

간편고지보험 고지항목별 무사고기간에 따른 암 발생 및 치료 상대위험도 예측*

Predicting Relative Risk of Cancer Occurrence and Treatment Based on Simplified Issue Disclosure Period

전 희 주** (Heuiju Chun)

인 태 교*** (Taekyo Leen)

〈목 차〉

- I. 서 론
- II. 연구배경 및 연구방법
- III. 실증분석
- IV. 결론 및 시사점

〈국문초록〉

본 연구는 표본코호트2,0DB를 활용하여 2002~2019년도에 건강보험 자격을 유지하고 있는 사람을 대상으로 2017년 초 시점을 기준으로 연령군 및 중대질병1~9, 입원, 수술, 계속투약, 계속치료 무사고기간을 1년 단위로 측정한 설명변수에 기반하여 로지스틱회귀모형을 이용하여 성별 암 발생 및 암 치료 여부를 예측하고자 한다.

무사고기간 증가에 따른 암 발생 상대위험도를 산출한 결과, 남성의 경우 협심증, 간경화증, 고혈압, 당뇨, 입원, 수술, 계속투약, 계속치료 무사고기간이 증가함에 따라, 여성의 경우에는 간경화증, 고혈압, 당뇨, 계속투약, 계속치료 무사고기간이 증가함에 따라 상대위험도가 감소하는 것으로 나타났다. 무사고기간 증가에 따른 암 치료 상대위험도를 산출한 결과, 남성의 경우 간경화증, 고혈압, 당뇨, 계속투약, 계속치료 무사고기간이 증가함에 따라, 여성의 경우 간경화증, 고혈압, 당뇨, 수술, 계속투약, 계속

* 제1저자, 동덕여자대학교 정보통계학과 부교수 (E-mail: hjchun@dongduk.ac.kr)

** 교신저자, 한양대학교 박사과정 (tkleen@naver.com)

투고일: 2024.01.29, 심사시작일: 2024.01.31, 최종수정일: 2024.04.01, 게재확정일: 2024.04.26.

치료 무사고기간이 1년 증가함에 따라 상대위험도가 감소하는 것으로 나타났다.

본 연구를 통해 전체 고지항목 조합의 경우의 수를 포괄하는 일반화된 모형을 수립했다는 점에 의의가 있다. 또한 암 발생 및 암 치료에 유의한 영향을 끼치는 각 고지항목의 무사고기간을 식별할 수 있고, 각 고지항목의 무사고기간이 1년 변화함에 따라 나타나는 상대위험도 변화량을 실증적으로 산출하여 제시하였다.

국문색인어 : 상대위험도, 암발생, 암치료, 일반화선형모형, 표본코호트2.0DB

〈ABSTRACT〉

This study calculated the relative risks of cancer incidence and cancer treatment according to the duration of accident-free periods for various SI (simplified issue) notification items. Using the National Health Insurance Service National Sample Cohort (NHIS-NSC) data, individuals maintaining health insurance eligibility from 2002 to 2019 were selected. The study measured accident-free periods according to age group, major diseases 1-9, hospitalization, surgery, continuous medication, and continuous treatment, based on the starting point in early 2017, using one-year intervals as explanatory variables. Logistic regression models were employed to predict gender-specific cancer incidence and cancer treatment.

Results of the calculated relative risks for cancer incidence with increasing accident-free periods indicated that, for males, the risk for cancer incidence decreased with an increase in accident-free periods related to angina, liver cirrhosis, hypertension, diabetes, hospitalization, surgery, continuous medication, and continuous treatment. For females, the risk for cancer incidence decreased with an increase in accident-free periods related to liver cirrhosis, hypertension, diabetes, continuous medication, and continuous treatment. Regarding the relative risks of cancer treatment with increasing accident-free periods, for males, the risk of cancer treatment decreased with an increase in accident-free periods related to liver cirrhosis, hypertension, diabetes, continuous medication, and continuous treatment. For females, the risk of cancer treatment decreased with an increase in accident-free periods related to liver cirrhosis, hypertension, diabetes, surgery, continuous medication, and continuous treatment.

