문함으로써 만들어지고, 창 안의 합산된 인구가 R에 도달할 때까지 주변 위치를 포섭해간다. 이 창 안에서 발생한 질병수는 합쳐진 지역 안에서의 모든 질병수의 합이다. 가장 먼 지역의 경우, 창 안에 해당 지역의 일부분만이 포함된다면, 이와 비례하여 해당 지역의 질병수의 일부분만 포함된다.

MR의 유의수준은 몬테-칼로 임의화(Monte Carlo Randomization)를 통해 경험적으로 발견된다.

참조 분포는 상대적인 지역에 국한된 인구 크기로부터 추정된 다항분포에 기초하여 위험인구(population at risk) 사이의 질병을 임의적으로 분포시킴으로써 생성된다.

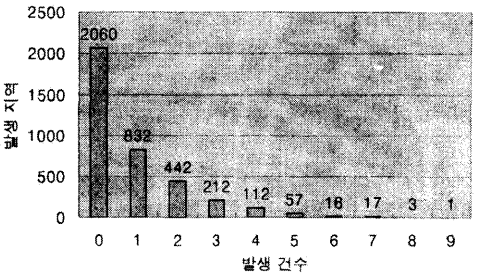
3. 우리나라의 소아암의 분포

1) 소아암 발생률

(1) 소아암 발생률에 대한 기술통계

1993년부터 1999년까지 전국의 소아암은 전체 95년의 0~14세 인구 10,235,504중 3,333 건이 발생하였다. 즉, 자료상에서 인구 1만명당 약 3.3건이 발생한 꼴로, 0~14세 인구에서 소아암의 발생율은 매우 낮다.

소아암 발생을 지역적으로 개괄해보면, 소아암 발생의 기준 시점으로 설정한 95년의 전국의 읍면동은 3,752개 지역이다. 이는 평균적으로 한 지역당 0.9건의 소아암이 발생한 수치이나, 이 중, 55%에 해당하는 2,060개 지역에서는 93년부터 99년, 6년의 기간 동안 소아암이 한 건도 발생하지 않았다. 또한, 전체 발생건수, 3,333건 중 1~3건이 발생한 지역은 1,468지역으로 소아암 발생 지역 2,060개 지역의 87%를 차지하고 있다. 이러한 발생은 한 건이 올라갈수록 거의 반 비례적으로 증가하고 있다. 이러한 자료의 범위는 최대값 9건까지 이른다.



〈그림-1〉 소아암 발생 지역

이는 소아암이 지역적으로도 매우 드문 질병임을 나타내는 동시에 나머지 다른 발생건수의 지역적 집중을 의심하게 한다.

〈표-2〉 소아암 발생지역의 기술통계치

최소값	0
1사분위수	0
평균	0.888326
중위수	0
3사분위수	1
최대값	9 (서울시 노원구 상계 8동)
분산	1.72
표준편차	1.31
왜도	1.9
첨도	4.2

실제로 소아암 건수의 왜도(비대칭도)는 3.52로 소아암 자료가 적은 값을 향해 치우쳐 있음을 알 수 있다. 다음의 〈표-2〉은 소아암 발생에 대한 기

술통계값이다.

그러나 이러한 발생건수는 해당 지역의 인구의 영향을 받기 때문에, 큰 값에 대한 왜곡을 확정적으로 진단할 수 없다. 최대값을 나타내는 송파구 상계 8동의 경우, 전체 0~14세 인구 순위 중 1% 안에 드는 지역으로, 상계 8동의 소아암 발생이 그 지역의 인구에 비례함을 부인하기 어렵다.

질병자료에서 인구요소를 제외하는 가장 기본적인 방법은 그것을 전체 모집단으로 나눠 상대적인 비율로 표시하는 것이다. 즉, 연구 대상의 모집단을 분모로 설정하는 것이다.

이는 질병의 기본 발생 지표인 발생율로 대상기간 동안에 새로 발생한 질병을 위험인구로 나눠 단위인구를 곱한 것이다. 발생율의 수식은 아래와 같다.

〈표-3〉 소아암 발생률의 기술통계치

최소값	0.00
1사분위수	0.00
평균	3.5
중위수	0.00
3사분위수	4.7
최대값	250.0
표준편차	8.11
왜도	11.93
첨도	276.50

로 떨어졌다. 이는 도수분포표의 확률밀도함수인 좌하단의 밀도그림에서도 확인 할 수 있다. 밀도그림에서 0~5개로 떨어질 확률이 매우 높음을 알 수 있다.

$$\text{발생률(Incidence rate)} = \frac{\text{일정기간위험에폭로된인구중 새로발병한환자수}}{\text{일정기간위험에폭로된중앙인구(population at risk)}} \times 10^{x} \text{ (단위인구)}$$

읍면동별 단위인구를 1만 명으로 설정하여 발생율에 대한 기술통계치를 산출한 결과, 왜도는 10으로 더더욱 커졌으며, 그 값의 범위 역시 그다지 줄어들지 않았다.

평균 역시 중위값을 훨씬 상회하고 있어, 전체 자료 중 높은 값을 향한 이상점이 존재함을 추론할 수 있다(〈표-3〉).

자료의 분포와 이상점을 시각적으로 탐색하기 위해, 도수분포표과 Boxplot, Density plot, QQ-plot과 누적 확률 분포도를 작성하였다.

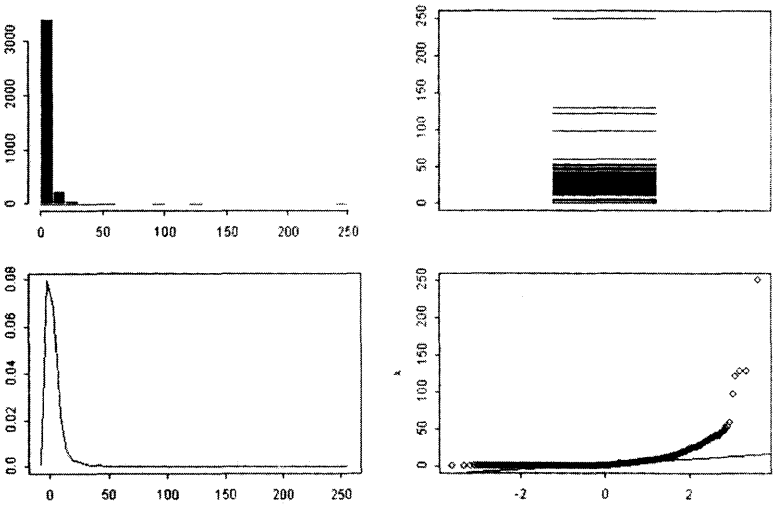
〈그림-2〉에서 좌상단은 소아암 자료에 대한 도수분포표이다. 값 간의 간격을 1로 놓고서 자료의 발생 빈도를 보았을 때, 대부분이 평균 5건의 이하

이는 소아암의 발생의 중심적인 경향이 0~5건에 분포함을 알 수 있다.

