

Reliability of self-reported data for prevalence and health life expectancy studies: comparison with sample cohort DB of National Health Insurance Services

Tae Yeon Kwon^a · Yousung Park^{b,1}

^aInstitute for Economic Research, Korea University;

^bDepartment of Statistics, Korea University

(Received August 24, 2016; Revised October 20, 2016; Accepted October 20, 2016)

Abstract

Korea Health Panel (KHP) data and Korea National Health and Nutrition Examination Survey (KNHANES) data are collected by self-assess and self-report for individual's health status and medical use. Previous studies have claimed that the reliability for prevalence rates and health life expectancies obtained from these data should be validated. National Health Insurance Services in Korea recently released a sample cohort DB that contain all data related to the use of medical facilities for all entire Korea citizens. It has been shown that disease-specific prevalence rates calculated from these data are representative and reliable for the entire population. In this paper, we evaluate the reliability of prevalence rates derived from self-reported data such as KHP and KNHANES by comparing to the prevalence rates from the sample cohort DB. We found that both KHP and KNHANES underestimate prevalence rates and in turn overestimate health life expectancies. Moreover, the general trends of health life expectancies might be distorted (except for the sample cohort DB) because of sampling and non-sampling errors.

Keywords: sample cohort DB, Korea National Health and Nutrition Examination Surveys, Korea Health Panel Data, prevalence rate, health life expectancy

1. 서론

유병률 연구(prevalence study)는 각종 의학, 보건학 등의 학술연구에 기본이 된다. 각 질병에 대한 유병률은 한국인 고유의 질병에 관한 위험 요인 및 예방과 치료에 중요한 정보를 제공하며, 이와 관련하여 이미 여러 분야에서 의학적 연구들이 진행되어 왔다 (Kim 등, 2005; Hur 등, 2008; Kim 등, 2013). 의학 연구 뿐 아니라 유병률 연구를 통하여 국가건강검진과 같은 현행 관리체계에 대한 시사점을 도출할 수 있고 보건의료분야 관련 국가 정책수립에 중요하게 사용될 수 있다. Laslett 등 (2012)은 사회경제적, 정치적 측면에서의 시기적절한 정보를 제공하기 위해 유병률을 비롯한 국민건강 통계의 중요성에 대

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013S1A5B8A01054750).

¹Corresponding author: Department of Statistics, 145, Anam-ro, Seongbuk-gu, Seoul 02841, Korea.
E-mail: yspark@korea.ac.kr

하여 지적하였다. Jung (2007)는 보건의료부문 연구사업 관련 재원 배분을 위한 우선순위를 설정하는 모델에 유병률을 이용하였다. 또한 Park 등 (2015)은 질병별 유병률과 유병률 추이에 대한 추정 및 예측을 건강보험재정추계에 이용하였다.

유병률 연구는 건강기대수명(health life expectancy) 연구에 있어서 무엇보다 중요한 역할을 한다. 단순 생존기간에 대한 지표인 기대수명에서 나아가 질병이나 부상으로 고통받는 기간을 제외한 건강한 삶을 유지하는 기간에 대해 나타내는 건강기대수명(health life expectancy)은 현대사회에서 삶의 질을 나타내는 주요한 지표로 역할을 하고 있다 (Lee 등, 2006). 이에 Sullivan (1971)은 사망률(mortality)과 유병률(morbidity)를 모두 고려한 단일지수로의 건강기대수명을 산출하는 방법을 제안하였다. 이후 2009년을 기준으로 미국은 207종류, 캐나다는 74종류, 일본은 51종류, 중국은 29종류, 대만은 11종류의 건강기대수명을 작성하여 공식적으로 공표하고 있다 (Jagger 등, 2012).

우리나라의 건강기대수명에 대한 연구는 Jung (2012)에서 소개되고 있으며, 국가별 건강기대수명을 연구하는 해외 연구와 (Mathers 등, 2001; Salomon 등, 2013), OECD 및 세계보건기구 통계에서도 산출된 값을 확인할 수 있다. 그러나 현재까지 산출된 우리나라의 건강기대수명 수치는 모두 국민건강영양조사자료 혹은 의료패널조사 자료에 근거하고 있으며, 이 자료의 주요한 부분은 개인의 건강상태에 대한 자가보고(self-report)에 근거하고 있다. 그러나 이러한 자료에 대해서는 건강상태 및 의료기관 이용에 대한 개인의 보고에 의존하기 때문에 실제 건강상태와 다를 수 있다는 비판이 있어왔다 (Crossley 등, 2002; Buratta와 Egidi, 2003; Cambois 등, 2007). 즉 이러한 자료들은 모집단에 대한 대표성이 충족된다 할지라도 응답한 자료의 신뢰도에 대한 검증이 추가되어야 유병률에 대한 대표성이 충족되고 이를 통해 추정된 기대여명 역시 신뢰할 수 있다고 할 수 있다.

최근 건강보험공단은 건강보험 및 의료급여권자 전체에 대한 국민건강정보 DB로부터 추출된 표본코호트 DB를 공개하였다. 우리나라의 건강보험제도는 전 국민을 대상으로 법에 의해 강제되기 때문에 전 국민의 의료기관 및 약국사용에 관련된 모든 자료가 국민건강보험공단에 저장 및 관리되고 있다. 자가보고형식(self-reported)이 아닌 측정형식(examination-based)의 자료로 구축된 국민건강정보 DB로부터 개체의 대표성이 확보되는 표본 추출방법에 의해 추출된 표본코호트 DB는 개인건강상태에 대한 자료의 대표성 및 신뢰도가 보장됨이 알려져 있다 (Lee 등, 2012; Lee 등, 2016).

이에 본 논문은 자가 응답에 기초한 국민건강영양조사 및 의료패널 자료로부터 산출한 유병률의 신뢰도에 대한 검증을 건강보험공단 표본코호트 DB의 결과와 비교함으로써 논의하고자 한다. 이들 유병률에 대한 비교는 당뇨병, 고혈압, 암, 그리고 심장질환과 뇌혈관질환의 질병별로 나누어 비교하였다. 그리고 산출된 유병률을 이용하여 각 질병으로부터의 건강기대수명을 산출하고 이들을 비교함으로써 유병률 연구를 위한 자료의 신뢰성에 대한 검증의 중요성을 논의하고자 한다.

Shenker 등 (2010)은 자가응답식 자료인 미국의 National Health and Nutrition Survey(NHANES)와 측정자료인 임상실험 자료에 각각 기초한 유병률들을 비교하여 응답식 자료에 기초한 유병률이 과소추정됨을 보인 바 있다. 본 논문에서는 한국의 국민건강영양조사와 의료패널 자료들로부터 구한 유병률이 측정형식의 자료인 표본코호트 DB로부터 구한 유병률에 비하여 역시 과소추정됨을 보이고 이러한 과소추정을 야기한 자료의 조사오차에 대하여 논의하고자 한다. 그리고 국민건강영양조사, 한국의료패널조사, 그리고 표본코호트 DB 각각에 근거하여 산출된 유병률을 이용하여 특정 질병에 걸리지 않고 삶을 영위하는 기대 수명인 질병별 건강기대수명을 추정하여 비교하였다. 자가응답자료에 근거하여 과소추정된 유병률은 건강기대수명의 과대추정의 문제로 이어졌음을 확인할 수 있었다. 또한 연도별 건강기대수명의 트렌드를 해석함에 있어 역동적 코호트구조를 유지하는 표본코호트 DB를 제외한 다른 자료는 문제가 있음을 확인할 수 있었다.

Table 2.1. Summary of sample cohort DB (NHIS) (unit: thousand people)

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Male	513	509	508	508	501	510	501	500	501	503	506	507
Female	512	508	508	508	501	511	500	500	501	503	506	507
Total	1,025	1,017	1,016	1,016	1,002	1,021	1,000	1,000	1,002	1,006	1,012	1,014

본 논문은 총 5장으로 구성되어 있다. 2장에서는 비교 분석에 사용되는 모든 데이터(표본코호트 DB, 국민건강영양조사, 한국의료패널조사)에 대하여 설명하였다. 3장에서는 국민건강영양조사 그리고 한국의료패널 조사로부터 구한 유병률과 그 신뢰성이 입증된 표본코호트 DB로부터 구한 유병률을 비교하였다. 4장에서는 그들 자료로부터 구한 건강기대수명을 비교하여 논의하였으며 5장에는 결론과 향후 과제를 제시하였다.

2. 건강정보에 관한 데이터

본 장에서는 표본코호트 DB에 대하여 요약하고 이와 비교대상이 될 국민건강영양조사 자료와 한국의료패널조사 자료에 대하여 간략히 소개하고자 한다. 그리고 각 자료의 비교를 위한 질병구분방법에 대하여 논의하고자 한다.

2.1. 표본코호트 DB(sample cohort DB of National Health Insurance Service; NHIS)

표본코호트 DB는 모집단인 국민건강정보 DB로부터 추출된 자료이다. 국민건강정보 DB는 2002년부터 건강보험 및 의료급여권자 전체에 대한 진료명세서와 진료내역, 상병내역, 처방전내역 등을 포함한 자료이다. 우리나라의 건강보험 가입은 의무가입으로 2010년 기준, 최근 3년간 우리나라 전체인구의 98% 이상이 건강보험 적용 대상자 (Park과 Lee, 2010)라는 점에서 이 DB는 사실상 장기간의 전 국민 전향적 코호트 연구 또는 중재의 효과를 검증할 수 있는 국내외 유일한 자료원이라 할 수 있다. 크게 자격 DB, 진료 DB 및 검진 DB 등으로 구성되어 있다.