This study is significant in that it established a generalized model that

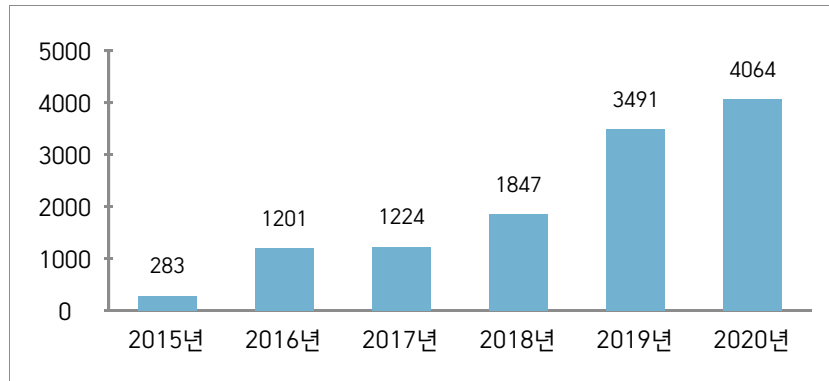
encompasses the number of cases of combinations of all notification items. In addition, this study identified the effect of accident-free periods for each SI notification item on cancer incidence and cancer treatment. It empirically showed that the changes in relative risks associated with a one-year increase in accident-free periods for each notification item can be calculated.

Key Words : Cancer Incidence, Cancer Treatment, Generalized Linear Model, NHIS Cohort2.0DB, Relative Risk

I. 서 론

현재 우리나라는 빠른 속도로 고령화되고 있으며, 통계청이 발표한 ‘2022 통계청 고령자 통계’에 따르면 2025년에는 65세 이상 인구가 전체의 20%를 넘어설 전망이다. 이는 의료기술의 발달과 더불어 질병을 가진 상태로 살아가는 유병인구의 증가를 일으킬 것으로 전망된다. 그러나 대부분의 유병자는 표준체를 대상으로 한 기존 보험상품에 가입하기 어렵다. 이로 인해 유병인구의 증가는 간편고지보험 시장의 확대를 가져왔다. <그림 1>에 의하면, 2020년 간편고지보험 신계약 건수는 406만 건으로 2015년 28만 건 대비 약 14배 증가한 것을 확인할 수 있다.

간편고지보험은 표준청약서 대비 고지항목을 축소한 상품으로, 고지항목을 통상 3개 정도로 제한하여 질병을 갖고 있더라도 해당 고지항목에 모두 해당되지 않으면 가입이 가능하도록 개발한 상품이다. 가장 전통적인 간편고지보험은 이른바 “3.2.5 보험”으로 “1) 3개월 이내 입원, 수술 소견 2) 2년 이내 입원, 수술 여부 3) 5년 이내 중대질병이력”의 3가지 고지항목에 대하여 질문하는 형태로 주로 판매되었다. 그러나 최근에는 기존의 3.2.5 보험 형태에서 고지기간의 변경, 중대질병 종류의 변경 등 고지항목의 세부적인 내용이 변형된 다양한 상품들이 출시되고 있다.



〈그림 1〉 연도별 간편고지보험 신계약 건수 현황 (단위: 천 건)

(자료출처: 보험개발원)

보험회사가 특정한 가입대상 집단을 설정하고, 해당 집단의 상대위험도를 파악하는 것은 적절한 보험료 수준을 결정하는 데에 있어 필수적이며, 수익성과 위험관리에 있어 핵심적인 요소이다. 그러나 간편고지보험 가입 집단의 질병 발생 상대위험도를 “입원, 수술, 치료, 투약, 중대질병 이력” 등 각 고지항목 요소별로 세분화하여 심층적으로 분석한 연구는 아직 제한적인 것으로 보인다.

간편고지 대상이 되는 질병 중에서도 암은 나머지 질병에 비해 큰 사회적 비용을 초래한다. 현경래 외(2017)에 의하면, 2015년 기준 암으로 인한 사회경제적 비용은 18조 3,369억 원에 달하며, 2006년 이후 2015년까지 사회경제적 비용의 연평균 증가율은 4.9%에 달한다. 따라서 암 발생 및 치료여부에 대한 예측모형을 제시하는 것은 사회적 비용의 예방 측면에서 매우 중요하다.

