

## 소아 천식을 통해서 본 의료보험 상병 자료의 완전성 추정 Capture-Recapture 분석방법의 적용

하미나<sup>1</sup>, 권호장<sup>1</sup>, 강대희<sup>2</sup>, 조수현<sup>2</sup>, 유근영<sup>2</sup>, 주영수<sup>2</sup>, 성주현<sup>2</sup>, 강종원<sup>2</sup>, 김대성<sup>2</sup>, 이상일<sup>3</sup>

단국대학교 의과대학 예방의학교실<sup>1</sup>, 서울대학교 의과대학 예방의학교실<sup>2</sup>, 삼성의료원 소아과<sup>3</sup>

= Abstract =

### Completeness Estimation of the Korean Medical Insurance Data in Childhood Asthma : Using Capture-Recapture Method

MN Ha<sup>1</sup>, HJ Kwon<sup>1</sup>, DH Kang<sup>2</sup>, SH Cho<sup>2</sup>, KY Yoo<sup>2</sup>  
YS Joo<sup>2</sup>, JH Sung<sup>2</sup>, JW Kang<sup>2</sup>, DS Kim<sup>2</sup>, SI Lee<sup>3</sup>

*Dept. of Preventive Medicine, Dankook University College of Medicine<sup>1</sup>,*

*Dept. of Preventive Medicine, Seoul National University College of Medicine<sup>2</sup>,*

*Dept. of Pediatrics, Samsung Medical Center<sup>3</sup>*

**Objectives :** The purpose of this paper is to estimate the completeness of the Korean Medical Insurance Data in childhood asthma.

**Methods :** Capture-recapture method was used to estimate the prevalence of childhood asthma and case ascertainment rate(completeness) of Korean Medical Insurance Data using two source model, 'Korean Medical Insurance Committee Data (KMICD)' and 'Nationwide Study of Asthma and Allergies in Korean Children'. The asthma cases were restricted to those who were born from 1981 to 1989 and were identified by their Resident Register Number. Asthma cases in Korean Medical Insurance Data were defined as cases coded by ICD-9 493 and ICD-10 J45. In 'Nationwide Study of Asthma and Allergies in Korean Children', asthma cases were defined as the children who had been diagnosed asthma and had experienced symptoms of asthma during the past 12 months. The defined cases in two data sources were matched by 13 digits Resident Register Number. The numbers of matched patients in two data sources were 245 of 32,825 eligible total subjects. Chapman and Wittes' nearly unbiased estimation was used for capture-recapture analysis of two data sources.

**Results :** Observed prevalence rate of childhood asthma was 5.3% and estimated prevalence rate by capture-recapture analysis was 11.6%. The highest prevalence rate was observed in 6-7 age group and the older the rate decreased. The completeness (the proportion of cases ascertained by KMICD to the total observed cases by two data

sources) was 20.6%, and ranged from 10.8% to 28.8% by area.

Conclusions : Invalid diagnosis of cases might overestimate the prevalence of childhood asthma and might underestimate the completeness of Korean Medical Insurance Committee Data in this study.

Key words : Childhood asthma, Capture-recapture method, Completeness, Korean Medical Insurance Data

## I. 서 론

환경성 질환을 모니터링하는데 기존에 생산되고 있는 각종 상병 및 사망자료를 효과적으로 활용하는 방안은 감시체계의 주요한 한 요소이다. 우리나라에서 생산되고 있는 상병 및 사망자료는 여러 가지가 있으나 그 중 의료보험연합회에서 생산되는 수진자료는, 성격상 전국적인 범주에서 전산화된 형태로, 또 거의 대부분의 국민이 가입해 있다는 점에서, 감시자료로서의 잠재적 가치가 매우 높은 것으로 보인다. 따라서 감시체계에서 이 의료보험 수진자료를 활용하기 위해서는 이 자료에 대한 질평가(quality evaluation)가 우선적으로 이루어져야 한다. 자료에 대한 질평가는 크게 두가지 측면에서 이루어져야 한다. 하나는, 이 자료에서 나타난 질병의 진단명에 대한 타당성(validity) 평가이고, 또 다른 하나는 이 자료가 해당 질병을 어느만큼이나 포괄해 내고 있는가하는 완전성(completeness)에 대한 평가이다. 본 연구는 이 중 완전성 평가에 관한 것이다.

