

## 잠재적 위험요인의 탐색에 관한 단일표본분석과 복합표본분석의 비교

변해원<sup>1,2\*</sup>

<sup>1</sup>대림대학교 언어재활과, <sup>2</sup>아주대학교의료원 노인보건연구센터

### Comparative Analysis of Unweighted Sample Design and Complex Sample Design Related to the Exploration of Potential Risk Factors of Dysphonia

Haewon, Byeon<sup>1,2\*</sup>

<sup>1</sup>Department of Language Therapy, Daelim University Collage

<sup>2</sup>Institute on Aging, Ajou University Medical Center

**요 약** 본 연구는 잠재적 위험요인을 탐색하는 방법으로 단순임의추출분석(unweighted sample design), 빈도 가중치를 적용한 단일표본분석(frequency weighted sample design), 가중치를 총화하여 적용한 복합표본분석(complex sample design)을 비교하고, 도출된 결과에 통계적인 차이가 있는지를 파악하고자 수행되었다. 자료원은 2009 국민건강영양조사(2009 Korea National Health and Nutrition Examination Survey)의 이비인후과 검진 자료를 이용하였다. 분석 방법은 피어슨의 교차검정(Pearson chi-square test)과 라오-스콧교차검정(Rao-scott chi-square test)을 이용하였다. 분석 결과, 빈도 가중치만을 적용한 단일표본분석의 경우에는 모든 변수가 유의한 위험요인으로 과대 예측되었고, 가중치를 적용하지 않은 단순임의추출 분석과 복합표본분석은 유의수준 및 결과에 차이가 있었다. 국가통계자료를 이용할 때, 연구의 결과가 전체 인구집단을 대표할 수 있도록 의미를 부여하기 위해서는 총화변수와 집락변수를 사용하여 가중치를 적용하는 복합표본분석이 필요하다. 나아가, 빈도 가중치만을 적용하는 경우에는 연구 결과에 대한 과잉해석의 가능성이 높기 때문에 각별한 주의가 요구된다.

**Abstract** This study compared the unweighted sample design, frequency weighted sample design and complex sample design to using 2009 Korea National Health and Nutrition Examination Survey in an effort to identify whether or not there is any difference in potential risk factors. Pearson chi-square test and Rao-scott chi-square test were applied to the analytic methods. As a result of analyses, all the variables were overestimated as significant risk factors in case of the unweighted sample design to which only the frequency weights were applied. In addition, there were differences in the confidence levels and results from the simple random sampling analysis and complex sample design to which no weight was applied. It is necessary to carry out the complex sample design rather than the analysis to which the frequency weights are applied, in order to ensure the findings to represent the whole population when our national statistics data is used.

**Key Words** : Complex sample, Dysphonia, National Survey, Risk factor, Rao-scott chi-square test

#### 1. 서론

건강 및 질병 관리의 책임은 개인적인 차원에서 점차 사회적 차원으로 확대되고 있으며, 질병에 대한 연구 또

한 임상학적 관점에서 예방학적 관점으로 초점이 맞추어지고 있다. 예방적 관점에서 볼 때, 질병은 특정 단일 원인에 기인한 것이 아닌 여러 가지 위험요인(risk factor)의 복합체이다. 따라서 효율적인 예방을 위해서는 질병 발생

\*Corresponding Author : Haewon, Byeon

Tel: +82-10-7404-6969 email: byun@ajou.ac.kr

접수일 12년 02월 17일

수정일 (1차 12년 03월 07일, 2차 12년 03월 13일)