따라서 본 연구에서는 간편고지보험의 각 고지항목 요소가 암 발생률 및 치료 발생률에 미치는 영향에 대한 모형을 수립하고자 한다. 이를 통해 각 고지항목의 구성요소가 보험회사의 위험에 어떠한 영향을 미치는지에 대한 통찰력을 얻고, 고령화로 변화하는 인구 구조에 대응하는 데 기여할 것으로 기대된다.

Ⅱ. 연구배경 및 연구방법

1. 연구배경

(1) 간편고지보험

전통적인 간편고지보험인 ‘3.2.5 보험’ 외에도 최근에는 고지기간 및 고지대상을 변경한 다양한 형태의 간편고지보험이 출시되고 있다. 삼성생명은 2020년 7월에 보험업계 최초로 “입원, 수술여부”에 대한 질의 대신 “6일이상 입원여부, 30일이상 투약여부”를 질의하는 상품을 출시했다. 이에 따라 장기 입원 및 장기투약이 없는 유병자가 가입할 수 있는 간편고지보험이 등장하였으며, 해당 상품에 대해 배타적 사용권 6개월을 부여받았다.

2022년 2월 삼성화재는 입원, 수술 이력에 대한 고지기간을 2년에서 5년으로 변경한, 이른바 ‘3.5.5 보험’을 출시했다. 고지기간이 길어짐에 따라 가입자 집단의 상대위험도를 기존 ‘3.2.5 보험’에 비해 낮출 수 있었고, 기존 간편고지보험에 비해 할증 수준을 대폭 낮춰 선풍적인 인기를 끌었다.

2022년 7월 DB손해보험에서는 입원, 수술 이력에 대한 고지기간을 1년부터 5년으로 다양화하여 하나의 상품으로 운영하는, 이른바 ‘3.N.5 보험’을 출시했다. 매년 무사고 시 더 저렴한 보험료의 간편고지보험으로 계약을 전환할 수 있는 기능을 탑재하여 배타적 사용권 6개월을 부여받았다.

2023년 9월 현대해상은 입원, 수술과 중대질병 이력에 대한 고지기간을 모두 1년으로 변경한, 이른바 ‘3.1.1 보험’을 출시했다. 입원, 수술 이력에 대한 고지기간을 축소할 뿐만 아니라 암/뇌졸중증/협심증/심근경색증/심장판막증/간경화증의 6대질병 이력 역시 고지기간을 1년으로 축소하여 기존 ‘3.2.5 보험’에서 가입이 거절되는 소비자 역시 가입할 수 있도록 출시되었다.

이렇듯 현재 간편고지보험은 짧은 시간 내에 입원 및 수술 고지기간, 중대질병의 범위, 중대질병 고지기간 등의 고지사항을 변경해 가입 고객을 세분화하는 방향으로 진화하고 있다. 앞으로도 이처럼 가입자의 필요에 따라 고지사항을 다양화한 간편고지보험이 출시될 것으로 기대된다.

(2) 선행연구

다양한 간편고지 상품의 출시와 성공에 더불어 고지항목과 질병 발생위험 간의 관계를 실증적으로 규명하고자 하는 시도 역시 진행되었다. 심현우 외(2021)는 국민건강보험 표본코호트 DB를 활용하여 보험계약상 청구경력 고지 기간의 간소화 정도에 따라 가입자의 의료 이용이 증가하는지 실증분석을 진행하였으며, 권혁성 외(2021)는 건강 수준 평점화 모형을 적용한 간편고지 상품의 유형별 차등화 방법을 제안하였다. 황지연 외(2020) 역시 국민건강보험 표본코호트 DB를 활용하여 간편고지 상품의 담보별 최적 언더라이팅 기준을 제시하였으며, 백혜연 외(2018)는 6개의 유병자집단을 정의하여 각각에 대한 질병별 상대위험도를 비교하였다.

