

건강상태에 따른 사망률 및 유병기간 분석과 건강여명을 활용한 건강나이 산출에 관한 연구

전희주*, 문기태**, 인태교***

국문초록

본 연구는 국민건강보험공단의 표본코호트DB를 이용하여 건강검진 정보에 기초한 개인의 건강상태에 따른 사망률 및 유병기간과 예측된 건강여명을 활용하여 건강나이를 산출하는 방법을 제안하고자 한다. 개인별 건강상태(BMI, 혈압, 혈당 등)에 따른 사망과 유병기간에 대하여 각각 로지스틱 회귀모형과 음이항 회귀모형의 통계적 예측 모형을 개발하였다. 또한, WHO, EUROSTAT과 일본 후생노동성 등 해외 통계산출기관 뿐만 아니라 통계청에서 발표하는 건강여명의 산출방식인 Sullivan 방법으로 개인별 건강여명 예측모형을 개발하고, 전체 집단 평균 건강여명과 비교를 통해 최종적으로 건강나이를 산출하였다. 2017~2019년의 최근 2년 이내 건강검진을 수검한 이력이 있는 40대를 대상으로 한 시뮬레이션 결과, 건강나이가 개선된 경우 보험사고율이 감소하는 등 건강나이 개선 여부에 따른 보험 사고율 증감여부가 유의미함을 확인하였다. 이로써 최종적으로 산출된 건강나이가 개인의 종합적인 건강상태를 나타내는 객관적인 지표로서 활용될 수 있는 가능성과 이를 활용한 보험상품 개발 가능성을 제시하였다.

□ 주제어: 건강나이, 건강여명, 표본코호트DB, 일반화선형모형, Coale and Kisker 모형

□ JEL: J32

투고일: 2023.10.25.

1차 수정일: 2023.12.19., 2차 수정일: 2024. 1.25.

게재확정일: 2024. 4.17.

* 제1저자, 동덕여자대학교 정보통계학과 부교수(hjchun@dongduk.ac.kr)

** 공저자, 연세대학교 의과대학 예방의학교실 외래교수(tubemed@hanmail.net)

*** 교신저자, 한양대학교 금융보험학과 박사(tkleen@naver.com)

I. 서론

통계청에 따르면 우리나라 국민의 기대수명(life expectancy)은 1970년 남성 58.7세, 여성 65.8세, 남녀 평균 62.3세에서 2022년 남성 79.9세, 여성 85.6세, 남녀 평균 82.7세로 지난 52년간 의료기술의 발전과 건강증진의 효과로 평균수명이 평균 20.4세 증가하였다. 그러나 수명의 연장은 국민 개개인에게는 긍정적인 면도 있으나 연장된 기간에 대한 유지비용 증가와 삶의 질을 고려하지 않은 건강하지 않은 질병기간의 연장은 개인의 노후 치료비용과 사회적 비용의 증가를 가져오는 또 다른 사회문제를 가져오고 있다. 그러므로 단순히 기대수명만 늘어나기 보다는 아프지 않고 질병에서 자유로운 건강수명의 연장이 무엇보다 중요하다 할 것이다.

기대여명은 어떤 특정 연령의 개인이 앞으로 생존할 것으로 기대되는 평균 생존년수를 말하고, 건강나이 또는 건강수명은 몸이나 정신에 특별한 이상 없이 튼튼한 상태로 활동을 하며 살 수 있는 기간으로 평균수명에서 질병 또는 장애를 가진 기간을 제외한 기간을 말한다.

본 논문에서 건강나이(health age)는 단순히 개인이 출생부터 현재까지 살아 온 기간을 계산하는 달력나이(chronological age)와 달리 개인의 종합적인 건강상태를 고려하여 계산하는 나이로 정의한다. 건강나이는 개인의 과거와 현재의 건강상태를 고려하여 산출함으로써 미래의 건강상태까지 예측할 수 있는 지표로써 활용될 수 있다. 다만, 2가지 객관적인 정보(출생 일자와 현재 일자)만 주어진다면 모두가 같은 나이를 계산하는 달력나이와 달리, 건강나이는 어떤 건강 정보를 고려하느냐에 따라 다양하게 산출될 수 있다. 건강나이에 대한 학술적으로 통일되고 일관된 개념 정의가 확립되지는 않았지만, 생체나이(biological age) 등 달력나이와 구분되는 나이를 정의하고자 하는 시도가 있었다. Diebel and Rockwood(2021)는 생체나이를 ‘생물 생리학적 지표를 활용하여 개인의 연령 관련 위험을 보다 정확하게 결정하는 것’이라 하였으며, Jackson *et al.*(2003)은 생체나이를 ‘개인의 동일 연령 평균수명과 주관적 수명(perceived life expectancy)의 차이(shortfall)를 사용하기 위해 느슨하게 사용되는 개념’이라 하였다.

건강나이의 잠재적인 활용가치가 높음에도 불구하고, 위와 같은 건강나이의 특성으로 인해 신체나이, 생체나이 등 유사한 용어를 명확히 구분하기 어려운 상태로 혼용되고 있다. 또한, 세계보건기구(WHO)에서 1980년대 건강여명 개념을 소개한지 약 40년이 흘렀음에도 불구하고 최근에는 관심이 증가하고 관련 산업의 활용이 증가하고 있다.

미국은 「Healthy People 2020」과 같이 매 10년 마다 건강수명 향상을 위한 국가차원의

전략을 추진하고 있다. 영국은 2000년 「The Health of the Nation」, 2010년 「Our Healthier Nation」을 추진하였고, 일본은 「국민건강가꾸기 운동」, 「활력있는 인생 80 건강 계획」, 「건강일본 21」과 같이 국민들의 건강을 향상시키기 위한 정책을 시행하고 있다.

우리나라 또한 국가차원의 건강수명 향상을 위해 2002년 「제1차 국민건강증진종합계획」을 세워 질병의 사전 예방과 건강증진서비스 제공을 위한 중장기 정책 방향을 제시하였으며, 2005년 「제2차 국민건강증진종합계획」에서는 건강수명 연장을 목표로 하는 정책 방향을 제시하였다. 또한, 10년 주기의 국민건강증진종합계획, 5년 주기의 만성질환 종합계획 등을 수립하여 국민건강증진을 도모하고 있다. 그리고 통계청은 2012년부터 건강수준별 기대여명(유병기간 제외 기대여명, 주관적 건강평가 기대여명)을 매 2년마다 발표하고 있다. 국민건강보험공단은 2007년부터 「건강나이가 알아보기」서비스를 개시하였으며, 몇 차례 고도화 사업을 거쳐 현재까지 운영 중이다. 이처럼 건강나이 또는 건강여명에 대한 당국의 관심은 지속적으로 증가하고 있다.

현재 국내에서 건강나이를 가장 적극적으로 활용하고 있는 산업은 보험업이라고 할 수 있다. 보험회사는 상품개발, 인수심사, 리스크관리, 보험료 산출 등 핵심 업무에 건강나이를 적용할 수 있다. 구체적으로 일부 보험회사는 건강증진형 보험상품 개발을 통해 건강나이 또는 건강등급 개선 시 보험편익(보험료 할인, 보험금 증액 등)을 제공하고 있으며, 전통적으로 보험나이 기준으로 보험료를 산출하는 방식에서 벗어나 건강나이를 기준으로 보험료를 산출하는 방식이 꾸준히 시도되고 있다.

산업에 건강나이를 적용하고 활용하기 위해서는 객관적인 건강정보와 국제적으로 통용되는 방법론을 활용한 건강나이 산출 방법론의 필요성이 대두되고 있으며, 그에 대한 객관적인 방법론을 제시하는데 본 연구의 목적이 있다. 본 연구는 국제 통계기관 및 국내 통계청에서 공표하는 건강여명(healthy life expectancy) 산출 방법론과 체질량지수(BMI), 혈압, 혈당, 콜레스테롤 수치 등 개인의 객관적인 건강정보를 통해 산출하는 시점의 건강지표에 기반한 건강나이 산출 방법론을 제시하고자 한다.

II. 선행연구와 연구 방법

1. 선행연구

국내에서 기대여명에 대한 연구는 초기에 생명표(life-table) 중심으로 진척되어 왔다. 생명표 관련 대표적 초기 연구로는 권태환·김태현(1990)으로 1971~1985년 기간의 사망력 유형 분석에 기초하여 1970~1991년에 걸친 22년간의 생명표를 작성하였다. 현재까지 우리나라의 생명표 작성은 대체적으로 통계청에서 발표하는 특정시점의 주민등록인구와 사망신고 관련 인구동태통계 자료를 기초로 작성되고 있다.

전통적인 횡단면 자료를 이용하는 Sullivan 방법은 어떤 특정 시점에서 연령별(연령구간별) 건강/불건강 상태에 있는 개인들의 비율과 일반 기간생명표에서 도출된 연령별 사망률 정보에 기초하여 건강여명을 구한다(Robine *et al.*, 2006). Sullivan 방법에 기초로 한 건강여명에 관한 국내 연구들로는 윤병준(1995), 남정자 외(1996), 윤병준·김정근(1996), 강은정·김나연(2007), 강은정 외(2008), 한소현·이성국(2012)이 있다. 우해봉(2009)은 고령화연구패널조사(KLoSA) 자료를 이용하여 우리나라 중고령층의 성별 및 교육수준별 인지장애 기대여명을 구하였고, 우해봉(2012)은 고령화연구패널조사 자료를 이용한 기대여명추정에서 건강측정치의 효과를 구하였다. 우해봉(2012)은 건강여명의 산출에 활용되는 건강측정치를 일상생활 수행능력, 수단적 일상생활 수행능력, 신체적 기능 수행능력, 자기평정적 건강 등의 설문에 의한 주관적 건강 측정 항목을 이용하여 생명표를 작성하는 방식을 취하였다.