게재확정일 12년 05월 10일

에 영향을 미치는 잠재적 위험요인을 명확하게 규명하고 이를 관리하는 것이 중요하다. 최근에는 질병 예방 및 건강에 대한 국가적인 관심이 고조되면서 국민건강영양조사(Korea National Health and Nutrition Examination Survey) 등의 국가통계조사가 수행되고 있으며, 이를 이용한 연구도 급증하고 있다. 질병관리본부의 조사에 따르면, 2001년부터 2011년까지 국민건강영양조사의 데이터를 이용한 학술연구는 총 463건에 달하는 것으로 보고되었다[1]. 이와 같은 연구들에서는 조사된 표본이 지역사회 인구를 기초로 추출되었기 때문에, 질병의 위험 요인 파악에 있어서 특정 지역이나 병원 데이터를 이용한 연구 보다 신뢰성이 높다. 그러나 국가통계조사를 이용하였음에도 불구하고, 많은 수의 국내 역학연구들이 가중치를 이용하지 않은 단일표본분석 방법으로 수행되었기 때문에 실제 인구집단의 현상이 아닌 편향(bias)된 결과를 해석할 가능성이 있다. 또한, 가중치를 적용한 연구의 경우에도 자료원인 국가통계조사가 층화(stratification), 집락 추출(cluster sampling) 등이 복합적으로 사용된 복합표본설계(complex sample design)에 의해 표본이 추출되었기 때문에, 기존의 단일표본분석의 빈도 가중치 방법만으로는 추정값에 오류가 발생할 수 있다[8]. 복합표본설계에 관한 선행연구[2, 3, 15, 16]에서 가중치 적용 방법에 따른 오류의 발생 가능성이 제기되어왔지만 이론적인 연구에만 국한되어 있으며, 실증적인 사례를 통해서 분석 방법에 따른 결과의 차이를 입증한 연구는 미흡한 실정이다. 본 연구에서는 국가통계조사의 음성질환 진단 자료를 이용하여 가중치를 적용하지 않은 단순임의추출분석, 빈도 가중치를 적용한 단일표본분석, 가중치를 층화하여 적용한 복합표본분석을 비교하고, 도출된 잠재적 위험 요인에 통계적인 차이가 있는지를 분석하였다.

## 2. 연구방법

### 2.1 연구자료

2009년 질병관리본부의 주관으로 조사된 국민건강영양조사를 이용하였다. 원시 자료는 질병관리본부에 연구 목적을 밝히고, 공식적으로 승인을 받은 후 사용하였다. 국민건강영양조사는 건강설문조사, 검진조사, 영양조사의 3개 영역으로 구분되어 있다[9]. 건강설문조사는 흡연, 음주 등의 건강행태에 대한 내용이 포함되어 있으며, 검진조사는 이비인후과 검진 설문 등이 포함되어 있다. 건강설문조사에서 건강설문조사의 교육수준조사와 경제활동 항목은 대상자와의 개별 면접을 통하여 조사되었고,

흡연, 음주 등의 건강행태 및 주관적 건강상태와 이비인후과 검진 설문의 본인 인지 음성장애 여부는 자기기입 방법으로 조사되었다. 본 연구에서는 건강면접조사와 검진조사의 자료를 병합하여 이용하였다.

### 2.2 연구대상

건강설문조사와 이비인후과 검진을 모두 완료한 만 19세 이상 만 60세 미만의 지역사회 성인 4,746명을 연구 대상으로 선정하였다. 연령을 만 19세부터 만 60세 미만으로 한정하는 이유는 첫째, 국민건강영양조사의 이비인후과 검사가 만 19세 이상의 인구를 대상으로 실시되었기 때문이다. 둘째, 설명 변수인 직업의 경우 만 19세 미만은 취업활동을 하는 인구가 존재하더라도 대다수가 직업이 확정되지 않은 경우가 많고, 60세 이상은 이미 은퇴한 대상이 많기 때문이다. 이비인후과 검진을 모두 완료한 대상자 중에서 음성 문제 설문 무응답자와 후두내시경 미검사와 및 판정불가자가 제외되어 총 4,601명(남 1,988명, 여 2,613명)이 최종 분석 대상으로 선정되었다.

