

연구논문

알레르기 질환 유병률 조사의 표본설계와 설계효과 분석*

Sampling Design and Design Effect Analysis for the Epidemiologic Study of Allergic Diseases in Korean Children

이기재^{a)} · 권호장^{b)} · 안강모^{c)} · 채유미^{d)} · 함명일^{e)} · 김지현^{f)} · 박용민^{g)}
· 이소연^{h)} · 한만용ⁱ⁾ · 김우경^{j)}

Kee-Jae Lee · Ho-Jang Kwon · Kangmo Ahn · Yoomi Chae · Myung-Il Hahm
· Jihyun Kim · Yongmean Park · So-Yeon Lee · Manyong Han · Woo Kyung Kim

2010년 「알레르기 질환 유병률 조사」는 국제적으로 비교 가능한 알레르기 질환 유병률 지표 산출을 목적으로 ISAAC 프로그램에 기반한 조사수행 체계에 따라 진행되었다. 본 연구는 새로운 「알레르기 질환 유병률 조사」를 위한 표본설계와 2010년 조사결과를 분석하여 얻은 설계효과와 급내상관계수를 제시한 것이다.

새로운 표본설계는 외국자료와의 비교, 유병률 추정의 정확도 향상, 그리고 서울과 지방의 추정결과 비교 등이 가능하도록 설계하였다. 이 조사는 학교를 1차추출단위로 하고 학급을 2차추출단위로 하는 층화2단계확률비례추출법에 따라 추출되었다. 1차추출단위인 학교에 대한 층화기준은 지역(7개 권역)과 학교 소재지의 특성(대도시, 고밀집, 중밀집, 저밀집)

* 이 연구는 질병관리본부 학술연구용역사업(과제번호 2010E3303400)으로 수행되었음.

a) 교신저자(corresponding author): 한국방송통신대학교 정보통계학과 교수 이기재.

E-mail: kjlee@knou.ac.kr.

b) 단국대학교 의과대학 의예과 교수

c) 성균관의과대학 삼성서울병원 소아청소년과 교수

d) 단국대학교 의과대학 직업환경의학과 조교수

e) 순천향대학교 의료과학대학 보건행정경영학과 조교수

f) 성균관의과대학 삼성서울병원 소아청소년과 임상강사

g) 건국대학교 의학전문대학원 소아과학교실 부교수

h) 한림대학교 의과대학 소아청소년과 조교수

i) 차의과학대학교 분당차병원 소아과학교실 교수

j) 인제대학교 서울백병원 소아청소년과 부교수

에 따른 구분을 이용하였다. 2010년 조사를 완료한 후 추출확률, 무응답 보정, 사후층화 조정 등을 반영한 가중치를 산출하여 추정에 활용하였다. 또한 2010년 조사데이터에 대한 설계효과와 급내상관계수를 분석하고, 향후 연구방향을 제시하였다.

주제어: 층화2단계확률비례추출법, 설계효과, 급내상관계수

To investigate the prevalence of asthma, atopic dermatitis, and allergic rhinitis, we did a cross-sectional survey using the ISAAC(International Study of Asthma and Allergies in Childhood) protocol to enhance the comparability of the prevalence estimates. The purpose of this article is to present the sample design, design effect and intra-cluster correlation analysis for the epidemiologic study of allergic diseases in Korean children.

The sampling frame was based on a comprehensive national list of schools in Korea. The participants of this survey were selected using a stratified two-stage cluster sampling design. For the first stage of sampling, schools were stratified by geographic regions (7 regions) and the type of school location (metropolitan city, urban areas, rural areas).

Because the participants for this survey were selected using a stratified two-stage cluster sampling design, we calculated the sampling weights for this study to take into account differential selection probabilities, non-response and post-stratification. The results and practical issues of the design effect analysis were presented.

Key words: stratified two-stage probability proportional to size sampling, design effect, intra-cluster correlation

I. 서론

사회경제 및 의학의 발달에 따라 19세기까지 주요 질환이었던 감염성 질환들이 점차 감소하고 있는 가운데 20세기 이후에는 비감염성 질환인 알레르기 질환의 유병률이 증가하고 있다. 천식, 알레르기 비염, 아토피 피부염 및 식품알레르기 등 대표적인 알레르기 질환들은 유전적 요인과 환경적 요인에 의해 발생한다고 알려져 있으나 아직도 정확한 기전은 밝혀져 있지 않다(안강모 외 2011).

알레르기 질환은 우리나라뿐 아니라 세계 여러 나라에서 높은 유병률로 인한 사회적 부담이 크게 문제가 되었지만 과거에는 유병률을 객관적으로 평가할 수 있는 방법이 없었다. 이러한 배경 하에서 알레르기 질환의 유병률을 정확히 산출하여 국가 간 비교를 목적으로 ‘국제 소아천식 및 알레르기 질환의 역학조사(International Study of Asthma and Allergies in Childhood: ISAAC)’ 프로그램이 1991년도에 시작되었다. ISAAC연구는 크게 세 단계로 계획되었는데 첫 번째 단계(Phase I)에서는 유병률을 조사하여 비교하는 것을 목표로 하였고, 두 번째 단계(Phase II)에서는 알레르기 질환의 위험요인을 조사하는 것을, 세 번째 단계(Phase III)에서는 첫 번째 단계 조사를 반복해서 유병률의 변화를 추적하는 것을 목표로 하고 있다(Asher et al, 1995).

우리나라에서는 대한 소아알레르기 및 호흡기학회가 주관이 되어 ISAAC 프로토콜에 따라 1995년도에 첫 번째 ‘한국 어린이, 청소년의 알레르기 질환에 관한 전국적 역학조사(Nationwide Study of Asthma and Allergies in Korean Children)’를 실시하였으며, 5년 뒤인 2000년도에 두 번째 전국적 역학조사를 실시하여 천식의 유병률 및 위험인자에 관한 전국적인 자료를 확보하게 되었다. 2회에 걸친 역학조사를 통해 우리나라 어린이·청소년에서 알레르기 질환 유병률이 증가함을 알 수 있었으나 2000년 이후 이와 같은 전국적 역학조사는 이루어지지 않았다.