표본코호트 DB는 모집단으로부터 추출된 개체들에 대한 자료조사(survey)가 아닌 각 개체의 의료기관 이용 시 자동으로 측정되어 누적되는 국민건강정보 DB로부터의 직접추출방식으로 구성된 데이터라는 점에서 기존의 건강자료들과는 큰 차이를 갖는다.

표본코호트 DB의 표본추출방식은 성별과 연령(18개 연령군), 소득분위(41개 구간)의 층화변수를 근거한, 비례배분(proportional allocation)에 의한 층화임의추출(stratified random sampling)방식이다. 최종 표본은 연령대-성별-소득분위 수준 조합에 따른 1,476개 층에서 비례배분된 표본수를 계통추출(systematic sampling)하였다. 표본의 대표성을 평가하는 목표변수로는 연간 총 의료비가 이용되었지만, Lee 등 (2012)와 Lee 등 (2016)의 모집단 정보에 근거한 연구에 의하면 표본코호트 DB는 이미 거주 지역과 보험료, 주요 질병의 유병률에서도 모집단을 적절하게 대표하는 것으로 나타났다.

표본코호트 DB는 2002년 기준으로 추출된 표본의 개체를 계속 유지하는 코호트자료이며, 이 때 시간의 변화에 따라 사망이나 추적실패의 이유로 발생하는 표본 개체감소 문제를 해결하기 위해 역동적 코호트(dynamic cohort) 구조를 사용하였다. 이 구조는 표본의 자연감소분 즉, 사망만큼의 새로운 표본을 기존 조사대상에 추가함으로써 표본의 크기를 유지하는 방식으로 이루어져 있다. 이러한 방식을 통한 표본의 구성은 우리나라의 인구구조의 변화를 그대로 반영할 수 있다는 점에서 장점을 갖는다고 하겠다. Table 2.1은 연도별 표본코호트 DB의 크기이다.

Table 2.2. Summary of Korea National Health and Nutrition Examination Survey (KNHANES) (unit: person)

	4 th cycle			5 th cycle			6 th cycle	
	2007	2008	2009	2010	2011	2012	2013	2014
Male	2,097	4,370	4,843	4,115	3,867	3,634	3,645	3,385
Female	2,497	5,374	5,690	4,843	4,651	4,424	4,373	4,165
Total	4,594	9,744	10,533	8,958	8,518	8,058	8,018	7,550

Table 2.3. Summary of Korea Health Panel (KHP) (unit: person)

	2008	2009	2010	2011	2012
Male	10,416	9,359	8,725	8,301	7,708
Female	10,867	9,794	9,160	8,734	8,164
Total	21,283	19,153	17,885	17,035	15,872

2.2. 국민건강영양조사(Korea National Health and Nutrition Examination Survey; KNHANES)

국민건강영양조사는 보건복지부 질병관리본부에서 수행하고 있는 조사로 국민의 건강수준, 건강행태, 식품 및 영양섭취 실태에 대한 국가 및 시도 단위의 대표성을 갖춘 통계를 산출하는 것을 목적으로 하고 있다. 이 데이터는 건강설문조사, 영양조사, 검진조사의 세 가지 조사로 이루어져 있다. 특히 세계보건기구(WHO)와 경제협력개발기구(OECD) 등에서 요청하는 흡연, 음주, 신체활동, 비만 관련 통계자료를 포함한다는 특징을 가지고 있다.

이 조사는 1998년부터 2005년까지 3년 주기의 단기조사 체계로 운영되었으며, 2007년부터 순환표본조사 방식에 근거한 연중조사 체계로 개편되어 현재 2014년도 자료까지 공개되어 있다. 최종 추출단위는 가구이며 추출된 가구 내 가구원은 전수조사한다. 표본 가구의 추출방법은 확률비례 층화집락추출과 계통추출방법이다. 자료의 구성은 Table 2.2와 같다.

2.3. 한국의료패널 데이터(Korea Health Panel; KHP)

한국의료패널은 한국보건사회연구원과 국민건강보험공단이 공동으로 수행하는 조사로, 의료이용형태와 의료비 지출 규모에 관한 정보뿐만 아니라 의료이용 및 의료비 지출에 영향을 미치는 요인들을 포괄적, 심층적으로 분석할 수 있는 패널 데이터를 구축하는데 주요 목적을 두고 있다. 이 자료는 크게 가구와 가구원에 대한 정보, 민간의료보험, 의료이용, 부가조사, 노인 장기요양에 관한 부분으로 이루어져 있다. 특히 아직 데이터 생산이 미흡한 영역인 비급여 부분, 의약품지출, 민간의료보험지출 등에 관한 정보를 포함하고 있다는 특징을 가지고 있다.

2008년 약 8,000가구를 추출하여 그 가구에 속해 있는 가구원을 대상으로 계속 조사하였다. 표본의 추출틀은 2005년에 실시된 인구주택총조사의 90%에 해당되는 전수자료이고, 지역(16개 시도)과 동부 및 읍면부(2개)를 층화변수로, 확률비례 2단계 층화집락추출 방식을 이용하여 표본을 추출하였다. 조사 대상의 사망 및 조사 거부 등의 이유로 이탈가구가 발생하여 2012년 조사가구가 5,856가구로 감소하였고 전국적으로 2,500가구를 추가 추출하여 2013년 조사부터 추가하였다. 현재 원자료는 2012년 자료까지 공개하고 있다.

의료패널 데이터는 연간데이터를 기본으로 하며, 1년 동안의 의료이용 자료를 모두 수집할 수 있었던 가구원 및 그 가구원이 속한 가구를 기준으로만 구성된다. 자료를 구성하는 응답자 수(가구원 수)는 Table 2.3과 같다.

Table 3.1. Prevalence rate for over 20 years old Korean (unit: %)

		2007	2008	2009	2010	2011	2012	2013	2014
Diabetes	NHIS	7.9	8.7	9.2	9.3	10.6	11.1	11.7	
	KHP		4.9	5.2	5.8	6.3	6.2		
	KNHANES	5.7	5.5	5.8	5.8	6.4	5.8	7.1	6.3
	KNHANES M	7.9	8.3	8.2	8.5	8.8	8.3	10.0	9.2
Hypertension	NHIS	15.4	17.0	17.9	17.8	19.5	20.2	20.6	
	KHP		13.0	13.6	15.3	16.3	15.9		
	KNHANES	13.6	14.9	15.6	16.5	15.8	16.4	16.2	15.6
	KNHANES M	20.5	23.0	29.8	29.8	26.5	26.7	25.7	24.3
Heart	NHIS	3.7	4.0	4.2	4.2	4.5	4.6	4.7	
	KHP		1.6	1.6	1.9	2.1	2.0		
	KNHANES	1.4	1.4	1.2	2.1	1.7	1.6	1.6	1.5
CVD	NHIS	2.7	3.3	3.4	3.4	4.0	4.1	4.2	
	KHP		1.6	1.6	2.0	2.1	2.1		
	KNHANES	1.2	1.1	1.0	0.7	0.9	0.9	1.5	1.2
Cancer	NHIS	2.2	2.6	2.8	3.0	3.3	3.6	3.9	
	KHP		1.8	2.0	2.5	2.7	2.7		
	KNHANES	0.6	1.0	1.0	1.2	1.4	1.3	1.4	1.9

CVD = cardiovascular disease.

2.4. 각 자료의 질병 구분

한국의료패널 자료와 표본코호트 DB의 질병구분은 한국표준질병사인분류(KDC-6)에 근거하여 정의하였다. 국민건강영양조사의 경우 당뇨와 고혈압의 측정자료에 대해서는 자료이용지침서(Centers for Disease Control & Prevention, 2014)를 따랐으며, 설문조사자료에 대해서는 건강설문조사의 설문지에 명시된 질병명에 의하여 정의하였다.

대표적 만성질환인 당뇨(KDC-6 code: E10-E14, KNHANES 질병명: 당뇨병)와 고혈압(KDC-6 code: I10-I13, KNHANES 질병명: 고혈압), 치명적 질병인 암(KDC-6 code: C00-C95, KNHANES 질병명: 암), 뇌혈관질환(KDC-6 code: I60-I69, KNHANES 질병명: 뇌졸중(중풍))과 심장질환(KDC-6 code: I20-I25, KNHANES 질병명: 심근경색, 협심증)에 대하여 살펴보았다.

이때, 고혈압의 경우 한국의료패널조사 자료는 고혈압성 질환(I11-I13)에 대한 분류가 따로 되어있지 않은 문제점이 있었다. 한국의료패널 자료의 이러한 한국표준질병사인분류에 의한 질병분류의 한계점은 이미 자료이용지침서(National Health Insurance Service(NHIS)와 Korea Institute for Health and Social Affairs(KIHASA), 2014)에 지적되어 있다.

3. 유병률 추정의 비교

Table 3.1은 각각의 자료에 근거한 20세 이상 인구의 질병별 유병률이다. 당뇨병과 고혈압의 경우 국민건강영양조사에는 의료인의 측정에 의한 결과(KNHANES M)와 자가응답 자료(KNHANES)가 모두 존재하여 이를 구분하여 산출하였다.