### 2.3 설명변수

설명변수는 선행연구[4-7]를 참고하여 음성장애의 위험요인으로 확인된 성별, 연령, 교육수준, 소득수준, 직업, 흡연 여부, 음주 행위, 갑상선장애, 주관적 음성문제, 주관적 건강상태를 포함하였다. 연령은 만 나이로 조사되었고, 연구자에 의해 19세 이상 40세 미만과 40세 이상 60세 미만으로 재분류 되었다. 교육수준은 무학, 초등학교 졸업이하, 중학교 졸업, 고등학교 졸업, 대학교 졸업 이상으로 분류하였다. 소득은 자료에 이미 형성된 4분위 수에 따라 분석하였다. 직업은 6차 한국표준직업분류[10]의 대분류에 비경제활동인구인 무직자, 주부 및 학생이 추가되어 분류되었고, 자료 해석의 간소화를 위해서 김기혜 등[11]의 연구를 참고하여 비경제활동(무직, 주부, 학생), 비육체노동자(관리자, 전문가 및 관련 종사자, 사무 종사자, 서비스 종사자, 판매 종사자), 육체노동자(농림어업 숙련 종사자, 기능원 및 관련 기능 종사자, 장치·기계 조작 및 조립 종사자, 단순노무 종사자, 군인)로 재분류 하였다. 흡연여부는 현재 흡연자, 과거 흡연자, 비흡연자로 분류하였다. 음주빈도는 최근 1년간 음주행위를 기준으로 월 1회 미만과 월 1회 이상으로 재분류 하였다. 갑상선장애 유병과 주관적 음성문제 인지는 ‘있다’와 ‘없다’로 분류하였다. 주관적 건강상태는 ‘매우 좋음’부터 ‘매우 나쁨’까지 5개의 척도로 조사되었고, 연구자에 의해 ‘ 좋음’, ‘보통’, ‘나쁨’으로 재분류 되었다.

## 2.4 결과변수

결과변수는 음성장애 유무로 정의하였다. 음성장애는 이비인후과 검진조사에서 이비인후과 전공의에 의해 70도 내시경을 이용하여 19세 이상의 지역사회 남녀 성인을 대상으로 후두의 비정상 소견(성대결절, 성대용종, 라인케부종, 성대마비, 성대구증, 성대각화증, 성대낭종, 후두염, 유두종, 육아종, 후두악성종양의증)이 조사되었고 질관리위원의 관리를 거쳐서 진단 자료가 생성되었다. 후두내시경 검진에서 음성질환으로 분류된 질환명 자료는 연구자에 의해 이분형의 음성장애 유무로 재분류 되었다.

## 2.5 자료처리 및 통계분석

음성장애의 잠재적 위험 요인은 가중치를 부여하지 않은 단순임의추출분석과 빈도 가중치를 적용한 단일표본분석, 가중치를 층화하여 적용한 복합표본분석을 이용하여 각각 분석하였다. 국민건강영양조사의 가중치는 조사에 참여한 개인이 우리나라 전체 인구를 대표하도록 부여되었으며 가중치에 대한 자세한 수식과 방법은 [9]에 소개되어 있다. 본 연구에서 가중치는 건강 설문과 검진의 자료를 병합하여 이용하였기 때문에 별도로 제시되어 있는 건강 설문 및 검진의 연관성 가중치를 적용하였다. 단순임의추출분석은 가중치를 적용하지 않고, 조사된 표본을 그대로 이용하여 분석하였다. 단일표본분석은 빈도 가중치만 지정하여 분석하였다. 복합표본분석은 국민건강영양조사의 표본설계를 고려하여 1차 추출단위인 집락 변수, 층화변수, 가중치를 지정한 후 분석하였다. 잠재적 위험 요인은 단순임의추출분석과 단일표본분석에서는 피어슨 교차검정(Pearson chi-square test)으로, 복합표본분석에서는 라오-스콧 교차검정(Rao-scott chi-square test)으로 각각 확인하였다. 모든 분석은 IBM SPSS version 20.0(IBM Inc., Chicago, Illinois)을 이용하였고, 유의수준은 양측검정에서 0.05로 설정 하였다.