2010년 「알레르기 질환 유병률 조사」는 국제적으로 비교 가능한 알레르기 질환 유병률 지표 산출을 목적으로 ISAAC 프로그램에 기초한 조사수행 체계에 따라 진행됐다. 조사대상은 외국자료와의 비교를 위해 6~7세군과 13~14세군으로 국한했으며 유병률 추정의 정확도를 확보하고 서울과 지방의 비교가 가능하도록 각 연령군의 표본크기를 4,000명으로 정했다.

조사에서는 설문조사지(한국어판 ISAAC 설문지, 비디오 설문지, 환경 설문지, 식품 알레르기 설문지), 검사도구(피부단자시험, 폐기능검사, 혈액검사), 식품알레르기 설문지 등의 조사도구를 사용하였다. 6~7세 아동을 대상으로는 부모응답 설문조사, 아토피 피부염 진찰, 피부단자시험, 혈액검사 등을, 13~14세의 청소년을 대상으로는 설문 및 비디오 설문조사, 아토피 피부염 진찰, 피부단자시험, 혈액검사 등이 이루어졌다.

본 연구는 2010년 「알레르기 질환 유병률 조사」를 위한 표본설계와 조사데이터에 대한 설계효과 분석결과를 제시하는 것이 목적이다. 표본설계 단계에서는 2000년 「알레르기 질환 유병률 조사」 결과를 분석하여 새로운 표본설계를 위한 기초자료로 활용하였고, 2010년 「알레르기 질환 유병률 조사」를 완료한 후 조사데이터에 대한 설계효과와

급내상관계수를 분석하였고, 이를 바탕으로 향후 연구방향을 제시하였다.

II. 기존 조사데이터에 대한 분석

1. 조사현황

전 세계적으로 천식 및 알레르기 질환이 증가함에 따라 1991년도부터 선진국을 중심으로 알레르기 질환에 대한 유병률의 국가 간 비교와 원인 규명을 위한 ISAAC 연구가 시작되었다. 우리나라도 대한 소아알레르기 및 호흡기학회 주관으로 1995년도와 2000년도에는 초등학생과 중학생을, 2003년도에는 유치원생을 대상으로 전국적인 천식, 아토피 피부염 등의 알레르기 질환에 대한 유병률 조사가 이루어졌다.

1995년과 2000년에 실시한 역학조사는 지역을 서울과 지방으로 구분하고, 나이 대는 초등학생 6~12세군과 중학생 13~15세군으로 나누어 실시되었다. 서울에서는 초등학교 10개, 중학교 10개를, 지방에서는 수원, 춘천, 청주, 울산, 창원, 전주, 제주, 안산 등 8개 도시의 초등학교 24곳, 중학교 24곳을 무작위로 선정하여 조사하였다.

2000년도 조사에서 서울의 초등학생 응답자는 8,028명, 중학생 응답자는 4,409명이고, 지방의 초등학생 응답자는 20,022명, 중학생 응답자는 10,394명으로 전체 42,653명이었다(오재원 외 2003).

2. 설계효과

설계효과(design effect : DEFF)는 어떤 특정한 복합표본설계에 의한 추정량의 분산과 단순임의추출법에 의한 추정량 분산을 비교하기 위하여 구한다. 어떤 특정한 표본추출방법 D에서 모수 θ 에 대한 추정량으로 $\hat{\theta}$ 를 사용할 때 설계효과(DEFF)는 다음과 같이 정의된다(Kish 1995).

$$DEFF(\hat{\theta}) = \frac{V_D(\hat{\theta})}{V_{SRS}(\hat{\theta})} \quad (1)$$

여기서, $V_{SRS}(\hat{\theta})$ 는 같은 표본크기를 가정할 때 단순임의추출법에서 구한 추정량의 분산이다. 설계효과 $DEFF(\hat{\theta})$ 에 대한 추정은 각각의 분산식에 추정치를 대입하여

$$def(\hat{\theta}) = \frac{\hat{V}_D(\hat{\theta})}{\hat{V}_{SRS}(\hat{\theta})} \text{로 구한다.}$$

복합표본설계에 의한 설계효과는 가중치 효과와 집락화 효과의 곱으로 표현되며 다음 식과 같이 나타낼 수 있다(Verma et al. 1980).

$$DEFF(\bar{y}) = (1 + cv^2(w_j)) \times (1 + (\bar{b} - 1)\bar{\rho}) \quad (2)$$

여기서 $cv(w_j)$ 는 가중치의 변동계수(coefficient of variance)이고, $\bar{b}$ 는 각 표본집락에 대한 응답자 수의 평균, $\bar{\rho}$ 는 집락 내 응답값의 유사성의 정도를 나타내는 값이다.

설계효과는 여러 가지 다양한 목적으로 활용된다. 예를 들어 대안이 될 수 있는 조사 설계와 비교할 때나 통계분석을 위해서 유효 표본크기(effective sample size)를 구할 때 또는 복합표본설계를 반영한 신뢰구간이나 가설검정을 수행할 때 사용된다(Lacher et al. 2004).

새로운 표본설계를 위한 기초자료를 얻기 위해 2000년 「알레르기 질환 유병률 조사」 조사결과에 대한 설계효과를 분석하였다. 2000년도에 실시된 「알레르기 질환 유병률 조사」에서 1차추출단위는 학교, 2차추출단위는 학급이었다. 각 표본학교에서는 학년별로 3~5개 학급을 랜덤추출하여 표본학급 내 전체 학생을 조사하였다.

<표 1>과 <표 2>는 2000년 「알레르기 질환 유병률 조사」 결과를 분석한 것으로 각각 초등학생과 중학생 대상 조사에 대한 천식과 아토피 피부염 유병률에 대한 설계효과를 분석하여 정리한 것이다. 천식과 아토피 피부염 유병률 추정에 대한 설계효과는 9개 도시를 층으로 간주하여 분석한 것으로 식 (1)에 따라 계산한 것이다.