우상단의 Boxplot과 우하단의 QQ-plot은 소아암 발생의 이례적 값들을 감지하였다.

Boxplot의 몸체 부분을 통해, 소아암의 사분위수 간 거리가 매우 좁아 자료의 분포가 집중적인데 반해, 3사분위수로부터 매우 떨어진 이상점들이 대거 발견되었다.

특히, 상위의 5지역은 매우 극단적인 이상점 지역으로, 값의 크기 순으로 부산시 해운대구 좌동, 대전광역시 중구 대흥 1동, 경기도 고양시 송포동 경기도 고양시 화정동, 강원도 원주시 행구동의 5개 지역이었다. 정규성 함수를 이용해 그려진 QQ-



〈그림-2〉 소아암 발생률에 대한 시각적 탐색

plot에서도 자료에서 이상점이 발견되었다.

질병의 분포를 위한 이론적인 분포로 포아송 분포가 사용된다. 이는 질병의 발생이 매우 드문 사건이기 때문이다. 자료에서 관측되는 소아암의 발생률이 이상적인 포아송 분포를 따르는지를 확인하기 위해 소아암 발생률을 포아송 분포을 따르는 QQ-plot 및 누적확률분포곡선과 비교하였다.

〈그림-3〉의 상단은 소아암 분포대 포아송 분포의 QQ-plot이고, 하단은 소아암 분포와 포아송 분포의 누적확률함수에 대한 비교도이다. 소아암의 전체 평균 발생률 3.3을 포아송 분포의 모수값으로 주어 포아송 분포와 소아암 발생률을 QQ-plot을 통해 대비해본 결과, 소아암 발생은 약 4건을 대칭으로 그 이하 값과 이상 값 모두에서 포아송 분포를 따르지 않음이 드러났다.

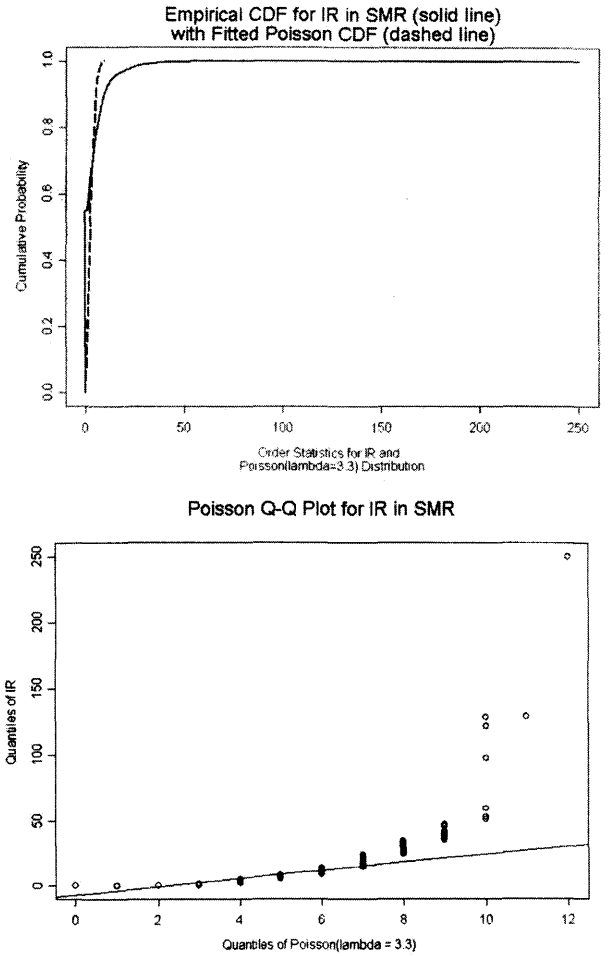
즉, 4건 이하에서는 소아암 발생률이 이론적인 포아송 분포로 예상되는 확률보다는 많은 수치로 발생하다, 4건을 기점으로 점점 소아암 발생률이 포아송 분포로 예측되는 건수보다 적게 발생함을 알 수 있다. 이는 소아암이 포아송 발생률로 예측되는 질병의 고른 분포에서 벗어나, 지역적으로 매우 불균등한 발생률을 보임을 나타낸다.

(2) 전국의 소아암 발생률의 분포

〈그림-4〉는 앞서의 〈그림-2〉 중 상자그림의 값을 지도화한 것이다. 상자그림에서 나타난 각 값의 범위와 이상점들을 지도화여, 각 지역별 이상지역들을 확인할 수 있다.

소아암의 전국의 발생률을 전국적으로 살펴봤을 때, 그 값들이 여전히 수도권, 대전, 대구, 부산 등 전통적으로 인구가 많은 지역들에서 높은 발생률을 보임을 알 수 있다. 이것이 인구에 영향을 받은 것인지, 아니면 대도시 지역이 가진 환경적 요인때문인지는 아직 알 수 없다.

자료를 탐색하기 위해서, 인구 크기의 차이를 상쇄해서 자료값의 비교하기 위해 수를 비율화하는 것은 두 가지 면에서 충분하지 않다(Haining, 2003). 첫번째 이유는 적은 인구에 기반한 비율은



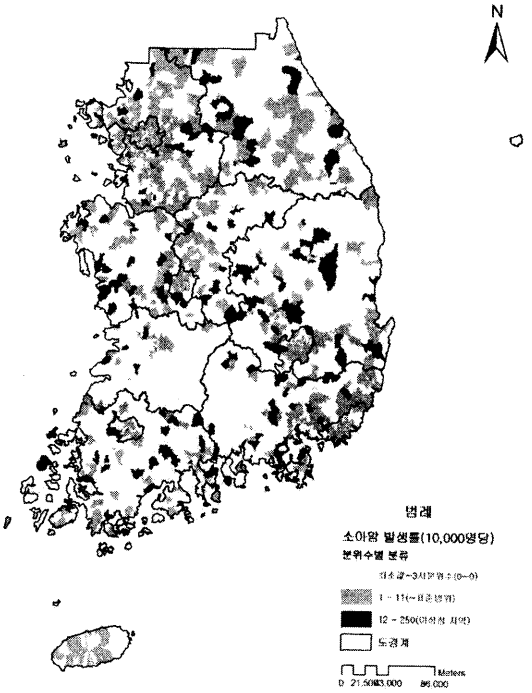
〈그림-3〉 소아암 발생률과 포아송 분포

큰 인구로부터 계산된 비율에 비해 자료 오류에 덜 안정적이라는 것이다. 자료 수집의 오류 등으로 인해 질병 건수를 더하거나 빼내는 것은 분모인 인구가 클 때보다 작을 때 더 큰 영향을 줄 수 있다.