# 3. 연구결과

## 3.1 가중치 여부에 따른 모집단의 특성

가중치 여부에 따른 모집단의 특성은 표 1에 제시하였다. 기술분석 결과, 모든 변수에서 가중치를 적용하지 않은 단순임의추출분석과 가중치를 적용한 복합표본분석은 비율의 차이가 있었고, 특히 성별과 연령은 특성이 반대로 도출되었다. 단순임의추출분석의 경우 여성의 비율이 더 높았지만, 실제 모집단을 대표하는 복합표본분석에서는 남녀의 비율이 유사하였다. 또한, 연령의 경우에도 단

순임의추출분석에서는 40세 이상 60세 미만의 비율이 더 높았지만, 복합표본분석에서는 19세 이상 40세 미만의 비율이 더 높았다.

복합표본분석 결과, 인구집단은 고졸(47.0%), 비육체노동자(45.6%), 비흡연자(52.6%), 월 1회 이상의 음주자(64.0%)의 비율이 높았다. 또한, 지역사회 인구 100명 당 약 3명은 갑상선장애, 약 6명은 음성장애가 있었으며, 약 16명은 주관적으로 건강상태가 나쁘다고 인지하였다.

## 3.2 가중치 여부에 따른 유병률 및 특성

가중치 여부에 따른 음성장애 유병률 및 특성은 표 2에 제시하였다. 가중치를 적용하지 않은 단순임의추출분석과 가중치를 적용한 복합표본분석은 유병률의 차이가 있었다. 교육수준에서 초등학교 졸업 이하의 음성장애 유병률은 단순임의추출분석에서 더 높았으며, 단순임의추출분석에서는 고등학교 졸업의 음성장애 유병률이 5.5%로 가장 낮았지만, 복합표본분석에서는 고등학교 졸업과 대학 졸업 이상의 유병률이 각각 6.3%로 동일하였다. 복합표본분석 결과, 남성(7.7%), 중학교 졸업(8.1%), 육체노동자(7.1%), 현재 흡연자(9.3%), 월 1회 이상 음주자(6.9%), 주관적으로 음성문제를 인지하는 사람(24.5%), 주관적으로 건강 상태가 나쁘다고 인지하는 사람(8.1%)에서 음성장애 유병률이 높았다.

## 3.3 잠재적 위험 요인 탐색을 위한 분석 방법의 비교

음성장애의 잠재적 위험 요인 탐색을 위한 분석 방법의 비교는 표 3에 제시하였다. 교차분석결과, 가중치를 적용하지 않은 단순임의추출분석과 빈도 가중치를 부여한 단일표본분석, 가중치를 층화하여 적용한 복합표본분석은 도출된 음성장애의 잠재적 위험요인에 차이가 있었다. 단순임의추출분석의 교차검정 결과, 음성장애의 잠재적 위험요인은 연령, 성별, 흡연, 주관적 음성문제, 주관적 건강상태 이었다( $P < 0.05$ ). 단일표본분석의 교차검정 결과, 모든 변수가 음성장애의 잠재적 위험요인으로 도출되었다( $p < 0.05$ ). 복합표본분석의 교차검정 결과, 음성장애의 잠재적 위험요인은 연령, 성별, 흡연, 주관적 음성문제 이었다( $p < 0.05$ ). 주관적 건강상태는 단순임의추출분석에서는 음성장애 여부에 따른 집단 간에 유의한 차이가 있었지만( $p = 0.019$ ), 복합표본분석에서는 유의한 차이가 없었다( $P = 0.230$ ).

