

국민건강보험 표본코호트2.0DB를 활용한 건강상태에 따른 암발생과 암수술건수 상대위험도 연구*

A Study on Relative Risk of Cancer Occurrence and the Number of Cancer Surgeries Based on Health Conditions Using National Health Insurance Cohort 2.0DB

전희주** · 인태교***

Heuiju Chun · Taekyo Leen

본 연구는 국민건강보험 표본코호트2.0DB를 사용하여 일반대상자와 간편고지대상자를 대상으로 남녀 각각에 대해 향후 1년 안에 예측된 암 발생률과 암수술건수를 적용하여 일반대상자와 간편고지대상자 간의 암발생 상대위험도와 암수술건수 상대위험도를 산출하였다. 일반대상자는 설명변수로 BMI, 혈압 등의 건강검진 항목들을 이용하고 간편고지대상자는 건강검진 항목과 직전 1년 장기투약 여부, 3가지 질병 치료 유무를 설명변수로 포함하였다. 암발생 예측은 로지스틱회귀모형을 이용하고, 암수술건수 예측은 음이항회귀모형을 이용하였다.

일반대상자와 간편고지대상자 간의 암발생 상대위험도와 암수술건수 상대위험도를 계산한 결과, 남녀 모두 암발생과 암수술건수 모두 일반대상자 대비 간편고지대상자의 위험도는 1보다 작게 나타났다. 이것은 가입조건을 완화한 간편고지만으로 유병자들을 포함한 보험가입대상자들의 상품영역 확대의 가능성을 보여주며, 또한 간편고지만으로도 암발생과 수술건수 담보의 상대위험도가 일반대상자 대비 확실히 구분됨을 실증적으로 제시하였다.

한글 색인어 : 건강검진, 상대위험도, 암발생, 암수술건수, 일반화선형모형, 표본코호트2.0DB

* 이 논문은 한국리스크관리학회 2022 연구지원사업의 지원을 받아 연구되었음.

** 제1저자, 동덕여자대학교 정보통계학과 부교수, email: hjchun@dongduk.ac.kr

*** 교신저자, 한양대학교 보험계리학과 박사과정, email: tkleen@naver.com

논문투고일: 2022. 11. 22 논문수정일: 2022. 12. 11 게재확정일: 2022. 12. 15

I. 서론

최근 전세계적으로 고령화 및 만성질환자 증가에 따른 국민 의료비 부담완화가 사회적인 현안으로 부상하고 있다. 우리나라도 빠른 고령화, 만성질환 증가 등으로 GDP 대비 의료비 지출 비중이 OECD 국가에 비해 빠르게 증가하고 있다. 이에 따라 국민의 관심은 기대수명이 아닌 건강수명의 연장, 치료보다 사전 예방으로 변화하고 있다. 이러한 사회적 수요 증가와 4차 산업혁명에 따른 혁신 기술이 결합되면서, 헬스케어 산업 또한 빠르게 성장하고 있다. 특히, 의료기관 뿐 아니라, IT시스템, S/W, 장비 제조사 등 다양한 신산업이 융복합된 스마트 헬스케어 산업은 빠르게 확장되고 있다.

빅데이터의 축적과 통계적 모형, 머신러닝의 발달에 따라 해외 주요국에서는 ‘보험’과 ‘헬스케어’가 결합된 건강증진형 보험상품 개발이 활성화되고 있다. 이러한 건강증진형 보험상품은 보험계약자와 보험회사 양 측면에서 서로 도움이 되는 유망한 인슈어테크 분야이다. 보험계약자는 스스로의 건강관리 노력을 통해 건강수명을 연장하고 보험료 할인 등 금융비용도 절감하며, 보험회사도 보험 계약자의 질병발생 확률, 조기 사망확률 등 사고위험이 낮아지면서 손해율이 하락하는 등의 혜택을 누릴 수 있다.

보험업계는 그동안 의료법상 등의 이유로 접근할 수 있는 데이터에 제한이 있어 관련 상품 및 서비스 제공에 한계가 있었다. 그러나 최근 정부 주도하에 보험업권 헬스케어 활성화를 추진하면서 관련 규제개선, 보험업권 공공데이터 활용, 헬스케어업계-보험업계 협업 방안 등의 논의가 활발하게 이루어지고 있다.

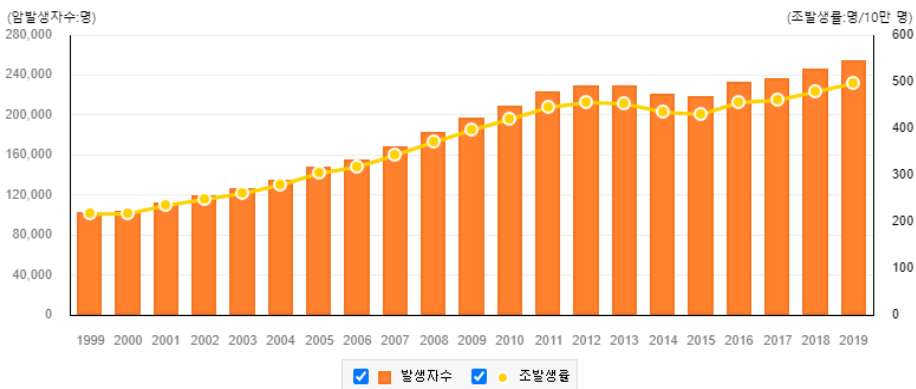
공공데이터 활용과 관련해서 건강보험공단에서는 연구자에게 건강보험DB에서 표본 추출한 약 100만 가량의 건강상태, 사망 및 질병의 발생, 의료이용 등의 정보를 가진 표본코호트2.0DB를 제공하고 있다. 표본코호트2.0DB는 질병의 진단, 입원, 수술 및 사망에 관한 자료와 함께 건강검진 시 측정된 다양한 건강 지표들을 포함하고 있어, 건강상태를 측정하고 분류할 수 있는 통계적 질병 발생 모형을 설계하는 데 유용하게 이용될 수 있다.

이러한 관점에서 본 연구는 건강보험공단에서 제공하는 표본코호트 2.0DB를 이용하여 건강검진에 의한 개인별 건강상태를 활용하여 한국인의 주요 질병인 암발생과 암수술건수를 예측하는 모형을 구해 일반대상자와 간편고지대상자의 암발생과 암수술건수 상대위험도 계산을 통해 간편고지의 효율성을 실증적으로 보이고자 한다.

II. 이론적 배경 및 연구 방법

1. 이론적 배경

2019년 현재 한 해 동안 암 진단을 받은 암발생자는 254,718명으로 남성은 134,180명, 여성은 120,538명으로 2010년 대비 22.2% 증가했고, 2019년 기준 가장 많이 발생한 암으로는 폐암, 위암, 대장암, 유방암, 간암 순이다¹⁾.