그러나 설문 패널조사에 의한 건강여명 계산의 문제점은 사회조사 표본으로 대상이 선택되어 건강 관련 정보를 설문을 통해 수집하는 과정에서 모집단의 특성을 정확히 반영하지 못하는 편의(bias) 현상이 나타날 수 있다. 또한 충분한 표본(특히, 사망건수)을 가지고 있지 못할 경우 기대여명 관련 추정치에 수반된 불확실성은 더욱 커질 수 있다는 점이다. 이러한 점에서 패널조사 자료에 기초하여 건강여명을 산출하는 것은 커다란 불확실성을 갖는 한계를 가지고 있다.

건강여명을 산출하는 방법론적 관점에서의 차이 또는 건강여명을 산출하는데 사용되는 건강 측정치들에 있어서도 기존 연구들은 매우 다양한 접근 방법을 취하고 있다. 기존 연구들은 건강 상태를 다양한 측정치들을 통해 평가하고 있으며, 대표적인 측정치들로는 자기평정적 건강(self-rated health), 장애(disability), 신체적 기능 제한(functional limitation), 만성질환

등이다(Crimmins and Cambois, 2003; Deeg *et al.*, 2003). Ritchie and Polge(2003)은 건강상태 측정치로 인지적 능력이나 우울증과 같은 정신건강을 고려하였고, Yang(2008)은 건강상태뿐만 아니라 행복과 같은 추상적인 영역으로 확대하기도 하였다.

국내의 건강여명에 관한 연구들 또한 다양한 측정치를 사용하여 건강상태를 측정하고 있다. 건강여명 산출을 위한 건강 측정치로 남정자 외(1996)와 윤병준·김정근(1996)은 활동제한일수와 이환기간을 사용하였고, 우해봉(2009)은 자기평정적 건강을 사용하였으며, 강은정·김나연(2007)과 강은정 외(2008)는 건강 관련 삶의 질 척도를 사용하였다. 한소현·이성국(2012)은 특정한 건강지표를 사용하는 대신 신체적 및 인지적 활동과 관련된 다양한 측정 문항들 중 어느 한 항목에서 제약이 있는 경우를 활동장애로 분류하여 건강여명을 산출하였다. 전희주·최경진(2020)은 건강나일로 생체연령을 구하는 방법을 제시하고 이를 이용하여 생존확률을 구하였으며, 전희주·최경진(2021)은 생체연령을 고려한 퇴직연금 프로그램 인출방식을 제시하였다.

2. 연구 데이터 및 연구 대상

본 연구는 국민건강보험공단에서 제공하는 표본코호트DB를 사용하였다. 표본코호트DB는 2006년 1년간 건강보험 가입자 및 의료급여 수급권자 자격을 유지한 전 국민을 기준으로 성, 연령, 가입자, 보험료, 지역을 층화하여 무작위 추출된 100만명(모집단의 2%) 표본의 2002년부터 2019년까지의 출생 및 사망, 자격 및 보험료, 의료이용 현황 정보를 포함하고 있는 자료이다. 해당 DB에서 자격 및 보험료 테이블, 출생 및 사망 테이블, 진료 테이블, 건강검진 테이블의 자료를 사용하였다.

본 연구의 분석대상자는 최신 자료를 활용하기 위해 2011~2019년 기간 중에 건강보험 자격을 유지하고 있고 당해 연도를 제외하여 최근 2년간 건강검진 이력이 있는 자를 대상으로 하였다. 건강검진 항목 중 미수검 항목이 있는 경우에는 최근 2년간 건강검진 결과 중 최신 결과값을 해당 미수검 항목에 기입하는 방식으로 결측치를 처리하여 분석을 위한 충분한 데이터를 확보하였다. 연령은 기준연도에서 출생연도를 차감하여 산출하고 대상자의 연령은 25세 이상으로 설정하였다.

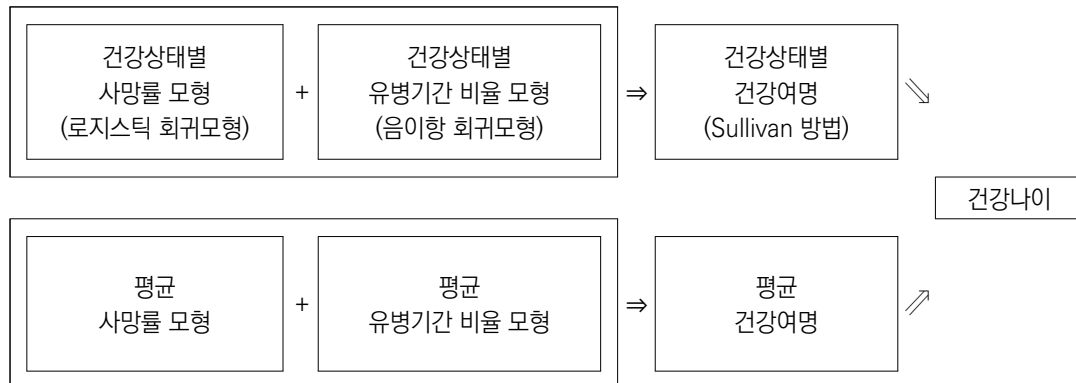
3. 연구 방법

(1) 연구 설계

본 연구에서는 개인의 건강상태별 건강여명과 평균 건강여명의 비교를 통해 최종적으로 건강나이를 산출한다. 건강상태별 건강여명 모형을 통해 산출한 개인의 건강상태별 건강여명과 평균 건강여명을 비교하여 개인의 건강상태별 건강여명이 평균 건강여명에 가장 가까운 값에 대응하는 실제연령을 개인의 건강나이라고 정의한다.

개인의 건강상태별 건강여명을 산출하기 위해서는 건강상태별 사망률 모형과 건강상태별 유병기간 비율 모형이 필요하다. 로지스틱 회귀모형을 이용하여 사망률을 예측하는 모형과 음이항 회귀모형을 이용하여 유병기간을 예측하는 모형을 구한다. [그림 1]은 사망률 예측모형과 유병기간 비율 예측모형을 이용하여 최종적으로 건강나이를 구하는 도식이다.

[그림 1] 건강나이 산출 모형



건강여명을 산출하는 방법으로는 Sullivan 방법, Multidecrement 방법, Multistate 방법 등 다양한 방법이 있다. 본 연구에서는 통계청뿐만 아니라, 일본, 캐나다, WHO, EUROSTAT 등 선진국과 국제기관에서 산출하는 방법인 Sullivan 방법을 활용하였다. 통계청의 건강여명 산출방법은 [표 1]에 제시되어 있다.

[표 1] 통계청 건강여명 산출방법

종류	유병기간 제외 기대여명	주관적 건강평가 기대여명
의미	실제 질환이나 장애 등으로 고통받는 기간을 제외한 기대수명(여명)	전체 여명(수명) 중 주관적으로 건강하다고 평가하는 기간
기초 자료	2020년 사회조사(유병기간)	2020년 사회조사(건강평가)
조사문항	귀하는 지난 2주일 동안 질병이나 사고로 아팠던 적이 있습니까? ① 있다 ② 없다 ※ '① 있다'고 응답한 경우, - 아팠던 일수 ()일 - 입원하거나 누워서 지냈던 일수 ()일	귀하의 전반적인 건강상태는 어떻습니까? ① 매우 좋다 ② 좋은 편이다 ③ 보통이다 ④ 나쁜 편이다 ⑤ 매우 나쁘다
작성 공식	$E'_x = \frac{1}{l_x} \sum_{i=x}^w (L_i \times (1 - \pi_i))$ - x : 연령 - i : x세 연령구간 - w : 총 연령구간수 - L_i : i연령대 정지인구 - π_i : i연령대 장애유병률 - l_x : x세 생존자수	
장애유병률 기준	응답한 사람들의 2주간 아팠던 일수 비율의 평균	건강상태를 '나쁜 편이다' 및 '매우 나쁘다'라고 응답한 사람 비율

Sullivan 방법은 Sullivan(1971)에서 건강수준을 나타내는 지표로써 사망률(mortality)과 장애율(disability rates)을 처음으로 모두 사용하는 방법으로 제시되었다. 사망률과 장애율을 개별적인 데이터를 통해 계산할 수 있다는 점과 전통적인 생명표 작성 방식으로 계산이 용이하다는 점 때문에 현재까지 통계청뿐만 아니라 국제기관에서 사용 중이다.

Sullivan 방법은 생명표 작성 방식에 유병기간 비율을 추가로 고려하여 건강여명을 산출하는 방식이다. 건강여명 산출을 위해서는 사망률과 유병기간 비율이 가장 중요한 요소이다. 본 연구에서는 유병에 대한 주관적인 평가에서 벗어나고자 통계청 기준에 의거하여 설문이 아닌 객관적인 데이터를 통해 확인할 수 있는 입원 및 내원기간을 유병기간으로 정의하였다. 참고로 유럽과 일본의 통계기관은 산출목적에 따라 유병기간 비율을 [표 2]와 같이 정의하고 있다.

[표 2] 산출기관별 건강여명 및 유병기간 비율 정의

산출기관	명칭	유병기간 비율 정의
EUROSTAT	DFLE (disability-free life expectancy)	'과거 6개월 이전의 기간 중 건강문제로 일상생활에 영향이 있다'로 응답한 비율
일본 후생노동성	DFLE without activity limitation	'현재 건강문제로 일상생활에 영향이 있다'로 응답한 비율
	DFLE without care need	국가 간병보험 통계의 요보호 2~5 상태의 비율
	life expectancy with self-perceived health	건강상태를 '별로 좋지 않다' 및 '좋지 않다'로 응답한 비율

(2) 분석대상자 정의

본 연구는 데이터 입력오류로 인한 영향을 최소화하고 모델의 안정성을 위하여 신상훈 외(2016)와 임현선 외(2020)의 기준에 의거하여 데이터 유효성 기준을 [표 3]과 같이 설정하였다.