Capture-recapture 분석방법은, 일찌기 동물생태학(animal ecology)에서 개발되어 사용되어 온 것으로, 차차 인구학 등의 영역에서도 사용하다가, 비교적 최근에 들어서야 역학분야에서 도입하여 활용하고 있는 분석방법이다(Wittes와 Sidel, 1968; Hook 과 Regal, 1995). 이 방법은 일반적으로 서로 다른 자료원에 나타나 있는 환자의 명단(patient lists)에서 겹치는 부분에 대한 정보를 이용하여 각 자료가 불완전하게 파악하고 있는 환자의 수를 보정된 수치로 파악하고자 시도되는 방법이다(Hook 과 Regal, 1995). 국제 역학회

의 정의에 따르면, 여러 가지 자료원에서 빠뜨리고 있는 환자의 수를 추정하여 한 인구집단에서의 유병률을 보정된 수치로 추정하는 방법으로 설명되어 있다 (Last, 1995). 역학 분야에서 사용하는 목적은 여러 가지 경우가 있으나 모두 이환된 인구집단의 수를 추정하고자하는 것과 관련되어 있다.

본 연구는 '환경성 질환의 감시체계 개발' 연구의 일환으로 시도된 것이다. 소위 환경성 질환이라고 하면, 다른 많은 질환이 이에 포함될 수 있겠지만 대표적으로는 천식, 그 중에서도 소아의 천식을 첫 번째로 꼽을 수 있다. 소아 천식은 흡연 등의 강력한 교란변수의 영향을 배제할 수 있기 때문에, 환경오염 특히 대기오염에 민감하게 반응하는 질환으로 알려져 있다 (Bates, 1996). 뿐만 아니라 최근에 수행된 전국적인 소아천식의 유병률조사에 대한 자료를 얻을 수 있어서 의료보험 상병 자료의 완전성을 평가할 수 있었다.

본 연구의 구체적인 목적은 capture-recapture 분석방법을 적용하여 첫째, 의료보험 수진자료에서 포괄하고 있지 못한 소아 천식환자수(missed cases)를 파악하여 보정된 유병률(adjusted prevalence)을 추정하고 둘째, 이에 대한 정보를 토대로 하여 이 자료가 포괄하고 있는 소아천식환자의 비율을 제시함으로써, 의료보험 수진자료의 완전성을 평가하는 것이다.

## II. 연구 재료 및 방법

### 1. 환자 명단에 관한 자료원들 (Data Sources of Asthma Cases)

Capture-recapture 분석방법은 환자 명단을 파악해 낼 수 있는 두 개 이상의 자료원을 가지고 수학적 모델에 적용하여 각각의 불완전한 자료원에서 파악하지 못한 환자수를 추정해 내는 방법으로, 본 연구에서는 두가지 자료원을 이용한 소위 'two source model'을 이용하였다. 의료보험연합회에서 생산하는 의료보험 수진 자료(이하 의료보험 상병 자료)이외에 사용한 또 다른 자료는, '한국 어린이/청소년의 알레르기질환에 관한 전국적 역학조사'(이하 역학조사자료)이다(이상 일 등, 1995). 이 자료는 우리나라 9개 지역을 서울/지방(공업지역 및 일반지역)으로 층화한 다음, 초등학교 및 중등학교를 무작위 추출하여 설문지를 통하여 증상 호소율을 조사한 것이다. 1995년 9월 21일에서 11월 14일에 걸친 조사기간 동안에 9개 지역 34개 초등학교 학생 25,361명과 34개의 중학교 학생 15,068명으로 총 40,429명을 대상으로 하고 있다. 천식 증상에 대한 설문은 초등학교 학생들의 경우는 학부모가 기입하도록 하였고, 중학교 학생들의 경우는 천식증상을 일으키는 환자의 video를 보여준 다음, 자신이 이와 유사한 증상을 경험한 적이 있는 지를 묻는 방식으로 조사하였다.

의료보험 상병 자료는 우리나라 의료보험 연합회의 자료로서, 공무원 및 사립학교 교직원 의료보험 관리공단 자료와 지역 및 직장 의료보험 연합회의 자료 모두가 포함되어 있는 자료이다.

