

Capture-recapture 방법을 이용한 1998년 제주도 볼거리 유행시 보고 자료의 완전성 평가

김명희, 박진경, 기모란¹⁾, 허영주²⁾, 최보율, 김정순³⁾

한양대학교 의과대학 예방의학교실, 을지의과대학 예방의학교실¹⁾, 국립보건원 방역과²⁾, 서울대학교 보건대학원³⁾

Evaluation of the Completeness of Case Reporting during the 1998 Cheju-do Mumps Epidemic, Using Capture-recapture Methods

Myoung Hee Kim, Jin Kyoung Park, Mo Ran Ki¹⁾, Young Joo Hur²⁾, Bo Youl Choi, Joung Soon Kim³⁾

Department of Preventive Medicine, College of Medicine, Hanyang University;

Department of Preventive Medicine, Eulji Medical College¹⁾;

Division of Communicable Disease Control, National Institute of Health²⁾; School of Public Health, Seoul National University³⁾

Objectives : To estimate mumps incidence during the study period and to evaluate the completeness of case reporting.

Methods : Capture-recapture methods, originally developed for counting wildlife animals, were used. The data sources were 1) the National Notifiable Communicable Disease Reporting System (NNCDRS; 848 cases), 2) the School Health Reporting System, temporarily administered by the Division of Education (SHRS; 1,026 cases), and 3) a survey of students (785 cases). We estimated the number of unobserved mumps cases by matching the three data sources and fitting loglinear models to the data. We then determined the estimated total number of mumps cases by adding this to the number of observed cases. Completeness was defined as the proportion of observed cases from each source to the total of estimated cases.

Results : The total number of observed cases was 1,844 and the total number of estimated cases was 1,935 (95% CI: 1,878~2,070). The overall completeness was 43.8% of the NNCDRS, 53.0% of the SHRS, and 40.6% of the survey. However, completeness varied by area and age.

Conclusion : Although the completeness of NNCDRS data appeared higher than in the past, it is difficult to generalize this result. In Korea, it is possible to estimate the size of health hazards relatively cheaply and quickly, by applying capture-recapture methods to various data using a multiple data collection system.

Korean J Prev Med 2000;33(3):313-322

Key Words: Capture-recapture, Completeness, National Notifiable Communicable Disease Reporting System

서론

질병 발생과 유병 상태 등 어떤 건강 상태의 정확한 규모를 파악하는 것은 보건학 관련 연구와 정책 결정의 첫 단계이자 가장 중요한 단계라고 할 수 있다. 그래서 많은 연구자들과 정책 결정자들은 정확하고 대표성 있는 자료를 수집하기 위해 각종 보고체계(reporting system)와 감시체계(surveillance system), 등록제도

(registration), 역학조사(epidemiologic survey), 지역사회 연구(community research) 등 여러 방법을 이용해 왔다(김정순, 1990; Thacker와 Stroup, 1994; 김정순, 1996; Stroup과 Teutsch, 1998).

대표적인 자료수집 방법이라 할 수 있는 법정전염병의 신고나 출생·사망 신고자료 등은 수동적 보고체계로서 비교적 적은 비용과 노력으로 자료를 수집할 수 있다는 장점이 있는 반면 자료의 정확

성이나 완전성(completeness)이 높지 않다는 문제가 있다(LaPorte, 1994). 그래서 이 자료만으로 연구를 진행하거나 정책 결정을 내리는 것은 심각한 오류를 낳을 수 있으며(Modesitt 등, 1990), 이러한 문제점을 보완하기 위하여 보고율이나 자료의 정확도, 자료 수집의 효율성 등을 평가하고 다른 자료를 통해 이를 보완하기도 한다.

우리 나라의 대표적 국가보고체계인 법정전염병 신고 자료(Nationally Notifiable Communicable Disease Reporting

System; NNCDRS)도 국내의 여러 연구를 통해 정확성과 완전성이 낮다는 것이 지적되어 왔다(김윤구 등, 1985; 김정순, 1987; 정태화, 1987; 맹광호, 1996; 신의철 등, 1996)

이렇듯 지역사회에서 발생하는 건강장해의 수를 완전히 헤아린다는 것(complete enumeration)은 매우 비용이 많이 들고 비효율적일 뿐 아니라 어떤 빼어난 방법을 사용해도 실질적으로 불가능한 일이다. 동물생태학 분야에서는 야생동물의 서식 규모를 알기 위해 모든 동물을 잡아 숫자를 헤아릴 수 없다는 점을 인정하고, 확보한 자료만으로 전체를 추정하는 방법을 모색해 왔는데 그 중의 하나가 capture-recapture 방법이다(LaPorte 등, 1995). 이는 1894년에 Peterson이 연못의 물고기 수를 추정하는데 처음 사용한 이래 생태학 분야에서 널리 활용되어 왔으며, 1949년 Sekar와 Deming이 인구집단의 출생률과 사망률 추정에 이를 적용하였고, 1960년대 말 Wittes 등에 의해 역학 분야에서 적극적으로 이용되기 시작하였다(Wittes 등, 1968; Wittes 등, 1974; International Working Group for Disease Monitoring & Forecasting, 1995). 보건학 분야에서는, 반복된 포획을 통해 동물의 숫자를 추정하는 것과는 달리, 불완전한 두 가지 이상의 자료원을 통하여 한 자료원에서만 파악된 사람들의 숫자, 둘 이상의 자료원에서 동시에 파악된 사람들의 숫자를 이용하여 전체 수를 추정한다. 특히 일반적 보고자료나 조사자료에서 누락되기 쉬운 약물 남용과 정신 질환, 노숙자 등의 규모를 추정하는데 유용하다고 알려져 있다(McKeganey 등, 1992; Fisher 등, 1994; Neugebauer, 1994). 또한 추정된 전체 건수를 토대로 하여 각종 보고·등록·감시체계 자료의 완전성을 평가하고, 가장 효율적으로 완전성 있는 자료를 산출할 수 있는 자료수집체계의 조합을 찾아내는데 활용할 수 있다(Hook와 Regal, 1992; McCarty 등, 1993; LaPorte 등, 1995).

capture-recapture 방법을 활용하기 위해서는 다음과 같은 몇 가지 기본 가정이

전제되어야 한다. 첫째, 연구 기간 동안 인구의 변화가 없어야 한다(closed population). 둘째, 고유한 개인 식별자가 존재하여(identifiability) 자료를 짝짓기할 수 있어야 하며, 셋째, 특정 자료원에 대하여 각 개인이 표집될 확률은 동일해야 한다(homogeneous catchability of individuals). 넷째, 각 자료원들은 독립적이어야 한다(International Working Group for Disease Monitoring & Forecasting, 1995).

하지만 이상의 조건들을 현실에서 모두 만족시키기는 매우 어려운데, 이러한 문제점을 해결하기 위해 도입된 여러 수학적 모형 중 가장 일반적으로 쓰이는 것이 로그 선형 모형이다(Hook와 Regal, 1982; Agresti, 1994; Hook와 Regal, 1995b). 로그 선형 모형은 포와손(Poisson) 분포를 따르는 자료의 일반선형 모형으로서, 다차원 분할표에서 각 cell 도수를 모형화하는데 흔히 쓰이는 방법으로 범주형 자료 사이의 연관성을 기술한다(Agresti, 1996).