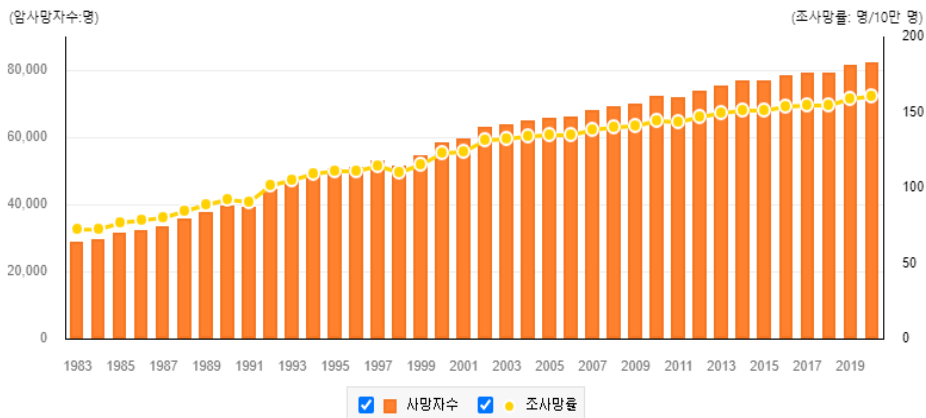


〈그림 1〉 암 발생자수와 인구 10만 명당 발생률

1) 보건복지부, 2021년 보도자료

우리나라의 암 조발생률²⁾은 인구 10만명당 1999년 215.9명 이후 2012년 254.0명까지 꾸준히 증가한 이후 2015년 428.3명, 2016년 455.1명부터 2019년 496.2명(남자 308.1명, 여자 297.4명)으로 매년 꾸준히 증가하고 있다(그림 1 참조). 2019년 현재 우리나라 국민들이 기대수명 83세까지 생존할 경우 암에 걸릴 확률은 37.9%이며, 기대수명이 80세인 남성은 5명 중 2명(39.9%), 기대수명이 87세인 여성은 3명 중 1명(35.8%)에서 암이 발생할 것으로 추정된다.³⁾ 미국의 경우 2019년 현재 남성은 5명 중 2명(40.5%), 여성은 8명 중 3명(38.9%)이 암이 발생하는 것으로 조사되었다(Siegel et al., 2021).

우리나라 암에 의한 조사망률⁴⁾은 꾸준히 증가 추세에 있으며 2020년 인구 10만명당 암사망자수는 160.1명으로 2010년의 10년 전 대비 15.7명 증가하였다(그림 2 참조).



〈그림 2〉 암 사망자수와 인구 10만명당 사망률

2) 조발생률(암발생률)은 연간 발생자수를 연앙인구로 나눈 수치를 100,000 분비로 나타낸 수치임

3) 국가암정보센터, 2022년도 보도자료

4) 조사망률(암사망률): 연간 사망자수를 연앙인구로 나눈 수치를 100,000 분비로 나타낸 수치

2020년 기준 인구 10만명당 암사망률이 높은 순으로는 폐암(36.4명), 간암(20.6명), 대장암(17.4명), 위암(14.6명), 췌장암(13.2명) 순이며, 2010년의 10년 전보다 폐암, 췌장암, 대장암 사망률은 증가하였고, 위암, 간암 사망률은 감소하였다⁵⁾.

암의 조기진단은 환자의 높은 생존율과 낮은 사망률을 가져오며 또한 의료비용을 현저히 줄일 수 있는 방법으로 환자 개인뿐만 아니라 국가재정과 기업들에게도 긍정적인 영향을 줄 수 있는 방법이다. 2021년 국민건강보험공단의 암환자의 요양급여비용은 무려 5조 2,609억⁶⁾이며, 민간보험회사의 암보험금 지급금액은 2018년 생명보험회사는 4조 9천억원, 손해보험사는 1조 1천억원(김석영, 2020)으로 국가와 보험사들에게 엄청난 재정적 영향을 미치고 있다. 또한 암환자와 환자의 가족들이 속한 기업들에게도 업무의 공백과 가족의 결근으로 기업의 경쟁력약화를 초래할 수 있다(이찬희 외, 2015).

특히 초기 암환자 진단의 중요성은 비용적인 측면에서 감히 상상을 초월한다. 미국의 예를 들면, 암 진단 후 처음 24개월 동안 1인 환자 비용을 보면 단계 0에서는 \$72,000, 단계 I/II \$97,000, 단계 III \$159,000, 단계 IV에서는 \$182,000로 암단계 진전시 비용의 증가는 더욱 커지게 된다(Alfayez al., 2021). 미국의 암 초기 진단의 비용 절감액은 무려 \$260억이 된다(Kakushadze et al., 2017).

개인에 대한 암질병 단계 또는 암에 걸릴 위험을 확률로 나타내는 암질병 예측모형 개발은 실로 매우 중요하다. 암질병 예측모형은 개인들의 암 방지를 위한 라이프스타일, 다양한 행동의 개인 건강관리의 방향 또는 지침을 제공해 줄 수 있다. 고전적인 암 예측모형은 성별, 연령과 같은 기본적인 인구통계적 정보, 흡연상태, 음주습관, 신체 정보 등의 라이프스타일 요인 또는 환경노출 정보, 병원 이력정보, 측정된 임상지표(biomarker), 이미지 정보 등의 임상 위험요소와 가족력 등이 포함된다(Wang et al., 2022; Britt et al., 2020; Gray et al.,

5) 보건복지부, 2021년 보도자료

6) 국민건강보험심사평가원, 2022년도 보도자료

2016; Win et al. 2012).

대부분의 암질병 모형 예측은 일반적으로 코호트 연구에 의한 임상데이터를 이용하여 로지스틱회귀모형 또는 Cox-비례위험회귀모형에 의해 개발된다(Usher-Smith et al., 2015; Colditz et al., 2000). 최근 해외에서는 암 생존자들의 생활 습관 및 삶의 질 정보를 이용하여 머신러닝에 의한 암질병 예측모형 개발 연구들도 증가하는 추세이다(Christodoulou et al., 2019; Stark et al., 2019; Kourou et al., 2015).

최근 국내 보험사들은 보험 시장 확대를 위해 유병자들을 위한 시장 확대 보험상품들을 개발하고 있으며, 특히 이를 위해 경증유병자, 노령자들에게 간편고지를 하는 것만으로 보험을 가입할 수 있는 간편고지 암보험 상품들을 적극적으로 판매하고 있다. 유병자보험은 무심사, 간편심사, 간편고지(간편가입)의 3가지 형태가 있고, 무심사보험은 모든 질병이나 치료내역에 대한 고지사항과 건강검진 절차 없이 보험가입이 가능하지만 보장 내용이나 보장금액을 축소하여 판매하고 있다. 간편심사보험은 고혈압이나 당뇨 치료병력에 대한 고지항목을 면제한 상품으로 주로 고혈압, 당뇨병을 가진 사람들이 가입한다. 간편고지보험은 청약서 질문 항목 중 일부 항목에 대해 심사하지 않는 상품으로서, 고지항목을 3개 이내로 제한하여 질병을 갖고 있더라도 해당 고지항목에 해당되지 않으면 가입이 가능하도록 심사 항목을 줄인 상품이다. 다음은 보험마다 질문내용은 상이하지만 가장 보편적인 간편고지 상품의 고지 질문의 예이다.

1. 최근 3개월 이내에 의사로부터 진찰 또는 검사(건강검진 포함)를 통하여 다음의 의료행위를 받은 사실이 있습니까?
1) 입원 필요 소견 2) 수술 필요 소견 3) 추가검사(재검사) 필요 소견
2. 최근 2년 이내에 질병이나 사고로 인하여 입원 또는 수술(제왕절개 포함)을 받은 사실이 있습니까?
3. 최근 5년 이내에 의사로부터 진찰 또는 검사를 통하여 암으로 진단받거나 입원 또는 수술을 받은 적이 있습니까?

국내 최초 유병자보험은 2007년 금호생명 무심사 상품의 최초 개발로 시작되었으며, 2012년 고혈압·당뇨 환자의 고지를 축소한 라이나생명의 실버암 상품의 판매 후 2018년 기준 38만여 건의 높은 판매 실적을 기록했다. 이에 따라 보험회사들은 고령자와 유병자를 대상으로 간편고지 상품을 개발하여 만성질환자인 유병자를 대상으로 다양한 간편고지 상품을 판매하게 되었다. <표 1>은 2015년~2020년 최근 연도별 간편고지, 간편심사, 무심사 등 유병자보험 신계약 건수 현황을 보여준다.