## 2. 환자 정의 및 짝짓기 과정(Case Definition & Matching Procedure)

분석대상은 이 두 자료에서 1981년에서 1989년 사이에 출생하였고, 주민등록번호 13자리수가 모두 확인 되는 소아로 국한하였다. 역학조사자료에서는 천식을 진단 받은 경험이 있으면서, 지난 12개월안에 천식 증상을 경험한 적이 있는 경우를 천식환자로 정의하였다. 의료보험 상병 자료에서는 주상병명 및 부상병명이 천식(ICD-9 : 493, ICD-10 : J45)인 건수만을 환자군으로 정의하였고, 역학조사 자료에서의 환자 파악 시

기와 일치되도록 하기 위해서, 1994년 9월부터 1995년 12월에 걸쳐서 수진된 경우(진료개시일 기준)에 한하여 분석하였다.

이렇게 정의된 천식환자들의 주민등록번호 13자리를 이용하여 두 자료원에서 동시에 확인될 수 있는 환자를 짝짓기하였다(그림 1).

## 3. 보정 유병률과 완전성 추정(Estimation of Adjusted Prevalence and Completeness)

Capture-recapture 분석방법의 two source model에서는 소위 'Peterson estimate'라고 불리는 maximum likelihood estimator(MLE)가 가장 흔히 사용되어 왔다. 그러나 Chapman과 Wittes 등은 MLE방법이 비교적 적은 규모의 표본에서는 bias를 가진다하여 좀 더 적절한 추정을 하게 해 주는 방법으로 'nearly unbiased estimator'(NUE)방법을 제안하였는데(Hook과 Regal, 1992), 본 연구에서는 이 방법을 사용하였다. 그리고 추정된 환자수에 대한 95% 신뢰구간을 구하고 유병률을 산출하였다. 그리고 의료보험 상병 자료의 완전성은 추정된 전체 환자수 중에서 실제 의료보험 상병자료에서 파악된 환자수의 백분율(%)로 나타내었다(표 1).

## Ⅲ. 연구 결과

### 1. 전체 환자수의 추정

역학조사자료에서 1981년에서 1989년 사이에 출생한 어린이는 총 33,607명이었고 이 중에서 주민등록번호가 정확히 기재되어 있지 않아 신원확인이 불가능한 사람을 제외한 32,825명이 본 연구의 대상 모집단이 되었다. 이들 중에서 설문조사 결과 천식으로 진단 받은 경험이 있으면서 동시에 지난 12개월 동안에 천식 증상을 경험한 적이 있는 경우가 1,190명으로서 이는 역학조사자료에서 파악된 천식 환자수가 되었다. 한편, 의료보험 상병 자료에서 1995년에 천식으로 수

Table 1. Two-source model

		Source 1		
		yes	no	
Source 2	yes	a	b	a+b
	no	c	d = ?	
		a+c		$T_{NUE} = a+b+c+d$

	Maximum likelihood estimator (MLE)	Nearly unbiased estimator (NUE)
Unobserved cell	$d = bc/a$	$d = bc/a + 1$
Total population	$T = (a+b)(a+c)/a$	$T = [(a+b+1)(a+c+1)/(a+1)] - 1$
Completeness of source 1	$a/(a+b)$	$(a+1)/(a+b)$
Completeness of source 2	$a/(a+c)$	$(a+1)/(a+c)$

$Var(T_{NUE}) = (a+b+1)(a+c+1)bc / (a+1)^2(a+2)$   
 95% C.I. of  $T_E = T_{NUE} \pm 1.96 \{Var(T_{NUE})\}^{1/2}$