[표 3] 데이터 유효성 기준

설명변수명	영문명	유효성 기준	단위
키	HGHT	100 < ≤ 220	cm
몸무게	WGHT	30 < ≤ 200	kg
체질량지수	BMI	10 < ≤ 40	kg/m ²
수축기혈압	BP_SYS	60 < ≤ 240	mmHg
이완기혈압	BP_DIA	40 < ≤ 160	mmHg
식전혈당(공복혈당)	FBS	40 < ≤ 300	mg/dL
총콜레스테롤	TOT_CHOL	40 < ≤ 400	mg/dL
HDL 콜레스테롤	HDL	10 < ≤ 200	mg/dL
LDL 콜레스테롤	LDL	200 < ≤ 370	mg/dL
혈청지오티AST	SGOT	≤ 200	U/L
혈청지피티ALT	SGPT	≤ 200	U/L
혈청크레아티닌	CRTN	≤ 10	mg/dL

사망률 모형은 고연령 구간의 연령 증가에 따른 사망률 증가가 확실하고, 사망률이 1이 되는 종국연령을 설정하는 Coale and Kisker(1990) 모형을 활용한다. 해당 모형에서는 특정 연령

이상 구간을 합산하여 산출하지 않아야 하므로 대상자의 연령 기준을 [표 4]와 같이 25세~84세로 설정하였으며, 해당 연도 사망자를 포함하였다.

한편, 유병기간 비율 모형은 고연령 구간의 연령 증가에 따른 유병기간의 단조 증가가 확실하지 않으므로 사망률 모형과 달리 85세 이상을 별도의 그룹으로 구분하고 대상자의 연령 기준을 [표 4]와 같이 25세 이상으로 설정하였다. 또한, 해당 연도 사망자를 포함하는 경우, 관찰기간이 1년 미만인 대상자가 포함되어 기간 불일치가 발생하므로 관찰기간을 1년으로 통제하기 위하여 사망자는 제외하였다.

[표 4] 모형별 연령 기준

모형	연령
건강상태별 사망률 모형	25세~84세
건강상태별 유병기간 비율 모형	25세 이상

본 연구에서 사용된 표본코호트DB의 관찰기간은 2011년부터 2019년으로 각 연도별 대상자 합계는 [표 5]와 같이 남성 4,767,346명, 여성 4,763,064명이다. 이중 최근 2년 이내 건강검진 수검이력이 있는 대상자를 한정하는 경우 남성 1,876,730명, 여성 1,813,238명이다.

[표 5] 분석대상자 선정과정

구분	남성	여성
2011~2019 연도별 대상자 합계	4,767,346명	4,763,064명
최근 2년 이내 건강검진 수검자	1,876,730명	1,813,238명

(3) 건강상태별 건강여명 모형

① 모형 설계

2011년~2019년의 매년 사망여부 또는 유병기간을 종속변수로 설정하고 그 직전 연도의 건강검진 결과 자료를 설명변수로 설정하였다. 다만, 모형의 통계적 유의성 확보에 필요한 건강검진 결과 데이터 확보를 위해 데이터 집계 기간을 확대하여 최근 2년 이내 수검한 건강검진 결과 중 각 항목별 최신 데이터를 최종 설명변수로 설정하였다.

사망여부는 기준년도(STD_YYYY)와 사망연도(DTH_YYYY)가 일치하는 경우 해당연도에 사망한 것으로 정의하였다. 유병기간은 입원 또는 내원하여 의료기관을 이용하는 기간으로 정의하였으며, 의료기관에는 병·의원으로 한정하여 치과 또는 한의원은 제외하였다.

사망률과 유병기간 비율에 영향을 미칠 수 있을 것으로 예상되는 건강검진 항목들을 [표 6]과 같이 우선 선정하였다. [표 6]의 각 항목들은 의학적으로 개별 항목이 개인의 건강상태에 영향을 미치는 수준과 기존 연구 결과(신상훈 외, 2016; 임현선 외, 2020)를 참조하여 국민건강보험공단에서 관리하는 전체 항목 중 대표 항목을 선정하였다.

체질량지수, 수축기혈압, 이완기혈압의 경우 선형으로 증가하거나 감소하는 형태로 건강상태를 나타내는 것이 아닌, 표준 대비 낮은 구간과 높은 구간으로 벗어나는 경우 건강상태에 영향을 미친다. 이 점을 보완하기 위하여 해당 세 개 변수는 각 변수의 제곱⁽²⁾ 변수를 추가하였다.

[표 6] 건강상태에 따른 건강지표 변수

설명변수명	영문명
연령	AGE
체질량지수	BMI
(체질량지수) ²	BMI ²
수축기혈압	BP_SYS
(수축기혈압) ²	BP_SYS ²
이완기혈압	BP_DIA
(이완기혈압) ²	BP_DIA ²
공복혈당	FBS
총콜레스테롤	TOT_CHOL
HDL 콜레스테롤	HDL
LDL 콜레스테롤	LDL
혈청지오티AST	SGOT
혈청지피티ALT	SGPT
혈청크레아티닌	CRTN

② 통계적 모형

본 연구의 사망률 예측모형의 반응변수는 사망여부의 이항자료로써 로지스틱 회귀모형이 적합하였다. 로지스틱 회귀모형은 일반화 선형모형(generalized linear model) 중의 하나로

반응변수가 0과 1인 이항값을 갖는 이항분포를 갖고(전희주·최경진, 2022), 반응변수 평균의 함수 $p_i = E(Y|x_i)$ 와 선형예측식 $x_i'\beta$ 를 연결하는 연결함수는 로짓(logit)을 사용하며 식 (1)과 같이 나타낼 수 있다.

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = x_i'\beta \dots\dots\dots (1)$$

여기서 x_i 는 설명변수 벡터($x_i = (1, x_{i1}, \dots, x_{ik})$, $i = 1, \dots, n$)이고 β 는 $(k+1) \times 1$ 개의 모수를 갖는 회귀계수 벡터이다.

유병기간 예측모형은 반응변수가 2011년~2019년도에 발생한 건수에 대한 빈도자료(count data)로 일반화 선형모형의 포아송 회귀모형 또는 음이항 회귀모형을 적용할 수 있다. 포아송 회귀모형은 반응변수가 양수인 정수 값을 취하고 평균과 분산이 비슷할 때 사용되는 가장 기본적인 모형인 반면에, 음이항 회귀모형은 평균에 비해 분산이 매우 큰 과대산포(over-dispersion)가 발생하는 경우에 적합되는 모형이다(Gurmu, 1991; 전희주·최경진, 2022). 선형예측식 $x_i'\beta$ 은 로그 연결함수(link function)를 통해 평균 모수 $\mu = E(y|x)$ 와 식 (2)와 같이 연결이 된다(전희주·최경진, 2023).

$$\log(\mu) = x_i'\beta \dots\dots\dots (2)$$

여기서 x_i 는 설명변수 벡터($x_i = (1, x_{i1}, \dots, x_{ik})$, $i = 1, \dots, n$)이고 β 는 $(k+1) \times 1$ 개의 모수를 갖는 회귀계수 벡터이다. 음이항분포의 평균과 분산은 식 (3)과 같이 평균과 과대산포 모수 D 의 함수로 표현이 가능하며, $D=0$ 인 경우 포아송분포가 된다(전희주·안철경, 2012).

$$E(y|x) = \mu, \quad \text{Var}(y) = \mu + D^2 \dots\dots\dots (3)$$

모형선택은 AIC(Akaike information criteria)와 후진제거법(backward elimination method)에 선택되었으며 AIC는 식 (4)와 같이 정의된다. 후진제거법은 모든 변수들을 가장 작은 영향을 주는 변수를 하나씩 제거해 가면서 더 이상 제거할 수 없을 때의 모형이 최적모형

이 되며, 최적모형을 찾는데 모든 변수들의 정보를 이용할 수 있다는 것이 장점이다.

$$AIC = -2\log\text{Lik} + 2 \times k \dots\dots\dots (4)$$

여기서 k 는 추정될 모수의 수이다.

③ 건강상태별 사망률 모형

Sullivan 방법 같이 생명표를 활용한 건강여명 산출방식은 특정 연령별 사망률(age-specific mortality)이 필요하다. 특정 연령별 산출을 위해 우선 로지스틱 회귀모형을 통해 현재의 건강정보(BMI, 수축기혈압, 이완기혈압 등)가 장래에도 동일하다는 가정하에 연령(AGE) 변수만을 변화하여 군단연령별 사망률을 산출한다.

군단연령별로 산출하는 이유는 사망자와 생존자 표본을 합산하여 변동성을 축소하고 연령별 사망률 변화의 전반적인 경향을 충분히 나타내기 위함이다. 분석의 기초가 되는 표본코호트DB가 전국민의 약 2%에 해당하는 약 100만명을 대상으로 작성된 자료이고, 특히 저연령 구간 사망자 표본이 적기 때문에 1세 단위로 사망률을 산출하는 경우 연령 증가에 따라 사망률이 감소하는 등 일반적인 상식에 어긋나는 결과가 도출될 수 있기 때문이다. 군단연령은 기본적으로 5세 단위로 하되, 사망자 표본이 특히 적은 저연령 구간(25세 이상 44세 이하)은 10세 단위로 하였다.