<표 1> 연도별 유병자보험 신계약 건수 현황(단위 :천건)

연도	간편고지	간편심사	무심사	계
2015년	283	379	84	786
2016년	1201	256	84	1541
2017년	1224	158	79	1461
2018년	1847	229	70	2147
2019년	3491	447	62	4000
2020년	4064	535	78	4677

(자료출처: 보험개발원)

국민건강보험 표본코호트2.0DB와 간편고지에 대한 연구로는 권혁성 외(2021)와 황지연 외(2020)이 있다. 권혁성 외(2021)은 국민건강보험 표본코호트DB에 기반한 건강 수준 평점화 모형을 활용한 간편고지 상품의 유형별 보험요율 차등화 방법을 제안하였고, 황지연 외(2020)는 국민건강보험 표본코호트DB를 활용하여 간편고지 상품의 담보별 최적 언더라이팅 기준을 산출하였다. 전희주, 최경진(2022)은 국민건강보험 표본코호트2.0DB를 기반으로 치아우식증과 치주질환 담보로 일반화선형모형의 예측값을 이용한 치아보험 요율 차등화 방법을 제안하였다.

2. 연구데이터

본 연구에서 사용되는 데이터는 국민건강보험공단에서 제공하는 표본코호트 2.0DB를 사용하였다. 표본코호트2.0DB는 2006년 1년간 건강보험 가입자 혹은 의료급여수급권자 자격을 유지한 한국 국적 보유자를 기준으로 성, 연령, 지역구분, 가입자 구분 및 보험료 분위를 층화하여 무작위 추출된 100만명(모집단의 2%)의 2002년부터 2019년까지의 출생 및 사망, 자격 및 보험료, 진료, 건강검진, 요양기관, 장기요양 등 의료이용 현황 정보를 포함하고 있는 자료이다. 본 연구는 표본코호트2.0DB의 정보 중 출생 및 사망, 건강보험 자격유지, 건강검진 및 진료명세서 정보를 활용하였다.

분석대상자는 2015~2019년 기간 중에 건강보험자격을 유지하고 있는 자 중 당해연도를 포함하여 직전 3년간 건강검진 이력이 있는 자를 대상으로 하였다.

3. 연구설계

본 연구에서는 로지스틱회귀모형을 이용하여 암발생률을 예측하는 모형과 음이항회귀모형을 이용하여 1년 내 암발생률과 암수술건수를 예측하는 모형을 산출하고자 한다.

개인의 암발생과 암수술건수는 남녀 각각에 대해 일반대상자와 간편고지대상자로 구분하여 각각의 모형을 구하고자 한다.

- 일반대상자: 2015년~2019년 기간 중에 건강보험자격을 유지하고 있고 당해연도를 포함하여 직전 3년간 건강검진 이력이 있는 자
- 간편고지대상자: 일반대상자 중에서 직전 5년간 7대질병⁷⁾으로 치료이력이 없는 자

모형에 적용되는 암발생과 암수술건수 분석대상자는 20세~79세로 정하였다.

7) 암, 협심증, 심근경색, 심장판막증, 간경화증, 특정뇌혈관질환, 투석중인 만성신장 질환

다음의 가설1과 가설2는 본 연구에서 일반대상자와 간편고지대상자의 상대위험도를 실증적으로 비교하고자 하는 가설이다.

가설1: 간편고지대상자는 일반대상자 대비 암발생 상대위험도는 1보다 작을 것이다.

가설2: 간편고지대상자는 일반대상자 대비 암수술건수의 상대위험도가 1보다 작을 것이다.

〈표 2〉는 암을 건강사고로 정의하고, 건강사고의 범위를 정의하는 한국표준질병 및 사인분류(Korean Standard Classification of Diseases ; KCD)이다.

〈표 2〉 암 분류표

대상질병	분류번호
입술, 구강 및 인두의 악성신생물	C00-C14
소화기관의 악성신생물	C15-C26
호흡기 및 흉곽내 기관의 악성신생물	C30-C39
골 및 관절연골의 악성신생물	C40-C41
흑색종 및 기타 피부의 악성신생물	C43-C44
중피성 및 연조직의 악성신생물	C45-C49
유방의 악성신생물	C50
여성 생식기관의 악성신생물	C51-C58
남성 생식기관의 악성신생물	C60-C63
요로의 악성신생물	C64-C68
눈, 뇌 및 중추 신경계통의 기타 부분의 악성신생물	C69-C72
갑상선 및 기타 내분비선의 악성신생물	C73-C75
불명확한, 이차성 및 상세불명부위의 악성신생물	C76-C80
림프, 조혈 및 관련조직의 악성신생물	C81-C96
독립된(원발성) 여러 부위의 악성신생물	C97
진성 적혈구 증가증	D45
골수 형성이상 증후군	D46
만성 골수증식질환	D47.1
본태성(출혈성) 혈소판혈증	D47.3
골수섬유증	D47.4
만성 호산구성 백혈병[과호산구증후군]	D47.5
달리 분류되지 않은 랑게르한스 세포 조직구증	D76.0
림프종모양구진증	L41.2

〈표 2〉의 분류를 바탕으로 암발생유무, 암으로 인한 수술건수를 반응변수로 정의하였다. 또한 암발생에 영향을 미칠 수 있을 것으로 예상되는 건강검진 항목들을 〈표 3〉의 설명변수들로 고려하였다. 고지혈증(총콜레스테롤, 고밀도콜레스테롤, 저밀도콜레스테롤, 중성지방) 검사는 검사주기가 4년에 1번이고, 검사율이 저조해 항목에서 제외하였으며, 간편고지대상자에 대해서는 위의 건강검진 항목 외에 간편고지 조건의 항목을 추가하였다.

〈표 3〉 건강검진 항목으로 고려된 설명변수

설명변수명		영문명
건강검진 항목	연령	AGE
	체질량지수, 체질량지수 ²	BMI, BMI2
	수축기혈압, 수축기혈압 ²	SYS, SYS2
	이완기혈압, 이완기혈압 ²	DIA, DIA2
	요단백	URN_PROT
	혈색소	HGB
	식전혈당(공복혈당)	FBS
	혈청크레아티닌	CRTN
	(혈청지오티)AST	SGOT
	(혈청지피티)ALT	SGPT
	감마지티피	GGT
간편고지 조건	직전 1년 장기투약※ 여부	MED
	직전 1년 고혈압 치료이력 유무	BLOOD
	직전 1년 고지혈증 치료이력 유무	HYPER
	직전 1년 당뇨병 치료이력 유무	DANG

※ 30일 이상 투약(30일 이내 동일상병으로 재투약시 동일 투약으로 간주)

4. 통계적 모형

본 연구의 예측을 위해 사용된 반응변수는 암발생과 암수술건수이다. 암발생은 이항자료로 일반화선형모형 중의 하나인 로지스틱회귀모형을 적합하였다. 로지스틱회귀모형은 반응변수가 0과 1인 이항값을 갖는 이항분포를 갖고, 선형예측식 $x'\beta$ 은 로짓(logit) 연결함수를 통해 평균 모수인 $p = E(y|x)$ 와 연결되어

식 (1)과 같이 표현된다(전희주, 최경진, 2021; ; 박복희, 전희주, 2018b).

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}_i' \boldsymbol{\beta}. \quad (1)$$

또 다른 반응변수인 암수술건수는 건수자료(count data)로써 일반화선형모형의 포아송회귀모형과 음이항회귀모형(negative binomial regression model)으로 적합이 가능하다(박복희, 전희주, 2018a; Gumu, 1991). 포아송회귀모형은 반응변수가 양의 정수이고 평균과 분산이 같다는 가정 하에 적용되는 가장 기본적인 모형이다. 그러나 음이항회귀모형은 분산이 평균보다 크게 나타나는 과대산포가 존재하는 경우에 적합한 모형으로 연결함수는 로그함수를 사용하여 식 (2)와 같이 표현된다.

$$\log(\mu) = \mathbf{x}' \boldsymbol{\beta}, \quad (2)$$

여기서 $\boldsymbol{\beta}$ 는 $(k+1) \times 1$ 모수 벡터를 나타내고, k 는 설명변수의 수이다.