자료의 평균화에 있어서 비율법이 충분하지 않은 두 번째 이유는 위해 작은 분모(인구)로 계산된 비율들은 큰 분모(인구)로 계산된 비율보다 큰 분산을 가지며, 비율값 자체 안에 평균-분산 의존성이 있다는 사실이다. 따라서, 발생률은 인구 요소를 효과적으로 제외시키는 척도가 될 수 없다. 지역별로 다양한 인구 크기의 변화를 제외시키고, 질병의 위험에 대해 더 안정적으로 예측할 수 있는 척도가 필요하다.

2) 소아암 발생에 대한 표준화이환비의 분포

질병의 발생률을 산출하는 데 있어서, 단순히 전체 인구와 전체 질병발생수간의 비율만을 고려한



〈그림-4〉 93~99년 우리나라의 소아암 발생률의 분포(사분위수별 분류)

조율(粗率: crude rate)은 질병발생의 특수성을 무시할 수 있다. 예를 들어, 어떤 질병은 특정 연령층에서 호발하며, 어떤 지역의 위험에 노출된 인구(population at risk)에 있어 이 연령층의 구성비율은 높을 수 있다. 따라서, 이러한 점을 고려하지 않은체, 전체 질병과 인구의 비율만으로 그 지역의 발생률을 산출하는 것은, 해당 지역의 질병발생정보를 왜곡시킬 수 있다[3].

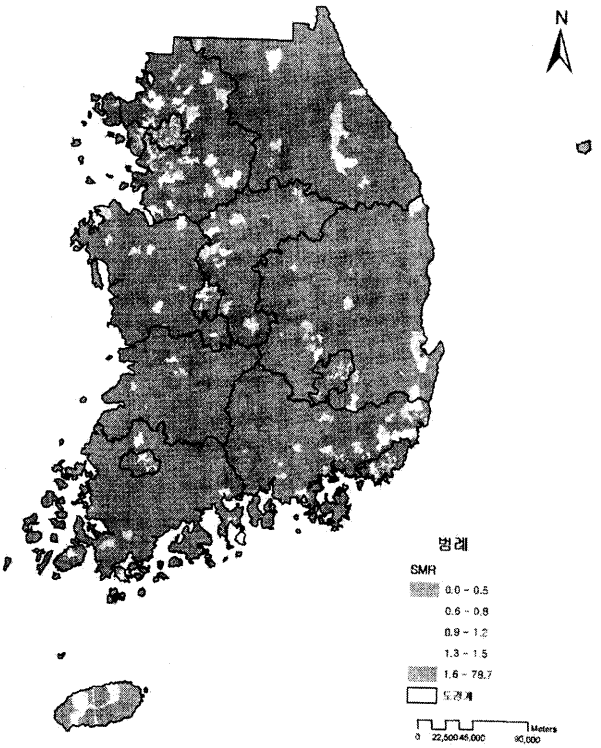
따라서, 해당지역의 인구구조를 고려한 발생률을 산출할 필요가 있다. 이에, 소아암 발생에 있어서 연령별로 표준화를 실시해 그 이환율을 평가하는 표준화 이환비(Standardazed Morbidity Ratio: SMR)를 산출하여 그 전국적인 분포를 살펴보았다.

SMR은 해당지역의 인구구조를 고려하여, 그 지역의 기대값을 산출하고, 이를 실제로 관측된 질병과 비교한다. SMR의 수식은 다음과 같다.

$$SMR(S\ tan\ dardizedMorbidity\ Ratio) = \frac{O}{E}$$

따라서, SMR의 값이 1 이상이면 예측치보다 과대발생을 의미하며, 1이하는 예측치보다 적게 발생했음을 의미한다. SMR 값이 1 이면, 실제 관측치와 예상치가 일치한다. SMR의 통해, 질병의 과대발생을 측정할 수 있으며, 인구구조의 보정을 통해

3) 여기서 말하는 '인구구조'는 연령 뿐만 아니라, 성비, 지역소득 등 생물학적, 사회적인 인구구조를 모두 포함한다.



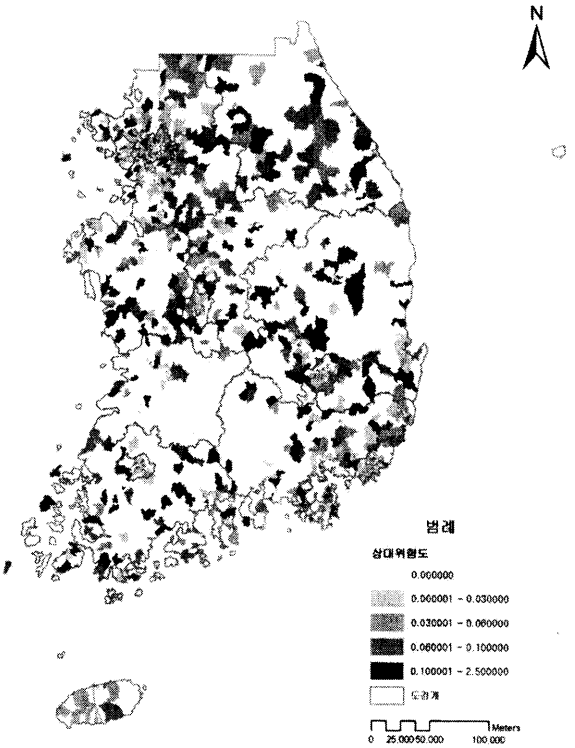
〈그림-5〉 93~99년 우리나라 소아암의 표준화 이환비의 분포

지역간의 비교가 가능해진다. 그러나 SMR은 그 값이 적은 관찰값(질병 건수)인지 아닌지를 고려하지 않는다(Jacquez and Greiling, 2003). 즉, 관찰값이 적음에 따라 덜 안정적이다, SMR의 안정성을 추정하여 유의한 군집을 찾아내기 위해, Wald 신뢰한계법(Wald confidence limits)을 사용하여, SMR의 신뢰구간을 계산하였다. 이 신뢰구간이 1을 포함하고 있으면, 그 값은 통계적으로 예측치와 다르지 않음을 의미한다.

SMR을 발생율과 비교해 봤을 때, 그러나 그 분포에 있어서 큰 차이는 나타나지 않고 있다. 이는 소아암이 특수 연령에 제한된 질병분포이기 때문이다. SMR에서 각 5세 인구의 발생율이 0~5세 0.00058, 5~9세 0.00061, 10~14세 0.00048로 큰 차이가 없기 때문이다.