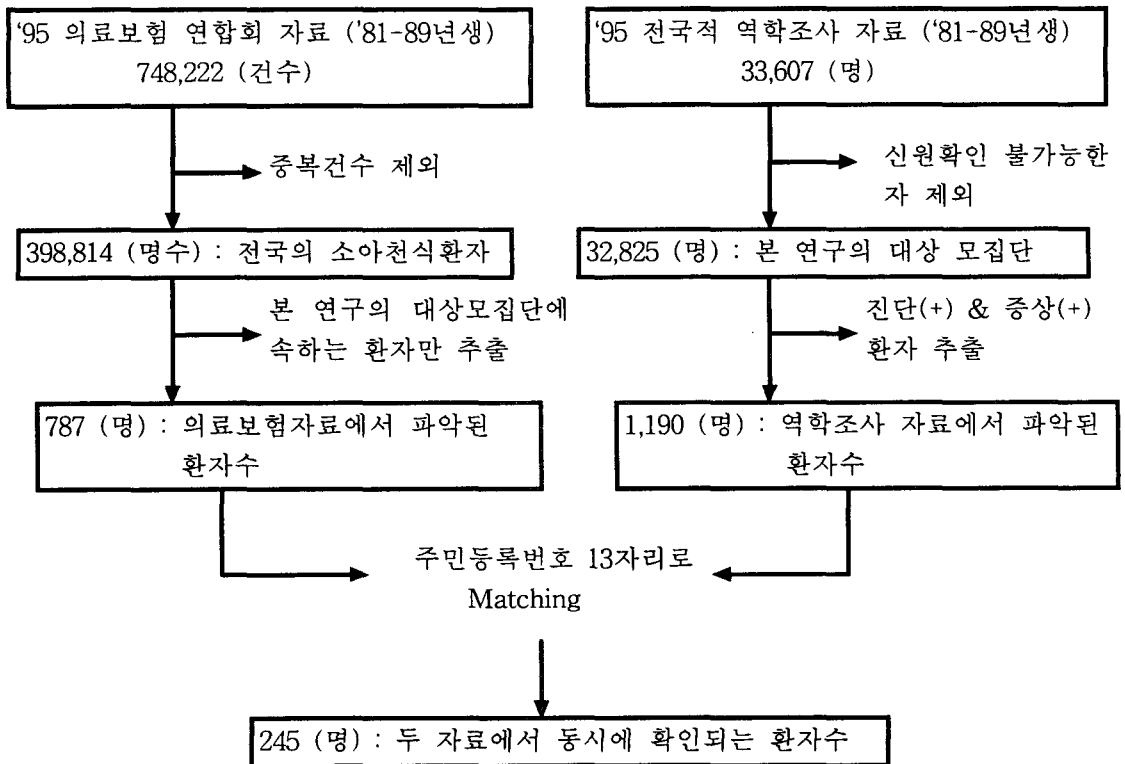


Fig 1. Flow of matching procedure.

**Table 2. Capture-Recapture Estimate of total cases**

	Both sources (a)	Source 1 only <sup>1)</sup> (b)	Source 2 only <sup>2)</sup> (c)	Total observed (T <sub>o</sub> )	Total estimated <sup>3)</sup> (T <sub>E</sub> )	95% C.I. of T <sub>E</sub> <sup>4)</sup>
All cases	245	542	945	1,732	3,814	3,462 - 4,165
Birth years						
'88 - '89	58	173	230	403	907	736 - 1,078
'83 - '87	168	306	645	1,117	2,298	2,049 - 2,548
'81 - '82	21	63	128	212	578	390 - 766
Region						
Seoul	77	141	340	558	1,172	985 - 1,359
Suwon	36	96	91	223	459	354 - 563
Ansan	30	22	80	127	180	146 - 214
Ulsan	12	32	80	124	320	189 - 452
Chunju	23	55	63	141	285	205 - 365
Jeju	30	55	91	176	337	256 - 418
Changwon	16	78	102	196	664	406 - 921
Chungju	15	46	69	130	328	206 - 449
Chunchun	6	17	34	57	139	64 - 214

1) Korean Medical Insurance Data

2) Nationwide study of Asthma and Allergies in Korean Children

3)  $T_E = [(a+b+1)(a+c+1)/(a+1)] - 1$

4) 95% C. I. of  $T_E = T_{NUE} \pm 1.96 \{Var(T_{NUE})\}^{1/2}$   
 $Var(T_{NUE}) = (a+b+1)(a+c+1)bc/(a+1)^2(a+2)$

**Table 3. Adjusted Prevalence Rate of Childhood Asthma**

	Population	Observed cases		Estimated cases		95% C. I. of Adjusted rate <sup>1)</sup>
		No.	Rate	No.	Rate	
All cases	32,825	1,732	5.3	3,814	11.6	11.3 - 12.0
Birth years						
'88 - '89	4,266	403	9.4	907	21.3	20.0 - 22.5
'83 - '87	19,330	1,117	5.8	2,298	11.9	11.4 - 12.3
'81 - '82	9,229	212	2.3	578	6.3	5.8 - 6.8
Region						
Seoul	10,230	558	5.5	1,172	11.5	10.8 - 12.1
Suwon	3,316	223	6.7	459	13.8	12.7 - 15.1
Ansan	3,155	127	4.0	180	5.7	4.9 - 6.5
Ulsan	2,832	124	4.4	320	11.3	10.1 - 12.5
Chunju	2,833	141	5.0	285	10.0	9.0 - 11.2
Jeju	2,684	176	6.6	337	12.6	11.3 - 13.8
Changwon	2,985	196	6.6	664	22.2	20.8 - 23.7
Chungju	2,863	130	4.5	328	11.5	10.3 - 12.6
Chunchun	1,927	57	3.0	139	7.2	6.1 - 8.1

1) An approximate 95% confidence interval for the binomial parameter p based on the normal approximation to the binomial distribution is given by  $(p - 1.96\sqrt{pq/n}, p + 1.96\sqrt{pq/n})$ , when  $pqn \geq 5$ .