Ⅲ. 실증분석

1. 관찰대상 집단의 일반적인 특성

〈표 4〉는 분류기준을 적용한 일반대상자와 간편고지대상자의 성별 연령별 분포이다.

〈표 4〉 일반대상자 및 간편고지대상자 특성

구분		일반대상자	%	간편고지대상자	%
성별	남성	1,243,253	50.1	1,088,134	50.3
	여성	1,239,293	49.9	1,074,398	49.7
연령	20~29	202,098	8.1	199,845	9.2
	30~39	387,409	15.6	377,571	17.5

구분	일반대상자	%	간편고지대상자	%
40~44	298,111	12.0	283,389	13.1
45~49	316,869	12.8	293,788	13.6
50~54	304,887	12.3	270,464	12.5
55~59	309,425	12.5	259,751	12.0
60~64	257,052	10.4	201,495	9.3
65~69	180,824	7.3	131,055	6.1
70~74	133,305	5.4	87,960	4.1
75~79	92,566	3.7	57,214	2.6
합 계	2,482,546	100.0	2,162,532	100.0

먼저 일반대상자를 보면, 모든 분류기준에 대해서 성별 남녀 비중은 각각 50.1%, 49.9%로 거의 동일하며, 집단 내 연령대별 비중은 30~39세(15.6%), 45~49세(12.8%), 55~59세(12.5%), 50~54세(12.3%), 40~44세(12.0%) 순 등으로 나타나 60세 미만의 연령대가 집단의 73.3%를 차지하고 있다. 간편고지대상자 또한 성별 남녀 비중은 각각 50.3%, 49.7%로 거의 동일하며, 집단 내 연령대별 비중은 30~39세(17.5%), 45~49세(13.6%), 40~44세(13.1%), 50~54세(12.5%), 55~59세(12.0%) 순 등으로 나타나 60세 미만의 연령대가 집단의 77.9%를 차지하고 있다.

2. 일반대상자 대비 간편고지대상자 상대위험도

본 연구는 국민건강보험 표본코호트2.0DB의 국민건강검진 데이터와 설계된 분석모형을 바탕으로 성별, 대상자별로 암발생유무, 암수술건수에 대한 일반대상자 대비 간편고지대상자의 상대위험도를 제시하고자 한다.

일반대상자와 간편고지대상자 각각에 대해 검진항목 등을 설명변수로 하는 암발생유무, 암수술건수에 적합한 예측모형으로 각각 로지스틱회귀모형과 음이항회귀모형을 적용하였다. 가장 먼저 반응변수인 암발생유무, 암수술건수에 대해 건강검진 관련 설명변수들을 각각에 대해 로지스틱회귀모형과 일변량 음이항회

귀모형을 적합하여 유의수준 0.05에서 유의한 설명변수들을 선택하였고 이들 유의한 설명변수들을 대상으로 후진제거법(backward elimination method)을 사용하여 AIC가 가장 작은 모형을 최적모형으로 선택하였다. 그리고 모형적합도 검증을 위해 이탈도(deviance) 통계량의 p-값 >0.05 인 모형인지를 확인하였다. 로지스틱회귀모형의 경우에는 예측치와 실제 관측치의 차이를 보여주는 적합도 통계량인 Hosmer and Lemeshow 카이제곱 통계량이 p-값 >0.05 인지를 추가적으로 확인하였다.

(1) 일반대상자

1) 암발생

〈표 5〉와 〈표 6〉은 일반대상자 남성과 여성 각각 암발생유무를 적합하기 위한 최적 로지스틱회귀 예측모형 결과이다. 남성의 경우 연령(AGE), 혈색소(HGB), 감마지티피(GGT), 혈청지피티(SGPT), 요단백(URN_PROT), 수축기혈압(SYS)이 유의한 설명변수로 선택되었으며, 여성의 경우 연령(AGE), 혈색소(HGB), 요단백(URN_PROT), 혈청지피티(SGPT), 감마지티피(GGT), 이완기혈압(DIA)이 설명변수로 선택되었다. 연령은 모든 남성과 여성 모두 연령이 증가할수록 암발생률은 높아지는 경향을 보인다.

〈표 5〉의 일반대상자 남성 암발생 모형 적합결과를 보면, 혈색소(HGB)는 한 단위 증가할 때 암발생 오즈는 $0.854(=\exp(-0.1577))$ 배 감소하고, 감마지티피(GGT)는 한 단위 증가할 때 암발생 오즈는 $1.001(=\exp(0.0011))$ 배 증가한다. 혈청지피티(SGPT)는 한 단위 증가할 때 암발생 오즈는 $1.001(=\exp(0.0010))$ 배 증가하고, 요단백(URN_PROT)은 양성 대비 음성이 암발생 오즈가 $1.292(=\exp(0.2565))$ 배 높다. 수축기혈압(SYS)은 1단위 증가할 때 암발생 오즈는 $1.003(=\exp(0.0032))$ 배 증가하는 것으로 나타난다.

〈표 5〉 일반대상자 암발생(남성) 로지스틱회귀모형

설명변수		자유도	추정치	표준오차	p-value
Intercept		1	- 0.6209	0.1890	0.0010
AGE	20~29	1	- 3.4723	0.1340	< 2e-16
	30~39	1	- 2.7049	0.0723	< 2e-16
	40~44	1	- 2.2667	0.0738	< 2e-16
	45~49	1	- 1.9720	0.0674	< 2e-16
	50~54	1	- 1.4916	0.0614	< 2e-16
	55~59	1	- 1.0296	0.0565	< 2e-16
	60~64	1	- 0.6961	0.0556	< 2e-16
	65~69	1	- 0.3260	0.0558	0.0000
	70~74	1	- 0.0231	0.0569	0.6848
75~79	0	0.0000	0.0000	-	
SYS		1	- 0.0032	0.0009	0.0007
URN_PROT	0(음성)	0	0.0000	0.0000	-
	1(양성)	1	- 0.2565	0.0464	0.0000
HGB		1	- 0.1577	0.0101	< 2e-16
SGPT		1	0.0010	0.0002	0.0000
GGT		1	0.0011	0.0002	0.0000
통계량			자유도	값	p-value
Hosmer and Lemeshow 카이제곱			8	5.0841	0.7485
AIC			66.639		

〈표 6〉의 일반대상자 여성 암발생 모형 결과를 보면, 혈색소(HGB)는 한 단위 증가할수록 암발생 오즈는 $0.951(=\exp(-0.0503))$ 배 감소하고, 요단백(URN_PROT)은 양성 대비 음성인 경우 암발생 오즈는 $1.166(=\exp(0.1535))$ 배 높게 나타난다. 혈색소(HGB)는 한 단위 증가할 때 암발생 오즈는 $0.951(=\exp(-0.0503))$ 배 감소하고, 감마지티피(GGT)는 한 단위 증가할 때 암발생 오즈는 $1.001(=\exp(0.0009))$ 배 증가한다. 혈청지피티(SGPT)는 한 단위 증가할 때 암발생 오즈는 $1.001(=\exp(0.0010))$ 배 증가하고, 이완기혈압(DIA)은 1단위 증가할 때 암발생 오즈는 $1.003(=\exp(0.0026))$ 배 증가한다.