3) 소아암 발생의 상대위험도의 분포

해당속성이 비율이나 백분율이라면, 지리적인 변화를 나타내기 위해 비율에 대해 탐 색적으로 지도화하는 것은 분석에 있어서 첫 번째 단계이다(Bailey and Gatrell, 1995). 그러나 앞서의 역학적 지수값들에서도 나타났듯, 질병비율의 변동은 관련된 인구크기의 변동과 함수 관계에 있고 이 값은 지역에 걸쳐 다양하게 변한다. 이 때문에, 이 비율값들은 잘못 해석될 수 있다. 이러한 문제를 해결



〈그림-6〉 소아암 발생의 상대위험도 [분위수별 분류]

하기 위해서 각 지역의 인구의 변동을 고려하면서 이례적인 지역을 부각시키기 위한 여러 방법들이 개발되었다.

비율과 관련하여 역학과 보건지리에서 일반적으로 사용하는 간단한 접근법 중 하나로 상대 위험도(Relative Risk[4])가 있다. 상대위험도는 지역 전체의 평균 발생률로 인해 해당 지역에 예측되는 값과 실제 발생수와의 비를 통해 측정될 수 있다. 이를 통해 그 지역의 전체 지역에 대한 상대적인 위험도를 추정할 수 있다.

상대위험도의 수식은 다음과 같다.

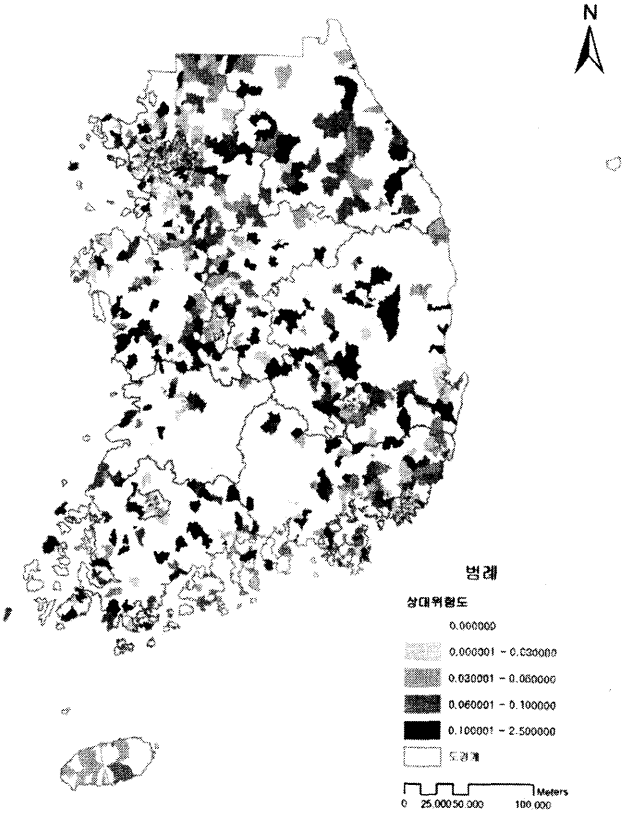
$$\text{상대위험도}(Relative\ Risk : PR) = \frac{O}{E} \times 100$$

$$E = n_i \left(\frac{\Sigma y_i}{\Sigma n_i} \right)$$

n_i , i지역의 인구수

y_i , i지역의 질병 발생수

4) 여기에서 사용하는 개념인 상대위험도는 역학의 코호트(cohort)연구에서 사용하는 지표중 하나인 상대위험도(Relative Risk)와는 그 수식에서 다소차이가 있다.



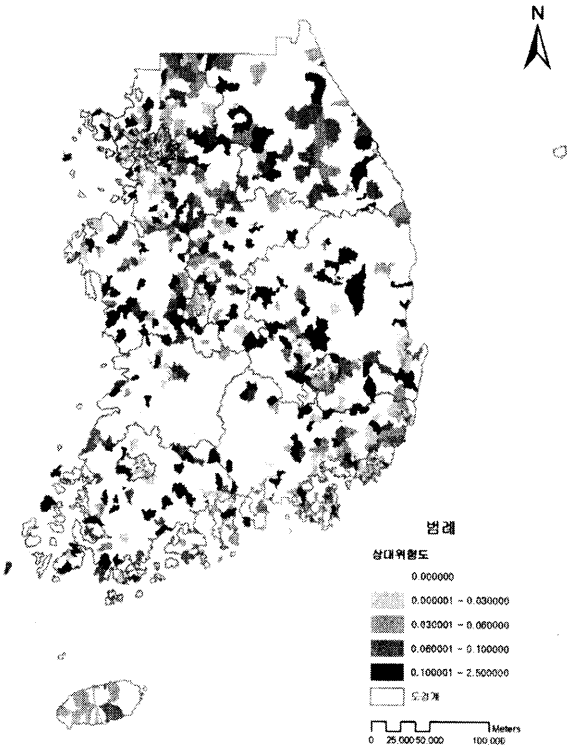
〈그림-7〉 소아암 발생의 확률값

4) 소아암의 발생에 대한 확률지도

앞서에서 발생수를 발생률로 치환하였을 때, 값의 변동값은 도리어 더 커졌다. 소아암 자료와 같이 드문 질병은 애초에 인구집단에 비해 매우 적은 수로 발생하기 때문에, 이러한 집계 하나하나에 따라 그 변동은 매우 크다. 따라서 숫자 하나의 차이로 해당 지역의 발생수는 과대평가될 수 있다. 이러한 문제를 '작은 수의 문제(Small Numbor Problem)' 라 한다. 수간의 변동에 신뢰도를 높이기 위해, 발생률이나 기대값에 대한 관찰값의 비율 대신 극단적인 값이 나타날 확률을 지도화할 수 있다. 확률지도에서는 질병 발생의 통계적인 유의성을 지도화한다. 통계적 유의성은 해당 지역에서 일반적으로 일어날 질병의 확률로부터 벗어난 정도를 통해 측정된다. 확률값은 해당 지역에서 일어난 질병이 우연적인 것일 확률로 질병의 분포의 가정에 쓰이는 포아송 분포를 이용해 산출된다.

$$P(x \geq k) = 1 - \sum_{x=0}^{k-1} P(x)$$

$$P(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$



〈그림-8〉 베이즈 추정에 의한 소아암의 표준화 이환비의 분포도

즉, 수 자체보다는 그 상대적인 위험도를 나타내기 때문에, 질병 분포에 있어서 지나치게 심한 변동을 줄일 수 있다.〈그림-8〉은 우리나라 소아암 발생에 대한 확률지도로, p-값이 유의한 지역(0.05 이하인 지역)과 아닌 지역을 2등분하여 작성하였다.