표본코호트 DB를 제외한 국민건강영양조사(KHANES, KHANES M)와 한국의료패널조사(KHP)에 근거한 본 논문에 제시된 모든 연도별 유병률 산출은 각각의 자료이용지침서에 따라, 자료에 제시된 사후층화 및 결측 보정을 위한 가중치를 고려하였으며, SAS의 SURVEYFREQ 프로시저를 사용하여

산출된 결과임을 밝혀둔다.

모든 질병에 대해 자가응답 자료인 국민건강영양조사(KHANES)와 한국의료패널자료(KHP)는 측정자료인 표본코호트(NHIS) 자료에 비하여 유병률을 과소추정함을 확인할 수 있다. 이는 자가응답에 의한 유병률의 추정은 직접 측정된 자료에 근거할 때보다 과소추정하는 경향이 있다는 기존의 연구와 일치한다 (Shenker 등, 2010). 그러나 국민건강영양조사 중 의료인의 측정자료(KNHANES M)는 표본코호트에 비해 고혈압의 경우 유병률을 과대 추정하였다.

표본코호트 DB의 고혈압 유병자는 의료기관으로부터 해당 질병의 진단을 받은 기록이 있는 환자이다. 혈압은 측정환경, 측정부위, 임상상황에 따라 변동성이 크기 때문에 여러 번 측정해야 한다. 이에 명확한 고혈압 환자인 경우를 제외하고 의심 환자의 경우 여러 차례에 걸쳐 혈압을 측정하여 최종 고혈압의 유병을 정의하도록 하고 있다 (Korean Society of Hypertension, 2013). 그러나 국민건강영양조사의 측정 자료에 의한 고혈압의 정의는 검진센터 방문을 통한 단 한 번의 혈압 측정에 근거하여 이루어지기 때문에 일시적으로 혈압이 상승한 사람까지 유병자로 정의할 수 있다. 이 때문에 국민건강영양조사의 측정자료(KNHANES M)가 다른 유병률은 과소추정하는 데에 반해 고혈압의 유병률을 과대 추정하는 현상이 발생하는 것으로 생각할 수 있다.

서로 다른 자료에 근거한 이러한 유병률 추정결과의 차이는 표본조사에 있어 발생하는 조사오차의 측면에서 설명할 수 있다. 조사오차는 크게 표집오차(sampling error)와 비표집오차(nonsampling error)로 구분되며, 비표집오차는 일반적으로 커버리지 오차(coverage error), 무응답 오차(non response error) 그리고 측정오차(measurement error)의 요소로 구성되어 있다 (Groves, 1989; Särndal 등, 2003; Groves 등, 2004; Lee, 2010). 표본구성의 대표성과 관련된 오차가 표집오차와 커버리지 오차이며 표집된 표본으로부터 자료의 값을 얻는 과정에서 발생할 수 있는 오차가 무응답 오차와 측정오차이다. 본 장에서는 각각의 자료에 근거한 유병률의 차이와 그 이유에 대하여 조사오차의 측면에서 보다 구체적으로 논의해보도록 하겠다.

3.1. 국민건강영양조사와 표본코호트 DB결과의 비교

Figure 3.1은 당뇨병 유병률의 추정에 관하여 표본코호트 결과, 국민건강영양조사의 의료인의 측정에 의한 결과 그리고 국민건강영양조사의 자가응답 자료에 근거한 결과를 비교한 결과이다. 국민건강영양조사는 3년 단위의 순환표본조사방법을 도입하고 있기 때문에 연도별 유병률은 전체 표본의 오직 1/3만을 사용하여 커버리지 오차에 영향을 받게 된다. 이러한 문제는 유사한 순환조사방법을 채택하고 있는 미국의 NHANES자료에서도 발생하는 문제이며 US Department of Health and Human Services (2014)는 이러한 순환자료를 이용한 연 단위 통계의 불안정성에 대하여 지적하고 자료의 조사주기인 2년 단위의 전국 통계를 산출할 것을 권장하고 있다. 이에 Figure 3.1(a)는 연도별 유병률을 Figure 3.1(b)는 순환표본조사는 국민건강영양조사의 특성에 따라 3년 평균 유병률(즉, 기수별 유병률)을 표본코호트의 당뇨병 유병률과 비교하고 있다. 3년 평균 유병률의 산출을 위하여 국민건강영양조사 자료는 자료에 제공된 기수 내 통합 가중치를 고려하여 산출하였고, 표본코호트 자료는 연도별로 산출된 유병률의 평균값을 취하였다.

Figure 3.1에서 국민건강영양조사, 특히 자가응답 자료에 근거한 경우의 유병률 추정에 있어 과소추정 현상뿐 아니라 그 과소추정 정도가 시간이 감에 따라 커지고 있음을 확인할 수 있었다. 국민건강영양조사는 동일한 조사방법을 매년 그리고 매 조사기수에 독립된 표본에 대하여 적용하기 때문에 이러한 유병률의 과소추정 정도의 심화는 표본의 대표성과 관련된 커버리지오차 혹은 표집오차의 문제일 가능성이 높다. 국민건강영양조사의 제 4기(2007년-2009년)와 6기(2013년-2015년)의 표본추출틀은 각각 2005년,

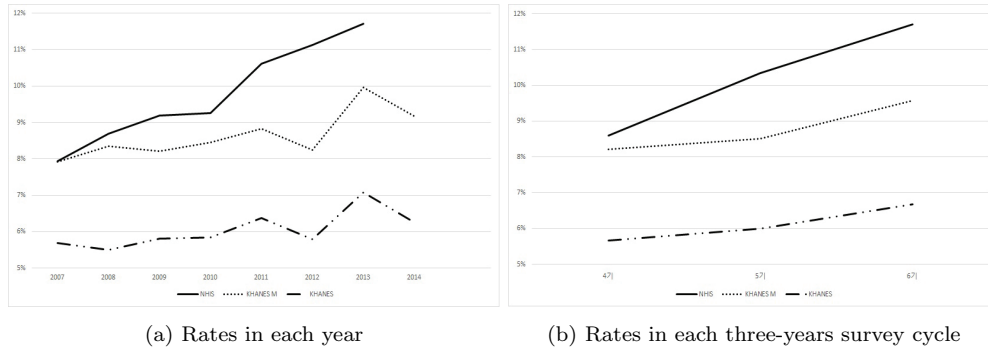


Figure 3.1. Trend of diabetes prevalence rates (over 20 years old people).

Table 3.2. Preference Rates of KHANES M and KHANES (unit: %)

Disease	Sex	Age	4 th cycle			5 th cycle			6 th cycle		
			NHIS	KHANES M	KHANES	NHIS	KHANES M	KHANES	NHIS	KHANES M	KHANES
Dia- betes	M	40-44	5.1	6.2	3.1	5.5	5.3	2.4	6.0	8.3	2.7
		70-74	26.5	17.1	12.7	30.3	21.0	16.9	33.8	26.4	19.5
	F	40-44	3.0	4.4	2.3	3.2	3.8	1.3	3.7	4.1	1.7
		70-74	27.1	21.1	18.0	31.4	21.5	19.0	34.4	27.6	25.3
Hyper- tension	M	40-44	8.4	25.5	8.4	9.2	28.1	7.5	9.5	26.4	7.6
		70-74	50.5	52.4	39.2	55.2	52.9	43.3	58.6	58.6	47.0
	F	40-44	5.0	10.5	4.9	4.7	9.9	4.3	4.5	6.5	3.2
		70-74	57.7	61.2	49.9	61.9	67.3	57.1	64.8	58.9	49.5

2010년 인구주택총조사 인구이고, 5기(2010년-2012년)의 경우 2009년 주민등록 인구이다. 이와 같은 표본추출틀의 변화는 변화하는 인구구조를 빠르게 반영하기 위함이다. 인구주택총조사는 5년 단위로 실시되는 조사이기 때문에 3년 단위로 실시되는 국민건강영양조사에서는 두 번의 인구주택총조사 사이에 시행된 조사를 위해 주민등록 인구를 표본추출틀로 설계하였다. 그러나 Park 등 (2010)는 인구주택총조사 인구와 주민등록 인구의 불일치를 지적한 바 있다. 즉, 주민등록 인구가 인구주택총조사 인구보다 더 크게 나타나며 연령대 그리고 성별 인구의 구성비 역시 두 자료가 크게 다르다는 것을 보여주고 있다. 이러한 점에 비추어 볼 때, 국민건강영양조사는 각 조사 기수별 표본추출틀이 일정한지 않은 문제가 발생하며 이는 유병률의 시계열적 추세분석에 있어 Figure 3.1에서 제기된 것과 같은 유병률의 과소 추정추세의 변화로 나타나고 있다.

Figure 3.1에 제시된 표본코호트 자료와 비교한 당뇨병 유병률의 차가 시간이 감에 따라 커지고 있는 현상을 보여주는 국민건강영양조사의 5기 및 6기 표본의 대표성을 재검토할 필요성을 제기하고 있다. 국민건강영양 조사의 당뇨병과 고혈압의 경우, 같은 조사 개체에 대하여 의료진에 의한 측정자료와 개별 설문에 의한 자료가 모두 존재하여 표본의 대표성 문제 및 무응답 오차를 동일하게 고정한 상태에서 자가진단에 따른 건강에 대한 응답 성향 및 이와 관련한 비표집 오차에 대하여 살펴볼 필요가 있다.