따라서, 우선 산출한 군단연령별 사망률의 직선보간을 통해 특정 연령별 사망률을 산출한다. 직선보간은 각 연령구간의 중앙값을 대표연령으로 하여 보간하므로 직선보간을 통해 산출되지 않는 저연령 구간(29세 이하; 25세~34세 구간의 중앙값 29.5 보다 작은 연령 중 가장 큰 정수)과 고연령 구간(83세 이상; 80세~84세 구간의 중앙값 82세 보다 큰 연령 중 가장 작은 정수)의 사망률을 추정하는 과정이 필요하다.

저연령 구간의 경우, 건강상태를 반영하지 않은 [표 5]의 2011~2019 연도별 대상자 합계에 해당하는 평균 사망률의 연령감소에 따른 변화율과 건강상태별 사망률의 변화율과 동일하다는 가정 하에 추정하였으며, 그 식은 식 (5)와 같이 나타낼 수 있다.

$$q'_{x-1} = q'_x \left(\frac{q_{x-1}}{q_x} \right) \dots\dots\dots (5)$$

x : 연령
 q'_x : x 세의 건강상태별 사망률
 q_x : x 세의 평균 사망률

고연령 구간의 경우, 연령 증가에 따른 사망률의 지수 증가율이 선형적으로 감소한다는 가정으로 도출된 Coale and Kisker(1990) 모형을 활용하여 추정하였다. Coale and Kisker 모형은 85세 이상 고연령의 사망률을 추정함으로써 생명표를 마감(close the life tables)하는데 사용된다. 통계청 역시 변형된 Coale and Kisker 모형을 통해 85세 이상 사망률을 추정하고 있다. 다만, 본 연구에서 군단연령 사망률의 직선보간을 통해 산출된 특정 연령별 사망률의 상한이 82세이므로 83세 이상 사망률을 Coale and Kisker 모형을 활용하여 추정하였으며, 그 식은 식 (6)과 같이 나타낼 수 있다.

$$q'_x = q'_{x-1} \times \exp(k_{82} + (x-82) \times s), x = 82, 83, \dots, w-1 \dots\dots\dots (6)$$

$$s = \frac{\ln\left(\frac{q'_{81}}{q'_w}\right) + (w-x_0+1) \times k_{82}}{\frac{(w-x_0) \times (w-x_0+1)}{2}}$$

$$\text{단, } k_{82} = \ln\left(\frac{q'_{82}}{q'_{81}}\right), w = 110(\text{남자}) \text{ 또는 } 112(\text{여자}), q'_w = 1, x_0 = 82$$

사망률이 1이 되는 연령인 중국연령(w)은 보험개발원에서 공표하는 제9회 경험생명표를 참고하여 남성 110세, 여성 112세로 설정하였다.

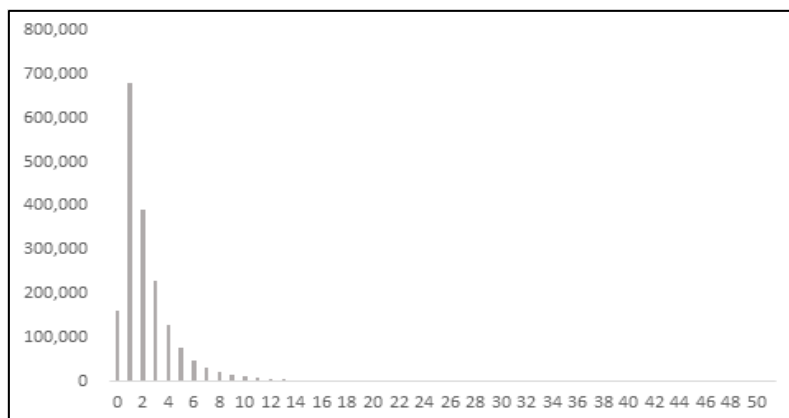
④ 건강상태별 유병기간 비율 모형

유병기간 모형은 반응변수가 빈도자료이고 평균에 비해 분산이 큰 과대산포를 나타내므로 음이항 회귀모형을 적용하였다. 주단위 유병기간의 평균과 분산을 살펴보면, 남자의 경우 각각 평균 2.5298, 분산 9.2960이었으며, 여자의 경우 각각 평균 3.4401, 분산 12.5026으로

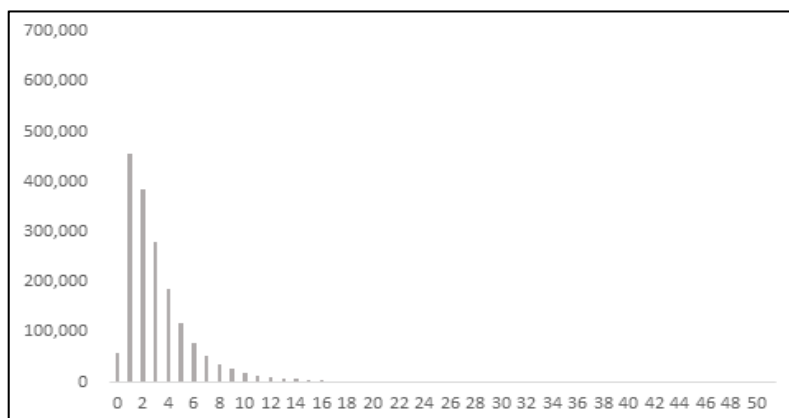
모두 분산이 평균보다 큰 과대산포를 나타냈다.

또한, [그림 2]와 [그림 3]과 같이 각각 남성과 여성의 유병기간별 분포도를 살펴보면, 0에 해당하는 표본이 많지 않아 영과잉음이향분포보다는 음이향분포가 더 적합함을 알 수 있다.

[그림 2] 남성 유병기간별 분포도



[그림 3] 여성 유병기간별 분포도



예측 모형의 안정성과 성능을 향상하고 평균적인 변화 예측을 위해 일단위로 표시된 유병기간을 주 단위로 환산하고 52주 이상 데이터는 이상치(outlier)로 판단하여 제외하였다. 따라서, 음이향 회귀모형의 종속변수를 0, 1, 2, ..., 50, 51로 통제하였다. 또한, 관찰 대상자는 사망

모형과 달리 사망자를 제외한 생존자로 한정하였다. 그 이유는 연중 사망자의 경우 생존자와 달리 관찰기간(exposure)이 각각 상이하므로 이를 1년으로 통일하기 위함이다.

건강상태별 사망률 모형과 마찬가지로 건강상태별 건강여명 산출을 위해 특정 연령별 유병기간 비율이 필요하다. 특정 연령별 유병기간 비율 산출을 위해 음이항 회귀모형을 통해 현재의 건강정보(BMI, 수축기혈압, 이완기혈압 등)가 장래에도 동일하다는 가정 하에 연령(AGE) 변수만을 변화하여 군단연령별 유병기간을 산출한다. 추정된 유병기간은 주 단위이므로 (추정값(주)*7-3)의 과정을 통해 일 단위 유병기간으로 환산하고 최종적으로 365를 나눠 연단위 유병기간 비율을 산출하였다. 여기서 일 단위로 환산하는 이유는 1주(7일) 단위로 산출함으로써 유병기간이 과대 산출될 수 있는 가능성을 관리하고자 3일을 차감한 평균일수(4일)를 적용하여 주 단위에서 일 단위로 환산하기 위함이다.

군단연령별 유병기간 비율의 직선보간을 통해 특정 연령별 유병기간 비율을 산출한다. 군단연령별로 산출하는 이유는 사망률 모형과 마찬가지로 표본을 합산하여 변동성을 축소하고 연령별 유병기간 비율 변화의 전반적인 경향을 충분히 나타내기 위함이다. 직선보간은 각 연령 구간의 중앙값을 대표연령으로 하여 보간하므로 직선보간을 통해 산출되지 않는 저연령 구간(26세 이하; 25세~29세 구간의 중앙값 27세 보다 작은 연령 중 가장 큰 정수)과 고연령 구간(86세 이상)의 유병기간 비율을 추정하는 과정이 필요하다.

저연령 구간의 경우, 건강상태별 사망률 모형과 마찬가지로 건강상태를 반영하지 않은 [표 5]의 2011~2019 연도별 대상자 합계에 해당하는 평균 유병기간 비율의 연령감소에 따른 변화율과 건강상태별 유병기간 비율의 변화율과 동일하다는 가정하에 유병기간 비율을 추정하였으며, 그 식은 식 (7)과 같이 나타낼 수 있다.

$$\pi'_{x-1} = \pi'_x \left(\frac{\pi_{x-1}}{\pi_x} \right) \dots\dots\dots (7)$$

x : 연령
 π'_x : x 세의 건강상태별 유병기간 비율
 π_x : x 세의 평균 유병기간 비율

다만, 고연령 구간의 경우 86세 이상 유병기간 비율은 85세 유병기간 비율과 같은 값으로 설정하였다. 이는 연령 증가에 따라 유병기간 비율이 증가한다는 뚜렷한 증거가 없음을 [표 7]과 같이 군단연령별 평균 유병기간 비율 산출을 통해 확인하였다. 남성의 경우 85~89세 구간

까지 유병기간 비율이 계속 증가하나, 90세 이상 구간에서는 오히려 감소한다. 여성의 경우 35~39세 구간에서 위험률이 소폭 감소하며 이후 80~84세 구간까지 유병기간 비율이 계속 증가하나, 85~89세 구간부터 감소한다.

또한, 사망률 모형과 달리 유병기간 비율 모형은 중국연령과 같은 개념이 없는 점에 기인한다.(즉, 유병기간 비율이 반드시 1이 되는 나이를 설정하여 그 연령까지 유병기간 비율이 단조 증가하는 가정을 설정하지 않는다.)