확률지도를 통해 여전히 서울과 대전, 부산 등의 대도시 지역과 중심으로 그 발생률이 높으며 그것이 발생할 확률은 매우 낮아, 이 지역의 높은 값이 유의한 것임을 알 수 있다. 또한 대도시 지역 외에 경상북도 지역 전체에 소아암의 분포가 높게 나타남을 알 수 있다. 이러한 낮은 확률값을 나타내는 읍면동은 밀집하여 있으며, 이 지역의 발생이 군집을 이루고 있는 것에 대한 의심을 가지게 한다.

그러나 확률지도는 그것이 전체 평균으로부터 극단적인 값인지를 나타내는 것이지 그 이웃들간의 값의 진정한 비교는 아니다. 또한 포아송분포 자체가 독립적이며 상호 독립적인 변수들을 위한 값이기 때문에, 확률지도상의 확률값은 각 지역의 값들 간의 차이라기 보다는 포아송 분포 모델에서 얼마나 적합하지 않은지를 측정한다 할 수 있다.

5) 경험적 베이즈 추정

경험적 베이즈 추정은 적은 관찰값 때문에 일어나는 수의 큰 변동을 전체 지역에 대해 가지고 있는 지식으로 평활화(smoothing)하는 지도평활 기

법이다. 이를 통해, 지역의 과대추정되거나 과소추정되는 값을 중심 경향값을 향해 조정해서, 좀 더 진정한 의미의 이상점을 찾을 수 있다.

이 때 사전분포로 이용되는 지식은 자료의 전역적 양상인 연구지역 전체의 질병비율이 이용된다. 본 논문에서 이용된 베이즈 추정량은 최대우도법을 사용하였다. 이 때, 두 모수의 추정량 $\hat{\phi}$와 $\hat{\gamma}$ 은 자료로부터 적률법을 이용하여 직접적으로 도출된다.

$$\hat{\theta} = \hat{\gamma} + \frac{\hat{\phi}(r - \hat{\gamma})}{(\hat{\phi} + \hat{\gamma}/n_i)}$$

$$\hat{\gamma} = \frac{\Sigma y_i}{\Sigma n_i}$$

$$\varphi = \frac{\Sigma n_i (r - \gamma)^2}{\Sigma n_i} - \frac{\gamma}{\bar{n}}$$

y_i = 질병발생수
n_i = 인구수

베이즈 추정을 통해 극단적인 값들을 조정한 결과 여전히 경상북도의 중부 지방이 전국적으로 높은 소아암 발생률을 보였다. 경북지역의 암발생률은 도경계를 넘어 경상남도로 이어져, 부산광역시까지 이어지고 있다. 이에 반해, 전라북도지역의 낮은 소아암 발생률은 주목할만하다. 이러한 전라북도 지방의 낮은 소아암 발생률은 도경계를 넘어 경상남도 지역까지 이어진다. 이상의 양상에 따라 경상남도 지역은 소아암 발생이 높은 지역과 낮은 지역으로 양분되는 경향을 보인다.

전국 소아암 지역의 또 다른 양상은 그 발생이 대도시를 중심으로 확산되는 경향을 보인다는 것이다. 특히 이러한 양상은 대구에서 가장 심하게 나타나며, 광주에서 가장 약한 특징을 보인다.

베이즈 추정에 있어서 관찰값이 많은 지역은 평균값에 기여를 많이 하기 때문에, 평균으로의 이동을 적게 하고, 관찰값이 적은 지역은 그 값에 신뢰성을 가질 수 없기 때문에, 평균값을 향해 평활화 정도가 커진다.

즉, 관찰값이 많은 대도시 지역은 그 값에 조정이 적고, 관찰값의 변동이 큰 농촌 지역은 그 조정의 폭이 커진다. 그러나 이는 지역적인 차별라기 보다는 관찰값의 분포의 차이이다. 또한 값의 조정 문제 뿐만 아니라, 일개 연구 지역 전체 평균으로 조정하는 것은 비공간적이다(Gaterell, 1995). 따라서 대도시 지역의 기여도를 줄이고 베이즈 추정량을 좀 더 공간적으로 산출하기 위해서는 그 조정값을 연구지역 전체의 값으로 하기 보다는 인접지역을 고려하는 모수 측정이 필요하다.

4. 소아암의 공간적 군집 탐색

1) 전역적 군집 탐색

(1) 베삭과 뉴웰의 방법

전역적 군집검정을 위해 귀무가설을 설정하였다.

Ho : 한 읍면동의 소아암수는 동일한 비율을 가진채 포아송 분포를 따른다.

Ha : 어떤 읍면동 지역의 소아암수는 포아송 분포로 예측되는 값보다 초과 발생한다.

이상의 가설을 검증하기 위해 Besag and Newell의 방법을 사용하여 전역적 군집성을 검증하였다. Besag and Newell의 방법은 연구 지역 내 일정한 전체 평균 가진 포아송 분포를 가정하기 때문에,

〈표-3〉 K값의 변화에 따른 검정력의 변화

	K=10	K=12	K=15	K=18
R값	166	180	206	258
기대값	106	111	116	121
p-값	0.004	0.004	0.003	0.001

전체 0~14세 인구, 10,235,504명 중 총 발생건수 3,333건을 통해, 평균 위험도 0.000325631가 산출되었다. 임계치(k)를 달리하여 각각의 검정력의 변화를 살펴본 결과는 다음과 같았다.

k값을 10으로 정하기 시작했을 때부터 전체 지역에 대한 적절한 검정력이 생기기 시작하였다. 이때, 유의한 군집의 수(r)는 166으로 나와 귀무가설 아래의 예측값 106을 상회하였다.

이 값이 나올 확률(p-값)은 귀무가설 하에서 0.004로 지역 내에 군집이 존재하지 않는다는 귀무가설을 기각하게 하였다. 이후의 k값 역시 계속해서 귀무가설을 기각하는 결과값들이 나왔다.

즉, 전국에 소아암의 군집이 존재함을 지지하고 있다.