Table 3.2는 국민건강영양조사의 자가진단에 따른 응답 자료로부터 구한 당뇨병과 고혈압의 유병률과 전문의료진의 측정에 의한 자료로부터 구한 유병률들을 성별, 연령대별로 표본코호트 자료 결과와 비교한 표이다. 동일한 표본에 대한 자료임에도 불구하고 국민건강영양자료의 측정자료와 설문자료 각각에 근거한 당뇨병 및 고혈압 유병률은 큰 차이를 보이고 있다.

Table 3.2를 통하여 검진자료에 근거한 유병률과 설문자료에 근거한 유병률의 차이는 저연령일수록 그리고 여성에 비해 남성의 경우에 더 큼을 알 수 있다. 검진자료에 근거한 유병률이 설문자료에 근거한 유병률에 비해 40-44세의 경우 약 2.7배, 70-74세의 경우 약 1.2배가량 컸으며, 남성의 경우 약 2.3배, 여성의 경우 약 1.7배가량 컸다.

이는 실제로 당뇨병 혹은 고혈압이 의심되지만 실제 그 질병을 앓고 있다고 대답하기를 꺼려하여 거짓된 응답을 하거나 본인의 질병에 대해 인지하지 못하고 잘못된 응답을 하는 경향이 전 연령대에서 나타나고 있으나, 잘못된 응답의 정도는 성별, 연령별로 차이가 남을 의미한다. 연령대가 상대적으로 낮은 남성의 경우에 응답을 통한 유병과 측정을 통한 유병 간의 차이가 크게 나타났다.

이러한 자가응답 자료를 통한 유병률의 과소추정현상은 우리나라에만 국한된 문제가 아니다. Shenker 등 (2010)은 자가응답 자료인 미국의 NHANES 자료로부터 구한 유병률이 임상실험 자료로부터 구한 유병률에 비해 과소추정됨을 보인 바 있다. 또한 Crossley와 Kennedy (2002), Buratta와 Egidi (2003), Cambois 등 (2007)은 이러한 측정오차가 응답자의 특성에 따라 다를 수 있음을 보인 바 있다.

Table 3.2를 통해 확인한 또 다른 사실은 비교적 젊은 연령대인 40-44세의 당뇨병 고혈압 유병률에 대하여 건강검진 즉 측정 자료가 표본코호트에 비해 모두 과대 추정한다는 점이다. 특히, 젊은 연령대에서의 측정에 의한 고혈압 유병률의 과대추정 정도가 심하였다. 이는 앞서 설명한 바와 같이 국민건강영양 조사의 측정자료라 할지라도 실제 고혈압 및 당뇨병의 진단에서와 같이 의심환자에 대한 혈압 및 혈당의 반복측정은 불가능하기 때문에 발생하는 현상이라 할 수 있다.

이러한 현상은 유병률의 추정에 있어서 측정자료에 근거한다 하더라도 잘못된 측정방법에 의하여 오차가 발생할 수 있음을 의미한다. 이러한 이유 때문에 40-44세 여성과 남성의 고혈압의 경우, 모든 조사 기수에서 응답자료에 기초한 유병률이 측정자료에 기초한 유병률보다 오히려 표본코호트의 유병률과 비슷함을 알 수 있다.

3.2. 한국의료패널 조사와 표본코호트 DB 결과의 비교

Table 3.3은 성별, 질병별, 그리고 연령대 별로 표본코호트에 의해 산출된 유병률과 의료패널자료에 의해 산출된 유병률을 비교한 결과이다. 국민건강영양조사의 결과와 마찬가지로 의료패널 조사에 의한 유병률의 추정치가 실제 유병률을 과소추정하고 있다. 그러나 국민건강영양조사와는 달리 조사 자료가 코호트형식으로 누적되기 때문에 추정된 유병률의 추세 측면에서는 한국의료패널 데이터가 국민건강영양 조사에 비해 표본코호트와 유사한 패턴을 보이고 있다.

그러나 유사한 코호트 자료임에도 불구하고 표본코호트 DB가 매년 사망과 출생을 반영하여 연 표본의 크기를 일정하게 유지시키는 역동적 코호트 방식을 채택하고 있는 데에 반해 한국의료패널 조사는 2012년 조사까지 표본의 추가 없이 조사를 진행하였다. Table 2.3에서 제시된 바와 같이 의료패널 조사의 표본은 구성 후 5년차 조사인 2012년 조사에서 원 표본유지율이 70.56%에 불과하다. 이러한 표본의 마모(household sample loss)는 해당연도의 인구구조를 반영하지 못하고 커버리지 오차의 증가로 이어진다. Table 3.1의 결과를 보면, 유병률의 과소추정 정도는 당뇨병의 경우 2008년 3.8%에서 2012년 5.0%로 커지는 경향이 있다. 특히 가장 최근 조사인 2011년에서 2012년 사이의 당뇨병 유병률의 증가, 감소 여부에 대해 한국의료패널 자료는 표본코호트 자료와 정반대의 결론을 주고 있다.

또한 한국의료패널 자료는 표본추출 당시 지역 변수들만을 층화변수로 이용하여 기타 국민건강의 중요한 기준변수인 성별 및 연령대별 대표성은 결여되어 있다. 가중치의 부여를 통한 사후층화의 적용으로 이러한 문제를 보완하고 있으나 NHIS와 KIHASA (2014)은 광역시도 이하 지역단위의 통계산출을 제한하고 있다. 이는 사후층화에 의한 가중치는 성별 연령별 유병률의 표집오차를 해결할 수 없음을 의미한다.

Table 3.3. Disease, sex and age group specific prevalence rates calculated from KHP and NHIS (unit: %)

Disease	Sex	Age	2008		2009		2010		2011		2012	
			NHIS	KHP	NHIS	KHP	NHIS	KHP	NHIS	KHP	NHIS	KHP
Diabetes	F	40-44	3.0	0.2	3.1	0.4	3.0	0.5	3.4	1.0	3.4	1.0
		70-75	27.4	15.1	28.9	14.7	28.1	17.7	32.8	19.1	33.4	18.9
	M	40-44	5.1	1.1	5.3	1.4	5.1	1.8	5.6	1.8	5.9	1.9
		70-75	26.8	15.5	27.8	16.6	27.5	19.0	31.1	20.9	32.3	21.1
Hypertension	F	40-44	5.1	0.5	5.0	0.6	4.7	1.2	4.8	1.7	4.6	1.9
		70-75	59.1	42.6	61.2	45.0	56.9	51.0	64.1	53.2	64.6	51.2
	M	40-44	8.5	2.4	8.9	3.0	8.9	4.1	9.3	5.5	9.5	5.6
		70-75	51.0	34.5	52.9	36.9	51.7	41.3	56.2	43.1	57.6	44.2
Cancer	F	40-44	2.0	1.0	2.1	0.9	2.2	1.6	2.6	2.0	2.8	2.1
		70-75	5.1	4.7	5.5	5.4	5.4	7.0	6.3	7.4	6.4	7.9
	M	40-44	1.1	0.1	1.1	0.1	1.3	0.1	1.4	0.1	1.5	0.2
		70-75	10.2	4.1	11.2	3.8	10.7	6.3	12.1	7.3	12.5	7.3
Heart	F	40-44	1.1	0.0	1.1	0.1	0.9	0.1	1.0	0.1	0.9	0.2
		70-75	13.5	6.4	14.2	7.2	13.4	9.3	15.3	9.0	15.3	8.4
	M	40-44	1.8	0.0	1.8	0.0	1.8	0.1	1.8	0.0	1.7	0.3
		70-75	14.6	4.4	15.3	5.3	15.3	7.0	15.9	7.6	16.0	8.0
CVD	F	40-44	0.7	0.0	0.7	0.0	0.7	0.0	0.8	0.1	0.8	0.0
		70-75	13.7	5.6	14.6	6.0	13.8	7.0	15.6	7.7	15.7	6.9
	M	40-44	0.9	0.4	0.9	0.4	0.8	0.3	1.0	0.0	1.0	0.3
		70-75	14.4	4.9	14.5	5.0	13.7	5.4	15.7	5.4	15.2	6.5

CVD = cardiovascular disease.

이를 확인하기 위하여 Table 3.3에 질병별, 성별 그리고 연령별 의료패널 데이터에 의한 유병률과 표본 코호트자료에 근거한 유병률을 비교한 결과를 제시하였다. 모든 경우에 의료패널 자료는 유병률을 과소 추정하고 있음을 알 수 있다. 또한 당뇨병과 고혈압의 경우 젊은 연령대일수록 그리고 남성에 비하여 여성의 경우에 과소추정 정도가 더 심하게 나타나고 있다.

성별 연령대에 따른 그룹별로 만성질환인 당뇨와 고혈압의 경우와 비교하여 유병률이 상대적으로 낮은 치명적 질환인 암, 심장질환, 뇌혈관질환의 유병율의 과소추정 정도는 큰 문제가 있음을 볼 수 있다. 특히 상대적으로 젊은 연령대인 40-44세의 뇌혈관질환과 심장질환의 경우, 조사된 표본 내에 유병자가 한 명도 포함되지 않아 유병률을 산출할 수 없는 경우가 여러 해 반복적으로 나타남을 알 수 있다. 이러한 경우 사후층화에 의한 가중치로 성별 연령별 유병률의 표집오차를 해결할 수 없음이 자명하다고 하겠다.