이 연구에서는 1998년 제주도에서 볼거리의 대규모 유행 당시 역학조사를 통해 수집한 여러 가지 자료원을 capture-recapture 방법에 적용시켜 1) 미보고(underreporting)를 감안한 발생의 규모를 추정하고, 2) 법정전염병 신고의 완전성을 평가하며, 3) 앞으로 국내 질병감시에서 capture-recapture 방법의 활용에 대해 고찰하고자 한다.

연구 방법

1. 자료의 구성

1) 법정전염병 신고자료(Nationally Notifiable Communicable Disease Reporting System; NNCDRS)

볼거리는 제 2종 전염병으로 지정되어 있어 진단 즉시 해당 보건소장을 거쳐 광역자치단체장에게 신고하도록 규정되어 있다(전염병예방법 제 4조). 1998년 1년 동안 제주도의 4개 보건소를 통해 신고된 볼거리 환자 수는 모두 1,259명으로 이는 제주도 전 지역의 모든 연령군을 포괄하

고 있으며 대부분 의료기관을 통해 보고되었다. 보고변수는 이름, 주민등록번호, 주소, 발병일 혹은 신고일, 신고 기관 등인데, 결측값이 많았고 일부 사례에서는 학교와 학년, 반 혹은 백신 접종력 등이 부가적으로 조사되어 있었다.

2) 교육청 보고자료(School Health Reporting System; SHRS)

학교를 중심으로 대규모 유행이 발생하였기 때문에 도교육청에서는 유행 초기부터 발생 현황을 조사하였는데, 이는 제주도 전역의 초·중·고등학교 전체와 일부 부설 유치원들을 포함하고 있다. 이렇게 조사된 환자 수는 1,065명이었으며 담임교사에 의한 조사였기 때문에 의료기관에서 진단을 받은 경우와 의료기관을 찾지 않은 환자들까지 포함하고 있었다. 조사 변수는 이름, 성별, 연령, 학교와 학년·반, 증상 발현일 등이었으며 일부 사례에는 백신 접종력이 포함되어 있었다.

3) 학생 역학조사 자료(Survey for students; SURVEY)

유행이 시작된 것은 신학기 시작 전후로 추측되는데 비해 중앙역학조사반의 현지 조사와 자료수집이 이루어진 시점은 6월 10일 이후였다. 이미 앞서의 두 가지 자료를 확보하였지만, 임상적 특성과 백신 접종력 등 주요한 역학적 정보들이 대부분 누락되어 있었고, 두 자료를 대조한 결과 많은 환례들이 보고에서 빠져 있을 가능성이 있었기 때문에 학교를 대상으로 환례에 대한 재조사를 실시하게 되었다. 이 때 사례 정의는 의료기관에서 진료를 받은 경우와 의료기관을 찾지 않았더라도 볼거리 주증상을 경험한 경우를 모두 포함하였다. 이렇게 조사된 환자 수는 총 921명이었는데 조사 변수는 이름, 성별, 연령, 생년월일, 주소, 학교와 학년, 반, 증상 발현일, 주요 증상과 합병증, 가족 감염여부, 백신 접종력 등이었다.

2. 자료 처리 및 분석

1) 자료의 제한과 짝짓기

먼저, 세 가지의 자료원이 포함하고 있는 대상집단과 조사 시기에 차이가 있었

기 때문에 비교 가능성을 높이기 위하여 연령을 학령기인 7~18세로 제한하고, 대상 기간은 1998년 3월 3일부터 현지 자료수집 종료 시점인 1998년 6월 20일(결측값인 경우 신고일을 근사치로 함)까지의 자료만을 분석에 사용하였다.

자료의 짝짓기는 이름을 기준으로 하여 생년월일, 주소, 학교·학년·반 중 한 가지 이상 같으면 동일인으로 하였다. 또한 이름이 다른 경우라도 생년월일과 학적사항 등이 완전히 일치하고 명백한 표기 오류가 의심되는 경우는 동일인으로 간주하였다.

2) 기초 분석

변수들에 따라 각 자료원에 수집될 확률이 다르게 나타날 가능성(heterogeneous catchability)을 고려하기 위해 각각의 자료원에 대하여 성별, 연령, 지역, 발병일에 관한 분포를 살펴보고, 여기에서 이질성의 단서가 된 변수들은 층화 분석의 층화 기준으로 하였다. 이 때 비교 척도는 각 변수별로 연구기간 중 발생률(incidence rate during the study period; IR)을 이용하였다.

3) 전체 발생 건수(N)의 추정

A,B,C 세 개의 자료원에 대하여 관찰 여부에 따라 2×2×2 분할표를 그리면 Table 1과 같다.

이 분할표에서는 보통의 로그 선형 모형과 달리 n000을 알지 못하기 때문에 marginal log-odds를 이용하여 모수를 추정할 수 없으므로 다른 방식의 parameterization이 필요하다. 이 때 특정 자료원의 주효과(main effect)를 그 자료원에서는 관찰되지 않으면서 나머지 두 자료원에서 관찰되는 일종의 대조(contrast) 효과로 정의하여, 7개의 알려진 관찰값으로

부터 로그 선형모형을 적합시켜 7개(u, u1, u2, u3, u12, u13, u23)의 모수를 추정할 수 있다(식 1). 이 때 최고차 교호작용항인 u123이 0이 된다는 가정을 통해 l000를 구할 수 있으며, 이를 지수 변환하면 비관찰값의 기대도수를 구할 수 있다 (Cormack, 1989; Cormack, 1992; Agresti, 1994; International Working Group for Disease Monitoring & Forecasting, 1995).

$$\begin{aligned}
 l_{111} &= u = \log(m_{111}) \\
 l_{011} &= u + u_1 \\
 l_{101} &= u + u_2 \\
 l_{110} &= u + u_3 \\
 l_{001} &= u + u_1 + u_2 + u_3 \\
 l_{010} &= u + u_1 + u_3 + u_{13} \\
 l_{100} &= u + u_2 + u_3 + u_{23} \\
 l_{000} &= u + u_1 + u_2 + u_3 + u_{12} + u_{13} + u_{23} + u_{123}
 \end{aligned}
 \dots(\text{식 1})$$

* notations

- l_{ijk} : the logarithm of the expected no. in each cell(i: 1st source, j : 2nd source, k : 3rd source, 1 or 0 : observed or not observed in each source)
- m_{ijk} : the expected no. in each cell
- u : the logarithm of the no expected to be in all sources
- u₁, u₂, u₃ : log odds against not appearing in each source for individuals who appear in the others(main effect)
- u₁₂, u₁₃, u₂₃, u₁₂₃ : interaction terms

적합한 모형의 선택은 상위 모형과 하위모형(nested model)의 이탈도(deviance) 차이인 G² 값(residual deviance)을 이용하였는데, 이는 두 모형에 사용된 모수의 개수 차이를 자유도로 하는 카이제곱 분포를 따른다(Agresti, 1996). 분석에는 SAS® 6.12를 이용하였으며, 포와슨 분포에 로그를 링크한 "proc genmod"를 적용하였다(Stokes 등, 1995). 모형은 다

음의 식 (2)와 같이 표현된다.

$$\begin{aligned}
 \log(m_{ijk}) &= \\
 &u + u_1^{a_i} + u_2^{b_j} + u_3^{c_k} + u_{12}^{ab} + u_{13}^{ac} + u_{23}^{bc} + u_{123}^{abc}
 \end{aligned}
 \dots(\text{식 2})$$

if observed in SHRS, a = 0 otherwise a=1
 if observed in NNCDRS, b = 0 otherwise b=1
 if observed in SURVEY, c = 0 otherwise c=1
 m_{ijk} : the expected no. in each cell

이상의 과정을 거쳐 추정된 비관찰도수 n000을 관찰도수(n)에 더한 합을 전체 발생 건수에 대한 추정값 N으로 하였으며, 신뢰구간은 profile likelihood interval 추정방법을 사용하였다(Regal과 Hook, 1984; Cormack, 1989; Buckland와 Garthwaite, 1991; Regal과 Hook, 1991).