2) 지역적 군집 탐색

(1) 턴불의 방법

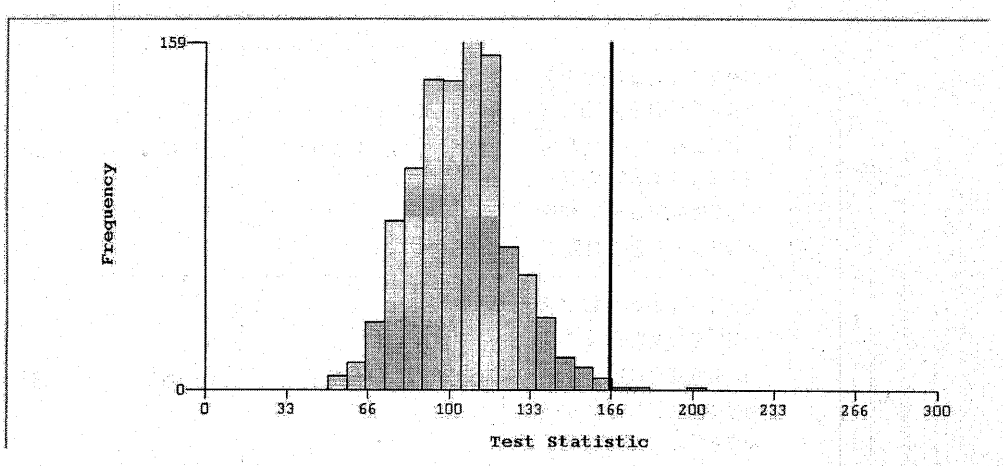
앞서의 전역적 군집 검정에서 전국에 소아암의 군집이 존재함이 도출되었기 때문에, 이에 따라 소아암 군집의 위치를 확인하는 검정을 실시하였다. 이것을 입증하기 위해 Turnbull의 방법을 이용하였다.

Turnbull의 방법론을 적용하기에 앞서 귀무가설을 설정하였다.

Ho : 동일한 인구를 가진 지역의 소아암의 수는 공통적인 비율을 가진 포아송 분포를 따른다.

Ha: 동일한 인구를 가진 지역의 소아암의 수는 공통적인 비율을 가진 포아송 분포로 예측되는 수보다 초과한다.

Turnbull의 방법에서 중요한 것은 권역을 확장하는 기준인 인구크기의 설정이다. 이러한 인구크기를 설정하는 방법은 앞서의 전역적 군집을 형성하



〈그림-9〉 K=10일 때의 몬테-칼로 임의화와 통계량

는 k값에 도달하기까지의 일반적인 인구크기를 고려하는 것이다. 연구지역 전체의 평균 0.000325631건으로 봤을 때, 질병군집이 10건까지 도달하기 위해서는 평균적으로 30,709명이 필요하다. 이러한 평균인구가 원형 권역에 포함될 때까지 포섭된 지역 내의 소아암수의 유의성을 검정한 결과는 다음과 같다.

인구를 30,000명으로 설정하여 유의 수준 0.05로 하여 검정한 결과, 군집이라 추정되는 세 지역이 선정되었으나, 이 값들은 모두 p-값이 0.2 이상으로 나와 통계적으로 유의하지 않았다.

(2) 국지적 모란 지수

군집을 측정의 방법론들을 분류하는 또 하나의 방법은 공간적 상관성의 측정기준을 거리 기반으로 두는가, 아니면, 두 지역 간 경계선의 공유 유무로 두는가로 볼 수 있다.

이 중 Turnbull의 방법이 중심점간의 거리를 통해 군집성을 측정했다면 국지적 모란 지수는 각 지역간의 인접성(경계선 공유 여부)을 통해 군집성을 측정한다. 모란 지수를 바탕으로 군집성을 측정한 결과는 〈표-4〉와 같다.

그러나 국지적 모란지수를 측정하기 위해서는 모든 지역을 방문한다. 따라서, "필연적으로 인접 지역의 LISA와 상관 관계를 갖게 마련이다. Ord와 Get is (1994)의 연구에서도 증명되었듯이, 연구 지역들이 근본적으로 밀집되어 분포하거나 이웃으로 정의된 지역이 많을 경우, 각 지역의 상관 지수

〈표-4〉 지역적 군집의 순위 지역

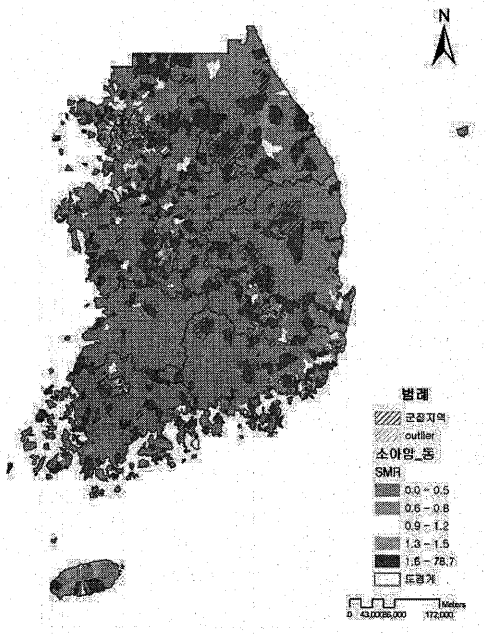
P=0.05

순위지역	해당지역	질병빈도	통계량	p-값
1순위 지역	경기도 군포시 금정동 경기도 군포시 산본1동 경기도 군포시 산본2동 경기도 군포시 재궁동 경기도 안양시 호계1동 경기도 안양시 호계3동	0.000766667	23.0000	0.232
2순위 지역	서울특별시 강동구 성내1동 서울특별시 강동구 성내2동 서울특별시 강동구 성내3동 서울특별시 송파구 오륜동 서울특별시 강동구 둔촌1동 서울특별시 강동구 천호3동	0.000753529	22.6059	0.279
3순위 지역	서울특별시 송파구 풍납2동 서울특별시 송파구 풍납1동 서울특별시 송파구 잠실4동 서울특별시 광진구 구의3동 서울특별시 강동구 성내1동 서울특별시 송파구 잠실6동	0.000733333	22.0000	0.485

는 근본적으로 상관 구조를 갖게 되며 이 결과 상관 지수에 대한 유의성은 필연적으로 오차를 수반한다(유은혜, 1999)."

모란 지수의 유의성을 측정하기 위해, 초기 유의수준을 0.05로 설정하였으나, 여러지역을 반복적으로 방문하기 때문에 발생하는 오류를 피하기위해 유의수준에 있어 본페로노이 보정(Bonferroni adjustment = 0.008757)과 시닥보정(Sidak adjustment = 0.008944)을 실시한 결과 새로운 유의수준 0.008이 도출되었다. 이 새로운 유의수준 이하에서 정규분포로 가정했을 때, 국지적 모란 지수 값이 유의한 지역 중 군집지역과 공간적 이상점 지역을 선정한 결과는 다음과 같다.

모란 지수가 양의 값을 나타내 공간적 자기상관(군집)을 나타내는 지역들은 주로 부산시 해운대구 일대가 주류를 이루며, 이 중 부산광역시 해운대구 좌동은 SMR의 값 역시 유의하게 다른 지역보다 높아, 이 지역에서 유의하게 높은 발생률이 군집을 이루고 있음을 발견할 수 있다.