한국의료패널 조사자료 역시 응답에 기초한 자료이기 때문에 유병률 산출에 있어 국민건강영양조사와 같은 비표집오차인 측정오차가 발생한다. 또한 코호트 형식의 자료이기 때문에 발생하게 되는 비표집오차 역시 존재하게 된다.

한국의료패널 조사는 가계부 기입방식과 조사원 면접조사의 조사방식을 이용한 의료코호트 자료를 제공하고 있다. 이 때 연중 발생하게 되는 입원, 응급, 외래서비스 이용 또는 의약품 이용에 대해서 기억을 보조할 수 있는 장치로 일지조사법(diary method)이 이용되며, 이를 통해 1년 동안의 모든 의료이용 자료를 수집할 수 있었던 사람에 대한 자료만으로 연간 의료패널데이터를 구성한다. 단, 사망의 경우는 연간의료이용 자료의 수집과 관계없이 포함한다. 이와 같은 조사의 특성상 무응답으로 인한 비표집오차가 발생하며 패널조사의 특성인 계속되는 응답에 대한 부담(respondent burden) 때문에 무응답률이 증가하고 무응답에 의한 비표집오차가 증가하게 된다 (Kalton 등, 1989).

또한, 한국의료패널 조사는 추출된 표본에 대하여 일정기간 단위로 반복된 조사를 시행하는 패널 조사로 표본으로 추출된 개체는 지난 1년간의 의료기관이용에 대한 이유와 시기까지 기억(혹은 기록)하여야 한다. 이로 인한 의료기관 이용 시점의 착각으로 인한 잘못된 응답(telescoping), 빠트린 응답(omission)에 의한 비표집오차가 발생하게 된다. 더불어, 일정 시간을 두고 반복적으로 같은 문항에 대답해야 하는 패널조사의 특성상 이전 응답 값에 영향을 받은 응답(panel conditioning), 반복된 조사로 인한 부담감에 의한 잘못되거나 성의 없는 응답(respondent burden) 그리고 조사대상이 패널에 포함된 기간에 따라 응답한 정보의 질이 달라질 가능성에 따른 오차(time-in-sample-error) 등의 측정오차의 발생을 야기할 수 있다 (Bailar, 1989; Kalton 등, 1989). 이와 같은 교체 없는 표본에 대한 조사에 근거한 코호트 자료의 비표집 오차의 우려 때문에 여러 패널 조사에서는 실제로 이미 다양한 형태의 교체표본조사(rotation sampling design) 방법을 도입하고 있다 (Bailar, 1989; Park 등, 2007).

4. 건강기대수명

각각의 자료로부터 산출된 유병률의 차이는 해당 질병에 걸리지 않고 사는 삶의 기간인 건강기대수명의 추정에 있어 차이를 발생시킨다. 이에 본 장에서는 국민건강영양조사, 의료패널조사 그리고 표본코호트 DB에 근거한 질병별 건강기대수명을 도출하고 이를 비교하여 논의하고자 한다.

4.1. 건강기대수명의 산출방법

건강기대수명은 몸이나 정신에 아무 탈 없이 튼튼한 상태로 활동을 하며 산 기간으로, 연령별 기대여명에서 질병으로 몸이 아픈 기간을 제외한 기간으로 정의한다 (Jung, 2012).

본 논문에서는 표본코호트 DB 자료에서 주어진 사망 사건을 이용하여 5세 단위의 간이 기대여명 및 건강기대수명을 산출할 것이다. 본 연구는 여러 건강 자료에 근거한 건강기대수명의 추정을 비교하기 위하여 표본코호트 자료를 제외한 자료들의 1세별 유병률을 산출에 있어서의 한계성 때문에 5세 단위로 건강기대수명 및 기대여명을 산출하였다. 이때 기대여명은 통계청의 5세 단위 사망률 자료를 이용하였으나 본 연구는 5세 단위의 간이생명표에 근거한 기대여명이기 때문에 1세 단위의 완전생명표에 근거한 결과와 약간의 차이가 발생할 수 있음 (Park 등, 2011)을 밝혀둔다.

기대여명 산출을 위해 먼저 다음의 식 (4.1)을 이용하여 실제 연령사망률(M_x)을 사망확률(q_x)로 변환한다.

$$q_x = \frac{nM_x}{1 + (n - A_n)M_x}, \quad \text{for } x = -0, \dots, k. \quad (4.1)$$

이때, $x = 0$ 이면

$$\text{Male : } A_n = \begin{cases} 0.330, & \text{if } M_x \geq 0.107, \\ 0.045 + 2.684M_x, & \text{if } M_x < 0.107, \end{cases}$$

$$\text{Female : } A_n = \begin{cases} 0.350, & \text{if } M_x \geq 0.107, \\ 0.053 + 2.800M_x, & \text{if } M_x < 0.107, \end{cases}$$

그리고 $x = 1, \dots, k - 1$ 이면 $A_n = n/2$ 이다.

이때 연령별 사망률(M_x)은 통계청의 연령별 사망률이고, k 는 전체 연령그룹의 수, 그리고 n 은 연령그룹을 나누는 단위로 본 논문에서는 $k = 20$ 개 $n = 5$ 세로 하였다. 사망확률(q_x)은 x 세의 사람이 $x + n$ 세에 도달하기 전 사망할 확률을 의미하며, x 세에서 $x + n$ 세 사이의 구간에서 가장 코호트(통상 10만명)

의 생존자 수 중 사망자 수의 비율로 정의된다 (Park과 Kim, 2011). 그리고 여기서 A_n 은 단위기간인 n 년 사이에 사망한 코호트가 평균적으로 사는 기간을 의미한다.

다음으로 x 세에서 $x+n$ 세까지의 사망률을 사용하여 해당 연령구간 내의 사망자 수와 생존자 수(l_x)를 구한다. 여기서 x 세의 생존자 수를 구할 때 시작점은 100,000명으로 한다. 그리고 x 세의 생존자 수가 $x+n$ 세에 이르기까지 살아가는 년 수인 총인년 수(L_x)를 구하면 다음 식 (4.2)에 따라 기대여명(e_x)를 도출할 수 있다. 특히 x 가 0세일 때의 기대여명을 기대수명(life expectancy)이라고 한다.

$$e_x = \frac{\sum_{y=x}^{\infty} L_y}{l_x}. \quad (4.2)$$

이때 x 세에서의 생존자수, l_x 는

$$l_x = \begin{cases} 100000, & \text{for } x = 0, \\ (1 - q_{x-n})l_{x-n}, & \text{for } x = 1, \dots, k \end{cases}$$

이고, x 세에서의 총인년수 L_x 는

$$L_x = \begin{cases} l_{x+1} + A_n(l_x - l_{x+1}), & \text{for } x = 0, \dots, k-1, \\ \frac{l_x}{M_x}, & \text{for } x = k \end{cases}$$

이다.

기대여명을 구하는 일련의 과정과 비슷한 방식으로 건강기대수명도 도출할 수 있다. 건강기대수명을 산출하는 방법은 크게 단일상태(single-state) 접근법과 다상태(multi-state) 접근법으로 나눌 수 있다 (Newnan, 1988; Rogers 등, 1989; Davis 등, 2001). 본 논문에서는 식 (4.3)에 제시된 Sullivan의 단일상태접근법을 통해 건강기대수명을 구할 것이다 (Sullivan, 1971).

$$he_x = \frac{\sum_{y=x}^{\infty} L_y (1 - d_y)}{l_x}. \quad (4.3)$$

이때, d_x 는 x 세에서의 유병률을 나타낸다.

4.2. 건강기대수명의 추정결과

식 (4.3)에 제시된 바와 같이 해당 연령대에서의 건강기대수명은 해당 연령 및 그 보다 고연령 대에서의 추정된 유병률을 모두 반영한 수치이기 때문에 건강기대수명의 비교를 통하여 유병률 추정의 전반적인 특징을 비교할 수 있다. 자가응답에 의해 산출된 유병률의 경우, 가구조사라는 점에서 20세 이하 연령그룹에서의 유병률은 그 조사의 특성상 정확하게 산출되기 어려울 수 있다. 이에 이러한 요인으로 인한 국민건강영양조사와 한국의료패널조사의 한계점을 제외한 자가응답 건강자료의 신뢰성에 대한 공정한 비교를 위하여 0세에서의 건강기대수명이 아닌 40-44세 연령구간에서의 기대여명 및 건강기대수명을 먼저 비교해보도록 하겠다. 4.1장에 제시한 바와 같이 사망률을 계산하는 데에는 통계청 연령별 사망률을 공통으로 이용하였기 때문에 자료에 상관없이 기대여명(life expectancy; LE)은 동일하며 건강기대수명 산출에 있어서는 각 자료의 유병률의 추정치에만 영향을 받는다.