각 질환 유병률에 대한 초등학교 전체의 평균 설계효과는 0.976에서 1.762 사이로 나타났다. 초등학생 전체에 대한 평균 설계효과는 1.338이고, 중위수는 1.308이다. 각 유병률에 대한 중학교 3개 학년의 평균 설계효과는 1.626에서 2.612 사이의 값을 보이고 있다. 전체 중학생에 대한 평균 설계효과는 1.983이고, 중위수는 2.409이다.

초등학생 조사에서는 천식에 비해서 아토피 피부염 유병률 추정에 대한 설계효과가 대체로 높고, 학년 변화에 따른 설계효과의 경향성은 없어 보인다. 반면 중학생 조사에서는 전체적으로 아토피 피부염에 비해서 천식 유병률 추정에 대한 설계효과가 높게 나타났다. 학년 변화에 따른 설계효과의 경향성은 없어 보이지만 천식의 경우에 학년이 높아짐에 따라 설계효과가 커지는 경향을 발견할 수 있다.

2010년 「알레르기 질환 유병률 조사」는 2000년도 조사와는 달리 초등학교 및 중학교 1학년생만을 대상으로 한다. 새로운 표본설계에서는 2000년도 조사결과의 초등학교 및 중학교 1학년생에 대한 설계효과의 중위수를 이용하였다. 먼저 식 (2)에 따라 급내상관계수(초등 및 중학교 1학년생에 대한 급내상관계수는 각각 0.0021과 0.0054이었음)를 구하고, 새로운 표본설계의 표본집락에 대한 평균 응답자 수를 산정하여 새로운 표본설계의 유효 표본크기(effective sample size)와 예상오차의 한계를 구하였다.

〈표 1〉 초등학생 조사의 설계효과 : 2000년 알레르기 질환 유병률 조사

구분	천식				아토피				평균	중위수
	증상(+)	지난 12개월(+)	진단(+)	치료(+)	증상(+)	지난 12개월(+)	진단(+)	치료(+)		
초1	1.243	1.073	1.346	0.958	1.694	2.483	1.790	1.014	1.448	1.403
초2	1.791	1.042	1.017	1.032	1.206	0.635	1.424	0.729	0.982	0.907
초3	1.768	1.549	1.395	1.101	1.096	1.186	1.867	1.449	1.297	1.239
초4	1.431	1.132	0.891	0.621	2.430	1.580	2.485	1.450	1.433	1.357
초5	1.066	0.705	0.962	0.979	0.939	1.376	1.138	1.021	1.035	1.000
초6	1.711	1.124	1.792	1.163	1.362	2.425	1.870	0.700	1.605	1.679
평균	1.502	1.104	1.234	0.976	1.455	1.614	1.762	1.061	1.338	1.344
중위수	1.571	1.099	1.182	1.006	1.284	1.478	1.829	1.018	1.308	1.233

〈표 2〉 중학생 조사의 설계효과 : 2000년 알레르기 질환 유병률 조사

구분	천식				아토피				평균	중위수
	증상(+)	지난 12개월(+)	진단(+)	치료(+)	증상(+)	지난 12개월(+)	진단(+)	치료(+)		
중1	1.846	2.186	1.287	1.679	1.736	1.396	2.800	1.916	1.856	1.791
중2	2.340	2.627	2.240	1.637	1.363	1.720	1.205	0.936	1.759	1.679
중3	2.302	3.022	2.935	1.561	2.470	2.002	2.213	2.170	2.334	2.258
평균	2.163	2.612	2.154	1.626	1.856	1.706	2.073	1.674	1.983	1.965
중위수	2.302	2.627	2.240	1.637	1.736	1.720	2.213	1.916	2.049	2.065

Ⅲ. 새로운 표본설계

1. 표본설계의 기본원칙

2010년 「알레르기 질환 유병률 조사」를 위한 표본설계는 국제적으로 비교 가능한 알레르기 질환 유병률 지표의 산출이 가능하도록 ISAAC 프로토콜에 따라 진행하는 것을 기본원칙으로 하였다. 새로운 표본설계의 기본방향을 정리하면 다음과 같다.

첫째, 우리나라 전국 단위의 모집단 대표성을 확보할 수 있도록 표본설계하였다. 우리나라에서 「알레르기 질환 유병률 조사」는 1995년도와 2000년도에 실시되었지만, 서울과 지방 8개 도시로 조사대상을 국한하여 우리나라 전체를 대표할 수 없었다. 2010년도 조사에서는 전국 단위의 모집단 대표성을 확보하도록 표본설계하여 전체 모집단을 대표할 수 있는 알레르기 질환 유병률 지표를 산출할 수 있도록 하였다.

둘째, 「알레르기 질환 유병률 조사」는 국제적 비교 목적뿐만 아니라 우리나라 내의 지역 간 비교와 도시규모간 비교 등도 중요하다. 따라서 층화방법과 표본배분법의 결정 과정에서는 정확도 높은 전국 추정 결과뿐만 아니라 지역 또는 도시규모간 비교도 가능하도록 하였다.

셋째, 복합표본설계(complex sampling design)를 반영한 가중치 작성 및 추정법을 제시하였다. 2010년 「알레르기 질환 유병률 조사」는 층화, 불균등추출확률, 다단계 추출 등이 복합적으로 사용된 복합표본설계에 따라 조사대상자가 선정된다. 따라서 복합표본설계를 반영한 가중치 작성과 추정법에 따라 알레르기 질환 유병률 지표가 산출되어야 한다.

2. 모집단 현황 및 층화

새로운 표본설계에서 1차추출단위는 학교이고, 2차추출단위는 학급이다. 각 표본학교에서 원칙적으로 3개 학급을 표본으로 추출하여 학급 내 전체 학생을 조사한다.