〈그림-10〉 소아암 군집지역

턴불의 방법과 모란 지수를 이용해 군집성을 측정한 결과 두 방법의 결과는 같지 않았다. 두 가지 방법론에서 차이가 난 것은, 두 방법이 지역간의 상관성을 측정하는데 있어서, 서로 다른 측정법을 사용하기 때문인 것으로 여겨진다. 지리적인 관계를 어떻게 측정하느냐는 지리적 변이의 패턴에 대한 기술과 군집 탐색과 바로 연결된다(Jacquez and Greiling, 2002). 즉, 주어진 기법은 그것이 탐지하도록 의도된 군집을 탐지한다. 따라서, 한 지역의 공간적 군집을 탐색하기 위해서는 한 가지 군집 탐색 기법만을 사용해서 그 지역의 군집여부를 있다 없다로 단정짓기 보다는, 여러 방법들을 통해 측정하여 그 의심도를 높여가는 것이 타당한 것으로 보여진다.

5. 결론

본 논문은 우리나라의 소아암의 분포를 1993년과 1999년 사이 진단자료를 바탕으로 전국의 읍면동 단위로 살펴보았다. 이를 위한 방법론으로는 탐색적 공간 자료분석(Exploratory Spatial Data Analysis)과 역학지수의 지도화, 질병발생자료에서 인구요소를 배제한 지도화 기법, 마지막으로 질병군집탐색법이 사용되었다.

먼저 소아암 진단자료의 속성에 대한 시각적 탐색을 통해 소아암 분포의 군집 가능성에 대한 단서를 포착하고, 이를 실제 역학지수를 통해 지도화하였다. 이러한 지도화 과정에서 크게 기존의 역학지수를 지도화(발생률과 표준화 이환율)하는 방법과

〈표-5〉 국지적 모란지수를 통한 군집

지역명	I값	P-값	상한값	하한값	SMR
부산광역시 해운대구 송정동	4.02	0.0000	15.2	0.3	2.1
대전광역시 동구 인동	1.02	0.0048	16.9	0.3	2.4
부산광역시 해운대구 중1동	2.49	0.0000	6.8	0.4	1.7
강원도 동해시 묵호동	1.86	0.0001	25.2	0.5	3.5
대전광역시 중구 대흥2동	2.70	0.0000	26.6	0.5	3.7
충청북도 제천시 남천동	1.40	0.0004	27.9	0.6	3.9
부산광역시 해운대구 우1동	1.05	0.0072	4.2	0.6	1.6
광주광역시 광산구 송정3동	1.22	0.0046	30.9	0.6	4.4
강원도 원주시 소초면	1.12	0.0009	10.0	0.6	2.5
서울특별시 마포구 상암동	0.82	0.0021	10.4	0.7	2.6
경기도 고양시 장항동	2.54	0.0000	33.6	0.7	4.7
강원도 강릉시 두산동	2.50	0.0000	34.3	0.7	4.8
경상남도 함안군 법수면	2.03	0.0000	34.9	0.7	4.9
경상북도 구미시 장천면	1.38	0.0001	35.1	0.7	4.9
경상남도 김해시 칠산동	2.45	0.0000	38.2	0.8	5.4
부산광역시 해운대구 반여1동	1.00	0.0006	4.8	0.8	2.0
부산광역시 해운대구 반송3동	3.57	0.0000	8.2	0.8	2.6
전라남도 여천시 삼일동	1.19	0.0056	14.3	0.9	3.6
충청북도 제천시 영전2동	2.36	0.0000	48.0	1.0	6.8
경상북도 안동시 도산면	1.66	0.0001	49.6	1.0	7.0
부산광역시 해운대구 반여2동	2.47	0.0000	6.0	1.0	2.5
경상북도 의성군 단북면	1.23	0.0042	55.4	1.1	7.8
강원도 태백시 화광동	1.58	0.0045	19.8	1.2	5.0
경상남도 거창군 마리면	2.30	0.0000	63.5	1.3	8.9
강원도 횡성군 강림면	8.06	0.0000	95.3	1.9	13.4
경기도 광주군 퇴촌면	2.06	0.0000	32.0	2.0	8.0
충청북도 제천시 명서동	2.02	0.0000	23.2	2.4	7.5
충청북도 제천시 화산2동	2.22	0.0000	30.8	3.2	9.9
광주광역시 광산구 평동출장소	1.77	0.0000	31.7	3.3	10.2
경상북도 안동시 녹전면	2.44	0.0000	57.7	3.6	14.4
강원도 원주시 행구동	9.70	0.0000	95.4	9.9	30.8
경기도 고양시 송포동	1.75	0.0000	86.8	17.5	39.0
부산광역시 해운대구 좌동	3.27	0.0000	165.1	37.5	78.7

〈표-6〉 이례지점 중 높은 발병률을 보이는 지역

지역명	I값	P-값	상한값	하한값	SMR
대전광역시 중구 선화2동	-1.65	0.0000	48.1	3.0	12.0
서울특별시 중구 소공동	-2.14	0.0000	115.0	2.3	16.2
경상남도 마산시 합포구 오	-1.22	0.0046	33.0	2.1	8.3
경상남도 진해시 태백동	-1.18	0.0026	21.5	3.0	8.1
경상남도 진해시 석동	-1.57	0.0011	41.5	2.6	10.4
경상남도 밀양시 청도면	-1.10	0.0012	65.7	1.3	9.3
경상남도 창녕군 장마면	-1.53	0.0004	83.3	1.7	11.7
경기도 시흥시 연성동	-1.03	0.0089	65.0	4.1	16.3
경상북도 포항시 북구 송라	-1.66	0.0029	43.4	2.7	10.9
경상북도 안동시 중구동	-2.97	0.0000	58.1	6.0	18.7
광주광역시 동구 산수동	-1.14	0.0081	37.6	3.9	12.1
강원도 춘천시 사북면	-1.96	0.0000	50.7	3.2	12.7
강원도 춘천시 죽림동	-1.17	0.0066	56.6	1.1	8.0
강원도 강릉시 저동	-1.66	0.0006	77.1	1.5	10.9
강원도 양양군 서면	-1.11	0.0005	42.5	2.7	10.6
전라남도 보성군 복내면	-1.10	0.0012	53.9	1.1	7.6
전라남도 장성군 동화면	-1.13	0.0084	55.6	1.1	7.8
충청남도 청양군 운곡면	-0.98	0.0069	49.2	1.0	6.9
충청북도 괴산군 사리면	-1.98	0.0000	51.2	3.2	12.8

〈표-7〉 이례지점 중 낮은 발병률을 보이는 지점

지역명	I값	P-값	상한값	하한값	SMR
대전광역시 동구 원동	-0.95	0.0088	0.0	0.0	0.0
대전광역시 중구 은행동	-1.30	0.0010	0.0	0.0	0.0
대전광역시 중구 선화동	-1.22	0.0008	0.0	0.0	0.0
대전광역시 중구 대흥3동	-1.03	0.0044	0.0	0.0	0.0
경기도 고양시 송산동	-0.95	0.0089	0.0	0.0	0.0
강원도 원주시 판부면	-0.81	0.0075	0.0	0.0	0.0
부산광역시 해운대구 우 2동	-1.71	0.0000	0.0	0.0	0.0
부산광역시 기장군 기장읍	-2.18	0.0000	0.0	0.0	0.0