Table 4.1에는 기대여명과 각각의 자료로부터 산출한 유병률에 근거하여 추정한 40-44세 연령에서의 건강기대수명(health life expectancy; HLE)이 성별에 따라 만성질환의 하나인 당뇨병과 치명적 질환에

Table 4.1. Sex specific diabete-free and cancer-free health life expectancy at age 40–44 years old

			2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
LE			42.1	42.5	42.9	43.4	43.7	44.3	44.8	44.9	45.2	45.2	45.7	46.2
F	Diabetes	NHIS	37.1	37.0	36.7	37.2	37.3	37.1	37.2	37.5	36.5	36.2	36.3	
	- free	KNHANES M					38.4	37.9	38.3	38.8	38.3	38.9	37.4	39.6
	HLE	KNHANES					39.8	39.4	40.1	40.4	39.3	40.3	39.7	41.0
		KHP							40.3	40.6	40.2	40.2	40.3	
	Cancer	NHIS	40.7	41.3	41.3	41.9	42.1	42.5	42.9	42.7	43.0	42.8	43.1	
	- free	KNHANES					43.4	43.6	44.0	44.1	44.3	44.3	44.8	45.0
	HLE	KHP							42.9	43.2	42.8	43.0		42.4
LE			35.7	36.0	36.5	37.0	37.4	37.8	38.2	38.4	38.7	38.9	39.5	40.0
M	Diabetes	NHIS	31.2	31.0	30.7	31.2	31.3	31.1	31.2	31.3	30.8	30.6	30.8	
	- free	KNHANES M					31.5	31.9	31.9	32.2	32.5	32.5	32.0	32.8
	HLE	KNHANES					32.5	34.0	33.7	33.8	34.2	34.6	34.3	34.7
		KHP							34.6	34.7	34.3	34.3	34.3	
	Cancer	NHIS	34.1	34.2	34.8	35.2	35.4	35.5	35.8	35.9	36.0	36.0	36.4	
	- free	KNHANES					36.8	37.0	37.5	37.6	38.0	38.1	38.7	38.7
	HLE	KHP							36.7	37.1	36.9	37.0	37.1	

하나인 암에 대하여 제시되어있다. 우리나라의 40–44세 인구의 기대여명은 증가하고 있으며 2003년부터 2014년까지 11년간 남/녀 모두에서 약 4세가량 증가하였다. 표본코호트 자료에 근거한 건강기대수명은 만성질환인 당뇨병의 경우 2010년까지 증가하다가 이후 감소하고 있다. 이는 평균적으로 과거에 비하여 오래 살지만 당뇨병에 걸리지 않고 건강하게 사는 기간은 줄어들고 있음을, 다시 말해 당뇨병의 투병과 관련한 한국인의 삶의 질은 나빠지고 있음을 보여준다. 그러나 치명적 질환인 암의 경우는 건강기대수명 역시 기대여명과 같이 증가하고 있다.

Table 4.1를 통하여 질병 및 성별에 관계없이 자가응답에 근거하여 과소추정된 유병률은 건강기대수명의 과대추정으로 이어짐을 확인할 수 있다. 국민건강영양조사와 한국의료패널 조사에서 모두 공통적으로 치명적 질환인 암에 비하여 당뇨병으로부터의 건강에 대한 건강기대수명의 차이가 더 큼을 볼 수 있다.

건강기대수명의 과대추정 문제 외에 시간에 흐름에 따른 건강기대수명의 추세에 대한 해석의 차이가 있는지를 살펴보기 위해 Table 4.1의 결과를 Figure 4.1에 제시하였다. 성별에 따라 결과의 차이에 대한 해석이 크게 다르지 않기 때문에 여성에 대한 결과만 제시하겠다.

국민건강영양조사 결과를 이용한 건강기대수명 연구의 한계점은 Figure 4.1(a)에 제시된 당뇨병에 대한 연도별 건강기대수명 결과에서 확인할 수 있다. 국민건강영양조사의 측정에 의한 결과와 설문에 의한 결과 모두 연도별 당뇨병 건강기대수명의 추세는 문제점을 드러내고 있다. 연도별로 표본코호트나 의료패널 조사에 근거한 결과에 비해 그 변동폭이 매우 크며 변동방향 역시 일정하지 않아, 한국인의 당뇨병에 관련한 삶의 질의 변화를 파악하는데 한계를 보여주고 있다.

이는 3장에 제기된 바와 같이 자가응답 자료의 측정오차에 의한 것일 뿐만 아니라 패널조사가 아닌 순환 자료를 이용한 연 단위 통계의 불안정성에 기인한 결과라 하겠다. 국민건강영양조사의 조사단위인 3년 단위로 건강기대수명을 구하여 비교하여도 이러한 문제점은 해결되지 않음을 Figure 4.1(c)를 통해 확인할 수 있다. 이때 국민건강영양조사의 3년 단위 건강기대수명은 자료에서 제공하는 조사기수 내 병합을 위한 가중치를 이용하여 3년 단위 유병률을 산출하여 계산하였다. 국민건강영양조사 설문에 근거한 당뇨병 건강기대수명 추세의 방향은 다른 자료에 근거한 추세의 변화 방향과 반대방향을 보여주고 있다. 조사기수별 일정하지 않은 표본추출틀 문제는 건강기대수명 연구에 그대로 반영되어 있다고 할 수 있다.

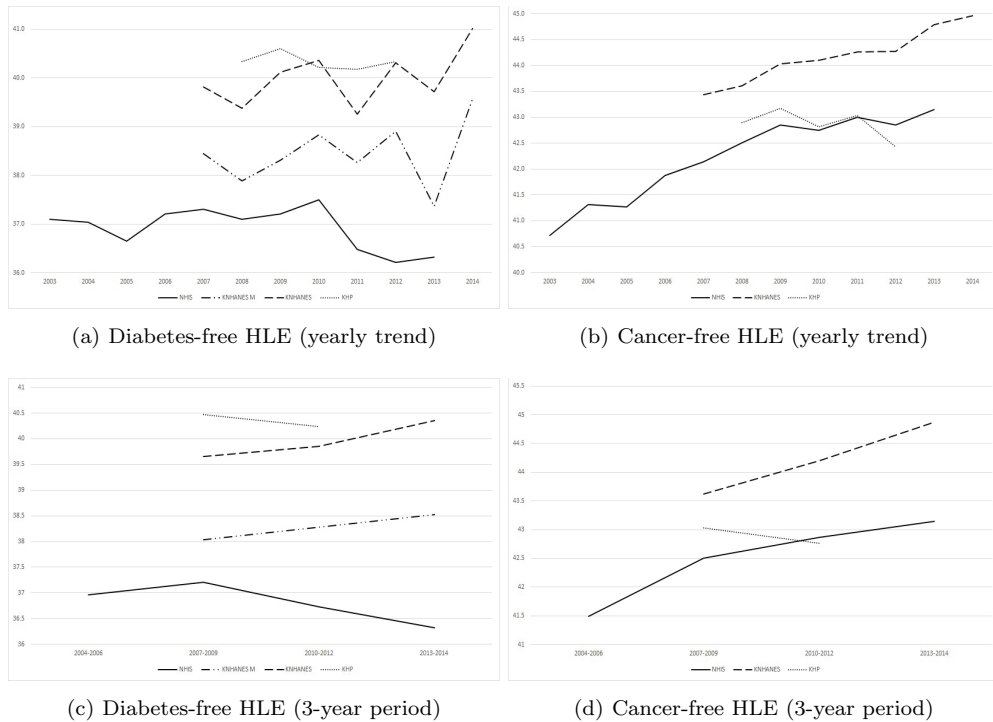


Figure 4.1. Trend of HLE (40-45 years old female).

의료패널 조사결과에 근거한 유병률 및 건강기대수명 연구의 문제점은 치명적 질환인 암의 경우에 있어 확연히 확인된다. 표본코호트자료의 암에 대한 유병률은 국립암센터에 등록되기 때문에 암유병률의 정밀도는 매우 높다고 할 수 있다. 국민건강영양조사와 표본코호트 자료에 근거한 결과는 암에 대한 건강기대수명이 증가함을 보여주고 있는데 반해 의료패널조사 결과에 근거한 암에 대한 건강기대수명은 시간이 갈수록 크게 변화하고 있지 않으며 오히려 감소하는 것으로 나타났다. 이는 의료패널 자료는 표본코호트 자료와 같은 코호트 자료이나 표본코호트 자료와는 달리 2013년 이전까지 표본의 추가가 없이 마모만 진행되어 표본의 크기가 점차 감소하여, 특히 유병률이 상대적으로 낮은 치명적 질환인 암의 경우, 단순한 분모의 감소로 인한 암유병률의 급증으로 나타나는 현상이라 하겠다.

Table 4.2에는 전 연령대에서의 사망률 및 유병률이 모두 고려되어 일반적인 기대여명 및 건강기대수명 비교연구에서 이용되는 0세에서의 기대여명과 건강기대수명 그리고 고령층의 유병률 및 사망률을 고려하기 위해 70-75세의 기대여명 및 건강기대수명을 제시하였다. 자가응답 자료에 근거한 건강기대수명의 과대추정문제와 그 추세해석에 있어서의 문제점은 40-45세에서와 마찬가지로 확인된다.

표본코호트에 근거한 건강기대수명의 연도별 변동률은 2011년을 제외하고 사망률만 고려한 기대여명의 연도별 변동폭과 유사하게 나타나고 있다. 그러나 국민건강영양조사 및 의료패널자료에 근거한 건강기대수명의 연도별 변동폭은 그 크기가 일정하지 않으며 변동방향 역시 해석하기 불가능할 정도로 일정하지 않다. 또한 변동크기 역시 기대여명의 연도별 변동크기에 비하여 매우 크게 나타나고 있으며, 특히 70-75세의 변동크기는 매우 크게 나타나고 있다.