**Table 4. Completeness (Ascertainment Rate) of Korean Medical Insurance Data**

	Total observed cases (T <sub>o</sub> )	Total estimated cases (T <sub>e</sub> )	Cases identified by Korean Medical Insurance data		
			No.	(% of T <sub>o</sub> )	(% of T <sub>e</sub> )
All cases	1,732	3,814	787	45.4	20.6
Birth years					
'88 - '89	403	907	231	57.3	25.5
'83 - '87	1,117	2,298	472	42.3	20.5
'81 - '82	212	578	84	39.6	14.5
Region					
Seoul	558	1,172	218	39.1	18.6
Suwon	223	459	132	59.2	28.8
Ansan	127	180	52	40.9	28.9
Ulsan	124	320	44	35.5	13.8
Chunju	141	285	78	55.3	27.4
Jeju	176	337	85	48.3	25.2
Changwon	196	664	94	48.0	14.2
Chungju	130	328	61	46.9	19.0
Chunchun	57	139	23	40.4	16.5

진된 건수 중 생년이 1981년부터 1989년까지인 경우가 총 748,222건이었고, 여기서 중복된 건수를 제외하고 난 398,814명이 1995년의 해당 연령의 전국 소아 천식환자가 되었다. 이 중에서 본 연구의 대상 모집단인 32,825명에 속하는 환자만을 추출하였을 때 787명이 되었는데 이들이 1995년 의료보험 상병자료에서 파악된 천식환자 수가 되었다. 이렇게 역학조사 자료에서 파악된 환자수 1,190명과 의료보험 상병자료에서 파악된 환자수 787명을 주민등록번호 13자리로 짝지은 결과 245명이 나왔고, 이들은 두가지 자료 모두에서 공통적으로 파악된 환자수가 되었다(그림 1).

이 결과를 이용하여 capture-recapture 분석방법의 two source model을 적용하여 추정, 산출한 환자수는 다음과 같다(표 2). 두가지 자료에서 단순히 관찰할 수 있는 환자수에 비하여, capture-recapture 분석방법을 이용하여 추정한 환자수는 연령별이나 지역별을 막론하고 모든 경우에서 2배가 넘는 것으로 나타났다.

## 2. 보정 유병률의 산출

Capture-recapture 분석방법을 이용하여 산출한 환자수에서 소아 천식의 유병률을 추정한 결과는 다음과 같다(표 3). 전체 환자 중에서 관찰유병률에 비하여 capture-recapture 분석방법을 이용하여 추정한 유병률이 5.3%에서 11.6%로 2배 이상 증가하였다. 가장 높은 유병률을 보인 연령은 6-7세군('88 - '89년생)으로서 관찰유병률은 9.4%, 추정유병률은 21.3%를 나타내었다. 관찰유병률과 추정유병률 모두 연령이 증가하면서 감소하는 경향을 나타내었다. 지역별로도 대체로 관찰유병률에 비하여 추정유병률이 약 2배의 증가를 보이고 있는데, 그 중 창원지역은 특이하게 3배 이상의 증가를 보이고 있으며 추정 유병률이 22.2%로서 가장 높은 수치를 나타냈다. 천식 유병률이 가장 낮은 지역은 안산지역으로 관찰유병률 4.0%, 추정유병률 5.7%이었다.

## 3. 의료보험 자료의 완전성 추정

추정된 환자수를 이용하여, 의료보험 상병 자료의 완전성을 평가하여 다음과 같은 결과를 얻었다(표 4).

전체적으로는 의료보험 상병 자료는 소아천식환자의 약 20% 정도만을 포괄하고 있는 것으로 나타났고, 연령별로는 6-7세군('88 - '89년생)이 25.5%로 가장 높은 완전성을 보이고 있다. 또 지역별로는 20%를 중심으로 약간의 변이를 보이고 있다. 특히 울산 지역은 13.8%로서 가장 낮은 수치를 보이고 있다.