질병 발생에 영향을 미치는 인구요소를 배제하는 지도화 기법들을 탐색하였다. 이 과정에서, 우리나라의 소아암의 분포는 농촌지역보다는 대도시 지역에 군집으로 집중되어 있는 경향을 보이며, 경상북도의 전체 일대에는 소아암의 분포가 고르게 높게 나타남을 알 수 있었다. 그러나, 아직까지 질병의 지도화 과정에 있어서, 작은 수의 문제와, 각 지역의 인구크기간의 변동으로 인한 수의 문제를 해결하기 위해서는 좀 더 지역적인 연관성을 반영할 수 있는 방법론이 필요하다. 마지막으로 지도화 과정에서 높게 나타나는 값들이 군집을 형성하는지 통계적으로 검정하기 위해 베삭과 뉴웰의 방법을 이용해 전역적 군집 여부를 확인하고, 그 위치를 찾아내기 위해 턴불의 방법을 사용하였으나, 통계적으로 유의하지 않았다. 이에 반해, 지역적 모란지수를 통해서는 경기도 고양시 일대와 부산시 해운대구 일대에서 소아암 군집을 발견할 수 있었다.

이 연구의 의의는 질병의 지리적 군집 분석에 있어서, 이러한 지리적 군집 분석이 분석역학(Analytical Epidemiology) 이전에 선행되어야 한다는 사실이다. 이미 미국 등의 선진국의 질병감시체계에서는 비용이 많이 들어가는 분석역학 이전에, 질병의 과다발생이 의심되는 지역에 있어서 이러한 종류의 군집분석을 실시한다. 이 군집분석의 결과에 따라, 후속의 연구의 필요성이 결정되는 것이고, 이것은 보건 자원과 역량의 효율적인 배치와 우선순위의 결정에 있어서 매우 중요하다. 또한, 질병의 지리적 군집과 여타의 사회경제적 변수나 환경적 변수와의 관계 여부를 검사하여, 역학적 인과성을 제공하는 단초가 될 수 있다.

참고문헌

강금님, 2002, " Bayesian Models에 대한 이해," 리서치 노트 가을호, 19-24.

김요은, 1999, "우리 나라 주요 암의 지역적 분포 현황," 대한지리학회지, 34(5), 543-560.

김정순, 1998, 역학원론, 신광출판사.

박기호, 1999, "세균성 이질의 탐색적 공간분석," 대한지리학회지, 34(5), pp. 473-491.

박미라, 1999, GIS를 활용한 국내 말라리아 발생에 관한 연구, 서울대학교 석사학위논문.

박정선, 백도명, 이경용, 이관형, 이기범, 1994, 우리나라 성인남자 암등록자의 지리적 분포에 관한 기술역학적 연구, 산업보건연구원.

유은혜, 1999, GIS의 통계적 공간분석에 관한 연구, 서울대학교 석사학위논문.

의사신문, 2002. 5. 2., "〈암정복에 도전한다〉소아암".

이덕희, 안돈희, 박강원, 신해림, 안윤옥, 1998, "소아암 발생률의 지역간 비교," 대한 암학회지, 30(3), 425-434.

한국전자통신연구원, 2001, 전자파 생체영향 및 표준화 연구, 정보통신부.

Bailey, T. C., Gatrell, A. C., 1995, *Interactive Spatial Data Analysis*, Longman Scientific & Technical.

Besag J., Newell, J., 1991, The Detection of Clusters in Rare Disease, *Journal of Royal Statistical Society Series A*, 154(1), 143-155.

Cliff, A., Haggett, P., 1988, *Atlas of disease distributions: analytic approaches to epidemiological data*, Blackwell Publishers.

Cromley, E. K., McLafferty, S. L, 2002, *GIS and Public Health*, The Guiford Press.

D'Aignaux, J. H., Cousen, Delasnerie-Lauprtre, N., Brandel, J. P., Salomon, D., Laplanche, J. L., Hauw, J. J., Alprovitch, A., 2002, Analysis of the Geographical Distibution of Sporadic Creutzfeldt-Jakob Disease in France between 1992 and 1998, *International Journal of Epidemiology*, 31, 490-495.

Elliott, P., Wakefield, J. C., Best, N. G. and Briggs, D. J., 2000, *Spatial Epidemiology: Methods and Applications*, Oxford University Press.

Fotheringham, A. S., Brunsdon, C., Charlton, M., 2000, *Quantitative Geography*, SAGE publications.

Gatrell,A. C., Lytnen, M., 1998, *GIS and Health*, Taylor & Francis.

Gatrell, A., 2002, *Geographies of Health: An Introduction*, Blackwell Publishers.

Haining, 2003, *Spatial Data Analysis*, Cambridge.

Insightful, 2001, *S-Plus 6 for Windows Guide to Statistics*, Volum 1, Insightful Corporation, Seattle, WA.

Jacquez, G. M., Greiling, D. A., 2003, Local Clustering in breast, lung and colorectal cancer in Long Island, New York, *International Journal of Health Geographics*, 2 : 3.

Le, N. D.,Petkau, A. J., Rosychuk, R., 1996, Surveillance of Clustering near Point Sources, *Statistics in Medicine*, 15, 727-740.

Meade, M. S., Earickson, R. J., 2000, *Medical Geography*, The Guilford Press.

O'Sullivan, D., Unwn, D., 2003, *Geographic Information Analysis*, John Wiley & Sons.

Rothman, K. J., Greenland, S., *1998, Modern Epidemiology*, Lippincott Williams & Wilkins.

Sankoh, O. A., Becher, H., 2002, Disease Cluster Methods in Epidemiology and Application to Data on Childhood Mortality in Rural Burkina Faso, *Control of Tropical Infectious Disease*, University of Heidelberg, Discussion Paper.

Turnbull, B. W., Iwano, E. J., Burnett, W. S., Howe, H. L., Clark, L. C., 1990, Monitoring for Clusters of disease: Application to Leukemia Incidence in Upstate New York, *American Journal of Epidemiology*, 132, S136-S143.

Waller, L. A., Turnbull, B. W., Clark, L. C., Nasca, P., 1994, Spatial Pattern Analyses to Detect Rare Disease Clusters, in Lange, N., Ryan, L., Billard, L., Brillinger, D., Conquest, L.,Greenhouse, J. (eds), *Case Studies in Biometry*, John Wiley & Sons, INC.