[표 기] 군단연령별 평균 유병기간 비율

연령	남성	여성
25~29	0.0153	0.0271
30~34	0.0175	0.0334
35~39	0.0211	0.0315
40~44	0.0249	0.0323
45~49	0.0288	0.0384
50~54	0.0356	0.0499
55~59	0.0436	0.0595
60~64	0.0545	0.0699
65~69	0.0701	0.0882
70~74	0.0860	0.1047
75~79	0.0995	0.1148
80~84	0.1088	0.1175
85~89	0.1121	0.1121
90+	0.1112	0.0984

Ⅲ. 실증분석

1. 관찰대상 집단의 일반적인 특성

본 연구는 국민건강보험공단에서 제공하는 표본코호트2.0DB를 사용하여 출생 및 사망 테이블의 개인 고유번호로 해당 코호트 전체 대상자의 출생연도 정보를 확인하였다. 출생연도 정보를 기준으로 이벤트 발생 당시의 연령(만 나이)을 산출하였고, 자격 및 보험료 테이블을 연결하여 대상자의 성(SEX)별 정보를 확인하였다.

[표 8]은 사망률 모형 대상자의 성별/연령별 분포이며, [표 9]는 유병기간 비율 모형 대상자의 성별/연령별 분포이다. II.3.(2) 분석대상자 정의에도 언급하였지만, 관찰기간을 통제하기 위해 유병기간 비율 모형은 연중 사망자를 제외하였기 때문에 사망률 모형 대상자 수와 차이가 있다.

[표 8] 사망률 모형 대상자의 성별/연령별 분포

구분	남성		여성	
연령	대상자수(명)	비중(%)	대상자수(명)	비중(%)
25~29	91,298	5.0	100,789	5.7
30~34	170,533	9.3	109,726	6.2
35~39	191,477	10.4	96,552	5.5
40~44	232,312	12.7	203,111	11.5
45~49	233,483	12.7	227,593	12.9
50~54	223,592	12.2	234,876	13.3
55~59	210,050	11.5	230,184	13.1
60~64	166,794	9.1	187,320	10.6
65~69	128,661	7.0	144,065	8.2
70~74	95,962	5.2	112,241	6.4
75~79	61,105	3.3	77,412	4.4
80~84	27,401	1.5	38,197	2.2
계	1,832,668	100.0	1,762,066	100.0

[표 9] 유병기간 비율 모형 대상자의 성별/연령별 분포

구분	남성		여성	
연령	대상자수(명)	비중(%)	대상자수(명)	비중(%)
25~29	91,257	5.0	100,772	5.7
30~34	170,443	9.3	109,694	6.2
35~39	191,365	10.5	96,523	5.5
40~44	232,000	12.7	202,986	11.5
45~49	232,994	12.7	227,413	12.8
50~54	222,837	12.2	234,582	13.3
55~59	209,005	11.4	229,825	13.0
60~64	165,591	9.0	186,869	10.6
65~69	127,456	7.0	143,511	8.1
70~74	94,420	5.2	111,471	6.3
75~79	59,365	3.2	76,313	4.3
80~84	26,090	1.4	36,995	2.1
85+	7,927	0.4	13,440	0.8
계	1,830,750	100.0	1,770,394	100.0

2. 건강상태별 건강여명 모형

(1) 사망 모형

[표 10]과 [표 11]은 각각 남성과 여성의 사망여부를 적합하기 위한 최적 로지스틱 회귀모형의 결과이다. 남성과 여성 각각의 사망모형에 대한 이탈도(deviance)에 대한 p-값=1.0으로 선택된 모형적합이 타당함을 보인다. 또한 Hosmer-Lemeshow 통계량 모두 유의수준 0.05 보다 크게 나타나 선택된 모형의 예측값이 잘 적합되었음을 보인다.

사망여부에 유의미한 영향을 미치는 변수를 중요도 순서대로 표기하였으며, 남성의 경우 연령(AGE), 혈청지오티AST(SGOT), 공복혈당(FBS), 크레아티닌(CRTN), 수축기혈압(BP_SYS), (수축기혈압)²(BP_SYS²), HDL콜레스테롤(HDL), (이완기혈압)²(BP_DIA²), 이완기혈압(BP_DIA) 순으로 나타났으며, 여성의 경우 연령(AGE), 크레아티닌(CRTN), 공복혈당(FBS), 혈청지오티AST(SGOT), HDL콜레스테롤(HDL), 수축기혈압(BP_SYS), (수축기혈압)²

(BP_SYS²) 순으로 나타났다.

[표 10]의 남성의 사망여부 예측 적합결과를 보면, 연령(AGE)이 증가할수록 사망률이 높아지는 경향을 보인다. 또한, 혈청지오티AST(SGOT)가 한 단위 증가할수록 사망의 오즈(odds)는 1.0145(=exp(0.0144))배 증가하고, 공복혈당(FBS)이 한 단위 증가할수록 사망의 오즈는 1.0049(=exp(0.0048))배 증가하는 것으로 나타났다.

[표 10] 남성 사망예측 로지스틱 회귀모형

요인	자유도	추정치	표준오차	p-value	LR χ^2	p-value
Intercept	1	2.1907	0.4465	<0.0001	-	-
AGE	25~34	-4.5661	0.0987	<0.0001	11,396.7	<0.0001
	35~44	-4.1547	0.0660	<0.0001		
	45~49	-3.5402	0.0647	<0.0001		
	50~54	-3.0786	0.0552	<0.0001		
	55~59	-2.5961	0.0481	<0.0001		
	60~64	-2.1362	0.0449	<0.0001		
	65~69	-1.7871	0.0437	<0.0001		
	70~74	-1.1493	0.0402	<0.0001		
	75~79	-0.5513	0.0391	<0.0001		
	80~84	0.0000	0.0000	-		
SGOT	1	0.0144	0.0006	<0.0001	510.5	<0.0001
FBS	1	0.0048	0.0004	<0.0001	152.2	<0.0001
CRTN	1	0.1785	0.0154	<0.0001	97.1	<0.0001
BP_SYS	1	-0.0619	0.0076	<0.0001	60.9	<0.0001
BP_SYS ²	1	0.0002	0.0000	<0.0001	56.2	<0.0001
HDL	1	-0.0062	0.0009	<0.0001	54.3	<0.0001
BP_DIA ²	1	0.0003	0.0000	<0.0001	21.4	<0.0001
BP_DIA	1	-0.0482	0.0105	<0.0001	19.8	<0.0001
Statistic	D.F.		Value		p-value	
Deviance	1,832,650		89,816		1.00	
Hosmer-Lemeshow χ^2	8		11.844		0.1583	
AIC			89,852			

[표 11]의 여성의 사망여부 예측 적합결과를 보면, 연령(AGE)이 증가할수록 사망률이 높아지는 경향을 보인다. 또한, 크레아티닌(CRTN)이 한 단위 증가할수록 사망의 오즈(odds)는

1.3240($=\exp(0.2806)$)배 증가하고, 공복혈당(FBS)이 한 단위 증가할수록 사망의 오즈는 1.0006($=\exp(0.0062)$)배 증가하는 것으로 나타났다.

[표 11] 여성 사망예측 로지스틱 회귀모형

요인	자유도	추정치	표준오차	p-value	LR χ^2	p-value
Intercept	1	-1.0961	0.5850	0.0610	-	-
AGE	25~34	-4.5997	0.1629	<0.0001	5004.2	<0.0001
	35~44	-4.0287	0.1055	<0.0001		
	45~49	-3.5912	0.0969	<0.0001		
	50~54	-3.1778	0.0794	<0.0001		
	55~59	-2.9997	0.0733	<0.0001		
	60~64	-2.5664	0.0664	<0.0001		
	65~69	-2.0230	0.0597	<0.0001		
	70~74	-1.4153	0.0532	<0.0001		
	75~79	-0.7303	0.0488	<0.0001		
	80~84	0.0000	0.0000	-		
CRTN	1	0.2806	0.0182	<0.0001	145.4	<0.0001
FBS	1	0.0062	0.0006	<0.0001	102.2	<0.0001
SGOT	1	0.0090	0.0012	<0.0001	48.2	<0.0001
HDL	1	-0.0085	0.0012	<0.0001	47.9	<0.0001
BP_SYS	1	-0.0500	0.0086	<0.0001	29.8	<0.0001
BP_SYS ²	1	0.0002	0.0000	<0.0001	28.5	<0.0001
Statistic	D.F.		Value		p-value	
Deviance	1,762,050		47,529		1.00	
Hosmer-Lemeshow χ^2	8		6.9683		0.5401	
AIC			47,561			

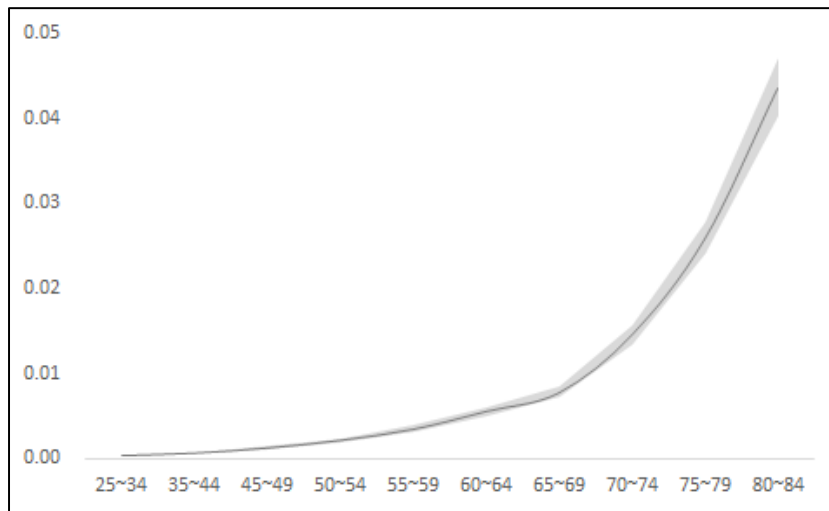
남성과 여성의 사망 로지스틱 회귀모형에 의한 연령별 평균 사망률 예측 결과는 아래 [표 12]와 같다.