〈표 6〉 일반대상자 암발생(여성) 로지스틱회귀모형

설명변수		자유도	추정치	표준오차	p-value
Intercept		1	-3.5947	0.1637	< 2e-16
AGE	20~29	1	-1.8733	0.0819	< 2e-16
	30~39	1	-1.1423	0.0565	< 2e-16
	40~44	1	-0.6563	0.0505	< 2e-16
	45~49	1	-0.5128	0.0479	< 2e-16
	50~54	1	-0.2801	0.0458	0.0000
	55~59	1	-0.2163	0.0454	0.0000
	60~64	1	-0.1497	0.0469	0.0014
	65~69	1	-0.1035	0.0515	0.0446
	70~79	0	0.0000	0.0000	-
DIA		1	0.0026	0.0013	0.0384
URN_PROT	0(음성)	0	0.0000	0.0000	-
	1(양성)	1	-0.1535	0.0509	0.0026
HGB		1	-0.0503	0.0102	0.0000
SGPT		1	0.0010	0.0003	0.0026
GGT		1	0.0009	0.0004	0.0101
통계량			자유도	값	p-value
Hosmer and Lemeshow 카이제곱			8	4.1477	0.8436
AIC			79,258		

〈표 7〉은 〈표 5〉와 〈표 6〉의 일반대상자 남성과 여성 각각의 로지스틱회귀모형의 추정값을 사용하여 각 개인들의 암발생 예측값을 구한 다음, 남성과 여성 각각의 연령군별 평균을 계산하여 얻은 예측 암발생률이다.

〈표 7〉 일반대상자 암발생 예측률

연령군	남성		여성	
	대상자수	암발생 예측률	대상자수	암발생 예측률
20~29	78,234	0.000818	79,328	0.002319
30~39	188,253	0.001811	105,386	0.004868
40~44	109,839	0.002859	92,201	0.008048
45~49	111,233	0.003902	97,561	0.009338
50~54	100,061	0.006376	94,575	0.011631
55~59	94,806	0.010221	93,996	0.012330
60~64	71,417	0.014436	75,237	0.013185
65~69	45,304	0.021301	49,335	0.013885
70~74	29,001	0.029378	56,028	0.015528
75~79	16,354	0.031491		

2) 암수술건수

〈표 8〉과 〈표 9〉는 일반대상자 남성과 여성 각각의 암수술건수를 적합하기 위한 최적 음이항회귀모형 결과이다. 남성의 경우 연령(AGE), 혈색소(HGB), 혈청지오티(SGOT), 감마지티피(GGT), 혈청지피티(SGPT), 요단백(URN_PROT), 수축기혈압(SYS), 이완기혈압(DIA), 식전공복혈당(FBS), 이완기혈압의 제공이 유의한 설명변수로 선택되었으며, 여성의 경우 연령(AGE), 혈색소(HGB), 요단백(URN_PROT), 혈청지피티(SGPT), 감마지티피(GGT), 식전공복혈당(FBS), 이완기혈압(DIA)이 선택되었다. 연령은 모든 남성과 여성 모두 연령이 증가할수록 암수술건수는 많아지는 경향을 보인다.

〈표 8〉 일반대상자 암 수술(남성) 음이항회귀모형

설명변수		자유도	추정치	표준오차	p-value
Intercept		1	0.1141	0.5625	0.8393
AGE	20~29	1	-3.6525	0.1671	$< 2e-16$
	30~39	1	-2.6026	0.0798	$< 2e-16$
	40~44	1	-2.1924	0.0802	$< 2e-16$
	45~49	1	-1.9560	0.0744	$< 2e-16$
	50~54	1	-1.3387	0.0669	$< 2e-16$
	55~59	1	-0.8761	0.0624	$< 2e-16$
	60~64	1	-0.4771	0.0611	0.0000
	65~69	1	-0.2482	0.0624	0.0001
	70~74	1	0.0302	0.0640	0.6375
	75~79	0	0.0000	0.0000	-
SYS		1	-0.0042	0.0014	0.0032
DIA		1	-0.0206	0.0135	0.1273
DIA2		1	0.0002	0.0001	0.0684
URN_PROT	0	0	0.0000	0.0000	-
	1	1	-0.1536	0.0528	0.0036
HGB		1	-0.2026	0.0110	$< 2e-16$
FBS		1	0.0011	0.0005	0.0221
SGOT		1	0.0047	0.0006	0.0000
GGT		1	0.0011	0.0002	0.0000
통계량			자유도	값	p-value
Deviance			1,237,312	29,991	1.0000
AIC			82,189		

〈표 8〉의 일반대상자 남성 암수술건수 모형 적합결과를 보면, 남성의 경우 혈색소(HGB)는 한 단위 증가할 때 평균 암수술건수는 $0.817(=\exp(-0.2026))$ 배 감소하고, 혈청지오티(SGOT)는 한 단위 증가할 때 평균 암수술건수는 $1.005(=\exp(0.0047))$ 배 증가한다. 요단백(URN_PROT)은 양성 대비 음성인 경우 평균 암수술건수는 $1.166(=\exp(0.1536))$ 배 많은 것으로 나타났다. 수축기혈압(SYS)은 한 단위 증가할 때 평균 암수술건수는 $0.996(=\exp(-0.0042))$ 배 감소하고, 식전공복혈당(FBS)은 한 단위 증가할 때 평균 암수술건수는 $1.001(=\exp(0.0011))$ 배 증가한다.

〈표 9〉 일반대상자 암수술건수(여성) 음이항회귀모형

설명변수		자유도	추정치	표준오차	p-value
Intercept		1	-3.6093	0.1854	$< 2e-16$
AGE	20~29	1	-2.2464	0.1117	$< 2e-16$
	30~39	1	-1.1465	0.0655	$< 2e-16$
	40~44	1	-0.5166	0.0559	$< 2e-16$
	45~49	1	-0.2458	0.0517	0.0000
	50~54	1	-0.2281	0.0515	0.0000
	55~59	1	-0.0883	0.0500	0.0775
	60~64	1	-0.0406	0.0516	0.4313
	65~69	1	-0.0061	0.0560	0.9138
	70~79	0	0.0000	0.0000	-
DIA		1	0.0026	0.0014	0.0625
URN_PROT	0	0	0.0000	0.0000	-
	1	1	-0.2527	0.0560	0.0000
HGB		1	-0.1101	0.0110	$< 2e-16$
FBS		1	0.0026	0.0006	0.0000
SGOT		1	0.0041	0.0008	0.0000
GGT		1	0.0028	0.0004	0.0000
통계량			자유도	값	p-value
Deviance			1,228,689	31,256	1.0000
AIC			91,040		

〈표 9〉의 일반대상자 여성 암수술건수 모형 적합결과를 보면, 혈색소(HGB)는 한 단위 증가할 때 평균 암수술건수는 $0.896(=exp(-0.1101))$ 배 감소하며, 요단백(URN_PROT)은 양성 대비 음성인 경우 평균 암수술건수는 $1.166(=exp(0.2527))$ 배 많은 것으로 나타났다. 식전공복혈당(FBS)은 한 단위 증가할 때 평균 암수술건수는 $1.003(=exp(0.0026))$ 배 증가한다. 혈청지피티(SGPT)는 한 단위 증가할 때 평균 암수술건수는 $1.004(=exp(0.0041))$ 배 증가하고 감마지티피(GGT)는 한 단위 증가할 때 평균 암수술건수는 $1.003(=exp(0.0028))$ 배 증가한다. 이완기혈압(DIA)은 한 단위 증가할 때 평균 암수술건수는 $1.003(=exp(0.0026))$ 배 증가한다.

〈표 10〉은 〈표 8〉과 〈표 9〉의 일반대상자 남성과 여성 각각의 음이항회귀모형 추정값을 사용하여 각 개인들의 암수술 예측 건수를 구한 다음, 남성과 여성 각각의 연령군별 전체 수술건수를 연령별 인원으로 나누어 얻은 암수술 예측률이다.