#### IV. 고 찰

본 연구에서 사용한 capture-recapture 분석방법은 두가지 이상의 자료원이 불완전한 것일 때 이를 함께 분석하여 특정한 인구집단의 수를 추정하고자 할 때나, 또는 실제 직접적으로 자세한 조사를 실시하여 얻은 유병률이나 발생률을 재검토하여 교정하고자 할 때, 또는 어떤 자료의 완전성(completenss)을 평가하고자 할 때, 또는 이환된 환자수 전체에 대한 타당한 최고 수치와 최저 수치의 범위를 도출하고자 할 때 사용된 분석방법이다. 그러나 이 방법을 적용하기 위해서는 몇가지 중요한 가정이 따라야 하는데, 그 중 사용하는 자료가 정확해야 한다는 점이 우선적이다. 즉, 진단의 정확성이 보장되어야 하며, 사용되는 여러 가지 자료에서 모두 분석대상이 된 진단된 환자의 시간적, 공간적인 위치가 같아야 하는 점 등이 그것이다 (Hook과 Regal, 1995).

이러한 capture-recapture 분석방법의 적용 시 전제되는 가정 중 진단의 정확성이라는 측면에서 본 연구에서 사용한 자료원들은 일정한 문제점을 가지고 있다. 의료보험 상병 자료는, 본 연구진이 수행한 다른 연구에서 밝혀진 바와 같이 천식의 경우 주상병 및 부상병명에 기입된 진단의 정확성이 60 - 70% 정도로 조사되었다. 이 정확성에 대한 조사는 해당 환자의 의무기록을 검토하여, 해당 질환에 대한 전문의의 판단을 거쳐서 새로이 내린 진단명과의 일치율을 본 결과이다. 또 역학조사 자료는 주로 설문조사를 통해 이루어진 것으로서, 엄격한 의미에서의 천식 진단에 필요한 기관지 유발시험이나 의사의 이학적 검진 등의 과정이 결여된 자료이다. 물론 환자에 대한 정의를 증상

이 있으면서 천식 진단을 받은 경험이 있는 경우만으로 엄격히 하였지만, 자기기입식 설문조사는 설문조사 대상자의 주관적인 판단이나 기억의 문제가 개입될 수 있고, 설문지 자체의 문제나 시행방법상의 문제, 면접자의 개별적 특성 등이 그 결과에 많은 영향을 미치게 된다.

한편, 알레르기 질환은 산업의 발달과 더불어 생활양식의 변화, 새로운 알레르겐의 유입 및 대기오염 등의 원인으로 발생이 증가하고 있는 추세이다. 설문조사를 통하여 얻은 소아의 천식 유병율은, 미국의 경우 6-16세 아동에서 1970년 3.2%에서 1978-1980년에 4.1%로 증가하였으며, 뉴질랜드에서는 11세의 아동에서 1969년 7.1%에서 1982년 13.5%로 약 2배 증가하였고, 영국에서도 매년 증가추세를 보이고 있다(김유영 등, 1996). 우리나라에서도 1983년 소아 천식 유병율을 5.6%(이혜란 등, 1983), 1990년 10.1%(신태순 등, 1990)으로 보고하였다. 그러나 메타콜린 기관지 유발 시험 등의 객관적인 방법을 결합해서 조사한 유병율은 이보다 낮아서 미국의 NHANES II 결과에서 6-8%였고, 우리나라에서도 1993년에 4-6%로 보고되었다(김유영 등, 1996).

본 연구에서 사용한 환자에 대한 정의가 하나는 설문조사에서 증상이 있으면서 진단받은 경험이 있는 경우만으로, 또 다른 하나는 병, 의원에서 치료받았던 기록을 근거로 한 것임에도 불구하고, 추정된 유병율이 전체적으로 11.6%라는 것은 이러한 천식유병률에 관한 국내의 연구결과들에 비추어 볼 때 과대한 수치라 할 수 있다. 연령별로는 6-7세군에서 다른 연령군에 비하여 가장 높은 추정치를 보였는데, 이것은 사춘기에 천식유병율이 감소한다는 일부의 보고(Martin 등, 1980)와 일치하는 소견이다. 그러나 21.3%라는 높은 수치는 두가지 자료 모두에서 진단이 정확하지 않은 경우가 있어서, 두 자료에서 동시에 확인할 수 있는 공통환자수가 실제보다 적어지고(Hook 과 Regal, 1995), 그 결과 유병율이 과대하게 추정(overestimate)된 것으로 판단된다( $TE = [(a+b+1)(a+c+1)/(a+1)]-1$ 에서 a가 변화해도 a+b와 a+c는 일정하므

로 a가 적어지면 TE 값은 더 커진다.). 과대 산출된 정도는 진단명이 부정확한 정도만큼 더 과대하게 산출되었다 할 수 있으므로(Hook과 Regal, 1995), 의료보험 상병 자료에서 천식 진단명의 정확성이 60 - 70% 정도라고 할 때 본 연구에서는 최소한 30 - 40% 정도의 과대추정이 있었다고 볼 수 있고, 실제의 참 유병율은 최소한 8%보다는 적을 것이라 생각할 수 있다.