[표 12] 평균 예측 사망률

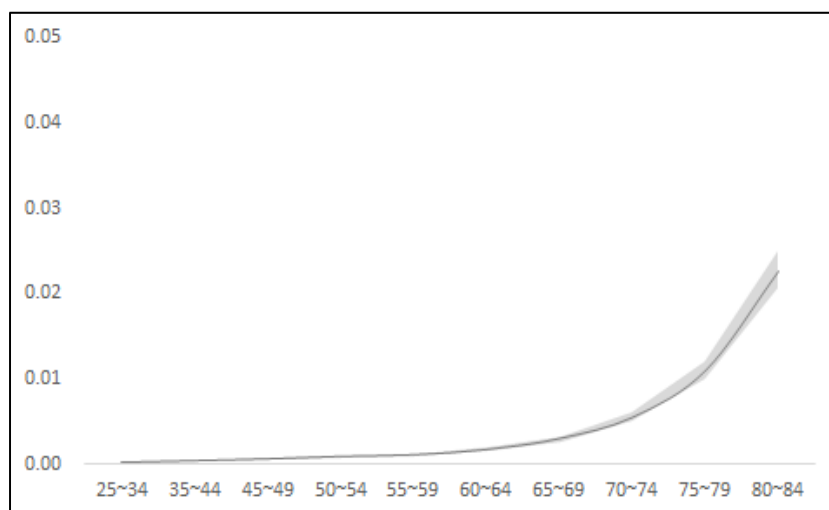
연령군	남성		여성	
	대상자수(명)	평균사망률	대상자수(명)	평균사망률
25~34	261,831	0.0004	210,515	0.0002
35~44	423,789	0.0007	299,663	0.0004
45~49	233,483	0.0013	227,593	0.0006
50~54	223,592	0.0022	234,876	0.0009
55~59	210,050	0.0035	230,184	0.0011
60~64	166,794	0.0056	187,320	0.0017
65~69	128,661	0.0078	144,065	0.0030
70~74	95,962	0.0146	112,241	0.0055
75~79	61,105	0.0259	77,412	0.0110
80~84	27,401	0.0436	38,197	0.0227

또한, 남성과 여성의 평균 예측 사망률의 95% 신뢰구간은 각각 [그림 4]와 [그림 5]와 같다.

[그림 4] 남성 평균 예측 사망률 및 95% 신뢰구간



[그림 5] 여성 평균 예측 사망률 및 95% 신뢰구간



(2) 유병기간 모형

[표 13]과 [표 14]는 각각 남성과 여성의 유병기간을 적합하기 위한 최적 음이항 회귀모형의 결과이다. 남성과 여성 각각의 유병기간모형에 대한 이탈도에 대한 p -값=1.0으로 선택된 모형 적합이 타당함을 보인다.

유병기간에 유의미한 영향을 미치는 변수를 중요도 순서대로 표기하였으며, 남성의 경우 연령(AGE), 크레아티닌(CRTN), 공복혈당(FBS), 혈청지오티AST(SGOT), (체질량지수)²(BMI²), HDL콜레스테롤(HDL), 체질량지수(BMI), 이완기혈압(BP_DIA), (이완기혈압)²(BP_DIA²) 순으로 나타났으며, 여성의 경우 연령(AGE), 크레아티닌(CRTN), 공복혈당(FBS), 혈청지피티 ALT(SGPT), (체질량지수)²(BMI²), 체질량지수(BMI), (수축기혈압)²(BP_SYS²), 수축기혈압(BP_SYS) 순으로 나타났다.

[표 13]의 남성의 유병기간 예측 적합결과를 보면, 연령(AGE)이 증가할수록 유병기간이 길어지는 경향을 보인다. 또한, 크레아티닌(CRTN)이 한 단위 증가할수록 유병기간은 $1.1080(=\exp(0.1026))$ 배 증가하고, 공복혈당(FBS)이 한 단위 증가할수록 유병기간은 $1.0016(=\exp(0.0016))$ 배 증가하는 것으로 나타났다.

[표 13] 남성 유병기간 음이항 회귀모형

요인	자유도	추정치	표준오차	p-value	LR χ^2	p-value	
Intercept	1	2.1959	0.0359	<.0001	–	–	
AGE	25~29	1	–1.5919	0.0083	<.0001	496,753	<.0001
	30~34	1	–1.5146	0.0080	<.0001		
	35~39	1	–1.3980	0.0079	<.0001		
	40~44	1	–1.2784	0.0078	<.0001		
	45~49	1	–1.1804	0.0078	<.0001		
	50~54	1	–1.0144	0.0078	<.0001		
	55~59	1	–0.8464	0.0078	<.0001		
	60~64	1	–0.6588	0.0078	<.0001		
	65~69	1	–0.4346	0.0078	<.0001		
	70~74	1	–0.2473	0.0079	<.0001		
	75~79	1	–0.1114	0.0081	<.0001		
	80~84	1	–0.0268	0.0086	0.0019		
85+	0	0.0000	0.0000	–			
CRTN	1	0.1026	0.0012	<.0001	7,431	<.0001	
FBS	1	0.0016	0.0000	<.0001	3,885	<.0001	
SGOT	1	0.0022	0.0000	<.0001	2,238	<.0001	
BMI ²	1	0.0008	0.0000	<.0001	387	<.0001	
HDL	1	–0.0009	0.0001	<.0001	334	<.0001	
BMI	1	–0.0311	0.0020	<.0001	234	<.0001	
BP_DIA	1	–0.0087	0.0006	<.0001	196	<.0001	
BP_DIA ²	1	0.0001	0.0000	<.0001	174	<.0001	
Statistic	D.F.		Value		p-value		
Deviance	1,830,729		1,691,178		1.00		
AIC ¹⁾			7,013,349				

[표 14]의 여성의 유병기간 예측 적합결과를 보면, 연령(AGE)이 증가할수록 유병기간이 길어지는 경향을 보인다. 또한, 크레아티닌(CRTN)이 한 단위 증가할수록 유병기간은 1.0867(=exp(0.0831))배 증가하고, 공복혈당(FBS)이 한 단위 증가할수록 유병기간은 1.0015(=exp(0.0015))배 증가하는 것으로 나타났다.

1) 포아송 회귀모형의 AIC는 7,631,214이다.

[표 14] 여성 유병기간 음이항 회귀모형

요인	자유도	추정치	표준오차	p-value	LR χ^2	p-value	
Intercept	1	1.7880	0.0305	<.0001	–	–	
AGE	25~29	1	–1.1270	0.0064	<.0001	330,753	<.0001
	30~34	1	–0.9764	0.0063	<.0001		
	35~39	1	–1.0298	0.0064	<.0001		
	40~44	1	–1.0198	0.0061	<.0001		
	45~49	1	–0.8975	0.0060	<.0001		
	50~54	1	–0.6906	0.0060	<.0001		
	55~59	1	–0.5542	0.0059	<.0001		
	60~64	1	–0.4198	0.0060	<.0001		
	65~69	1	–0.2141	0.0060	<.0001		
	70~74	1	–0.0587	0.0061	<.0001		
	75~79	1	0.0303	0.0062	<.0001		
	80~84	1	0.0584	0.0066	<.0001		
85+	0	0.0000	0.0000	–			
CRTN	1	0.0831	0.0013	<.0001	4,177	<.0001	
FBS	1	0.0015	0.0000	<.0001	2,699	<.0001	
SGPT	1	0.0022	0.0000	<.0001	2,530	<.0001	
BMI ²	1	0.0006	0.0000	<.0001	360	<.0001	
BMI	1	–0.0197	0.0016	<.0001	144	<.0001	
BP_SYS ²	1	0.0000	0.0000	<.0001	71	<.0001	
BP_SYS	1	–0.0026	0.0004	<.0001	49	<.0001	
Statistic		D.F.		Value		p-value	
Deviance		1,770,374		1,652,199		1.00	
AIC ²⁾				7,642,156			

[표 15]는 남성과 여성의 [표 13]과 [표 14]의 유병기간 음이항 회귀모형에 의한 연령별 평균 예측 유병기간 결과이다.

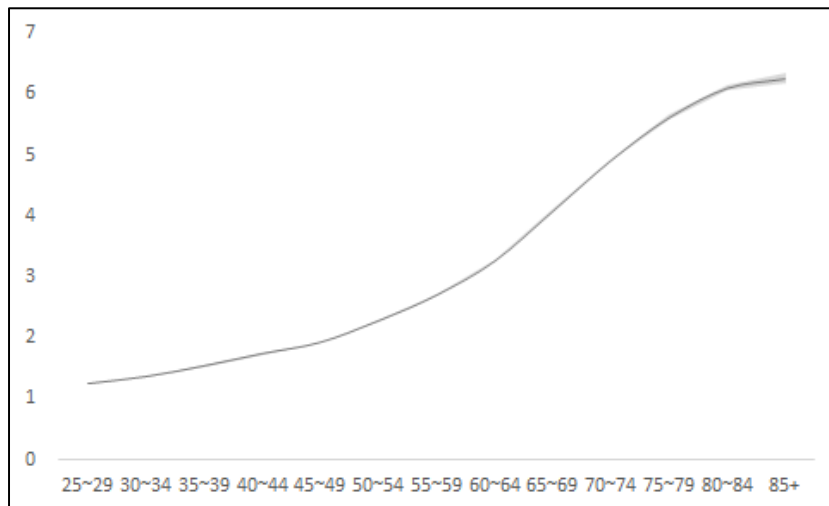
2) 포아송 회귀모형의 AIC는 8,405,808이다.