국민건강영양조사를 통해 산출된 70-75세 여성의 당뇨병에 대한 건강기대수명의 경우 측정에 의한 결과

Table 4.2. Sex specific diabete-free and cancer-free healthy life expectancy at age 0 and 70–75 years old

			2007	2008	2009	2010	2011	2012	2013	2014	
0 yr	F	LE	83.6	84.2	84.7	84.9	85.3	85.3	85.9	86.3	
		Diabetes	NHIS	77.0	76.9	76.9	77.3	75.3	75.1	75.2	
		-free	KNHANES M	78.2	77.6	78.0	78.6	78.2	78.9	77.2	79.5
		HLE	KNHANES	79.7	79.3	80.0	80.3	79.4	80.4	79.8	81.1
			KHP		80.3	80.6	80.2	80.2	80.5		
		Cancer	NHIS	81.9	82.3	82.6	82.6	81.8	81.7	82.0	
		-free	KNHANES	83.3	83.5	83.9	84.1	84.2	84.3	84.8	85.1
		HLE	KHP		72.9	73.1	72.9	83.0	77.6		
			LE	77.0	77.4	77.8	78.0	78.4	78.7	79.3	79.8
		Diabetes	NHIS	70.7	70.6	70.7	70.8	69.3	69.2	69.4	
	-free	KNHANES M	70.4	71.4	71.3	71.5	71.8	72.1	71.4	72.6	
	HLE	KNHANES	71.8	73.6	73.3	73.4	73.7	74.3	74.1	74.5	
		KHP		69.3	69.4	69.0	69.0	64.2			
	Cancer	NHIS	75.0	75.1	75.4	75.4	74.6	74.7	75.1		
-free	KNHANES	76.4	76.7	77.1	77.2	77.6	77.9	78.4	78.4		
HLE	KHP		71.4	61.9	61.7	71.7	71.8				
70–75 yr	F	LE	16.2	16.7	17.1	17.2	17.4	17.3	17.8	18.2	
		Diabetes	NHIS	12.9	12.8	13.0	13.1	12.4	12.0	12.1	
		-free	KNHANES M	13.5	13.5	13.5	14.0	12.9	13.7	12.5	14.1
		HLE	KNHANES	14.2	14.2	14.6	14.8	13.3	14.6	13.9	14.8
			KHP		13.7	14.1	13.8	13.9	14.0		
		Cancer	NHIS	15.5	15.9	16.3	16.2	16.5	16.4	16.7	
		-free	KNHANES	16.1	16.5	16.9	17.0	17.2	17.2	17.3	17.8
		HLE	KHP		16.0	16.4	16.2	16.4	15.7		
			LE	12.7	12.9	13.3	13.3	13.6	13.6	14.0	14.3
		Diabetes	NHIS	9.8	9.7	9.8	9.8	9.4	9.2	9.4	
	-free	KNHANES M	10.8	10.9	10.8	10.5	11.7	10.7	11.0	11.9	
	HLE	KNHANES	11.1	11.7	11.6	11.2	12.0	11.6	11.9	12.0	
		KHP		10.9	11.1	10.8	10.7	10.6			
	Cancer	NHIS	11.5	11.5	11.8	11.7	11.8	11.6	12.0		
-free	KNHANES	12.6	12.4	12.8	13.0	13.1	13.1	13.6	13.7		
HLE	KHP		12.1	12.4	12.2	12.0	12.0				

와 설문에 의한 결과 모두 2011년부터 2014년까지 -5%에서 +13%까지의 변동 폭으로 해마다 증가 감소가 반복되고 있다. 이는 앞서 제기된 국민건강영양조사의 측정오차를 포함한 비표집오차 뿐 아니라 표본의 대표성과 관련한 표집오차의 문제로 인한 표본의 안정성과 대표성에 문제와 연관되어 있다고 할 수 있다. 이러한 오차들로 인하여 국민건강영양조사 자료를 사용할 때, 연령별 건강기대수명의 추세연구, 고연령 건강기대수명, 그리고 최근 건강기대수명 연구에 있어서 조심스러운 접근이 필요하다고 할 수 있다.

의료패널 조사를 통해 산출된 여성의 0세에서의 암으로부터의 건강기대수명은 2011년 약 14% 증가, 2012년 약 17% 감소하였다. 남성의 경우에도 2009년 13% 감소, 2011년 약 16% 증가하였다. 동일한 자료를 이용한 40–44세 및 70–75세의 암에 대한 건강기대수명의 추세변화에서 이와 같은 급격한 변화가 보이지 않았다. 이는 특히 드문 유병자 및 그와 관련한 사망자가 발생하는 저연령의 치명적 질환에 대하여 표본의 추가 없이 마모만 발생하는 의료패널 조사를 통하여 건강기대수명에 대한 해석은 문제가 있음을 보여주고 있다.

5. 결론

본 논문은 표본의 대표성 및 유병률의 신뢰도가 검증된 건강보험공단의 표본코호트 DB를 이용한 유병률에 근거하여, 기존의 자가응답식 자료인 국민건강영양조사와 한국의료패널 데이터의 질병별 유병률 및 건강기대수명의 신뢰도에 대하여 검증하였다. 그 결과 이들 응답식 자료를 이용한 유병률은 과소추정되고 이로 인하여 건강기대수명은 과대추정되고 있음을 확인하였다. 이와 같은 현상은 모든 성별과 연령대에서 질병에 관계없이 확인되었다.

유병률의 변화 추세에 대한 분석에 있어서는 유일하게 패널조사방법이 아닌 국민건강영양조사의 경우 매 조사기수별 일치하지 않는 표본추출틀 문제와, 3년 단위의 순환조사방식을 이용한 조사라는 점에서 연 단위 유병률 추세추정이 안정적이지 못함을 확인할 수 있었다. 한국의료패널조사의 경우, 동일한 개체에 대한 패널조사가 가지는 telescoping, panel conditioning, time-in-sample-error, 그리고 respondent burden 등의 문제로 인한 비표집오차와 2012년까지 표본의 추가가 이루어지지 않아 표본의 마모로부터 야기되는 조사오차로 인해, 유병률이 낮은 질병 및 집단에 있어서의 유병률 및 건강기대여병의 추정에는 문제가 있음을 확인하였다.

국민건강영양조사를 이용할 경우 코호트자료가 아니라는 점에서 건강기대수명 추세를 통한 삶의 질의 변화에 대한 연구에는 한계가 있음을 확인하였고 의료패널조사는 코호트의 마모 및 성별 연령별 세부그룹별 유병률 산출의 문제점으로 특히 유병률이 낮은 저연령의 치명적 질환에 대한 건강기대수명 연구에는 한계를 드러내고 있음을 확인하였다. 이러한 점에서 최근 일반에 공개된 국민건강보험공단의 표본코호트 DB는 보다 정확한 유병률 및 건강기대수명에 관한 연구를 가능하게 하였다는 측면에서 그 의의가 크다고 할 수 있다. 표본코호트 자료는 전 국민이 의무가입 대상인 국민건강보험을 기초로 한 자료로 작성된 국민건강정보 DB에서 개인단위로 추출되어 구성된 자료이다. 이는 가구단위의 추출로 구성된 기존의 건강조사 표본들이 갖는 커버리지 오차 및 표집오차에 대한 우려가 적다고 할 수 있다. 또한 표본의 추출 역시 지역 변수들에만 근거한 다른 건강 데이터와 달리 성별 및 연령대를 고려한 표본을 추출하여 이들 변수에 따라 세분화된 그룹별 유병률 및 건강기대수명 비교에 대한 보다 정확한 연구를 가능하게 하였다.

무엇보다 표본코호트 DB의 경우 건강상태에 대한 자료는 진료 혹은 검진 자료로, 자가진단에 따른 응답형식 자료가 아닌 의료인의 측정에 의한 자료이고 개인이 의료기관을 이용하였을 경우 자동으로 자료가 누적되기 때문에 개인의 기억이나 자가진단에 의존한 다른 건강데이터에 비하여 무응답오차와 측정오차에 대한 우려가 적다는 장점이 있다.

현재까지 우리나라의 건강기대수명의 산출은 자가보고형식의 설문자료에 근거하였기 때문에 본 논문에서 밝힌 측정오차 및 커버리지 오차를 포함한 비표집오차와 일정하지 못한 표본추출틀 혹은 표본의 마모로부터 발생하는 표집오차의 문제들로부터 자유롭지 못하다. 이에 우리나라 인구구성의 특징을 잘 반영하여 추출된 개체들의 건강에 대한 의료인의 측정에만 근거하여 누적된 국민건강자료인 표본코호트 DB를 사용한 건강기대수명의 산출과 이용이 반드시 필요하다고 하겠다.

Acknowledgments

- 본 연구는 한국보건사회연구원과 국민건강보험공단이 공동으로 주관하는 한국의료패널 2008-2012년 연간데이터(Version 1.1)를 활용한 결과로 연구의 결과는 한국보건사회연구원과 국민건강보험공단과는 관련이 없음을 밝힙니다.
- 본 연구는 질병관리본부에서 제공하는 국민건강영양조사 2007-2014년을 활용하였으며 연구의 결과는 질병관리본부와는 관련이 없음을 밝힙니다.