4) 완전성 평가

완전성은 자료수집의 민감도(sensitivity)를 반영하는 것으로서 실제 발생 사례 중 보고 혹은 조사된 사례의 비율로 표시되며(Klaucke 등, 1988; Hilsenbeck 등, 1992), 각 자료원의 완전성은 식 (3)과 같이 정의하였다.

$$\text{completeness}(\%) = (\text{no. of individuals in each source}) \div N \times 100 \dots(\text{식 3})$$

5) 층화 분석

기초 분석에서 나타난 결과를 토대로 표집 확률의 이질성(heterogeneous catchability)의 원인이 될 수 있는 소집단으로 층화하고, 각 층에 대해 로그 선형모형을 다시 적합시켜 계층별 N을 추정하고 profile likelihood interval을 구하였다. 그리고 계층별 N의 합을 구하여 전체 발생 건수를 추정하였다.

Table 1. Contingency table according to three sources of data

	source B(+)		source B(-)		total
	source C(+)	source C (-)	source C (+)	source C (-)	
source A (+)*	n ₁₁₁ †	n ₁₁₀	n ₁₀₁	n ₁₀₀	n ₁₊₊
source A (-)*	n ₀₁₁	n ₀₁₀	n ₀₀₁	n ₀₀₀	n ₀₊₊
total	n ₊₁₁	n ₊₁₀	n ₊₀₁	n ₊₀₀	N

* (+) or (-); observed or unobserved in the source

† n_{ijk}; cell count of corresponding condition, ith, jth, & kth source(1 or 0; observed or unobserved)

연구 결과

1. 연구 대상 자료의 특성

연구 대상 인구인 초·중·고등학생의 숫자는 91,813명으로 제주도 전체 인구의 약 17.4%를 차지하고 있으며 초·중·고등학교 모두에서 군 지역에 비해 시 지역의 학생 수, 학교 당 재학생의 수가 많았다(Table 2).

각 자료원별로 성별, 연령, 지역, 발생일

분포를 살펴보면 포함 연령과 발생 시기의 분포가 다르고, 지역별 구성비도 다르다는 것을 알 수 있다(Table 3).

이를 7~18세까지의 연령군과 6월 20일까지의 발병자만을 대상으로 제한한 결과, 법정전염병 신고 자료 1,259건 중 848건(67.4%), 교육청 보고 자료 1,064건 중 1,026건(96.4%), 역학조사 자료 920건 중 785건(85.3%)이 분석에 포함되었으며 자료의 구성은 Fig. 1과 같다.

각 자료원으로부터 조사된 사례를 이용하여 대상기간의 지역별 발생률(/1,000)을 구한 결과, 제주시의 경우 학교를 통해 조사된 두 개 자료의 발생률에는 차이가 없었으며 법정전염병 신고자료에 의한 발생률이 나머지 두 가지에 비해 더 낮게 나타났다. 서귀포시의 경우 자료원마다 발생률간에 유의한 차이가 없었으며(p=0.000, χ^2 -test), 법정전염병 신고자료에 의한 발생률이 가장 높았다. 북

Table 2. Demographic characteristics by area

Area*	Cheju	Seoguipo	Buk-gun	Nam-gun	Total
Total population†	266,316	84,976	98,417	78,651	528,360
Elementary school†					
No of schools	26	16	43	32	117
No of students	24,409	6,778	5,677	5,244	42,108
No of students per school	939	424	132	164	360
Middle school†					
No of schools	12	7	14	8	41
No of students	12,578	3,799	3,344	2,840	22,561
No of students per school	1,048	543	239	355	550
High school†					
No of schools	11	6	6	4	27
No of students	14,374	5,434	5,072	2,264	27,144
No of students per school	1,308	906	845	566	1,005
Total					
No of schools	49	29	63	44	185
No of students	51,360	16,011	14,093	10,348	91,813

* Cheju and Seoguipo are cities. Buk-gun and Nam-gun are counties.

† data from Vital Statistics of '97 population registration, Cheju-do

‡ data from Division of Education, Cheju-do

Table 3. Demographic characteristics of population by sources*

Variables	NNCDRS	SHRS	SURVEY
Age(years)	N = 1,246	N = 1,064	N = 920
mean(±SD)	10.6(±6.0)	11.5(±3.0)	10.8(±2.9)
range	0~82	6~18	6~18
Sex(numbers, (%))	N = 1,119	N = 1,048	N = 899
male	578(51.7)	528(50.4)	465(51.7)
female	541(48.3)	520(49.6)	434(48.3)
Area(numbers, (%))	N = 1,259	N = 1,063	N = 917
Cheju	58(4.6)	162(15.2)	178(19.4)
Seoguipo	406(32.2)	77(7.2)	161(17.6)
Buk-gun	185(14.7)	408(38.4)	526(57.4)
Nam-gun	610(48.5)	416(39.1)	52(5.7)
Date of onset†	N = 1,247	N = 1,047	N = 790
median	May 1	May 7	May 10
mode	Apr 30	May 1	Apr 30
range	Mar 3~Nov 26	Mar 6~Jun 21	Mar 6~Jul 12

* NNCDRS, data from Nationally Notifiable Communicable Disease Reporting System; SHRS, data from School Health Reporting System(temporarily administered by Division of Education); SURVEY, data from questionnaire survey

† if record has no information on onset, it is substituted with date of reporting

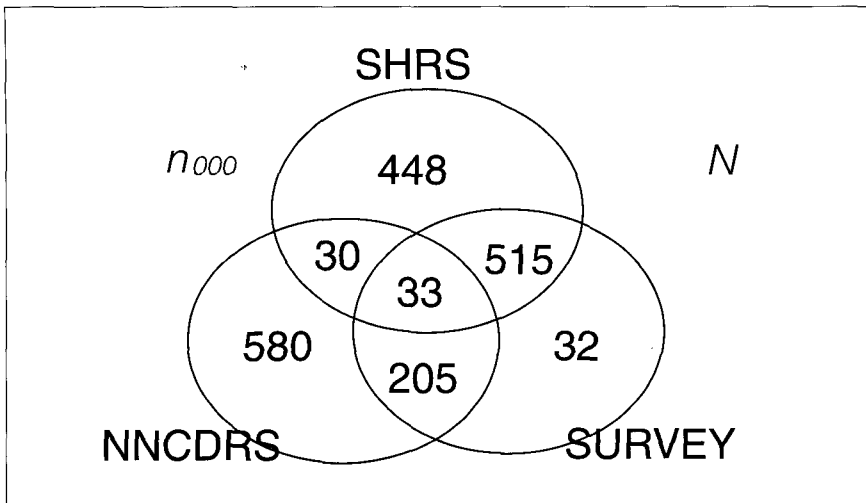


Fig. 1. Three-source combinations for estimating the size of a population.
 Note: *NNCDRS, data from Nationally Notifiable Communicable Disease Reporting System; SHRS, data from School Health Reporting System(temporarily administered by Division of Education); SURVEY, data from questionnaire survey