[표 1] 가중치 여부에 따른 인구 특성

[Table 1] General characteristics of Korean population according to sampling weight

Characteristics	Unweighted N (%)	Weighted N (weighted %)
Total	4,601	26,541,594
Sex		
Man	1,988 (43.2)	13,515,089 (50.9)
Woman	2,613 (56.8)	13,026,504 (49.1)
Age		
19~39	2,237 (48.6)	14,095,276 (53.1)
40~59	2,364 (51.4)	12,446,317 (46.9)
Education level		
Elementary school	485 (10.6)	2,360,226 (8.9)
Middle school	468 (10.2)	2,356,567 (8.9)
High school	2,065 (45.1)	12,392,293 (47.0)
≥ College	1,559 (34.1)	9,282,028 (35.2)
Income		
1st quartile	472 (10.4)	2,679,287 (10.2)
2ed quartile	1,023 (22.5)	5,822,237 (22.2)
3rd quartile	1,482 (32.5)	8,560,445 (32.6)
4th quartile	1,578 (34.6)	9,167,888 (35.0)
Occupation		
Economically inactive	1,495 (32.6)	8,457,795 (32.0)
Non-manual worker	2,003 (43.6)	12,061,732 (45.6)
Manual worker	1,091 (23.8)	5,932,687 (22.4)
Smoking		
Nonsmoker	2,674 (58.3)	13,914,524 (52.6)
Past smoker	731 (15.9)	4,629,069 (17.5)
Current smoker	1,184 (25.8)	7,908,621 (29.9)
Alcohol drinking		
<1 time per month	1,815 (39.6)	9,521,125 (36.0)
≥1 time per month	2,774 (60.4)	16,931,090 (64.0)
Thyroid disorders		
Yes	148 (3.2)	728,921 (2.8)
Dysphonia		
Yes	276 (6.0)	1,709,209 (6.4)
Self reported voice problem		
Yes	278 (6.0)	1,601,825 (6.0)
Subjective health status		
Good	2,080 (45.4)	12,289,889 (46.6)
Normal	1,687 (36.9)	9,781,093 (37.1)
Bad	811 (17.7)	4,322,887 (16.4)

[표 2] 음성장애 유병률 및 특성  
 [Table 2] Prevalence of Dysphonia and characteristics

Variables	Unweighted N (%)		Weighted N (weighted %)	
	ND	Dysphonia	ND	Dysphonia
<b>Age</b>				
19~39	1,841 (92.6)	147 (7.4)	12,475,456 (92.3)	1,039,634 (7.7)
40~59	2,484 (95.1)	129 (4.9)	12,356,929 (94.9)	669,576 (5.1)
<b>Sex</b>				
Man	1,841 (92.6)	147 (7.4)	12,475,456 (92.3)	1,039,634 (7.7)
Woman	2,484 (95.1)	129 (4.9)	12,356,929 (94.9)	669,576 (5.1)
<b>Education level</b>				
Elementary school	451 (93.0)	34 (7.0)	2,203,659 (93.4)	156,567 (6.6)
Middle school	434 (92.7)	34 (7.3)	2,165,540 (91.9)	191,027 (8.1)
High school	1,951 (94.5)	144 (5.5)	11,615,568 (93.7)	776,725 (6.3)
≥ College	1,466 (94.0)	93 (6.0)	8,701,332 (93.7)	580,697 (6.3)
<b>Income(home)</b>				
1st quartile	437 (92.6)	35 (7.4)	2,462,180 (91.9)	217,108 (8.1)
2en quartile	961 (93.9)	62 (6.1)	5,468,488 (93.9)	353,749 (6.1)
3rd quartile	1,406 (94.9)	76 (5.1)	8,062,833 (94.2)	497,612 (5.8)
4th quartile	1,478 (93.7)	100 (6.3)	8,546,685 (93.2)	621,203 (6.8)
<b>Occupation</b>				
Economically inactive	1,416 (94.7)	79 (5.3)	7,981,975 (94.4)	475,821 (5.6)
Non-manual worker	1,877 (93.7)	126 (6.3)	11,248,919 (93.3)	812,814 (6.7)
Manual worker	1,020 (93.5)	71 (6.5)	5,512,113 (92.9)	420,575 (7.1)
<b>Smoking</b>				
Nonsmoker	2,550 (95.4)	124 (4.6)	13,268,020 (95.4)	646,505 (4.6)
Past smoker	682 (93.3)	49 (6.7)	4,304,428 (93.0)	324,641 (7.0)
Current smoker	1,081 (91.3)	103 (8.7)	7,170,559 (90.7)	738,063 (9.3)
<b>Alcohol drinking</b>				
<1 time per month	1,712 (94.3)	103 (5.7)	8,974,043 (94.3)	547,082 (5.7)
≥1 time per month	2,061 (93.8)	173 (6.2)	15,768,963 (93.1)	1,162,127 (6.9)
<b>Thyroid disorders</b>				
No	4,175 (94.0)	266 (6.0)	24,060,664 (93.5)	1,662,631 (6.5)
Yes	138 (93.2)	10 (6.8)	682,343 (93.6)	46,579 (6.4)
<b>Self reported voice problem</b>				
No	4,116 (95.2)	207 (4.8)	23,622,627 (94.7)	1,317,141 (5.3)
Yes	209 (75.2)	69 (24.8)	1,209,757 (75.5)	392,069 (24.5)
<b>Subjective health status</b>				
Good	1,966 (94.5)	114 (5.5)	11,547,762 (94.0)	742,127 (6.0)
Normal	1,592 (94.4)	95 (5.6)	9,167,617 (93.7)	613,476 (6.3)
Bad	745 (91.9)	66 (8.1)	3,973,473 (91.9)	349,414 (8.1)