새로운 표본설계에서 고려한 층화변수는 권역 구분, 각 권역 내 도시규모(인구 및 인구밀도 고려), 아파트 가구 비율 등이다. 우선 각 권역별 통계생산을 위해서 전국의 7개 특별·광역시와 9개 도 지역을 7개 권역으로 층화하였다. 광역 생활권역과 인구규모 등을 고려하여 5개 권역(서울, 경기·인천, 충청, 호남, 영남)으로 구분하고, 다른 시·도와 지역적 특성이 다른 강원도와 제주도는 독립된 층으로 구성하였다.

서울은 인구 규모와 지역적 차이를 고려하여 4개 지역(북동, 남동, 남서, 북서)으로 세부 층화하였다. 각 권역에서 세부층은 해당 권역 내의 광역시를 하나의 층으로 구성하고, 나머지 시·군은 인구밀도에 따라 2개 또는 3개의 층으로 구성하였다. 다만, 광역시에 속한 군 지역이나 인구밀도가 낮은 구(부산 강서구와 기장군, 대구 달성군, 인천 강화군과 용진군, 울산 울주군)에 대해서는 해당 권역의 다른 시·군과 묶어 같은 기준으로 세부 층을 구성하였다. 강원도와 제주도는 전체 학생수가 적기 때문에 세부 층을 구성하지 않았다.

〈표 3〉 권역 및 세부 층별 초등학교 및 중학교 1학년생 현황

(단위: 개, 명)

권역	세부층	초등학교 1학년			중학교 1학년		
		학교수	학급수	학생수	학교수	학급수	학생수
서울	북동	171	840	24,501	105	991	33,388
	북서	129	575	16,558	84	657	21,693
	남동	113	541	15,861	80	736	25,853
	남서	171	844	24,856	104	977	33,572
경기·인천	인천광역시	199	886	25,129	111	942	36,377
	고밀집	644	2,582	83,791	344	3,098	117,191
	중밀집	388	1,056	31,137	165	1,003	35,979
	저밀집	151	244	6,333	63	249	7,531
강원	강원	418	590	14,298	164	632	19,397
충청권	대전광역시	139	520	15,516	86	633	22,003
	중밀집	179	567	16,882	84	696	23,582
	저밀집	549	789	17,501	240	796	23,474
호남권	광주광역시	147	529	15,815	84	618	23,087
	중밀집	318	749	20,331	137	874	29,680
	저밀집	646	811	15,064	326	769	20,542
영남권	부산, 대구, 울산	536	1,996	56,104	317	2,617	90,744
	중밀집	557	1,535	43,762	268	1,788	60,329
	저밀집	625	818	16,537	333	858	23,408
제주	제주	117	235	6,292	42	230	8,294
전국		6,197	16,707	466,268	3,137	19,164	656,124

〈표 3〉은 권역 및 세부 층별 초등학교 및 중학교 모집단 현황을 정리한 것이다. 초등학교 1학년생 조사에서 전체 조사대상 학교수는 6,197개교이다. 초등학교 1학년 학생수를 기준하여 볼 때 경기·인천의 고밀집 층과 영남권의 광역시 층이 각각 18.0%와 12.0%로 높은 비중을 보이고 있다. 전체적으로 인구밀도가 낮은 시·군의 1학년 학생의 학교당 학급수는 2개 미만이고, 학교의 1학년 학생수도 30명 미만인 경우가 상당수 있다.

중학생 대상 조사에서 조사대상 전체 학교수는 3,137개교이다. 중학교 1학년 학생수를 기준하여 볼 때 경기·인천의 고밀집 층과 영남권의 광역시 층이 각각 17.9%와 13.8%로 높은 비중을 보이고 있다. 학교수, 학급수 및 학생수에 대한 층별 분포 양상은 전체적으로 초등학교 모집단과 유사하다.

3. 표본크기 및 표본배분

2010년 「알레르기 질환 유병률 조사」의 목표 표본크기는 초등학교 1학년 4,000명, 중학교 1학년생 4,000명으로 전체 8,000명이다. 초등학교 1학년의 학급당 평균 학생수는 세부층에 따라 차이가 있다(〈표 3〉참고). 저밀집 층의 학급당 학생수는 25명 미만으로 적고, 광역시 층과 고밀집 층의 학급당 학생수는 28명 이상으로 상대적으로 많다. 초등학교 1학년 대상의 목표 표본크기 4,000명을 조사하기 위해서 초등학교 1학년생 대상 표본 학교수는 45개 학교로 결정하였고, 중학교 1학년생 대상 표본 학교수는 40개 학교로 결정하였다.

「알레르기 질환 유병률 조사」는 주요 알레르기 질환에 대한 유병률 추정을 목적으로 한다. 이 조사의 중요한 추정 모수인 모비율 추정에 대한 95% 신뢰수준의 최대 허용오차 한계는 초등학교 1학년생에 대한 전국 추정의 경우는 약 1.9%p, 중학교 1학년생에 대한 전국 추정의 경우는 약 2.2%p로 예상된다. 모비율에 대한 예상 최대 허용오차 한계는 〈표 1〉과 〈표 2〉의 설계효과를 기초로 산출하였다.

초등학교 및 중학교 1학년 대상 조사의 각 권역별 최종 표본 학교수는 〈표 4〉와 같다. 각각 초등학교와 중학교에 대해서 각 권역별 학급수의 제공근에 비례배정하는 것을 원칙으로 하되, 권역 내 세부 층의 수와 서울과 지방의 유병률 비교가 가능하도록 일부 층의 표본 학교수를 조정하였다. 이를 통해 강원과 제주를 제외한 나머지 각 권역에서 초등학교생과 중학생에 대해 각각 500명 이상의 응답자를 얻을 수 있었다. 각 권역에서 세부층별 표본배분은 세부층별 학생수에 비례배정하는 것을 원칙으로 하였다.