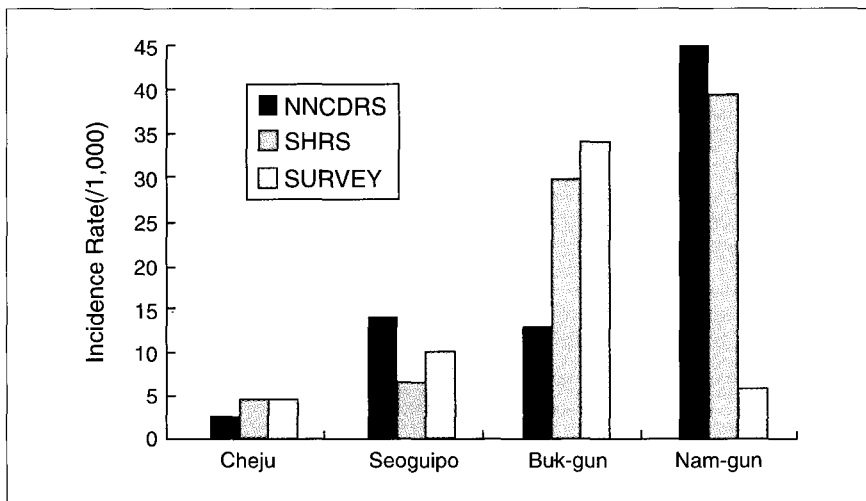


Fig. 2. Incidence rates/(1,000) during the study period from each source by area.
 Note: NNCDRS, data from Nationally Notifiable Communicable Disease Reporting System; SHRS, data from School Health Reporting System(temporarily administered by Division of Education); SURVEY, data from questionnaire survey

제주군에서도 자료원별 발생률이 다르게 나타났는데($p=0.000$, χ^2 -test), 법정전염병 신고자료에 의한 것이 가장 낮고 역학조사자료로부터 얻은 발생률이 가장 높았다. 남제주군도 마찬가지로 세 가지 자료원에 의한 발생률이 다르게 나타났으며($p=0.000$, χ^2 -test), 법정전염병 신고자료에 의한 발생률이 높고 역학조사 자료에 의한 것이 특히 낮게 나타났다(Fig. 2).

각 자료원으로부터 조사된 사례를 이용하여 대상기간의 학교 종류별 발생률(1,000)을 구한 결과 세 자료원에서 모두

초등학생이 가장 높았고, 중학생, 고등학생의 순서로 낮아졌으며($p=0.000$, χ^2 -trend test), 학령별로 자료원에 의한 발생률의 차이를 살펴보았을 때 초등학생, 중학생, 고등학생 모두 통계적으로 유의미한 차이를 보였다($p=0.009$, 0.000 , 0.002 , χ^2 -test)(Fig. 3).

자료원별로 일주일 간격의 환자 발생 분포를 살펴보았을 때 나머지 두 자료원에 비해 교육청 보고 자료에서 유행이 빨리 나타나는 것을 관찰할 수 있었다(Fig. 4).

2. 전체 발생 건 수 추정 및 완전성 평가

Table 4는 세 자료원에 대하여 로그선형모형을 적합시킨 결과이며, 증화하지 않은 상태로 자료 전체를 적합시킬 경우 최종 선택된 모형은 식 (4)의 형태로 나타났다.

$$\log(mijk) = u + u_i^a + u_j^b + u_k^c + u_{ij}^{ab} + u_{jk}^{bc} + u_{ki}^{ca} + u_{ijk}^{abc} \quad \dots(\text{식 } 4)$$

$$u = 3.5160, u_1 = 1.8070, u_2 = 2.7269,$$

$$u_3 = -0.1367, u_{12} = -4.5842, u_{13} = 1.1767$$

이 모형식에 근거하여 추정한 비관찰값은 91건이었으며 이를 관찰값과 합한 전체 발생 추정 건수는 1,935건(95% CI: 1,878~2,070)이었다.

이를 통해 자료원의 완전성을 평가한 결과는 Table 5와 같다. 법정전염병 신고자료의 경우 증화하지 않은 상태에서 43.8%의 완전성을 보였으며 지역별로 증화한 경우 시 지역이 49.8%로 군 지역의 42.3%와 유의미한 차이가 있었다($p=0.005$, χ^2 -test). 마찬가지로 초등학생과 중·고등학생으로 증화한 경우에도 완전성은 각각 47.4%와 36.6%로 유의미한 차이를 보였다($p=0.000$, χ^2 -test). 교육청 보고자료와 역학조사 자료의 경우 시 지역과 군 지역, 초등학생과 중·고등학생 계층의 완전성이 모두 유의미한 차이를 보였다($p=0.000$, χ^2 -test). 각 자료원 사이의 완전성을 비교해보았을 때 교육청 자료가 53.0%로 가장 높게 나타났으며 이어 보건소 자료 43.8%, 역학조사 자료 40.6%로 나타났는데, 이들 사이의 차이는 모두 통계적으로 모두 유의미한 것이었다($p=0.000$, χ^2 -test)(Table 5).

고 찰

자료의 완전성을 평가하기 이전에 자료 자체의 정확도(validity)에 대한 평가가 선행되어야 하는데 이 연구에서는 의료기관에서 진단받지 않은 경우도 환례에 포함시켰고, 이에 대한 확인을 하지 않았기 때문에 진단의 정확성이 문제될 수 있다. 그러나 볼거리의 증상은 비교적 특징적이어서 일반인도 쉽게 구분할 수 있

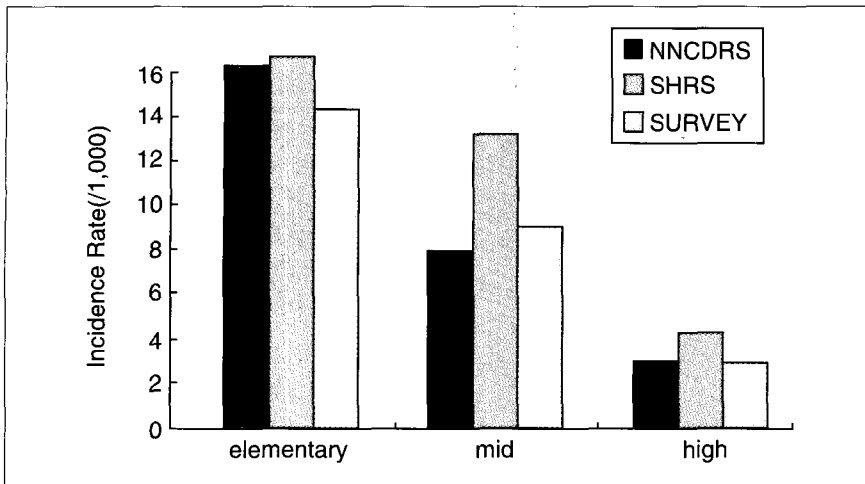


Fig. 3. Incidence rates(/1,000) during the study period from each source by school.