[표 15] 평균 예측 유병기간

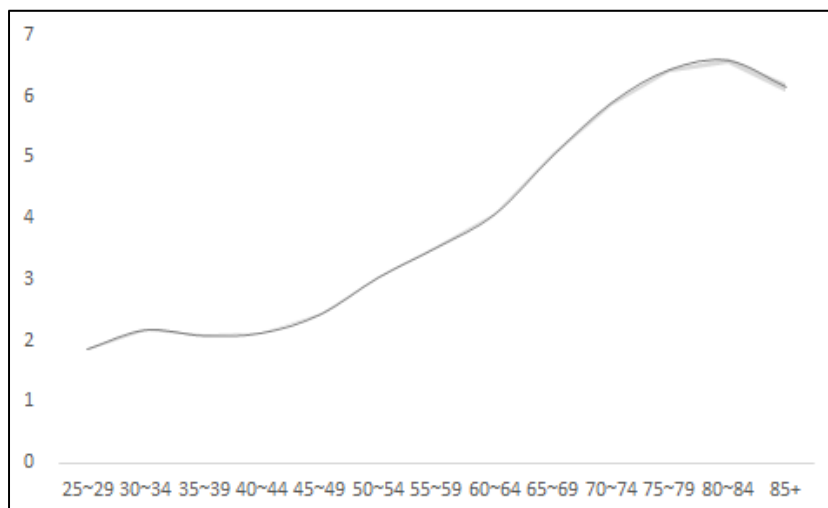
연령군	남성		여성	
	대상자수(명)	평균유병기간	대상자수(명)	평균유병기간
25~29	91,257	1.2434	100,772	1.8665
30~34	170,443	1.3597	109,694	2.1810
35~39	191,365	1.5370	96,523	2.0868
40~44	232,000	1.7386	202,986	2.1317
45~49	232,994	1.9234	227,413	2.4340
50~54	222,837	2.2789	234,482	3.0375
55~59	209,005	2.7019	229,825	3.5255
60~64	165,591	3.2614	186,869	4.0704
65~69	127,456	4.0815	143,511	5.0294
70~74	94,420	4.9178	111,471	5.8844
75~79	59,365	5.6240	76,313	6.4254
80~84	26,090	6.1110	36,995	6.5756
85+	7,927	6.2640	13,440	6.1404

또한, 남성과 여성의 평균 예측 유병기간의 95% 신뢰구간은 각각 [그림 6]과 [그림 7]과 같다.

[그림 6] 남성 평균 예측 유병기간 및 95% 신뢰구간



[그림 7] 여성 평균 예측 유병기간 및 95% 신뢰구간



(3) 평균 건강여명 모형

평균 건강여명 모형은 건강상태별 건강여명 모형과 달리 예측 모형이 불필요하며, 성별, 연령별 평균 사망률과 평균 유병기간을 산출하고, Sullivan 방법을 활용하여 평균 건강여명을 최종 산출하였다.

평균 사망률은 건강상태별 사망률 모형 대상자의 성별, 군단연령별 평균 사망률을 산출하고 직선보간 및 Coale and Kisker 모형을 활용하여 성별, 연령별 평균 사망률을 최종 산출하였다.

평균 유병기간 비율은 건강상태별 유병기간 비율 모형 대상자의 성별, 군단연령별 평균 사망률을 산출하고 직선보간을 활용하여 성별, 연령별 평균 유병기간 비율 모형을 최종 산출하였다.

[표 16]은 남성과 여성의 대표 연령별 평균 사망률, 평균 유병기간 비율, 평균 건강여명이다.

[표 16] 대표 연령별 평균 사망률, 유병기간 비율, 건강여명

연령군	남성			여성		
	평균 사망률	평균 유병기간 비율	평균 건강여명	평균 사망률	평균 유병기간 비율	평균 건강여명
25	0.0003	0.0144	57	0.0001	0.0246	62
30	0.0004	0.0167	52	0.0002	0.0309	57
35	0.0006	0.0197	47	0.0003	0.3220	52
40	0.0007	0.0235	43	0.0004	0.0321	48
45	0.0012	0.0272	38	0.0005	0.0360	43
50	0.0018	0.0328	33	0.0008	0.0453	38
55	0.0030	0.0404	29	0.0010	0.0557	33
60	0.0048	0.0501	24	0.0015	0.0657	29
65	0.0069	0.0639	20	0.0025	0.0809	25
70	0.0119	0.0797	16	0.0045	0.0980	20
75	0.0214	0.0942	13	0.0088	0.1107	16
80	0.0365	0.1052	10	0.0180	0.1165	13
85	0.0569	0.1120	8	0.0316	0.1165	10

3. 건강나이 산출 모의실험

건강상태별 건강여명 모형을 통해 산출한 건강상태별 사망률 모형과 건강상태별 유병기간 비율 모형에 각 개인별 실제 데이터(연령, 혈당, 혈압 등)를 대입하여 건강상태별 건강여명을 예측하고, 평균 건강여명과 비교하여 각 개인의 건강나이를 산출하였다.

모의실험(simulation) 데이터로는 표본코호트DB의 가장 최근 데이터인 기준년도 2017~2019년을 사용하였다. 그러나 성별에 따른 모든 연령구간의 건강상태별 건강여명 예측결과를 모두 제시할 수 없어 본 논문에서는 시뮬레이션을 위한 기초 데이터의 사례로 보험 가입율이 높은 연령구간인 40~49세의 데이터를 선정하여 그 중 최근 2년 이내 건강검진을 수검한 이력이 있으며, [표 3]의 데이터 유효성 기준을 만족한 자를 대상으로 선정하여 건강나이를 모의실험하였다.

[표 17]의 성별 건강나이 증감 분포를 살펴보면 비교적 좌측에 치우친 분포를 보여 상당수가 평균 대비 건강나이가 젊게 산출됨을 알 수 있다. 이는 모형 산출 시 현재 개인의 건강정보

(BMI, 수축기혈압, 이완기혈압 등)가 연령 증가에도 불구하고 장래에도 동일하다는 가정의 영향으로 보인다. 해당 문제를 해소하기 위해 연령 증가에 따른 건강정보 변화 예측 모델 개발하는 별도 연구를 고려해 볼 수 있다.

또한, 남성 44.4%, 여성 56.1%로 상당수가 건강나이가 실제나이 대비 -1세 또는 -2세를 나타내고 있어 이를 건강한 사람에게 혜택을 제공하는 보험 상품에 적용하는 경우 상당수가 혜택을 받을 수 있을 것으로 기대된다.

[표 17] 성별 건강나이 증감 분포(%)

구분	≤-5	-4	-3	-2	-1	0	+1	+2	+3	+4	≥+5
남성	0.6	3.3	11.6	21.6	22.8	16.1	9.5	5.3	3.1	1.9	4.2
여성	0.8	4.8	16.9	30.2	25.9	12.8	4.7	1.7	0.8	0.5	1.0

건강나이의 개선 또는 악화 여부에 따른 보험사고율을 분석한 결과는 [표 18]과 같다. 건강나이가 개선된 경우에 해당하는 대상자와 악화된 경우에 해당하는 대상자의 보험사고율을 비교하는 경우, 개선된 경우에 해당하는 대상자의 보험사고율이 낮은 점으로 미루어 건강나이가 개인의 건강상태를 나타내는 지표로 사용할 수 있는 근거가 될 수 있음을 확인하였다. 단, 기준연도와 대상 연령구간을 다르게 설정하면 [표 18]에서 제시된 성별 건강나이 증감여부별 보험사고율은 다르게 나타날 수 있다.

[표 18] 성별 건강나이 증감여부별 보험사고율

구분		사망률	수술률 (연간1회)	입원율 (입원일수)	3대질병* 유병률
남성	평균	0.0011	0.1585	1.1508	0.0337
	개선(-1세 이하)	0.0008	0.1552	1.0576	0.0291
	악화(+1세 이상)	0.0017	0.1649	1.4174	0.0425
여성	평균	0.0005	0.1599	1.0453	0.0378
	개선(-1세 이하)	0.0004	0.1566	0.8574	0.0353
	악화(+1세 이상)	0.0008	0.1814	2.2378	0.0494

* 3대질병 : 암, 허혈성심장질환, 뇌혈관질환

IV. 결론 및 시사점

본 연구는 국내외 통계기관이 공표하는 건강여명 산출 방법론을 적용하여 개인의 건강상태에 따른 건강여명을 예측하고 건강나이를 산출하는 방법을 제안하였다. 연구 기초자료로써 국민건강보험공단 표본코호트DB의 진료내역과 건강검진내역을 이용하였으며, 산출된 건강나이는 개인의 건강수준을 대표하는 객관적이고, 종합적인 지표로 활용할 수 있다.

건강나이 증감여부에 따른 보험사고율 분석 결과, 건강나이가 개선된 집단의 경우(실제나이보다 건강나이가 작게 산출된 경우) 보험사고율 역시 개선되었음을 확인하였다. 이를 통해 종래의 사람을 대상으로 하는 보험요율을 결정하는 위험의 구분단위로써 사용되는 달력나이 대신에 건강나이를 활용할 수 있는 가능성을 제시하였다.