4. 표본추출

2010년 「알레르기 질환 유병률 조사」의 표본추출법은 층화2단계확률비례계통추출법을 적용하였다. 먼저 1차추출단위인 표본학교는 각 층에서 해당 학교 내 학급수에 비례하는 확률비례계통추출법에 따라 추출하였다. 표본학교에서는 최대 3개 학급까지 표본학급을 선정하였고, 선정된 표본학급 내의 전체 학생을 조사하였다.

해당 초등학교 또는 중학교에서 조사 대상 학년의 전체 학생수가 20명 미만인 경우에는 표본추출 대상에서 제외하였다. 조사 대상 학년의 전체 학생수가 20명 미만인 학교를 표본추출에서 제외하더라도 전체 초등학교 1학년생에 대한 포괄비율(coverage rate)은 96.5%이고, 중학교 1학년생에 대한 포괄비율은 99.2%로 대단히 높다.

〈표 4〉 각 층별 표본크기 및 표본배분 현황

(단위: 개)

권역	세부층	초등학교 1학년 대상			중학교 1학년 대상		
		학교수	학급수	표본학교	학교수	학급수	표본학교
서울	북동	171	840	3	105	991	2
	북서	129	575	2	84	657	2
	남동	113	541	2	80	736	2
	남서	171	844	3	104	977	2
경기·인천	인천광역시	199	886	2	111	942	2
	고밀집	644	2,582	5	344	3,098	4
	중·저밀집	539	1,300	3	228	1,252	2
강원	강원	418	590	2	164	632	2
충청	대전광역시	139	520	2	86	633	2
	중밀집	179	567	2	84	696	2
	저밀집	549	789	2	240	796	2
호남	광주광역시	147	529	2	84	618	2
	중밀집	318	749	2	137	874	2
	저밀집	646	811	2	326	769	2
영남	부산,대구,울산	536	1,996	4	317	2,617	4
	중밀집	557	1,535	3	268	1,788	2
	저밀집	625	818	2	333	858	2
제주	제주	117	235	2	42	230	2
전국		6,197	16,707	45	3,137	19,164	40

IV. 2010년 「알레르기 질환 유병률 조사」의 설계효과 및 급내상관계수 분석

1. 표본대체 및 응답률 현황

초등학교 1학년생 대상의 조사에서 표본으로 선정된 45개 학교 중에서 8개교에서는 조사를 진행할 수 없어 같은 층의 예비 표본학교로 대체하였다. 표본대체 사유로는 조사 협조를 거부한 경우(7개교)와 표본학교가 소재한 지역 특성상 결손가정이 많아서 실제적인 조사가 어려운 경우(1개교)로 나뉜다. 표본학교에 대한 원표본 유지율은 82%로 높은 수준이다. 표본대체가 이루어진 지역은 서울, 경기, 인천 등 수도권 지역이었다. 중학교 1학년생 대상의 조사에서 전체 표본 학교수는 40개교이고, 이 중 8개교에서는 조사 협조를 거부하여 같은 층의 예비 표본학교로 대체하였다. 표본학교에 대한 원표본 유지율은 80%로 높은 수준이다.

표본학교에서 추출된 표본학급 내 전체 조사대상자에 대한 전체 응답률은 92.1%로 대단히 높은 수준이다. 지역별 현황을 보면 충청권의 대전광역시와 중밀집 층의 응답률이 각각 83.2%와 81.4%로 상대적으로 낮은 수준이지만 전반적으로 90% 이상으로 높은 수준이다 (<표 5> 참고).

중학교 대상 조사에서 표본으로 추출된 전체 조사대상자에 대한 전체 응답률은 93.8%로 대단히 높은 수준이다. 초등학교 1학년생의 응답률에 비해 전반적으로 높게 나타났다는데, 이는 초등학생의 경우에는 해당 학생의 학부모가 응답해야 하지만 중학교 1학년생의 경우에는 본인이 직접 응답하기 때문인 것으로 판단된다. 세부 층의 응답률 현황을 보면 전체적으로 높은 수준을 보이고 있으며, 상대적으로 호남권 중밀집 층의 응답률이 상대적으로 낮지만 대체로 높은 수준이다(<표 5> 참고).

2. 가중치 작성 및 추정

「알레르기 질환 유병률 조사」는 층화, 집락화 및 다단계추출 등이 적용된 복합표본 설계에 따라 추출되었다. 이 조사에 대한 가중치는 설계가중치, 무응답 조정, 사후층화 조정의 과정을 통해서 산출되었다.

〈표 5〉 각 층별 응답률 및 응답자 현황

(단위: 명, %)

권역	세부 층	초등학교 1학년 대상			중학교 1학년 대상		
		조사대상	응답자	응답률	조사대상	응답자	응답률
서울	북동	286	254	88.8	210	206	98.1
	북서	277	248	89.5	200	197	98.5
	남동	188	177	94.1	234	199	85.0
	남서	271	248	91.5	196	182	92.9
경기·인천	인천광역시	212	195	92.0	230	213	92.6
	고밀집	564	512	90.8	465	451	97.0
	중·저밀집	332	310	93.4	221	210	95.0
강원	강원	198	168	84.8	212	196	92.5
충청	대전광역시	238	198	83.2	239	218	91.2
	중밀집	199	162	81.4	218	215	98.6
	저밀집	163	162	99.4	233	223	95.7
호남	광주광역시	191	186	97.4	224	222	99.1
	중밀집	197	186	94.4	220	191	86.8
	저밀집	127	122	96.1	214	197	92.1
영남	부산, 대구, 울산	369	357	96.7	434	387	89.2
	중밀집	214	208	97.2	215	204	94.9
	저밀집	169	164	97.0	188	184	97.9
제주	제주	153	146	95.4	231	217	93.9
전국		4,348	4,003	92.1	4,384	4,112	93.8