- 본 연구는 국민건강보험공단의 NHIS-2014-2-006를 활용한 것으로 연구의 결과는 국민 건강보험공단과 관련이 없음을 밝힙니다.

References

- Bailar, B. A. (1989). Information needs, surveys, and measurement errors. In D. Kasprzyk et al. (Eds.), *Panel Surveys* (pp. 1–24), Wiley-Interscience, New York.
- Buratta, V. and Egidi, V. (2003). Data collection methods and comparability issues. In J. M. Robine et al. (Eds.), *Determining Health Expectancies* (pp. 187–202), John Wiley & Sons, New York.
- Cambois, E., Robine, J. M., and Ormiche, P. (2007). Did the prevalence of disability in France really fall in the 1990s? A discussion of questions asked in the French health survey, *Population*, **62**, 315–337.
- Crossley, T. F. and Kennedy, S. (2002). The reliability of self-assessed health status, *Journal of Health Economics*, **21**, 643–658.
- Davis, B. A., Heathcote, C. R., and O'Neill, T. J. (2001). Estimating cohort health expectancies from cross-sectional surveys of disability, *Statistics in Medicine*, **20**, 1097–1111.
- Groves, R. M. (1989). *Survey Errors and Survey Costs*, John Wiley & Sons, New York.
- Groves, R. M., Fowler, F. J., Couper, J. M., Lepkowski, E., Singer, R., and Tourangeau, R. (2004). *Survey Methodology*, Wiley-Interscience, New York.
- Hur, N. W., Choi, C. B., Uhm, W. S., and Bae, S. C. (2008). The prevalence and trend of arthritis in Korea: results from Korea national health and nutrition examination surveys, *Journal of the Korean Rheumatism Association*, **15**, 11–26.
- Jagger, C. and Robine, J. M. (2011). Healthy life expectancy. In R. G. Rogers & E. M. Crimmins (Eds.), *International Handbook of Adult Mortality* (pp. 551–568), Springer, Dordrecht.
- Jung, Y. (2007). The criteria and models of prioritization on health related research programs, *Health and Welfare Policy Forum*, **128**, 105–113.
- Jung, Y. (2012). The life expectancy and health-adjusted life expectancy of Koreans, *Health and Welfare Policy Forum*, **193**, 5–18.
- Kalton, G., Kasprzyk, D., and McMillen, D. B. (1989). Nonsampling Errors in Panel Survey. In: Panel Surveys. In D. Kasprzyk et al. (Eds.), *Panel Surveys* (pp. 249–270), Wiley-Interscience, Wiley, New York.
- Kim, H. and Kim, E. (2013). Prevalence of osteoarthritis and its affecting factors among a Korean population aged 50 and over, *Journal of Korean Public Health Nursing*, **27**, 27–39.
- Kim, S. A., Park, W. S., Ohrr, H. C., Kang, H. Y., Lee, D. H., Yi, S. W., Kwak, Y. H., and Song, J. S. (2005). Prevalence and management status of diabetes mellitus in Korea, *Korean Journal of Medicine*, **68**, 10–17.
- Korean Society of Hypertension(2013). Hypertension treatment guidelines, Available from: www.koreanhypertension.org/download/131107_01.pdf
- Laslett, L. J., Alagona, P., Clark, B. A., Drozda, J. P., Saldivar, F., Wilson, S. R., Poe, C., and Hart, M. (2012). The worldwide environment of cardiovascular disease: prevalence, diagnosis, therapy, and policy issues, *Journal of the American College of Cardiology*, **60**, S1–S49.
- Lee, J., Kim, K., and Lee, J. (2012). *Construction of an appropriate sampling design and a sample database using the National Health Information Database*. National Health Insurance Corporation, Available from: <https://nhiss.nhis.or.kr/>
- Lee, J., Lee, J. S., Park, S. H., Shin, S. A., and Kim, K. (2016). Cohort Profile: the National Health Insurance Service-National Sample Cohort (NHIS-NSC), South Korea, *International Journal of Epidemiology*, Available from: : <http://doi.org/10.1093/ije/dyv319>
- Lee, S. (2010). Understanding sample survey with the total survey error paradigm, *Journal of the Korean Official Statistics*, **15**, 44–74.
- Lee, T. W., Ko, I. S., and Lee, K. J. (2006). Health promotion behaviors and quality of life among community-dwelling elderly in Korea: a cross-sectional survey, *International Journal of Nursing Studies*, **43**, 293–300.

- Mathers, C. D., Sadana, R., Salomon, J. A., Murray, C. J., and Lopez, A. D. (2001). Healthy life expectancy in 191 countries, 1999, *The Lancet*, **357**, 1685–1691.
- Centers for Disease Control & Prevention (2014). Korea National Health and Nutrition Examination Survey raw data user guide, Available from: <http://www.cdc.go.kr>
- National Health Insurance Service (NHIS) and Korea Institute for Health and Social Affairs (KIHASA). (2014). User guide for Korea Health Panel data, Available from: <https://www.khp.re.kr:444/>
- Newman, S. C. (1988). A Markov process interpretation of Sullivan's index of morbidity and mortality, *Statistics in Medicine*, **7**, 787–794.
- Park, I. and Lee, D. (2010). Study on Health Insurance Financial Outlook, *National Health Insurance Corporation*, Available from: <http://www.nhis.or.kr/menu/retrieveMenuSet.xx?menuId=G9200>
- Park, Y. S., Choi, J. W., and Kim, K. W. (2007). A balanced multi-level rotation sampling design and its efficient composite estimators, *Journal of Statistical Planning and Inference*, **137**, 594–610.
- Park, Y., Kim, K., and Kim, S. (2010). Unreliability of official population in Korea, *Journal of the Korean Association for Survey Research*, **11**, 71–95.
- Park, Y. S. and Kim, S. Y. (2011). A method for construction of life table in Korea, *The Korean Journal of Applied Statistics*, **24**, 769–789.
- Park, Y., Park, H., and Kwon, T. Y. (2015). Financial projection for national health insurance using NHIS sample cohort data base, *The Korean Journal of Applied Statistics*, **28**, 663–683.
- Rogers, A., Rogers, R. G., and Branch, L. G. (1989). A multistate analysis of active life expectancy, *Public Health Reports*, **104**, 222–226.
- Salomon, J. A., Wang, H., Freeman, M. K., Vos, T., Flaxman, A. D., Lopez, A. D., and Murray, C. J. (2013). Healthy life expectancy for 187 countries, 1990–2010: a systematic analysis for the Global Burden Disease Study 2010, *The Lancet*, **380**, 2144–2162.
- Särndal, C., Swensson, B., and Wretman, J. (2003). *Model Assisted Survey Sampling*, Springer, New York.
- Shenker, N., Raghunathan, T. E., and Bondarenko, I. (2010). Improving on analyses of self-reported data in a large-scale health survey by using information from an examination-based survey, *Statistics in Medicine*, **29**, 533–545.
- Sullivan, D. F. (1971). A single index of mortality and morbidity, *HSMHA Health Reports*, **86**, 347–354.
- US Department of Health and Human Services (2014). National Health and Nutrition Examination Survey: sample design, 2011–2014, Available from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_162.pdf

자가 응답식 자료에 근거한 유병률 및 건강기대수명 연구의 신뢰도 분석: 건강보험 표본코호트 DB와의 비교

권태연^a · 박유성^{b,1}

^a고려대학교 경제연구소, ^b고려대학교 통계학과

(2016년 8월 24일 접수, 2016년 10월 20일 수정, 2016년 10월 20일 채택)

요약

한국 의료패널 데이터와 국민건강 영양조사는 추출된 표본의 자가 진단에 따른 건강상태(self-assessed health)와 그들의 의료기관 이용에 대한 자가응답식 자료(self-reported data)이다. 이러한 자료에 근거한 유병률 연구 및 그에 따른 건강기대수명 연구에 관하여 유병률의 신뢰도에 대한 검증이 선행되어야 한다는 주장은 이미 여러 연구에서 제기되었다. 반면 최근 공개된 건강보험공단의 표본코호트 DB는 전 국민을 대상으로 의료기관 이용에 관련된 모든 자료가 저장된 자료인 국민건강정보 DB로부터 추출된 객관적인 자료이다. 또한 추정된 질병별 유병률에 대하여 그 대표성 및 신뢰도가 확보되어 있음이 검증된 자료이다. 이에 본 논문에서는 우리나라 국민의 유병률에 대한 대표성 및 신뢰도가 확보되어 있는 표본코호트 DB와의 비교를 통하여 이들 응답식 데이터에서 도출된 유병률의 신뢰도에 대하여 논의하였다. 자가응답식 자료를 통한 유병률은 표본코호트 DB를 통한 유병률에 비교하여 보았을 때 과소추정되어 있고 이러한 과소추정은 건강기대수명의 과대추정 문제로 이어지고 있음을 확인할 수 있었다. 또한 표본코호트 DB를 제외한 우리나라 건강자료의 안정적이지 못한 표본의 문제는 추정된 건강기대수명의 트렌드를 왜곡하는 문제가 추가적으로 발생할 수 있음을 확인하였다.

주요용어: 표본코호트 DB, 국민건강영양조사, 한국의료패널조사, 유병률, 건강기대수명

이 논문 또는 저서는 2013년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2013S1 A5B8A01054750)

¹교신저자: (02841) 서울시 성북구 안암로 145, 고려대학교 통계학과. E-mail: yspark@korea.ac.kr