ND: Non organic dysphonia

[표 3] 음성장애의 잠재적 위험 요인 탐색을 위한 분석 방법의 비교  
 [Table 3] Comparing analysis for potential risk factors of dysphonia

Category variables	P-value*	P-value*	P-value*
	(unweighted sample)	(frequency weighted sample)	(complex weighted sample)
Age	0.012	<0.001	0.019
Sex	0.001	<0.001	0.001
Education level	0.385	<0.001	0.631
Income	0.261	<0.001	0.530
Occupation	0.341	<0.001	0.429
Smoking	<0.001	<0.001	<0.001
Alcohol drinking	0.434	<0.001	0.205
Thyroid disorders	0.699	0.012	0.974
Self reported voice problem	<0.001	<0.001	<0.001
Subjective health status	0.019	<0.001	0.230

\*Pearson chi-square test for unweighted sample or frequency weighted sample; Rao-scott chi-square test for complex sample

#### 4. 고찰 및 결론

본 연구는 인구집단을 대표할 수 있는 국가 통계 자료의 음성장애 진단 자료를 이용하여 단일표본분석과 복합표본분석 방법에 따른 잠재적 위험요인의 차이를 분석하였다.

본 연구에서 국민건강영양조사의 표본은 가중치의 적용 여부에 따라 도출된 인구 집단의 특성 및 유병률에 차이가 있었다. 국민건강영양조사의 표본 추출은 1차 추출단위인 동읍면, 2차 추출단위인 조사구, 3차 추출단위인 가구의 3단계 층화표본추출방법을 기초로 하고 있다[9]. 이러한 층화표본추출방법은 지역사회 건강조사 등의 보건의료조사 뿐만 아니라 소규모사업체 근로실태조사, 고령화연구패널 등의 대부분의 국가 통계 조사에서 동일하게 이용되고 있다[12-14]. Kalsbeek & Heiss[2]에 따르면, 층화표본으로 추출된 국가자료에서 모집단을 추정할 때, 가중치를 무시하고 분석하면 추정치에 심각한 편향이 발생할 수 있다고 하였다[2]. 이는 본 연구의 결과에서도 입증되었는데, 가중치 적용 여부에 따라 모집단의 비율에 차이가 있었고, 특히 성별과 연령은 모집단의 특성이 반대로 도출되었으며, 이에 따른 음성장애 유병률의 경향도 차이가 있었다. Lee[8]는 복합표본으로 추출된 자료의 경우 층화, 집락추출, 가중치 등의 표본설계를 분석단계에서 반영하지 않으면 모수의 추정치에 심각한 편향이 발생할 수 있고, 추정량의 분산이 과소평가될 수 있음을 지적하였다. 따라서 층화표본추출 방법에 의해 조사된 국가 통계자료를 연구에 이용할 경우, 복합표본의 방법을 적용

해야만 도출된 결과를 전체 인구 집단으로 확대하여 신뢰성 있는 결과를 해석할 수 있다. 특히, 모집단의 특성을 반영하고, 특정 질병의 편향되지 않은 유병률을 추정하기 위해서는 분석과정에서 반드시 가중치가 적용된 방법이 요구된다.