표본추출과정에서 1차추출단위인 표본학교는 해당 층에서 크기측도(measure of size)인 학급수에 비례하는 확률비례계통추출법에 따라 추출되었고, 각 표본학교에서는 3개 학급을 추출하여 표본학급 내 전체 학생을 조사하도록 하였다. 따라서 앞선 표본추출과정을 통해서 추출된 h 번째 층 내 i 번째 표본학교의 j 번째 표본학급에 속한 k 번째 초등학교 또는 중학교 1학년 학생에 대한 추출가중치(sampling weight)는 다음과 같다.

$$w_{hijk} = \left[\left(\frac{n_h M_{hi}}{\sum_{l=1}^{N_h} M_{hl}} \right) \times \frac{m_{hi}}{M_{hi}} \right]^{-1} = \frac{\sum_{l=1}^{N_h} M_{hl}}{n_h \times m_{hi}} \quad (3)$$

여기서 N_h 는 h 번째층의 모집단 전체 학교수, n_h 는 h 번째층의 표본 학교수, M_{hi} 는 h 번째층 내 i 번째 학교의 전체 학급 수, m_{hi} 는 h 번째층 내 i 번째 표본학교의 표본 학급 수이다.

본 연구에서 무응답 조정은 지역층, 학교, 학년, 학급 등의 차이에 따른 응답률의 차이를 보정하기 위해서 이루어졌다. 무응답 조정셀은 각 지역층의 표본학교 내 표본학급을 이용하였다.

2010년도 「알레르기 질환 유병률 조사」의 사후층화는 2010년 5월 기준으로 작성된 「학교 통계 현황(한국교육개발원)」을 이용하였다. 본 연구에서 사용된 사후층은 각각 초등학교 1학년과 중학교 1학년에 대하여 권역 × (남, 여) 구분을 이용하였다. 이 조사에서 사후층화 보정은 추정량의 편향을 줄여주고, 응답자의 인구사회학적 특성 분포를 최신 모집단 정보와 일치시키는 역할을 한다.

「알레르기 질환 유병률 조사」에서 추정대상은 대부분 모비율이다. 본 조사에서 각종 모비율 추정을 위해서 사용된 가중치를 이용한 추정량은 다음과 같이 정의된다.

$$\hat{p} = \frac{\sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}}{\sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}} = \frac{\sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}}{w_{...}} \quad (4)$$

여기서 w_{hij} 는 각 응답자에 부여된 가중치이고, y_{hij} 는 각 응답결과로 특정 속성을 갖고 있는 경우에는 1, 아니면 0의 값을 갖는다. L 은 층의 수, n_h 는 층 h 에서의 표본학교 수, m_{hi} 는 층 h 내 i 번째 표본학교의 응답자 수이다. $w_{...} = \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}$ 은 전체 응답자에 대한 가중값의 합계이다.

앞서 제시한 모비율 추정량에 대해서 층화와 2단집락추출 등의 표본설계를 반영한 추정분산은 다음과 같이 계산된다.

$$\text{var}(\hat{p}) = \sum_{h=1}^L \frac{n_h (1 - f_h)}{n_h - 1} \sum_{i=1}^{n_h} (e_{hi} - \bar{e}_{h..})^2 \quad (5)$$

여기서, $f_h = n_h/N_h$, $e_{hi.} = \sum_{j=1}^{m_{hi}} w_{hij} y_{hij} / w_{...}$, $\bar{e}_{h..} = \left(\sum_{i=1}^{n_h} e_{hi.} \right) / n_h$ 이다.

3. 주요 추정결과에 대한 설계효과와 급내상관계수 분석

〈표 6〉부터 〈표 8〉은 2010년 「알레르기 질환 유병률 조사」의 주요 알레르기 질환에 대한 유병률, 설계효과, 급내상관계수를 정리한 것이다. 본 연구에서 추정량의 분산 계산과 설계효과의 분석은 통계분석 소프트웨어인 SAS Ver. 9.1(SAS Institute, NC, USA)의 복합표본조사 분석용 모듈을 이용하였다. 한편, 급내상관계수(intraclass correlation coefficient)는 앞서 제시한 식 (2)에 따라 구하였다. 초등학교와 중학교 대상 조사에서 표본학교 당 평균 응답자 수는 각각 89.0명과 102.8명이고, 가중치에 대한 변동계수는 각각 51.8%와 47.1%이다. 권역별 알레르기 질환 유병률 추정결과는 안강모 등(2011)의 연구를 참고하기 바란다.

일반적으로 설계효과는 불균등 가중치나 표본학교에서 조사되는 평균 학생수 등의 영향을 받기 때문에 급내상관계수를 통해 각 알레르기 질환의 유병률 추정에서 학교 내 학생들의 유사성의 정도를 비교·평가할 수 있다.

〈표 6〉 천식(호흡기질환 평가 문항) 유병률 및 오차의 한계

항목(결과)	초등학교			중학교		
	유병률	설계효과	급내상관계수	유병률	설계효과	급내상관계수
현재까지 천식증상 여부(sx ever ¹⁾)	19.1%	1.33	0.0005	15.4%	1.04	- 0.0015
12개월 이내 천식증상 여부(12m sx ²⁾)	10.3%	1.26	- 0.0001	8.2%	1.05	- 0.0014
현재까지 천식 진단 여부(dx ever ³⁾)	10.5%	3.88	0.0234	7.5%	1.31	0.0007
12개월 이내 천식 치료 여부(12m tr ⁴⁾)	4.1%	1.65	0.0034	1.8%	1.13	- 0.0007
운동 후 쌹쌹소리 여부(12m exer ⁵⁾)	3.0%	1.75	0.0043	7.3%	1.36	0.0011
마른 기침 (12m dry ⁶⁾)	18.9%	1.86	0.0053	13.3%	2.05	0.0067