Note: NNCDRS, data from Nationally Notifiable Communicable Disease Reporting System; SHRS, data from School Health Reporting System(temporarily administered by Division of Education); SURVEY, data from questionnaire survey

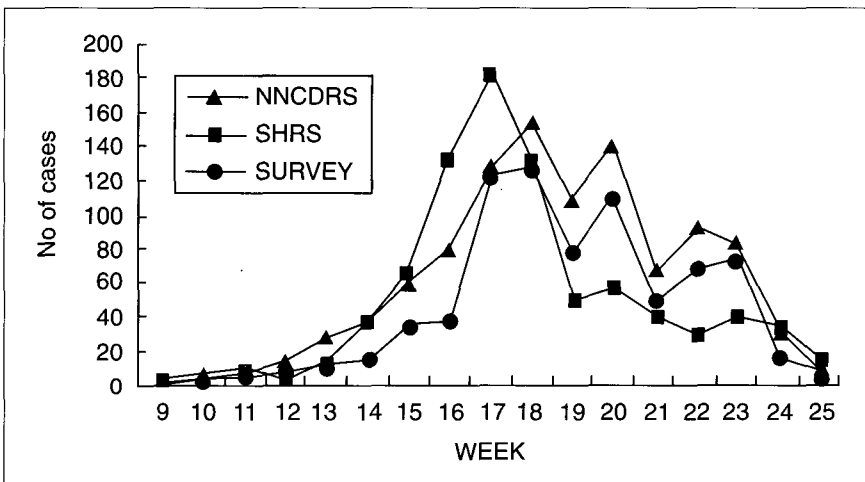


Fig. 4. Timely distribution by sources.

Note: NNCDRS, data from Nationally Notifiable Communicable Disease Reporting System; SHRS, data from School Health Reporting System(temporarily administered by Division of Education); SURVEY, data from questionnaire survey

Table 4. Log-linear models fitting the data with stratification by subgroups

Subgroups	model	residual deviance*	df	n [†]	N [‡] (95% CI [§])
Total	$u_{12} + u_{13}$	0.0286	1	1,844	1,935(1,878~2,070)
Area(1)				1,842	1,864
Cheju	u_{12}	1.3756	1	214	218(n.a.)
Seoguipo	$u_{12} + u_{13}$	1.1522	1	246	263(n.a.)
Buk-gun	$u_{12} + u_{23}$	0.6913	1	553	554(n.a.)
Nam-gun	$u_{12} + u_{13}$	2.1756	1	829	829(n.a.)
Area(2)				1,839	1,909
Urban	$u_{12} + u_{13}$	0.1707	1	460	484(472~527)
Rural	$u_{12} + u_{13}$	2.2995	1	1,379	1,425(1,403~1,511)
School				1,844	1,911
Elementary	$u_{12} + u_{13}$	0.0471	1	1,294	1,351(1,317~1,441)
Mid & High	$u_{12} + u_{13}$	0.4609	1	550	600(583~719)

* deviance0-deviance1, † observed no. of cases, ‡ estimated no. of cases, § profile likelihood interval estimates, || not available due to negative value

으며 특히 여러 원인의 이하선염 중 유행을 일으킬 수 있는 것은 볼거리 바이러스에 의한 것밖에 없기 때문에 유행 시기 발생하는 이하선염은 모두 볼거리로 생각해도 무방하다는 점에서 큰 문제는 되지 않을 것으로 보인다(Evans와 Kaslow, 1997). 하지만 유행 시작 후 3개월이 지난 시점에서 역학조사가 실시되었기 때문에 회상편견이 개입했을 가능성이 있다. 임상증상이 경미했다면 이환 사실을 망각할 가능성이 있는데, 한편으로는 사회적 이슈가 된 후에 본인의 증상을 과대 해석하는 것도 가능하기 때문에, 그 결과가 과대 추정되었는지 과소 추정되었는지 결론 짓기는 어렵다.

이 연구에 의하면 법정전염병 자료의 완전성은 43.8%로 그 동안 발표된 국내의 연구 결과에 비해 매우 높다. 김정순(1987)은 1986년 볼거리의 보고율을 1.9%로 추정하였으며, 신의철 등(1996)은 1994년의 전체 전염병 보고율(홍역, 볼거리, 풍진은 조사대상에서 제외)이 약 27%인 것으로 추정하였다. 미국의 경우 볼거리의 보고율은 10% 내외인 것으로 알려져 있는데, 볼거리 발병 시 25% 정도가 무증상이며, 의료기관을 찾는 환자는 전체 발병자의 약 27%, 의사에 의해 보고가 이루어지는 경우는 전체의 6% 정도라고 보고하기도 하였다(Evans와 Kaslow, 1997). 다른 연구로부터 간접적으로 추정해볼 수도 있는데, 1985~1988년에 미국

Table 5. Completeness(%) of each source

Subgroups	estimated N	NNCDRS		SHRS		SURVEY	
		n*	completeness (%)†	n	completeness (%)†	n	completeness (%)†
Total	1,935	848	43.8	1,026	53.0	785	40.6
Area	1,909	844	44.2	1,001	52.4	773	40.5
Urban	484	241	49.8	223	46.1	280	57.9
Rural	1,425	603	42.3	778	54.6	493	34.6
School	1,911	844	44.2	1,002	52.4	775	40.6
elementary	1,351	640	47.4	653	48.3	554	41.0
mid & high	560	204	36.4	349	62.3	126	22.5

Note : NNCDRS, data from Nationally Notifiable Communicable Disease Reporting System; SHRS, data from School Health Reporting System(temporarily administered by Division of Education); SURVEY, data from questionnaire survey

* observed no. of cases, † (observed no. of cases) ÷ (estimated N) × 100

에서 조사된 입원 백일해 환자의 보고율이 32%였고(Sutter와 Cochi, 1992), 영국의 경우 1982년 백일해 유행 당시 보고율이 18.7%로 추정되었는데(Jenkinson, 1983), 이보다 위중도가 낮은 불거리의 경우 더욱 낮을 것으로 짐작할 수 있다.