일반화선형모형을 활용하여 접근한 연구로는 전희주·최경진(2022)이 치아 우식증과 치주질환과 관련된 치아보험 요율 차등화 방법을 제안한 적이 있으며, 전희주·최경진(2023)은 국민건강보험 건강검진코호트DB와 표본코호트DB를 활용한 크라운, 치수치료 담보 치아보험 상품의 상대위험도를 제안하였다. 또한 전희주·인태교(2022)는 암에 대하여 건강검진 이력에 따른 암 발생 및 암 수술 상대위험도를 예측하기 위해 일반화선형모형을 활용하여 연구·분석하였다. 해외에서도 가입자별 상대도를 반영하여 보험료를 산출하기 위해 일반화선형모형을 활용한 방법을 다양한 연구를 통해 제시하였다(De Jong and Heller, 2008; David, 2015; Naufal et al., 2019).

그러나 최근 간편고지 시장의 트렌드인 입원 및 수술 고지기간, 중대질병 고지기간의 다변화를 반영한 통합된 상대위험도 모형은 연구를 통해 제시되지 않은 것으로 보인다. 따라서 본 연구에서는 현재 간편고지보험에서 적용하고 있는 여러 가지 고지항목 요소들에 대하여 각 요소가 실제 암 발생률 및 치료율에 미치는 영향에 대한 모형을 수립하고, 암 발생률 및 치료율에 유의한 영향을 미치는 고지항목 요소들을 판별하고자 한다.

2. 연구데이터

본 연구에서 사용되는 데이터는 국민건강보험공단에서 제공하는 표본코호트 2.0DB를 사용하였다. 표본코호트2.0DB는 국민건강보험공단에서 국민건강정보 DB를 기반으로 구축한 DB로, 2006년 1년간 건강보험 가입자 혹은 의료급여수급권자 자격을 유지한 한국 국적 보유자를 기준으로 약 100만 명을 표본으로 추출하였다. 또한 매년 사망, 이민 등 자격상실(자료 절단)로 인한 표본 수의 자연 감소를 감안해 매년 신생아 표본을 추가함으로써 표본 크기가 유지되도록 하는 반역동적 코호트 구조로 이루어져 있다. 본 연구는 표본코호트2.0DB의 정보 중 출생 및 사망, 건강보험 자격유지, 진료명세서 정보를 활용하였다.

분석대상자는 2007~2019년 기간에 건강보험자격을 유지하고 있는 사람 중 2017년 초 시점에 생존해 있는 사람을 대상으로 하였다. 그중 과거 5년간(2012~2016년)에 암 발생 또는 치료 이력이 있는 경우 분석 대상에서 제외하였다.¹⁾

3. 연구설계

본 연구의 관심 대상인 설명변수는 2017년 초 기준 각 대상자의 군단연령(10세 구간별) 및 일반적인 간편고지보험의 고지항목(입원, 수술, 중대질병1~9진단, 계속해서 7일 이상 치료, 계속해서 30일 이상 투약)별 무사고기간으로 구성된다. 예를 들어, 특정 환자가 2014년도에 수술을 받은 기록이 있고, 2015~2016년도에 수술을 받은 기록이 없다면 수술에 대한 무사고기간은 2년으로 계산된다.²⁾ 또한, 과거 관찰 기간의 제한으로 인해 무사고기간이 10년을 초과하는 경우에도 10년으로 처리하였다.

분석에 사용된 설명변수의 목록 및 조작적 정의는 <표 1>과 같다.

1) 5년간 과거 이력을 확인한 것은 보험사에서 일반적으로 최초발생 정의 시 사용하는 과거 관찰기간을 참고하였음.

2) 입원, 수술 등의 발생 시점은 명세서 최초 발급 일자를 기준으로 하며, 무사고기간은 1년 단위로 계산됨.