1) sx ever: 현재까지 한번이라도 천식증상(숨 쉴 때 가슴이 쌹쌹거리는 소리나 휘파람 소리) 여부

2) 12m sx: 최근 12개월 동안 한번이라도 천식증상(숨 쉴 때 가슴이 쌹쌹거리는 소리나 휘파람 소리) 여부

3) dx ever: 천식 진단 여부

4) 12m tr: 최근 12개월 동안 천식 치료 여부

5) 12m exer: 최근 12개월 동안 운동 후 쌹쌹소리 여부

6) 12m dry: 최근 12개월 동안 마른기침 여부

〈표 7〉 알레르기 비염(코질환) 유병률 및 오차의 한계

항목(결과)	초등학생			중학생		
	유병률	설계효과	급내상관계수	유병률	설계효과	급내상관계수
현재까지 알레르기 비염 증상 여부 (sx ever ¹⁾)	47.4%	2.06	0.0071	47.5%	1.31	0.0007
최근 12개월 이내 알레르기 비염증상 여부1(12m sx type I ²⁾)	44.3%	2.04	0.0069	41.8%	1.69	0.0038
최근 12개월 이내 알레르기 비염증상 여부2(12m sx type II ³⁾)	18.4%	2.93	0.0149	18.6%	1.53	0.0025
알레르기 비염 진단 여부(dx ever ⁴⁾)	37.7%	6.98	0.0512	29.9%	1.63	0.0033
최근 12개월 이내 알레르기 비염 치료 여부(12m tr ⁵⁾)	29.4%	5.89	0.0414	19.8%	2.07	0.0068

1) sx ever: 현재까지 한번이라도 감기나 독감 없이 재채기, 콧물 또는 코막힘 증상 여부

2) 12m sx type I: 최근 12개월 동안 감기나 독감 없이 재채기, 콧물 또는 코막힘 증상 여부

3) 12m sx type II: 최근 12개월 동안 재채기, 콧물, 코막힘 증상, 눈이 가렵고 눈물 나는 증상 여부

4) dx ever: 알레르기 비염 진단 여부

5) 12m tr: 최근 12개월 이내 알레르기 비염 치료 여부

〈표 6〉은 천식 관련 유병률 추정에 대한 설계효과와 급내상관계수 현황이다. 대체로 초등학생 조사에 비해서 중학생 조사의 설계효과가 다소 작게 나타나고 있다. 급내상관계수 현황을 보면 초등학교 1학년생의 ‘현재까지 천식 진단 여부’ 유병률 추정을 제외하면 전체적으로 0근처의 값을 보이고 있다.

〈표 7〉은 알레르기비염에 대한 유병률, 설계효과와 급내상관계수 현황이다. 초등학생 대상의 조사에서 알레르기 비염 진단 관련 유병률 추정에 대한 설계효과는 중학생 조사에 비해 크게 나타났다. 특히 초등학교 조사에서 ‘알레르기 비염 진단 여부’와 ‘최근 12개월 이내 알레르기 비염 진단 여부’에 대한 설계효과가 다른 조사항목에 비해 상당히 크다. 급내상관계수는 이들 두 조사항목 추정의 경우를 제외하면 0근처의 값을 보이고 있다.

일반적으로 설계효과가 크게 나타나는 것은 표집오차 구성요소 중 1차추출단위인 학교 간 변동은 크고, 학교 내 변동은 작은 경우에 발생한다. 이렇게 설계효과가 크게 발생하는 이유에 대해서는 다음 두 가지 관점에서 살펴봐야 한다.

첫째, 설계효과가 아주 큰 값이나 작은 값을 보이는 경우는 잠재적인 표본설계나 데이터 질 문제를 포함한 데이터 관련 문제가 있을 수 있음을 시사한다. 예를 들어 특이치

(outlier)의 문제나 가중치가 지나치게 큼으로써 발생하는 문제가 있을 수 있음을 시사한다(Lacher et al. 2004). 이와 같은 관점에서 조사항목 ‘알레르기 비염 진단 여부(dx ever)’와 ‘최근 12개월 이내 알레르기 비염 치료 여부(12m tr)’ 조사항목과 가중치의 관계를 살펴볼 때 가중치에 의한 특이치 발생은 없었다.

둘째, 알레르기 비염 질환의 유병률과 그 지역의 환경적 요인과의 관련 가능성에 대한 추가 연구가 필요하다. <표 7>의 분석결과를 보면 다른 질환에 비해서 초등학생의 알레르기 비염 유병률 추정에 대한 설계효과가 높게 나타나고 있다. 알레르기 비염 진단과 치료 여부는 알레르기 질환의 강도(심각성)를 나타낸다고 볼 수 있다. 급내상관계수 측면에서 보면 높은 수치는 아니지만, 학교별 진단 및 치료 여부에 대한 비율 추정결과와 지역적 특성의 연관관계에 대한 추가적인 연구가 요구된다.

<표 8>은 아토피 피부염에 대한 유병률, 설계효과, 급내상관계수 현황이다. 초등학교 조사에서 ‘아토피 피부염 진단 여부’에 대한 설계효과가 다른 조사항목에 비해서 상당히 크게 나타나고 있는 것을 제외하면 초등학생 조사와 중학생 조사는 설계효과 측면에서 비슷한 수치를 보이고 있다. 앞선 <표 6>과 <표 7>에서도 천식이나 알레르기 비염의 경우에는 진단 여부와 치료 여부 추정에 대한 설계효과는 증상 여부 추정에 대한 설계효과에 비해 상대적으로 큰 값을 보이고 있다. 이는 지역별로 병원이나 진료시설에 대한 접근 편리성과 생활수준에 차이가 있기 때문에 각 질환의 진단이나 치료에 대한 지역적 유사

〈표 8〉 아토피 피부염(피부질환) 유병률 및 오차의 한계

항목(결과)	초등학생			중학생		
	유병률	설계효과	급내상관계수	유병률	설계효과	급내상관계수
현재까지 아토피 피부염 증상 여부 (sx ever ¹⁾)	26.9%	1.85	0.0052	19.9%	1.70	0.0038
최근 12개월 이내 아토피 피부염 증상 여부(12m sx ²⁾)	20.5%	1.65	0.0034	12.6%	1.87	0.0052
아토피 피부염 진단 여부 (dx ever ³⁾)	35.5%	3.40	0.0191	24.0%	1.68	0.0037
최근 12개월 이내 아토피 피부염 치료 여부(12m tr ⁴⁾)	15.7%	1.65	0.0034	8.6%	1.53	0.0025

1) sx ever : 현재까지 한번이라도 아토피 피부염 증상 여부

2) 12m sx : 최근 12개월 동안 아토피 피부염 증상유형

3) dx ever : 아토피 피부염 진단 여부

4) 12m tr : 최근 12개월 동안 아토피 피부염 치료 여부

성이 증상 여부에 대한 유사성에 비해서 높게 나타난 것으로 해석할 수 있다. 앞서 제시한 <표 1>과 <표 2>의 아토피 피부염 관련 유병률 추정에 대한 설계효과 분석에서도 이와 유사한 경향을 발견할 수 있다.