이렇듯 기존 연구결과들에 비해 이 연구에서 완전성이 높게 추정된 이유로는 먼저 연구 시점의 차이를 지적할 수 있다. 그 동안 의사들의 신고의식 향상과 신고·기록 관리의 체계화 등으로 인해 신고율이 향상되었을 가능성이 있다. 하지만 이번 사례에서 보다 중요한 요인은 사회적 관심의 증대와 적극적인 보고의 독려이었던 것으로 생각된다. 어떤 새로운 진단법의 등장이나 사회적 관심 정도의 변화, 감시체계 운영방법의 변화 등이 발생하면 보고체계의 민감도는 변화하며(Klaucke 등, 1988; Halperin 등, 1989), 전염성 질환이 대규모로 유행하여 사회적 공론화가 되고 관심이 높아지면 보고율이 일시적으로 상승할 수 있다고 알려져 있다(Jenkinson, 1983). 1998년의 불거리 유행은 국가예방접종이 실시된 이후 가장 대규모의 유행이었으며(보건복지부, 1999), 특히 제주도의 경우 일부 학교에 휴교령이 내려질 정도로 규모가 컸기 때문에 언론 보도 등(한겨레 신문 “제주, 유행성 전염병으로 휴교사태...” 1998.4.22, 서울 경제신문 “불거리 감염자, 제주 전염병 사상 최다...” 1998.5.30, 경향신문 “전국 불거리 급속 확산, ‘비

상 ...” 1998.6.8) 사회적 관심을 크게 끌었다. 이에 따라 보건소를 중심으로 대대적인 예방접종 등 강력한 방역대책이 강구되는 상황이었으며, 이러한 요인들로 인하여 의사들의 보고율과 자료의 완전성이 높아졌을 가능성이 있다. 따라서 이 결과만을 가지고 법정전염병 신고의 완전성이 높아졌다고 일반화하기는 어려울 것이다.

또한 이번 연구에서의 완전성이 다른 연구들에 비해 높은 것은 사실이지만, 43.8%의 완전성이라는 것은 여전히 법정전염병 신고자료만으로 우리나라의 전염병 발생 규모를 측정하는 것에는 문제가 있음을 보여준다.

법정전염병 신고 자료에 비해 교육청 보고자료의 완전성이 더욱 높게 나타난 이유는 불거리의 증상이 위중하지 않기 때문에 의료기관을 찾지 않는 경우가 많고(Evans와 Kaslow, 1997), 학교를 중심으로 대규모 유행이 발생하였기 때문에 학교 보건 차원에서 적극적인 자발적 조사가 시행되었기 때문으로 볼 수 있다. 역학 조사의 완전성이 낮게 측정된 것은 역학 조사의 목적과 시행 방법에서 그 이유를 찾을 수 있다. 그 동안 발견 못했던 환례를 찾아내고, 과거 환례들에 대한 추가 정보를 얻기 위해 역학조사를 실시하였지만, 연구진이 직접 조사하지 못하고 조사서를 학교별로 배부하여 담임교사와 양호교사로 하여금 재조사를 하게 하는 과정에서 많은 탈락자가 생긴 것으로 보

인다. 특히 남제주군의 경우 역학조사를 통해 얻어진 사례가 기존 자료의 약 10%에 불과한데 이 지역은 초기의 대규모 유행 이후 여러 기관에서 수 차례 반복 조사를 하였고, 그 결과가 충분히 환류되지 않은 전례가 있어서 순응도가 더욱 낮았던 것이라 생각된다.

이러한 사실들은 역학조사 또한 완전한 환자 발견방법이 아니라는 것을 확인시켜주는 것이며, 이는 유행조사시에도 여러 가지 자료원들을 적극적으로 활용하여 사례를 발견해야 할 필요성을 제기한다.

이 연구 자료들이 capture-recapture 방법을 적용하기 위한 기본 가정에 부합되는가를 살펴보면 다음과 같다.

먼저 폐쇄형 인구집단에 대한 가정인데, 제주도는 비교적 고립되고 안정된 지역이며, 또한 불거리의 임상적 특성이 사망을 일으키거나 위중한 결과로 인해 다른 지역으로의 후송을 필요로 하지 않는다는 점에서 이 가정은 잘 만족시킨다고 할 수 있다.

짜짓기의 기준이 되는 고유한 식별자의 존재는 우리나라의 경우 단일한 주민등록번호가 있기 때문에 쉽게 해결될 수 있을 것으로 보이지만 실제 자료를 처리하는 데에서는 해결하기 어려운 문제 중의 하나이다. 이 연구에서도 주민등록번호가 기재되어 있지 않은 사례가 많아서 학교와 학년·반의 일치 유무, 혹은 생년월일, 주소지 등을 일일이 대조해야만 했

는데, 결측값과 표기 오류가 많아 동일인에 대한 확인이 불가능하여 비판찰치과 과대 추정되었을 가능성이 존재한다.

자료원에 표집될 확률이 개인들간에 동일인가 여부를 판단하기 위해 시행된 기초분석은 높은 이질성을 시사하였다. 지역별 · 학령별로 조사된 자료의 건수는 자료원마다 큰 차이를 보였는데 이는 대상 인구의 차이뿐만 아니라 조사자의 순응도, 조사 기관의 분포 등에 의해 일어난 것으로 생각된다. 이를테면 교육청 보고 자료의 경우, 학교 당 학생의 숫자가 적어 환자 파악이 비교적 용이한 군 지역의 완전성이 높았던 것으로 해석할 수 있으며 법정전염병 신고 자료의 경우 의료기관의 수가 많은 시 지역의 완전성이 높았던 것으로 해석할 수 있다. 법정전염병 신고 자료의 경우 중 · 고등학생 자료의 완전성이 초등학생에 비해 낮고 교육청 보고 자료에서는 이와 반대의 결과를 보여주는데, 이는 질병 이환 시 초등학생들은 비교적 쉽게 학교에 결석하고 의료기관을 찾는데 비해 중고생들은 이환되어도 학교에 출석하기 때문에 의료기관에서 환자가 발견되는 정도가 낮아서인 것으로 생각한다.

이러한 이질성 요인을 감소시키기 위해 층화 분석을 실시하였는데, 이 때 표본의 수가 작아져 적절한 모형식을 통해 추정값을 찾기가 어려웠다. Hook와 Regal (1993; 1995a)은 capture-recapture 방법 적용 시 가장 문제되는 것은 이질적인 표집 가능성이며, 이를 극복하기 위해 층화 분석하는 경우 표본의 숫자가 작아지는 문제가 있고, 층화할 수 있도록 자료가 모이지 않는 경우가 많으며, 층화에 의해 자료의 변이가 더 커지는 경향이 있다고 지적하였다. 그래서 계층을 크게 시 · 군 지역, 초등학생과 중 · 고등학생으로 분류하였으며 이렇게 얻어진 추정값들의 합계는 층화 없이 구한 추정값과 크게 다르지 않았다. 만일 이들이 큰 차이를 보였다면 생각하지 못한 높은 차원의 교호작용이 존재했다고 볼 수 있으며, 그런 경우 층화에 의해 얻어진 값을 사용하는 것이 바람직할 것이다(Hook와 Regal, 1995b).

자료원간의 독립성 가정 만족 여부는, 적합된 모형식에서 교호작용항으로 판단할 수 있는데, 층화하지 않은 경우와 층화한 경우 모두에서 각각 유의한 교호작용이 있는 것으로 나타났다. 이렇게 로그선형모형을 사용하였을 경우 비독립성 가정에 위배는 교호작용항을 모형에 포함 시킴으로써 해결할 수 있지만, 두 자료원들만 짝지는 경우 자료의 의존성 방향에 따라 매우 비뚤린 결과를 낳을 것이라고 예상할 수 있다(Wittes 등, 1974; Brenner 등, 1995). 따라서 부가적 정보가 없거나 독립성이 확실하지 않은 경우 두 자료원을 짝짓기하여 추정하는 것은 바람직하지 않을 것이다.