〈표 1〉 연령 및 고지항목 무사고기간별 설명변수

설명변수명		영문명
진료데이터 항목	군단연령	MAGE
	중대질병1 무사고 기간	SI1
	중대질병2 무사고 기간	SI2
	중대질병3 무사고 기간	SI3
	중대질병4 무사고 기간	SI4
	중대질병5 무사고 기간	SI5
	중대질병6 무사고 기간	SI6
	중대질병7 무사고 기간	SI7
	중대질병8 무사고 기간	SI8
	중대질병9 무사고 기간	SI9
	입원 무사고 기간	PAT
	수술 무사고 기간	SUR
	계속치료* 무사고 기간	TREAT
	계속투약** 무사고 기간	MED

* 7일 이상 치료(30일 이내 동일상병으로 내원시 동일 치료로 간주)

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

중대질병1: 진료DB명세서 중 진단코드가 협심증에 해당하는 경우

중대질병2: 진료DB명세서 중 진단코드가 급성심근경색증에 해당하는 경우

중대질병3: 진료DB명세서 중 진단코드가 뇌출혈에 해당하는 경우

중대질병4: 진료DB명세서 중 진단코드가 허혈성심장질환에 해당하는 경우

중대질병5: 진료DB명세서 중 진단코드가 기타 뇌혈관질환에 해당하는 경우

중대질병6: 진료DB명세서 중 진단코드가 심장판막증에 해당하는 경우

중대질병7: 진료DB명세서 중 진단코드가 간경화증에 해당하는 경우

중대질병8: 진료DB명세서 중 진단코드가 고혈압에 해당하는 경우

중대질병9: 진료DB명세서 중 진단코드가 당뇨병에 해당하는 경우

입원: 진료DB명세서 중 서식코드가 입원에 해당하는 경우

수술: 진료DB 명세서 중 수술 여부가 수술에 해당하는 경우

계속하여 7일 이상 치료: 동일 상병으로 누적하여 7일 이상 치료한 경우

계속하여 30일 이상 투약: 동일 상병으로 누적하여 30일 이상 투약한 경우

표준체 고지사항에 포함되는 10대 질병 중 암(백혈병 포함), 에이즈에 대한 과거 무사고기간은 설명변수에서 제외하였으며, 허혈성심장질환 및 기타 뇌혈관질환은 별도로 추가하였다. 암에 대한 과거 무사고기간은 본 연구주제인 간편고지 분석대상인 암 발생과 암치료의 최초발생 조건에 해당되어 설명변수에서 제외하였다. 백혈병의 경우 암과 속성이 유사하여 제외하였으며, 에이즈의 경우 표본코호트DB에서 민감정보로 분류되어 마스킹코드로 분류되어 과거 무사고기간을 측정할 수 없어 분석 대상에서 제외하였다. 또한, 일반적으로 표준체 고지사항에서는 뇌, 심장 혈관 관련 질환으로 협심증, 급성심근경색증, 뇌졸중증(뇌출혈, 뇌경색)이 포함되나 본 연구에서는 해당 질병을 보수적으로 판단하여 허혈성심장질환 및 기타 뇌혈관질환의 범위까지 설명변수에 포함시켰다.

위 설명변수에서 사용되는 중대질병은 제8차 한국표준질병·사인분류(KCD 코드)를 기준으로 정의하였으며, 그 범위는 다음 표와 같다.

〈표 2〉 협심증 분류표

대상질병	분류번호
협심증	I20

〈표 3〉 급성심근경색증 분류표

대상질병	분류번호
급성 심근경색증	I21
후속심근경색증	I22
급성 심근경색증 후 특정 현존 합병증	I23

〈표 4〉 뇌출혈 분류표

대상질병	분류번호
거미막하출혈	I60
뇌내출혈	I61
기타 비외상성 두개내출혈	I62
뇌경색증	I63

〈표 5〉 허혈성심장질환 분류표

대상질병	분류번호
기타 급성 허혈심장질환	I24
만성 허혈심장병	I25

〈표 6〉 기타 뇌혈관질환 분류표

대상질병	분류번호
출혈 또는 경색증으로 명시되지 않은 뇌졸중	I64
뇌경색증을 유발하지 않은 뇌전동맥의 폐쇄 및 협착	I65
뇌경색증을 유발하지 않은 대뇌동맥의 폐쇄 및 협착	I66
기타 뇌혈관질환	I67
달리 분류된 질환에서의 뇌혈관장애	I68
뇌혈관질환의 후유증	I69