2010년 「알레르기 질환 유병률 조사」에 대한 설계효과 및 급내상관계수 분석결과에 대한 시사점을 정리하면 다음과 같다.

첫째, 설계효과와 급내상관계수 분석결과를 보면 초등학교 조사의 ‘알레르기 비염 진단 여부’와 ‘최근 12개월 이내 알레르기 비염 진단 여부’에 대한 설계효과를 제외하면 전체적으로 낮은 수준이다. 이는 표본 학교에서 약 100명씩을 조사하는 현행 표본설계가 알레르기 질환 유병률 추정 측면에서 효율적임을 알 수 있다. 향후 표본설계에서는 앞서 구한 급내상관계수 분석결과를 이용하여 표본 학교 내 최적 조사대상자 수를 결정하여 활용할 수 있을 것이다.

둘째, 초등학교에 비해서 중학생 조사에서 설계효과와 급내상관계수가 낮게 나타나고 있다. 향후 표본설계에서도 현행 표본설계의 기본틀을 유지하여 초등학교 표본 학교수에 비해서 중학교의 표본 학교수를 줄이는 것이 추정의 정확도 측면에서 바람직하다는 점을 알 수 있다.

셋째, 천식이나 아토피 피부염 질환 유병률 추정에 대한 설계효과에 비해 알레르기 비염 질환 유병률 추정의 설계효과가 높게 나타나고 있다. 이는 알레르기 비염 질환의 유병률과 그 지역의 환경적 요인과의 관련 가능성에 대한 추가 연구가 필요함을 시사한다.

V. 결론

본 연구에서는 천식 및 알레르기 질환에 대한 전국적 역학조사를 가능하게 하는 표본설계와 2010년 「알레르기 질환 유병률 조사」 결과를 분석하여 현행 조사에 대한 점검과 향후 연구방향을 제시하였다.

알레르기 질환에 대한 성공적인 정책수립을 위해서는 현재의 알레르기 질환 유병률, 과거부터 현재까지의 유병률 변화 추이, 알레르기 질환 발생 및 유병률 변화의 원인 규명 등 국제적으로 비교 가능한 정확한 자료수집과 통계결과의 산출이 필요하다. 2010년 「알레르기 질환 유병률 조사」는 국제적으로 비교 가능한 알레르기 질환 유병률을 산출하고자 ISAAC 프로그램에 기초한 조사수행 체계에 따라 진행되었다. 이 조사를 통해서 알레르기 질환에 관한 국내의 현황을 파악하고 이를 통해 효과적인 정책 수립 및 적극적

인 대처방법을 모색할 수 있을 것이다. 아울러 이번 연구에서 확보되는 신뢰성 있는 자료는 앞으로 2000년 이후의 유병률 변화 추이 및 알레르기 질환의 발생과 관련한 위험인자 분석에 이용될 것이다.

2010년 「알레르기 질환 유병률 조사」 결과로부터 산출된 각종 알레르기 질환에 대한 유병률 추정결과는 우리나라 전국을 대표하고 국제적으로 비교 가능하다는 점에서 큰 의의를 갖는다. 또한, 국제적으로 공인된 도구를 사용함으로써 알레르기 질환에 관한 외국 자료와의 비교·분석뿐만 아니라 국제적 공동연구에 참여할 수 있고 이를 통해 관련 산업개발 및 지원을 위한 기초자료로 활용할 수 있게 될 것이다.

참고문헌

- 안강모 · 김지현 · 권호장 · 채유미 · 함명일 · 이기재 · 박용민 · 이소연 · 한만용 · 김우경. 2011. “2010년 한국 어린이·청소년의 천식, 알레르기비결막염, 아토피 피부염 증상 유병률 조사: 복합표본설계를 이용한 전국 서베이.” 《대한의사협회지》 54(7): 769-778.
- 오재원 · 김규언 · 편복양 외 18인. 2003. “1995년과 2000년 학동기와 2003년 학동전기 소아에서의 아토피 피부염의 역학적 변화에 관한 전국적인 연구.” 《소아알레르기 및 호흡기》 13(4): 227-237.
- Asher, M.I., U. Keil, H.R. Anderson, R. Beasley, J. Crane, F. Martinez, E.A. Mitchell, N. Pearce, B. Sibbald, A.W. Stewart, D. Strachan, S.K. Weiland, and H.C. Williams. 1995. “International Study of Asthma and Allergies in Childhood (ISAAC): Rationale and Methods.” *European Respiratory Journal* 8(3): 483-491.
- Kish, L. 1995. “Methods for Design Effects.” *Journal of Official Statistics* 11(1): 55-77.
- Lacher, A.A., L.R. Curtin and J.P. Hughes. 2004. “Why Large Design Effects Can Occur in Complex Sample Designs: Examples from the NHANES 1999-2000 Survey.” *Proceedings of the Section on Survey Research Methods, American Statistical Association* 3841-3846.
- Verma, V., C. Scott and C. O’Muircheartaigh. 1980. “Sample Design and Sampling Errors for the World Fertility Survey.” *Journal of the Royal Statistical Society Series A* 143(4): 431-473.