그 동안 국내에서 이루어진 법정전염병 신고자료에 대한 평가 연구는 자료의 정확성(validity)에 관한 연구와 완전성 혹은 보고율(reporting rate)에 관한 것으로 구분할 수 있다. 이 때 요약 통계량만을 비교하여 완전성을 평가하는 경우(김윤구 등, 1985; 김정순, 1987) 비교적 손쉽게 시행할 수는 있지만 자료의 변이와 분포에 관한 특성을 밝히기 어렵다는 단점이 있으며, 개별 의료기관의 의무기록을 직접 조사하는 방법의 경우(신의철 등, 1996) 자료의 정확도까지 확인할 수 있다는 장점은 있지만 시간과 비용이 많이 든다는 측면에서 상시적으로 활용하기 어렵다는 한계를 갖는다(Hilsenbeck 등, 1992). 또한 병의원 의무기록 조사의 경우 의료기관을 찾지 않은 환자에 대해서는 전혀 파악하지 못하기 때문에 위험도가 낮은 질환의 발생 규모를 추정하는데 어려움을 지닌다.

법정전염병 신고제도와 같은 수동적 국가보고체계의 경우 약간의 정확성을 훼손하더라도 비용을 감소시키고 보다 넓은 보고범위(coverage)를 얻는 것이 중요한 경우가 많다. 따라서 1차적으로 자료의 질을 높이기 위해 정확도와 보고율을 높이는 방법이 강구되어야겠지만 이와 더불어 자료의 불완전성을 인정하고 효과적으로 이를 보완할 수 있는 노력을 기울이는 것도 필요할 것이다.

이 때 capture-recapture 방법은 불완전

한 기존의 자료들을 이용하여 빠르게 적은 비용으로(Gutteridge와 Collin, 1994) 질병 발생의 규모를 측정할 수 있다는 점에서 쉽게 활용할 수 있는데, 이를 위해서는 최소 2가지 이상의 질병 보고 · 감시체계가 존재해야 하며, 반드시 질병 보고체계가 아니더라도, 이용이 가능한 다른 자료- 이를테면 경찰기록, 약물관대 자료 등을 활용할 수도 있다(Hardy 등, 1987; Prevots 등, 1994; Brenner 등, 1995; Domingo-Salvani 등, 1998; Sekikawa 등, 1999).

국내에서도 현재 개발 중에 있는 각종 질병 감시체계와 예시된 연구들에서처럼 대상인구와 질병 특성에 맞는 다양한 자료를 활용하여 capture-recapture 방법을 적용하면 질병 발생의 규모를 좀더 정확히 파악할 수 있을 것으로 보인다. 최근에 이를 이용하여 의료보험 청구자료의 완전성을 평가한 연구(하미나 등, 1997), 지역 암등록 자료의 완전성을 평가한 연구(임정수 등, 1999) 등이 발표되었는데, 앞으로 활용 범위가 확대될 수 있을 것이다. 일례로 이 연구에서처럼 학생 계층에 많이 발생하면서 위중도가 높지 않아 의료기관을 찾는 정도가 낮을 것으로 생각되는 호흡기 전염병의 경우 학교를 활용한 감시체계 자료를 이용하여 법정 전염병 신고자료를 보완하고 평가할 수 있고, 또한 지역사회의 각종 전염병 유행조사 시에 수집 가능한 여러 자료들을 활용하여 보정된 발생 추정값을 구할 수 있을 것이다.

하지만 이러한 2가지 이상의 질병 자료를 적절하게 활용하기 위해서는 먼저 각 체계마다 사례정의를 분명히 하는 것이 필요하며(Desenclos와 Hubert, 1994; Egeland 등, 1995), 자료원간의 정확한 연계(record linkage)를 위해 자료원마다 정확하고 고유한 식별코드를 부여하는 것, 자료의 연계 과정에서 나타날 수 있는 개인 정보의 보호 문제에 관하여 충분한 사회적 합의와 기술적 고안이 선행되어야 할 것이다(Stroup와 Teutsch, 1998).

요약 및 결론

1998년 제주도에서 발생한 불거리 유행 시기(3월 3일부터 8월 31일) 중 3월 3일부터 6월 20일까지의 유행에 대하여 법정전염병 신고 자료, 교육청 보고 자료 및 역학조사반 환자 조사 자료 등의 3가지 자료원으로부터 발생자를 파악하고, capture-recapture 방법을 적용하여 전체 환자 발생 규모를 추정하고 법정 전염병 신고자료의 완전성을 평가하고자 하였다.

유행기간 동안 법정전염병으로 신고된 사례는 840건이었으며, 교육청에 보고된 자료는 1,026건, 역학조사반의 환자 조사 자료는 785건이었고, 세 자료원을 통해 집계한 유행기간 동안의 환자 발생 건수는 총 1,844명이었다. 로그 선형모형을 이용한 capture-recapture 방법을 적용한 결과 추정 발생 건수는 1,935(95% CI:1,878-2,070)건이었고, 법정전염병 신고 자료의 완전성은 43.8%으로 추정되었다. 도시 지역(49.8%)이 농촌 지역(42.3%)에 비하여, 초등학교 학생의 경우(47.4%)가 중·고등학교 학생의 경우(36.4%)에 비하여 법정전염병 신고의 완전성이 높았다. 신고 자료의 완전성이 43.8%라는 이 연구의 결과는 기존에 알려진 법정전염병 신고율 보다 높은 것이었다. 이는 대규모 유행 시기이었다는 점과 여러 경로 즉, 유행에 대한 보건 기관의 독려, 역학조사반의 지역의 보건의료인에 대한 설명회, 언론 보도에 의한 인지도 향상 등으로 인하여 신고율이 평상시 보다 높았던 것으로 생각한다.

이 연구는 특정 전염병의 유행 시기에 3가지 자료원을 확보하여 capture-recapture 방법을 적용하여 신고 자료의 완전성을 파악하였다. 국내의 전염병 신고자료의 완전성이 낮다고 알려져 있는데, 외국의 경험들을 보더라도 신고율이 완벽한 경우는 없어 여러 보완적인 방법으로 지역사회에서의 전염성 발생 규모를 파악하여야 한다. 향후 유행 시기뿐만 아니라 평상시에도 2가지 이상의 자료원을 확보하여, 다른 전염병에도 적용할 수 있을 것으로 생각한다. 이는 신고의 완전

성을 파악하여 지역사회의 발생 규모를 보다 정확하게 추정하고, 나아가 신고의 완전성을 높이는 데 기여할 수 있을 것으로 생각한다.

참고문헌

김윤구, 정태화, 송 철, 이명원, 김정순 등. 입원장티푸스 환자의 신고에 관한 연구. *한국역학회지* 1985; 7(2): 240-255

김정순. 신고자료의 정확도 및 추정신고율. *한국역학회지* 1987; 9(2): 157-160

김정순. 역학원론, 증보 제 4판. 신광출판사; 1990. p.93-104

김정순. 전염병 연구의 실제. *한국역학회지* 1996; 18(1): 1-6

맹광호. 우리나라 법정 전염병 발생 신고현황과 신고체계 개선 방안. *한국역학회지* 1996; 18: 27-32

보건복지부, 보건복지 통계연보 1999.