〈표 7〉 심장판막증 분류표

대상질병	분류번호
류마티스성 승모판질환	I05
류마티스성 대동맥판질환	I06
류마티스성 삼첨판질환	I07
다발판막질환	I08
비류마티스성 승모판장애	I34
비류마티스성 대동맥판장애	I35
비류마티스성 삼첨판장애	I36
폐동맥판장애	I37
상세불명 판막의 심내막염	I38
달리 분류된 질환에서의 심내막염 및 심장판막장애	I39
인공심장판막의 기계적 합병증	T820
인공심장판막에 의한 감염 및 염증반응	T826
인공심장판막의 존재	Z952
이종심장판막의 존재	Z953
기타 심장판막대치물의 존재	Z954

〈표 8〉 간경화증 분류표

대상질병	분류번호
간의 섬유증 및 경변증	K74
알코올성 간섬유증 및 간의 경화증	K702
알코올성 간경변증	K703

〈표 9〉 고혈압 분류표

대상질병	분류번호
본태성(원발성) 고혈압	I10

〈표 10〉 당뇨병 분류표

대상질병	분류번호
당뇨병	E10-E14

본 연구의 관심대상인 반응변수는 〈표 11〉과 같이 정의되며, Y1과 Y2는 미 발생시 0, 발생시 1의 값을 가지는 이분변수(binary variable)로 측정된다. Y2의 경우 암 치료 여부를 판단하기 위해 암 발생 이후 1년의 기간을 추가로 관찰해야 한다. 표본코호트2.0DB의 최대 관측기간이 2019년도까지이기 때문에 Y1과 Y2 모두 암 발생에 대한 관찰기간을 2017~2018년으로 설정했다.

〈표 11〉 반응변수의 정의

반응변수명	정의
Y1	2017~2018년도 내 암 발생여부
Y2	2017~2018년도 내 암이 발생했으며, 발생 이후 1년 이내에 암 치료 발생여부

Y1과 Y2에서의 ‘암 발생’은 보험사에서 일반적으로 최초발생 정의시 사용하는 정의를 준용하여, 직전 5년간 암으로 진료받지 않은 대상이 암으로 진료받는 것으로 정의한다. 단, ‘암 진료’란 진료명세서에 기재된 질병코드가 〈표 12〉

암 분류표에 포함되는 경우를 말하며, 해당 분류는 제8차 한국표준질병·사인 분류를 기준으로 정의하였다.

〈표 12〉 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 관련 조직의 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

Y2에서 ‘치료’의 기준은 ‘암 수술’, ‘항암 약물치료’, 또는 ‘항암 방사선치료’를 받은 것으로 정의하였으며, 암 치료 및 경과기간에 대한 세부정의는 〈표 13〉과 같다.

〈표 13〉 암 치료 및 경과기간과 관련된 세부정의

용어	정의
암 수술 ³⁾	주상병이 〈표 12〉의 ‘암’에 해당하며, 진료DB 명세서(20T) 기준 수술(OPRTN_YN=‘9’)에 해당하는 경우
항암 약물치료 ⁴⁾	주상병이 〈표 12〉의 ‘암’에 해당하며, 진료DB 진료내역(30T) 및 처방전 교부상세내역(60T)에서 다음의 약물을 처방/투여한 이력이 있는 경우 1) 약효분류(EFMDC_CLSF_NO)상 421(항악성종양제) 또는 617(주로 악성종양에 작용하는 것)에 해당하는 경우 단, 면역기능증진을 목적으로 하는 상황균사체엑스 ⁵⁾ 단독사용은 제외 2) ATC코드분류상 ‘LOX’(Antineoplastic and immunomodulating agents)에 해당 ⁶⁾ 하는 경우
항암 방사선치료	주상병이 〈표 12〉의 ‘암’에 해당하며, 진료DB 진료내역(30T)상 영상진단 및 방사선치료(CLA_CD=‘10’)로 항암치료(ITEM_CD=‘02’) 이력이 있는 경우 단, 방사선치료계획에 관련한 수가코드 ⁷⁾ 는 제외하여 분석
n년 이내 암 치료 발생	암 발생자가 암 최초 발생 시점부터 $[365*(n-1), 365*n]$ 일에 속하는 시점에 암 치료행위가 있는 경우

4. 통계적 모형

본 연구에서는 설명변수인 각 고지항목별 무사고기간이 암 발생 및 암 치료에 미치는 영향을 분석하고자 한다. 본 연구의 종속변수는 대상자별 암 발생 여부와 암 치료 여부로 0 또는 1의 값을 가지는 이분 변수(binary variable)이기 때문에, 이에 적합한 로지스틱회귀모형을 사용하였다.