본 연구에서 단순임의추출분석, 빈도 가중치를 적용한 단일표본분석, 복합표본분석을 이용하여 음성장애의 잠재적 위험요인을 통계적으로 검정한 결과, 빈도 가중치만을 적용한 단일표본분석의 경우에는 모든 변수가 유의한 위험요인으로 과대 예측되었고, 가중치를 적용하지 않은 단순임의추출 분석과 복합표본분석은 유의수준 및 결과에 차이가 있었다. 이러한 차이는 가중치 여부에 따른 교차 분석 방법의 차이에서 기인한 것으로 사료된다. 전통적으로 범주형 자료의 교차 분석 방법으로는 주로 피어슨의 교차검정이 수행되어 왔는데, 집락표본추출의 경우 관찰값들 간에 상관관계가 존재하게 되어 범주형 자료 분석에 이용되는 통계량들이  $\chi^2$  분포를 따르지 않게 된다[15-16]. 또한, Lv 등[3]도 집락표본추출자료에서 단순히 빈도가중치만을 부여할 경우에는 표본의 크기가 커져서 모든 변수의 유의수준이 작아질 가능성이 있음을 지적하였다[2]. 그럼에도 불구하고, 많은 연구자들이 이러한 추출방법에 의해 얻어진 자료가 단순임의 표본추출에 의해 추출된 것으로 가정하여 통계적 분석을 수행하고 있기 때문에 정확한 분석 결과를 얻지 못하고 있다[15]. 이와 같은 문제의 대안으로 Lee[17]는 복합표본자료의 범주형 분석 방법으로 교차검정에 대한 수정 검정 방법인 라오-스캇 통계량을 이용할 것을 제안하였고, 미국의

제3차 국민건강영양조사(Third National Health and Nutrition Examination Study)를 이용하여 왈드 검정법(Wald test)에 비해 라오-스캇 검정법이 안정적인 통계량을 제시할 수 있음을 입증하였다. 이상을 종합하면, 향후 국가자료를 이용하여 범주형 자료에서 관심 모수에 대한 편향되지 않은 추정값을 도출하고자 할 때에는 라오-스캇 교차검정이 효과적인 것이다.

연구의 제한점은 다음과 같다. 첫째, 본 연구의 잠재적 위험요인은 다른 요인을 통계적으로 보정하지 않은 단변량 분석(unadjusted analysis)에 의해서만 도출되었다. 따라서 도출된 잠재적 위험 요인이 개연성이 있는 위험요인 인지 확인하기 위해서는 복합표본설계에 의한 다변량 분석이 추가적으로 필요하다. 둘째, 도출된 음성장애의 잠재적 관련 요인은 단면연구를 기초로 하였기 때문에 관련성이 확인되었을지라도 이를 인과관계로 확대하여 해석할 수는 없다. 시간적 선후관계에 따른 인과관계를 파악하기 위해서는 전향적 연구(prospective study)가 필요하다.

이상으로 본 연구를 종합하면, 층화표본추출방법으로 조사된 국가통계자료를 이용할 경우에 연구의 결과가 전체 인구집단을 대표할 수 있도록 의미를 부여하기 위해서는 복합표본분석에서의 검정방법을 이용해야 할 것이다. 또한, 빈도 가중치만을 적용할 경우에는 결과에 대한 과잉해석의 가능성이 높기 때문에 각별한 주의가 요구된다.