의료보험 상병 자료의 완전성 평가에 있어서도, 진단의 정확성 정도가 영향을 미치는 것은 마찬가지이다. 즉 역학조사 자료에서 진단이 100% 정확하지 않으면 두가지 자료에서 공통으로 발견되는 환자수가 참값에 비하여 적어지게 되어, 의료보험 상병자료의 완전성인 “두 자료에서 공통으로 발견되는 환자수/역학조사 자료에서 발견되는 총 환자수( $a/(a+b)$ )”는 참값에 비하여 낮은 수치를 가지게 된다. 따라서 본 연구에서 약 20%정도로 낮은 완전성을 보인 것은, 사용한 자료에서 진단의 정확성이 낮았던 점을 반영하고 있다.

의료보험 상병자료의 완전성에 영향을 미쳤을 것으로 생각되는 또다른 요소로는 실제 천식이 있었다 하더라도 병, 의원 진료를 받지 않고 약국이나 민간요법으로 해결하는 경우도 적지 않았을 것이라는 점이다. 이것이 설문조사를 통해서 얻은 환자수보다 실제 병, 의원에서 치료를 받은 환자수가 더 적게 나타나게 하였다고 볼 수 있다. 연령별로 완전성에 차이가 나타난 것은 주요하게는 역학조사가 나이별로 다른 조사방식을 취했던 것에서 비롯되는 것으로 보인다. 즉 6 - 7세 군에서 가장 높은 완전성을 보였는데, 이 연령군은 학부모가 설문지를 기입하도록 하게 하였던 초등학교 학생들 중에서 실제로 학부모가 기입하였을 가능성이 가장 높은 연령군이다. 따라서 이 6 - 7세군은 역학조사의 결과 중에 상대적으로 가장 믿을 만한 결과를 가지는 연령군으로 볼 수 있고, 이것이 실제의 의료기관에서 치료를 받은 것에 대한 자료인 의료보험 상병자료의 환자 명단과 일치하는 사례가 더 많아져서 나타난 결과로 해석할 수 있다.

이처럼 본 연구에서 사용한 자료는 capture-recap-

ture 분석방법이 전제하고 있는 자료의 정확성에 관한 가정, 즉 진단명의 타당성을 충분히 충족시키고 있지 못하다. 우리나라를 비롯한 개발도상국의 경우 보건과 관련된 자료들이 불완전한 경우가 많고 불완전한 자료는 대부분의 경우 부정확할 가능성이 큰 것이 현실이다. 그러나 이러한 제한점에도 불구하고 향후 자료의 부정확성의 정도를 감안하여 사용한다면, capture-recapture 분석방법은 불완전한 자료를 이용하여 환자수를 추정할 수 있는 현실적이고도 주요한 방법론의 하나가 될 것으로 생각한다.

## V. 결론 및 요약

우리나라 소아의 천식유병률을 파악하고, 의료보험 상병 자료의 완전성을 평가하기 위하여, ‘의료보험연합회의 수진 자료’와 ‘한국 어린이/청소년의 알레르기 질환에 관한 전국적 역학조사’의 두가지 자료를 이용하여 capture-recapture 분석방법을 적용, 다음과 같은 결과를 얻었다.

1) 소아 천식의 관찰유병률이 5.3%인데 반하여, capture-recapture 분석방법을 적용하여 추정된 유병률은 11.6%이었다.

2) 추정된 유병률은 연령별로는 6-7세군이 가장 높았고, 연령이 증가할수록 감소하는 경향을 나타내었다.

3) 의료보험 상병자료의 완전성은 20.6%로 추정되었고, 연령별, 지역별로 13.8 - 28.9%까지의 변이를 보였다.

4) 사용한 두가지 자료의 정확성(accuracy)의 문제로 인하여, 소아 천식 유병률은 실제 참값에 비하여 과대하게 추정(overestimate)되었고, 의료보험 상병자료의 완전성은 실제에 비하여 더 낮게 추정(underestimate)된 것으로 판단되었다.