〈표 10〉 일반대상자 암수술 예측률

연령군	남성		여성	
	대상자수	암수술 예측률	대상자수	암수술 예측률
20~29	94,427	0.000435	106,344	0.000912
30~39	235,286	0.001297	149,721	0.002858
40~44	153,106	0.002030	143,415	0.005699
45~49	159,306	0.002671	156,169	0.007594
50~54	147,983	0.005107	155,883	0.007613
55~59	146,886	0.008337	161,615	0.008753
60~64	120,147	0.012739	135,887	0.009300
65~69	83,249	0.016512	95,430	0.009739
70~74	59,261	0.022886	124,240	0.010021
75~79	37,679	0.023212		

(2) 간편고지대상자

1) 암발생

〈표 11〉과 〈표 12〉는 일반대상자 남성과 여성 각각 암발생유무를 적합하기 위한 최적 로지스틱회귀 예측모형 결과이다. 남성의 경우 연령(AGE), 혈색소(HGB), 장기투약여부(MED), 요단백(URN_PROT), 감마지티피(GGT), 혈청지피티(SGPT), 수축기혈압(SYS)의 요인이 나타났으며, 여성의 경우 연령(AGE), 장기투약여부(MED), 감마지티피(GGT), 혈색소(HGB), 요단백(URN_PROT), 혈청크레아티닌(CRTN), 수축기혈압(SYS)이 선택되었다. 연령은 모든 남성과 여성 모두 연령이 증가할수록 암발생률은 증가하는 경향을 보인다.

〈표 11〉의 간편고지대상자 남성 암발생 모형 적합결과를 보면, 혈색소(HGB)

는 한 단위 증가할 때 암발생 오즈는 $0.883(=\exp(-0.1248))$ 배 감소하며, 장기투약여부(MED)에 따른 암발생 오즈는 장기투약하지 않은 집단 대비 장기투약한 집단이 $1.239(=\exp(0.2145))$ 배 증가한다. 감마지티피(GGT)는 한 단위 증가할 때 암발생 오즈는 $1.001(=\exp(0.0013))$ 배 증가한다. 요단백(URN_PROT)은 양성 대비 음성이 암발생 오즈가 $1.315(=\exp(0.2741))$ 배 높다. 혈청지피티(SGPT)는 한 단위 증가할 때 암발생 오즈는 $1.001(=\exp(0.0008))$ 배 증가하는 것으로 나타났다.

〈표 11〉 간편고지대상자 암발생(남성) 로지스틱회귀모형

설명변수		자유도	추정치	표준오차	p-value
Intercept		1	-1.7743	0.2645	0.0000
AGE	20~29	1	-3.3036	0.1671	< 2e-16
	30~39	1	-2.4870	0.0971	< 2e-16
	40~44	1	-2.0939	0.0986	< 2e-16
	45~49	1	-1.8342	0.0920	< 2e-16
	50~54	1	-1.4239	0.0863	< 2e-16
	55~59	1	-1.0012	0.0810	< 2e-16
	60~64	1	-0.6873	0.0803	< 2e-16
	65~69	1	-0.3025	0.0811	0.0002
	70~74	1	-0.0471	0.0843	0.5762
	75~79	0	0.0000	0.0000	-
SYS		1	-0.0021	0.0012	0.0889
URN_PROT	0	0	0.0000	0.0000	-
	1	1	-0.2741	0.0633	0.0000
HGB		1	-0.1248	0.0141	< 2e-16
SGPT		1	0.0008	0.0003	0.0035
GGT		1	0.0013	0.0002	0.0000
MED	0	0	0.0000	0.0000	-
	1	1	0.2145	0.0375	0.0000
통계량			자유도	값	p-value
Hosmer and Lemeshow 카이제곱			8	3.4654	0.9019
AIC			41,523		

〈표 12〉의 간편고지대상자 여성 암발생 모형 적합결과를 보면, 혈색소(HGB)는 한 단위 증가할 때 암발생 오즈는 $0.958(=\exp(-0.0429))$ 배 감소하며, 장기투약여부(MED)에 따른 암발생 오즈는 장기투약하지 않은 집단 대비 장기투약한 집단이 $1.135(=\exp(0.1269))$ 배 증가한다. 감마지티피(GGT)는 한 단위 증가할 때 암발생 오즈는 $1.002(=\exp(0.0015))$ 배 증가하며, 요단백(URN_PROT)은 양성 대비 음성인 경우 암발생 오즈는 $1.162(=\exp(0.1503))$ 배 높은 것으로 나타났다.

〈표 12〉 간편고지대상자 암발생(여성) 로지스틱회귀모형

설명변수		자유도	추정치	표준오차	p-value
Intercept		1	-4.1339	0.2515	< 2e-16
AGE	20~29	1	-1.5043	0.1209	< 2e-16
	30~39	1	-0.8229	0.1017	0.0000
	40~44	1	-0.4562	0.0977	0.0000
	45~49	1	-0.2644	0.0944	0.0051
	50~59	1	-0.2318	0.0889	0.0091
	60~69	1	-0.1303	0.0898	0.1469
	70~74	1	-0.0044	0.1048	0.9668
	75~79	0	0.0000	0.0000	-
SYS		1	0.0018	0.0011	0.0978
URN_PROT	0	0	0.0000	0.0000	-
	1	1	-0.1503	0.0671	0.0252
HGB		1	-0.0429	0.0130	0.0010
CRTN		1	-0.2210	0.1048	0.0350
GGT		1	0.0015	0.0004	0.0007
MED	0	0	0.0000	0.0000	-
	1	1	0.1269	0.0351	0.0003
통계량			자유도	값	p-value
Hosmer and Lemeshow 카이제곱			8	4.8603	0.7727
AIC			50,608		

〈표 13〉은 〈표 11〉과 〈표 12〉의 간편고지대상자 남성과 여성 각각의 로지스틱회귀모형의 추정값을 사용하여 각 개인들의 암발생 예측값을 구한 다음, 남성과 여성 각각의 연령군별 평균을 계산하여 얻은 예측 암발생률이다.

〈표 13〉 간편고지대상자 암발생 예측률

연령군	남성		여성	
	대상자수	암수발생 예측률	대상자수	암수발생 예측률
20~29	77,577	0.000580	78,758	0.001905
30~39	185,071	0.001372	103,980	0.003818
40~44	106,288	0.002117	90,043	0.005631
45~49	105,538	0.002833	94,100	0.006918
50~54	91,611	0.004388	173,020	0.007248
55~59	82,638	0.006885		
60~64	58,672	0.009681	103,821	0.008274
65~69	34,564	0.014640		
70~74	20,264	0.019345	25,049	0.009621
75~79	10,819	0.021074	15,284	0.009749

2) 암수술건수

〈표 14〉와 〈표 15〉는 간편고지대상자 남성과 여성 각각의 암수술건수를 적합하기 위한 최적 음이항회귀모형 결과이다. 남성의 경우 연령(AGE), 혈색소(HGB), 혈청지오티(SGOT), 감마지티피(GGT), 장기투약여부(MED), 이완기혈압(DIA), 요단백(URN_PROT), 수축기혈압(SYS)이 설명요인으로 선택되었으며, 여성의 경우 연령(AGE), 혈색소(HGB), 감마지티피(GGT), BMI, 혈청지피티(SGPT), 식전공복혈당(FBS), 장기투약여부(MED) 요인이 선택되었다. 연령은 모든 남성과 여성 모두 연령이 증가할수록 암수술건수는 많아지는 경향을 보인다.

〈표 14〉의 일반대상자의 남성 암수술건수 모형 적합결과를 보면, 혈색소(HGB)는 한 단위 증가할 때 평균 암수술건수는 $0.845(=\exp(-0.1687))$ 배 감소하고, 혈청지오티(SGOT)는 한 단위 증가할 때 평균 암수술건수는 $1.004(=\exp(0.0040))$ 배 증가한다. 감마지티피(GGT)는 한 단위 증가할 때 평균 암수술건수

는 $1.001(=\exp(0.0012))$ 배 증가하고, 이완기혈압(DIA)은 한 단위 증가할 때 평균 암수술건수는 $1.007(=\exp(0.0068))$ 배 증가하는 것으로 나타났다.