## References

- [1] Oh, K. W., "Introduction to database for public health research", 2011 Epidemiology & Biostatistics Training Program, pp. 86, 2011.
- [2] Kalsbeek, W., & Heiss, G., "Building bridges between populations and samples in epidemiological studies", *Annu Rev Public Health*, Vol. 21, pp. 147-169, 2000.
- [3] Lv, J., He, P. P., Tu, W. X., & Li, L. M., "Estimation of sampling error on data from cluster sample survey", *Zhonghua Liu Xing Bing Xue Za Zhi*. Vol. 29, No. 1, pp. 78-80, 2008.
- [4] Byeon, H., "The association between duration of self-reported voice problems and voice disorders among adults", *Phonetics and Speech Sciences*, Vol. 3, No. 3, pp. 125-132, 2011.
- [5] Roy, N., Merrill, R. M., Gray, S. D., & Smith, E. M., "Voice disorders in the general population: prevalence, risk factors, and occupational impact", *Laryngoscope*, Vol. 115, No. 11, 1988-1995, 2005.
- [6] Preciado-López, J., Pérez-Fernández, C., Calzada-Uriondo, M., & Preciado-Ruiz, P., "Epidemiological study of voice disorders among teaching professionals of La Rioja, Spain", *J Voice*, Vol. 22, No. 4, pp. 489-508, 2008.
- [7] Bermúdez de Alvear, R. M., Barón, F. J., & Martínez-Arquerom, A. G., "School teachers' vocal use, risk factors, and voice disorder prevalence: guidelines to detect teachers with current voice problems", *Folia Phoniatrica et Logopaedica*, Vol. 63, pp. 209-215, 2011.
- [8] Lee, K. J., "Comparison of regression model approaches fitted to complex survey data. *Survey research*", Vol. 2, No. 1, pp. 73-86, 2000.
- [9] Korea Centers for Disease Control and Prevention, "The Third Korea National Health and Nutritional Examination Survey 2009", Seoul: Korea Centers for Disease Control and Prevention. 2010.
- [10] Korea National Statistical Office. "The Korean standard classification of occupations", Dejeon: Korea National Statistical Office. 2007.
- [11] Kim, K. H., Lee, K. H., Lee, S. M., Lee, S. Y., Lee, Y. S., Lim, K. R., Chang, J. E., Cho, S. W., Choi, E. H., Chung, S. T., Jin, E., & Son, M. "The proportional mortality ratios of specific-cause mortality by occupation and education among men aged 20-64 in Korea(1993-2004)", *JPMPH*, Vol. 40, No. 1, pp. 7-15, 2007.
- [12] Korea Centers for Disease Control and Prevention, "2009 Community Health Survey, Seoul: Korea Centers for Disease Control and Prevention", 2012.
- [13] Ministry of Employment and Labor, "2005 Survey report on labor conditions at small size establishments", Gwacheon: Ministry of Employment and Labor, 2006.
- [14] Korea Labor Institute, "2008 Korean Longitudinal Study of Ageing", Seoul: Korea Labor Institute, 2009
- [15] Nam, K., Cho, B., "χ<sup>2</sup> tests for categorical data in complex sampling surveys", *The Korean Economic Review*. Vol.20, No.1, pp. 277-288, 1993.
- [16] Rao, J. N. K., & Scott, A. J., "The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables", *Journal of the American Statistical Association*, Vol. 76, pp. 221-230, 1981.
- [17] Lee, S., "A study on varian stability of quadratic form test statistics under a complex survey design", *Journal of The Korean Official Statistics*, pp. 123-138, 1999.

**변 해 원**(Haewon Byeon)

[정회원]



- 2009년 8월 : 단국대학교 대학원 (이학석사)
- 2011년 8월 : 아주대학교 대학원 예방의학교실 박사 수료
- 2009년 9월 ~ 현재 : 아주대의료원 노인보건연구센터 연구원
- 2011년 3월 ~ 현재 : 대림대학교 언어재활과 교수

<관심분야>

음성 의학, 치매 예측 모형