신의철, 맹광호, 신호철, 박용규, 박기동 등. 우리나라의 법정전염병 보고율 추정. *한국역학회지* 1996; 18: 18-26

임정수, 권순석, 김상용, 박경수, 손석준 등. capture-recapture 방법을 이용한 광주광역시 지역암등록 자료의 완전성 추정. *예방의학회지* 1999; 33(1): 31-35

정태화. 전염병 신고자료의 Validity. *한국역학회지* 1987; 9(2): 168-169

하미나, 권호장, 강대회, 조수현, 유근영 등. 소아천식을 통해서 본 의료보험 상병자료의 완전성 추정: capture-recapture 분석방법의 적용. *예방의학회지* 1997; 30(2): 428-43

Agresti A. Simple capture-recapture models permitting unequal catchability and variable sampling effort. *Biometrics* 1994; 50: 494-500

Agresti A. An Introduction to categorical data analysis. New York: John, Wiley & Sons; 1996. p.145-204

Brenner H, Stegmaier C, Ziegler H. Estimating completeness of cancer registration: an empirical evaluation of the two source capture-recapture approach in Germany. *J Epidemiol Community Health* 1995; 49: 426-430

Buckland ST, Garthwaite PH. Quantifying precision of mark-recapture estimates using the bootstrap and related methods. *Biometrics* 1991; 47: 255-268

Cormack RM. Log-linear models for capture-recapture. *Biometrics* 1989; 45: 395-413

Cormack RM. Interval estimation for mark-recapture studies of closed populations. *Biometrics* 1992; 48: 567-576

Desenclos JC, Hubert B. Limitation to the

universal use of capture-recapture methods. *Int J Epidemiol* 1994; 23(6): 1322-1323

Domingo-Savany A, Hartnoll RL, Maguire A, Brugal MT, Albetin P et al. Analytical considerations in the use of capture-recapture to estimate prevalence: Case studies of the estimation of opiate use in the metropolitan area of Barcelona, Spain. *Am J Epidemiol* 1998; 148(8): 732-740

Egeland GM, Perham-Hester KA, Hook EB. Use of capture-recapture analysis in Fetal Alcohol Syndrome surveillance in Alaska. *Am J Epidemiol* 1995; 141(4): 335-341

Evans AS, Kaslow RA. Viral infections of Humans, 4th ed. New York and London: Plenum Medical Book Company; 1997. p.531-550

Fisher N, Turner SW, Pugh R, Taylor C. Estimating numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis. *BMJ* 1994; 308: 27-30

Gutteridge W, Collin C. Capture-recapture techniques-quick & cheap. *BMJ* 1994; 308: 531

Halperin SA, Bortoiussi R, MacLean D, Chisholm N. Persistence of pertussis in an immunized population: Results of the Nova Scotia enhanced pertussis surveillance program. *J Pediatr* 1989; 115: 686-93

Hardy AM, Starcher II T, Morgan M, Druker J, Kristal A et al. Review of death certificates to assess completeness of AIDS case reporting. *Public Health Rep* 1987; 102(4): 386-391

Hilsenbeck SGT, Kurucz C, Duncan RC. Estimation of completeness and adjustment of age-specific and age-standardized incidence rates. *Biometrics* 1992; 48: 1249-1262

Hook EB, Regal RR. Validity of Bernoulli Census, log-linear, and truncated binomial models for correcting for underestimates in prevalence studies. *Am J Epidemiol* 1982; 116(1): 168-176

Hook EB, Regal RR. Capture-recapture methods. *Lancet* 1992; 33: 742

Hook EB, Regal RR. Effect of variation in probability of ascertainment by sources ("Variable catchability") upon "Capture-recapture" estimates of prevalence. *Am J Epidemiol* 1993; 137(10): 1148-1166

Hook EB, Regal RR. Capture-recapture methods in epidemiology: Methods and limitation. *Epidemiol Rev* 1995a; 17(2): 243-264

Hook EB, Regal RR. Internal validity analysis: a method for adjusting capture-recapture

- estimates of prevalence. *Am J Epidemiol* 1995b; 142(supp 9): s48-s52
- International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation 1: History and theoretical development. *Am J Epidemiol* 1995; 142(10): 1047-1058
- Jenkinson D. Whooping cough: what proportion of cases in notified during an epidemic?. *BMJ* 1983; 287: 185-186
- Klaucke DN, Bueher JW, Thacker SB, Parish RG, Trowbridge FL et al. Guidelines for evaluating surveillance systems. *MMWR* 1988; 37(s-5): 1-18
- LaPorte RE. Assessing the human condition: capture-recapture techniques. *BMJ* 1994; 308: 5-6
- LaPorte RE, Dearwater SR, Chang YF, Thomas JS, Aaron DJ, Anderson RL et al. Efficiency and Accuracy of Disease Monitoring systems: Application of Capture-Recapture Methods to Injury Monitoring. *Am J Epidemiol* 1995; 142(10): 1069-1077
- McCarty DJ, Tull ES, Moy CS, Kwok CK, LaPorte RE. Ascertainment corrected rates: application of capture-recapture methods. *Int J Epidemiol* 1993; 22(3): 559-565
- McKeganey N, Barnard M, Leyland A Coote I, Follet E. Female streetworking prostitution and HIV infection in Glasgow. *BMJ* 1992; 305: 801-804
- Modesitt SK, Hulman S, Fleming D. Evaluation of active versus passive AIDS surveillance in Oregon. *Am J Epidemiol* 1990; 80: 463-464
- Neugebauer R. Voluntary and involuntary capture-recapture sample-problems in the estimation of hidden and elusive population. *Am J Public Health* 1994; 84: 1068
- Prevots DR, Sutter RW, Strelbel PM, Weibel RE, Cochi SL. Completeness of reporting for paralytic poliomyelitis, United States, 1980 through 1991. *Arch Pediatr Adolesc Med* 1994; 148: 479-485
- Regal RR, Hook EB. Goodness of fit based confidence intervals for estimates of the size of a closed populations. *Stat Med* 1984; 3: 287-91
- Regal RR, Hook EB. The effects of model selection on confidence intervals for the size of a closed population. *Stat Med* 1991; 10: 717-721
- Sekikawa A, Eguchi H, Tominaga M, Manaka H, Sasaki H et al. Evaluating the reported prevalence of type 2 diabetes mellitus by the Oguni diabetes registry using a two-sample method of capture-recapture. *Int J Epidemiol* 1999; 28: 498-501
- Stokes ME, Davis CS, Koch GG. Categorical data analysis using the SAS system. SAS® institute Inc.; 1995. p.425-462
- Stroup DF, Teutsch SM. Statistics in public health; quantitative approach to public health problems. New York: Oxford University Press; 1998. p.39-57
- Sutter RW, Cochi SL. Pertussis Hospitalizations and Mortality in the United States, 1985-1988 : Evaluation of the completeness of national Reporting. *JAMA* 1992; 267(3): 386-391
- Thacker SB, Stroup DJ. Future directions for comprehensive public health surveillance and health information systems in the United States. *Am J Epidemiol* 1994; 140(5): 383-397
- Wittes JT, Colton T, Sidel VW. Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *J Chronic Dis* 1974; 27: 25-36
- Wittes JT, Sidel VW. A generalization of the simple capture-recapture model with application to epidemiological research. *J Chronic Dis* 1968; 21: 287-301