## 참고문헌

김유영, 조상헌, 하미나, 안윤옥, 이상일, 장석일, 김우



- 경, 박재경, 김윤근, 송숙희, 지영구, 민경업. 설문지와 메타콜린 기관지 유발시험을 이용하여 조사한 한국의 소아 천식 유병율. *알레르기* 1996; 16(2): 175-184
- 신태순, 이금자, 윤혜선. 국민학교 아동에서의 알레르기 질환에 관한 조사. *알레르기* 1990; 10: 201-212
- 이상일, 손병관, 이준성, 이하백, 고영울, 김규언. 한국 어린이, 청소년의 알레르기 질환에 관한 전국적 역학조사. 1995년도 보건의료기술연구개발사업 보고서. 1995
- 이혜란, 홍동성, 손근찬. 소아 알레르기에 관한 조사. *대한의학협회지* 1983; 26: 254-262
- Bates DV. The effects of air pollution on children. *Environ Health Perspectives* 1996; 103 Supp 6: S49-S53
- Bobo JK, Thapa PB, Anderson JR, Gale JL. Acute encephalopathy and seizure rates in children under age two years in Oregon and Washington State. *Am J Epi* 1994; 140(1): 27-38
- Davis SF, Streble PM, Atkinson WL, Markowitz L, Sutter RW, Scanlon KS, Friedman S, Hadler SC. Reporting efficiency during a measles outbreak in New York City, 1991. *Am J Pub Health* 1993; 83(7): 1011-1015
- Egeland GM, Perham-Hester KA, Hook EB. Use of capture-recapture analyses in fetal alcohol syndrome surveillance in Alaska. *Am J Epi* 1995; 141(4): 335-341
- Fischer M, Leyland A, Cormack R, Goldberg DJ, Bloor M, Green ST, Taylor A, Covell R, Mckeganey N, Platt S. Estimating the population prevalence of injection drug use and infection with human immunodeficiency virus among injection drug users in Glasgow, Scotland. *Am J Epi* 1993; 138(6): 170-181
- Hook EB, Regal RR. Internal validity analysis : A method for adjusting capture-recapture estimates of prevalence. *Am J Epi* 1995; 142(9) Suppl: S48-S52
- Hook EB, Regal RR. Capture-recapture methods in epidemiology: Methods and limitation. *Epi Reviews* 1995; 17(2): 243-264
- Hook EB, Regal RR. The value of capture-recapture method even for apparent exhaustive surveys. *Am J Epi* 1992; 135(9): 1060-1067
- International working group for disease monitoring and forecasting. Capture-recapture and multiple-record systems estimation I : History and theoretical development. *Am J Epi* 1995; 142(10): 1047-1058
- International working group for disease monitoring and forecasting. Capture-recapture and multiple-record systems estimation II : Application in human diseases. *Am J Epi* 1995; 142(10): 1059-1068
- LaPorte RE, Dearwater SR, Chang YF, Songer TJ, Aaron DJ, Anderson RL, Oslen T. Efficiency and accuracy of disease monitoring systems : Application of capture-recapture methods to injury monitoring. *Am J Epi* 1995; 142(10): 1069-1077
- Martin AJ, McLennan LA, Landau LI. The natural history of childhood asthma to adult life. *Br Med J* 1980; 280: 1397-1400
- Mastro TD, Kitayaporn D, Weniger BG, Vanichseni S, Laosunthorn V, Uneklabh T, Uneklabh C, Choopanya K, Limpakarnjanarat K. Estimating the number of HIV-infection drug users in Bangkok: A capture-recapture method. *Am J Pub Health* 1994; 84(7): 1094-1099
- McCarty DJ, Tull ES, Moy CS, Kwoh CK, LaPorte RE. Ascertainment corrected rates: Applications of capture-recapture methods. *Int J Epi* 1993; 22(3): 559-286
- Regal RR, Hook EB. Goodness-of fit based confidence intervals for estimates of the size of a closed population. *Stat Med* 1984; 3: 287-291
- Walter SD, Birnie SE, Marrett LD, Taylor SM, Reynolds D, Davies J, Drake J, Hayes M. The geographical variation of cancer incidence in Ontario. *Am J Pub Health* 1994; 84(3): 367-376
- Wittes J, Sidel VW. A generalization of the simple capture-recapture model with applications to epidemiological research. *J chron Dis* 1968; 21: 287-301