〈표 14〉 간편고지대상자 암수술(남성) 음이향회귀모형

설명변수		자유도	추정치	표준오차	p-value
Intercept		1	-2.0107	0.3223	0.0000
AGE	20~29	1	-3.6150	0.2152	$< 2e-16$
	30~39	1	-2.4573	0.1179	$< 2e-16$
	40~44	1	-2.1696	0.1203	$< 2e-16$
	45~49	1	-1.9834	0.1151	$< 2e-16$
	50~54	1	-1.3877	0.1075	$< 2e-16$
	55~59	1	-0.9526	0.1029	$< 2e-16$
	60~64	1	-0.6532	0.1026	0.0000
	65~69	1	-0.3106	0.1053	0.0032
	70~74	1	-0.0620	0.1106	0.5752
	75~79	0	0.0000	0.0000	-
SYS		1	-0.0032	0.0021	0.1308
DIA		1	0.0068	0.0031	0.0287
URN_PROT	0	0	0.0000	0.0000	-
	1	1	-0.1505	0.0816	0.0650
HGB		1	-0.1687	0.0172	$< 2e-16$
SGOT		1	0.0040	0.0010	0.0000
GGT		1	0.0012	0.0003	0.0002
MED	0	0	0.0000	0.0000	-
	1	1	0.1129	0.0449	0.0119
통계량			자유도	값	p-value
Deviance			1,084,054	14,213	1.0000
AIC				43,168	

〈표 15〉의 일반대상자의 여성 암수술건수 모형 적합결과를 보면, 혈색소(HGB)는 한 단위 증가할 때 평균 암수술건수는 $0.912(=\exp(-0.0926))$ 배 감소하고, 감마티피페(GGT)는 한 단위가 증가할 때 평균 암수술건수는 $1.002(=\exp(0.0022))$ 배 증가한다. BMI는 한 단위 증가할 때 $1.017(=\exp(0.0168))$

배 증가하고, 혈청지피티(SGPT)는 한 단위 증가할 때 평균 암수술건수는 1.002($=\exp(0.0027)$)배 증가한다. 식전공복혈당(FBS)은 한 단위 증가할 때 평균 암수술건수는 1.002($=\exp(0.0020)$)배 증가하며, 장기투약여부(MED)에 따른 평균 암수술건수는 장기투약하지 않은 집단 대비 장기투약한 집단이 1.109($=\exp(0.1032)$)배 높게 나타났다. 요단백(URN_PROT)은 양성 대비 음성인 경우 평균 암수술건수는 1.174($=\exp(0.1602)$)배 많은 것으로 나타났다.

〈표 15〉 간편고지대상자 암수술(여성) 음이항회귀모형

설명변수		자유도	추정치	표준오차	p-value
Intercept		1	-4.4502	0.2469	$< 2e-16$
AGE	20~29	1	-1.9454	0.1457	$< 2e-16$
	30~39	1	-1.0241	0.1097	$< 2e-16$
	40~44	1	-0.3807	0.1016	0.0002
	45~54	1	-0.1498	0.0925	0.1053
	55~59	1	-0.0723	0.0965	0.4534
	60~64	1	-0.0096	0.0980	0.9219
	65~74	1	0.0166	0.0960	0.8624
	75~79	0	0.0000	0.0000	-
BMI		1	0.0168	0.0052	0.0012
URN_PROT	0	0	0.0000	0.0000	-
	1	1	-0.1602	0.0764	0.0361
HGB		1	-0.0941	0.0143	0.0000
FBS		1	0.0020	0.0008	0.0095
SGPT		1	0.0027	0.0009	0.0027
GGT		1	0.0022	0.0005	0.0000
MED	0	0	0.0000	0.0000	-
	1	1	0.1032	0.0396	0.0092
통계량			자유도	값	p-value
Deviance			1,066,111	18,698	1.0000
AIC			-57,191		

〈표 16〉은 〈표 14〉와 〈표 15〉의 간편고지대상자의 남성과 여성 각각의 음이 항회귀모형 추정값을 사용하여 각 개인들의 암수술발생 예측 건수를 구한 다음, 남성과 여성 각각의 연령군별 전체 수술건수를 연령별 인원으로 나누어 얻은 암수술 예측률이다.

〈표 16〉 간편고지대상자 암수술 예측률

연령군	남성		여성	
	대상자수	암수술 예측률	대상자수	암수술 예측률
20~29	93,374	0.000299	105,165	0.000789
30~39	229,824	0.001004	145,438	0.002077
40~44	146,262	0.001401	135,712	0.004185
45~49	148,379	0.001751	282,109	0.005433
50~54	131,705	0.003278		
55~59	123,116	0.005209	135,973	0.005956
60~64	92,991	0.007211	107,874	0.006480
65~69	8,744	0.010432	120,006	0.006852
70~74	37,562	0.013826		
75~79	22,114	0.015330		
			33,849	0.006892

(3) 일반대상자 대비 간편고지대상자의 상대위험도

〈표 17〉은 암발생과 암수술건수에 대한 일반대상자 대비 간편고지대상자의 상대위험도이다. 암발생 상대위험도는 각 연령군에서 간편고지대상자의 암발생 예측률을 일반대상자의 암발생 예측률로 나눈 값으로 정의하였고, 암수술 상대위험도는 각 연령군에서 간편고지대상자의 암수술 예측률을 일반대상자의 암수술 예측률로 나눈 값으로 정의하였다. 암발생 상대위험도와 암수술건수의 상대위험도는 모두 1보다 크게 작아 간편고지대상자의 위험도는 일반대상자 보다 상대적으로 낮음을 보여준다. 본 연구모형에서 제시한 가설인 “간편고지대상자는 일반대상자 대비 암발생과 암수술건수의 상대위험도가 1보다 작을 것이다.”는 채택된다.

〈표 17〉 일반대상자 대비 간편고지대상자 상대위험도

연령군	암발생		암수술건수	
	남성	여성	남성	여성
20~29	0.709046	0.821475	0.687356	0.865132
30~39	0.757592	0.784306	0.774094	0.726732
40~44	0.740469	0.699677	0.690148	0.734339
45~49	0.726038	0.740844	0.65556	0.715433
50~54	0.688206	0.594790	0.641864	0.713648
55~59	0.673613	0.587835	0.624805	0.680452
60~64	0.670615	0.549716	0.566057	0.696774
65~69	0.687292	0.595895	0.631783	0.703563
70~74	0.658486	0.619590	0.604125	0.683764
75~79	0.669207	0.627834	0.660434	0.687756

가입조건이 간소한 간편고지대상자가 일반대상자보다 암발생 및 암수술건수 상대위험도가 낮다는 것은 보험가입이 어려운 유병자들까지 포함하여 보험가입 대상의 확대되더라도 적절한 위험관리가 된다는 사례로 큰 의미가 있다 할 것이다.

IV. 결론 및 시사점

질병발생의 위험도를 측정하는 모형은 미국의 프레밍햄 연구가 전 세계적으로 상당히 오랫동안 사용되어 왔으나 그 결과가 유전적 배경 또는 환경이 다른 타 민족이나 타인종에게 적용하는 것은 문제가 있어 많은 국가들이 자체적인 모형 개발을 진행하였으며, 국내는 1990년대 후반부터 이와 유사한 모형을 개발하여 발표하고 있다.