또한, 통계청에서 발표하는 유병기간 제외 기대여명과 주관적 건강평가 기대여명은 통계청에서 실시하는 사회조사의 일부 설문문항을 차용하여 산출 중이다. 다만, 설문문항이 ‘지난 2주일 동안 질병이나 사고로 아팠던 적이 있습니까?’ 또는 ‘귀하의 전반적인 건강 상태는 어떻습니까?’와 같이 응답자의 기억 또는 주관적 평가에 의존함으로써 객관성이 다소 결여되어 있다고 할 수 있다. 본 연구에서는 이와 같은 문제를 해결하고자 유병기간을 병원이용기간으로 정의함으로써 건강여명 산출의 객관성을 부여하였다.

보건복지부 산하 한국건강증진개발원이 2021년 발표한 「제5차 국민건강증진종합계획(Health Plan 2030)」의 목표 중 하나가 2030년까지 건강수명 73.3세를 달성하는 것이다. 2005년 발표한 「제2차 국민건강증진종합계획」부터 건강수명 연장을 목표로 설정할 만큼 국민의 건강증진 수준을 파악하기 위한 지표로써 건강수명(건강여명)의 중요성을 가늠할 수 있다.

본 연구는 개인의 설문방식이 아닌 개인의 실제 건강 측정치를 나타내는 건강검진 결과에 따른 건강여명을 예측하는 방법론을 제시함으로써 건강여명 연장을 위한 중간 목표로 건강상태 변화에 따른 건강여명 변화를 예측할 수 있을 것이다. 또한, 예측된 건강여명을 바탕으로 건강나이를 산출하는 방법론을 제시하고, 건강나이 개선여부에 따른 보험사고율 증감여부가 유의미함을 제시함으로써 이를 활용한 보험상품 개발의 가능성을 제시하였다.

건강나이를 활용한 보험 상품으로 실제나이 대신 건강나이를 기준으로 보험료를 산출하는 상품 또는 일정 주기마다 건강나이를 재산출하여 건강나이가 개선된 경우 보험료 할인, 보험

금 증액 및 해약환급금 추가 적립 등 보험편익을 제공하는 상품을 개발할 수 있을 것이다.

이러한 상품 개발을 통해 동일 연령대 평균 대비 건강상태가 준수한 고객이 기존 보험 상품 대비 저렴한 보험료로 보험을 가입함으로써 개인과 사회의 안정적인 경제활동을 보장하는 보험의 순기능과 개별 위험(건강상태)에 비례한 공평한 위험 분담 시스템 구축을 기대할 수 있을 것이다. 또한, 2021년 7월부터 판매되고 있는 4세대 실손보험은 비급여 보험금 수령액에 따라 1~5등급으로 구분하여 비급여 특약 보험료를 할인 또는 할증하는 비급여 보험료 차등제도를 운영한다. 이는 보험계약자간 보험료 부담 형평성을 제고하고, 불필요한 비급여 의료이용 감소를 유도하기 위함이다.

이에 더하여 본 논문에서 제시한 건강나이로 대표되는 건강상태에 따른 보험료 차등 제도를 도입함으로써 사전적 예방 건강관리 활동 유인을 제공하여 비급여 의료이용 감소를 유도하고, 국민건강증진에 기여할 수 있을 것이다.

참 고 문 헌

- 강은정·김나연, 2007, 한국인의 건강관련 삶의 질과 건강보정 기대여명, 국민건강영양조사 제3 조사결과 심층분석연구: 건강면접 및 보건의식 부문, 질병관리본부 한국보건사회연구원, 66-95.
- 강은정·김나연·윤석준, 2008, 한국인의 건강보정 기대여명의 측정, 보건행정학회지, 제18권 1호, 108-126.
- 권태환·김태현, 1990, 한국인의 생명표: 1970-85년의 사망유형분석을 중심으로, 서울대학교 출판부.
- 남정자·조맹제·최은진, 1996, 한국인의 건강수준에 미치는 영향 분석, 한국보건사회연구원.
- 신상훈·전동운·오성진·윤세정·박종관·이한철·육태미·임현선, 2016, 표본 코호트를 이용한 한국형 심장나이 모델 생성, 국민건강보험 일산병원 연구소, 연구보고서 2016-20-004.
- 우해봉, 2009, 한국 중고령층의 성별 교육수준별 건강기대여명 차이, 한국사회학, 제39권 4호, 83-107.
- _____, 2012, 건강기대여명 추정에서 건강 측정치의 효과에 관한 연구, 한국사회학, 제43권 1호, 165-187.
- 윤병준, 1995, 건강기대수명에 의거한 한국인의 건강수준에 관한 연구, 서울대학교 대학원 보건학 박사학위 논문.
- 윤병준·김정근, 1996, 한국인의 활동장애가 없는 건강여명에 관한 연구, 한국인구학, 제19권 1호, 123-137.
- 임현선·오현철·장정현·윤소라·이재광·박소희·강혜정·장호열, 2020, 국민건강정보DB 활용 빅데이터 연구의 질 향상을 위한 분석방법 점검도구 개발 연구 - 국민건강정보DB 활용 문헌들의 방법론 고찰, 국민건강보험 일산병원 연구소, 연구보고서 2020-20-015.
- 전희주·안철경, 2012, 음이향 회귀모형을 이용한 보험설계사들의 조직성과에 관한 연구, 보험 금융연구 제23권 제4호, 29-60.

전희주·최경진, 2020, 생체연령을 고려한 생존확률 산정에 관한 연구, 연금연구, 제10권 제2호, 117-134.

_____, 2021, 생체연령을 고려한 퇴직연금 프로그램 인출방식에 관한 연구, 리스크관리 연구, 제32권 제2호, 101-134.

_____, 2022, 주택연금 가입자의 신탁방식 전환 의향 분석 및 시사점, 금융감독연구, 제9권 제2호, 171-200.

_____, 2022, 국민건강보험 빅데이터를 활용한 치아보험 요율 차등화에 관한 연구: 치아우식증·치주질환을 중심으로, 보험금융연구, 제33권 제3호, 119-147.

_____, 2023, 국민건강보험 건강검진코호트DB와 표본코호트DB를 활용한 크라운, 치수치료 담보 치아보험 상품의 상대위험도 연구, 금융감독연구, 제10권 제1호, 137-160.

한소현·이성국, 2012, 우리나라 지역별 건강수명과 관련요인, 한국인구학, 제35권 2호, 209-232.

통계청, 생명표(국가승인통계 제 101035호).

한국건강증진개발원, 2019, 국민건강증진종합계획:

<https://www.khealth.or.kr/board?menuId=MENU00733&siteId=null>

Coale A., and Kisker, E. E., 1990, Defects in Data on Old Age Mortality in the United States: New Procedures for Calculating Approximately Accurate Mortality Schedules and Life Tables at the Highest Ages, *Asian and Pacific Population Forum* 4, 1-31.

Crimmins, E. M., and Cambois, E., 2003, Social Inequalities in Health Expectancy. 111-125 in Robine, J., Jagger, C., Mathers, C. D., Crimmins, E. M., and Suzman, R. M.(eds.), *Determining Health Expectancies*, John Wiley & Sons, NJ.

Deeg, D. J. H., Verbrugge, L. M., and Jagger, C., 2003, Disability Measurement. 203-219 in Robine, J., Jagger, C., Mathers, C. D., Crimmins, E. M., and Suzman, R. M.(eds.), *Determining Health Expectancies*, John Wiley & Sons, NJ.

- Diebel, L. W. M. and Rockwood, K., 2021, Determination of Biological Age: Geriatric Assessment vs Biological Biomarkers, *Current Oncology Reports* 23, 104.
- Jackson, S. H. D., Weale, M. R., and Weale, A. R., 2003, Biological Age-what is It and Can It be Measured?, *Archives of Gerontology and Geriatrics* 36(2), 103-115.
- Gurmu, S., 1991, Tests for Detecting Overdispersion in the Positive Poisson Regression Model, *Journal of Business and Economic Statistics* 9, 215~222.
- Ritchie, K., and Polge, C., 2003, Mental Health Expectancy. 175-182 in Robine, J., Jagger, C., Mathers, C. D., Crimmins, E. M., and Suzman, R. M.(eds.), *Determining Health Expectancies*, John Wiley & Sons, NJ.
- Robine, J., Jagger, C., Oyen, H. V., Cox, B., Cambois, E., Romeiu, I., Clavel, A., and Roy, S. L., 2006, *Health Expectancy Calculation by the Sullivan Method: A Practical Guide*. Montepellier, France: Euro.
- Sullivan, D. F., 1971, A Single Index of Mortality and Morbidity, *HSMHA Health Rep* 86(4), 347-354
- Yang, Y., 2008, Long and Happy Living: Trends and Patterns of Happy Life Expectancy in the U.S., 1970-2000, *Social Science Research* 37(4), 1235-1252.

A study on calculating healthy age using mortality rate, disease period, and healthy life expectancy according to health status

Heujju Chun*, Kitae Moon**, Taekyo Leen***

Abstract

The purpose of this study is to analyze the mortality rate and duration of illness according to an individual's health status based on health examination information using the sample cohort DB, and to propose a method for calculating health age using the predicted health life expectancy based on the mortality rate and duration of illness. Statistical prediction models for mortality and duration of illness according to individual health status were developed using logistic regression models and negative binomial regression models, respectively. And an individual health life expectancy prediction model was developed using the Sullivan method based on life tables, and the health age was finally calculated by comparing it with the average health life expectancy of the entire population. A simulation was conducted on people in their 40s who had taken health examinations within the past two years as an example.

Key words: health life expectancy, health age, health examination, generalized linear model, sample cohort DBI

JEL Classifications: J32

* First author, Associate Professor, Department of Statistics and Information, Dongduk Women's university(hjchun@dongduk.ac.kr)

** Adjunct Professor, Department of Preventive Medicine, Yonsei University College of Medicine (tubemed@hanmail.net)

*** Corresponding author, Ph.D, Hanyang University(tkleen@naver.com)