본 연구는 2015년부터 2019년까지의 5년간의 국민건강보험공단에서 건강검진을 받은 코호트 대상자에 기반하여 일반대상자와 간편고지대상자에 대하여 남

녀 성별에 대하여 기존의 건강예측모형 연구에서 검증된 건강상태를 나타내는 BMI, 혈압 등의 설명변수들을 기반으로 각각 로지스틱회귀모형과 음이항회귀모형을 이용하여 암 발생률과 암수술건수를 예측하였다,

최종적으로 일반대상자와 간편고지대상자의 건강검진에 따른 각각 암발생과 암수술건수 담보별 통계적 모형에 기반하여 일반대상자 대비 간편고지대상자의 연령별 암발생 상대위험도와 암수술건수 상대위험도를 객관적으로 측정하였다. 암발생 상대위험도와 암수술건수 모두 간편고지대상자가 일반대상자 대비 상대위험도가 1보다 작게 나타났다. 이것은 가입조건을 완화한 간편고지만으로 유병자들을 포함한 보험가입대상자들의 상품영역 확대의 가능성을 보여주며, 또한 간편고지만으로도 암발생과 암수술건수 담보의 상대위험도가 일반대상자 대비 확실히 구분됨을 실증적으로 보여주었다.

본 연구의 결과를 통한 기대효과 및 시사점은 다음과 같다.

먼저, 국민건강보험 표본코호트2.0DB를 활용할 경우, 정교한 위험분석을 통한 차등적 보험료 부과가 가능하기에 가입자의 역선택 위험을 감소시켜 보험회사의 재무건전성 향상에도 도움이 될 것이며, 가입자간 보험료 부담의 형평성이 재고됨으로써 선량한 가입자의 부당한 보험료 부담 전가가 해소될 수 있을 것이다.

일반대상자 대비 간편고지대상자의 암발생 및 암수술건수의 상대위험도 비교를 통한 간편고지 상품의 정확한 평가는 보험사 입장에서 보험대상 확대를 가능하게 하고 보험소비자 입장에서는 보험가입이 어려운 유병자들의 보험혜택을 받을 수 있을 것으로 사료된다.

또한 국민건강보험 빅데이터와 같은 공공의료데이터의 이용 활성화는 기존상품의 시장 확대뿐만 아니라 사회적 변화 및 질병패턴 변화에 대응한 신규보험상품 개발에도 기여할 수 있을 것이다. 그리고 보험회사의 보험시장 확대 및 다양한 보험상품 개발이 이뤄진다면, 보험소비자의 다양한 위험에 대한 보장 서비스가 가능한 동시에 소비자의 보험상품 선택권은 증가될 수 있을 것이다.

〈참고문헌〉

- 김석영. “민간 암보험의 사회적 역할과 발전 방향”, 제2회 보험법 포럼 (2020).
- 이찬희, 전희주, 김석영. “암보험 언더라이팅 효과 실증분석 - 국민 암 발생률과 보험회사 암발생률 비교.” 보험학회지 102.15 (2015): 59-85.
- 권혁성, 오피재, 강민용, 우경석. “건강보험통계 기반 건강 수준 평점화 모형을 활용한 간편고지 상품 유형별 보험요율 차등화에 관한 연구.” 리스크관리연구 32.4 (2021): 99-147.
- 박복희, 전희주. “음이항회귀모형을 이용한 보험고객과 설계사의 관계설정이 보험 구매에 미치는 영향.” 보험금융연구 93.4 (2018a): 91-115.
- 박복희, 전희주. “제3보험의 가입요인: 간병보험을 중심으로.” 리스크관리연구 29.3 (2018b): 77-106.
- 전희주, 최경진. “국민건강보험 빅데이터를 활용한 치아보험 요율 차등화에 관한 연구: 치아우식증, 치주질환을 중심으로.” 보험금융연구 33.3 (2022): 119-147.
- 전희주, 최경진. “생체연령을 고려한 퇴직연금 프로그램 인출방식에 관한 연구.” 리스크관리연구 32.2 (2021): 101-134.
- 황지연, 이상엽, 주석훈. “건강보험통계를 활용한 간편고지 상품의 담보별 최적 언더라이팅 기준 산출.”, 보험학회지 124.20 (2020): 1-34.
- 보건복지부, 2020년 국가암등록통계, (2021).
- Alfayez, A. A, Kunz, H., and Lai A. G.. “Predicting the risk of cancer in adults using supervised machine learning:a scoping review.” BMJ Open 11 (2021): 1-11.
- Britt, K. L., Cuzick, J., and Phillips, K. A.. “Key steps for effective breastcancer prevention.” Nat Rev Cancer 20 (2020): 417-36.
- Christodoulou, E., Ma, J., Collinsb, G. S., Steyerberg, E. W., Verbakela,e,

- JU. Y., and Calst, B. V.. “A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models.” *Journal of Clinical Epidemiology* 110 (2019): 12-22.
- Colditz, G. A., Atwood, K. A., Emmons K, Monson, R. R., Willett, W. C., Trichopoulos, D., and Hunter, D. J.. “Harvard report on cancer prevention volume 4: Harvard Cancer Risk Index. Risk Index Working Group, Harvard Center for Cancer Prevention.” *Cancer Causes Control* 11 (2000): 477-488.
- Gray, E. P., Teare, M. D., Stevens, J., and Archer, R.. “Risk prediction models for lung cancer: a systematic review.” *Clin Lung Cancer* 17 (2016): 95-106.
- Kakushadze Z, Raghubanshi R, and Yu W.. “Estimating cost savings from early cancer diagnosis.” *Data* 2 (2017):30.
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., and Fotiadis, D. I., “Machine learning applications in cancer prognosis and prediction.” *Computational and Structural Biotechnology Journal* 13 (2015): 8-17.
- Siegel R. L., Miller, K. D., Fuchs, H. E., and Jemal, A., “Cancer Statistics, 2021” *CA: A Cancer Journal for Clinicians* 71 (2021): 7-33.
- Stark, G. F., Hart, G. R., Nartowt, B. J., and Deng, J. “Predicting breast cancer risk using personal health data and machine learning models.” *PLoS ONE*, 14.12 (2019): 1-17.
- Usher-Smith, J., Emery, J., Hamilton, W., Griffin, S. J., and Walter, F. M.. “Risk prediction tools for cancer in primary care.” *British Journal of Cancer* 113 (2015): 1645-1650.

- Wang, Y. Zhu, M, Ma, H. and Shen, H.. “Polygenic risk scores: the future of cancer riskprediction, screening, and precision prevention,” Med. Rev. 1.2 (2022)L 129-149.
- Win, A. K., Macinnis, R. J., Hopper, J. L., and Jenkins, M. A.. “Risk prediction models for colorectal cancer: a review.” Cancer Epidemiol, Biomarkers Prev 21 (2012): 398-410.

〈Abstract〉

This study predicted cancer incidence rate and the number of cancer surgeries within the next one year for general subjects and simple notification subjects using the National Health Insurance Service Cohort2.0 DB. Relative risk of cancer occurrence and the number of cancer surgeries between simple notification subjects and general subjects were calculated. For the general subjects, health checkup items such as BMI and blood pressure were used as explanatory variables, and health checkup items, long-term medication for the previous year, and treatment for three diseases were included as explanatory variables for the simple notification subjects. A logistic regression model was used to predict cancer occurrence, and a negative binomial regression model was used to predict the number of cancer surgeries.

As a result of calculating the relative risk of cancer occurrence and the number of cancer surgeries between the general subjects and the simple notification subjects, each relative risk of cancer occurrence and number of cancer surgeries in both men and women was less than 1 for the simple notification subjects compared to the general subjects. This shows the possibility of expanding the product range for those subject to insurance, including patients with pre-existing diseases, with only simplified notices that ease the subscription conditions. In addition, it empirically demonstrated that each relative risk of cancer occurrence and the number of cancer surgeries for simple notification subjects was clearly distinguished from that of general subjects.

Keywords: Cohort DB, health checkup, generalized linear model, the relative risk, cancer occurrence, the number of cancer surgeries