-
- 3) 수술 기준은 권용진 외(2020)과 Park et al.,(2021)의 유사한 기준을 준용하였음.
 4) 항암 약물치료 및 항암 방사선치료는 민간 보험에서 정의하는 치료 수준과 유사한 기준을 준용하였음.
 5) 주성분코드 100102ACH, 100101ALQ, 100130ALQ, 100101APD에 해당.
 6) 주성분코드 112401ATB, 134801BIJ, 485501ATB, 485502ATB, 485605ATB, 485606ATB, 568201BIJ, 485607ATB, 558902BIJ, 134830BIJ, 568230BIJ, 112402ATB, 485601ATB, 485602ATB, 485603ATB, 485604ATB, 485503ATB, 134801ATB에 해당.
 7) 행위분류코드(수가코드) HD010, HD011, HD012, HD013, HD014, HD015, HD016, HD017, HD018, HD019, HD020, HD021, HD022, HD023, HD041, HD410, HD411, HD412, HD413, HD414, HD415, HD416, HD418, HD419, HD420, HD441, HD040, HD060에 해당.

로지스틱회귀모형은 일반화선형모형(GLM)의 특수한 형태로, 반응변수 Y_i 가 0 또는 1의 값을 가지는 이항분포를 따르며, 그 모수는 p_i 이다. 평균모수 $p_i = E(y_i|\mathbf{x}_i)$ 와 선형예측식 $\mathbf{x}_i'\boldsymbol{\beta}$ 은 로짓(logit) 연결함수를 통해 다음과 같이 연결된다(전희주, 최경진, 2021, 전희주·최경진, 2022).

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

여기서 $\mathbf{x}_i$ 는 크기가 $(k+1) \times 1$ 인 설명변수의 i 번째 관측값 벡터, $\boldsymbol{\beta}$ 는 크기가 $(k+1) \times 1$ 인 회귀계수 벡터를, k 는 설명변수의 수를 나타낸다.

그러나 범주형 변수를 포함하여 회귀분석의 다중공선성을 진단하고자 하는 경우에는 각 설명변수에 대한 자유도가 달라져, 이를 고려한 조정된 분산팽창계수인 GVIF(generalized variance inflation factor)를 사용해야 한다(Fox and Monette, 1992). 다음과 같은 일반적인 다중회귀모형에서 k 개의 설명변수가 존재한다고 가정하자.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon \quad (2)$$

여기서 (2)의 모형에 범주형 변수가 하나 포함되어 있으며, 처음 r 개의 설명변수는 해당 범주형 변수가 가질 수 있는 값에 대한 지시변수(indicator variable)라 가정하자. 이 경우 (2)의 일반적인 다중회귀모형에 대하여 처음 r 개의 설명변수를 $\mathbf{X}_1$ 으로, 나머지 $k-r$ 개의 설명변수를 $\mathbf{X}_2$ 로 분리하여 (3)과 같이 나타낼 수 있다(단, $k > r$).

$$\mathbf{y} = \beta_0 + \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \epsilon \quad (3)$$

(3)의 모형에서 GVIF는 다음과 같이 정의된다.

$$GVIF_1 = \frac{\det(R_{11})\det(R_{22})}{\det(R)} \quad (4)$$

(4)에서 R_{11} 은 $\mathbf{X}_1$ 에 대한 상관계수행렬을, R_{22} 는 $\mathbf{X}_2$ 에 대한 상관계수행렬을, R 은 $\mathbf{X}$ 에서 상수항을 제외한 행렬에 대한 상관계수행렬을 나타낸다. Fox and Monette(1992)는 GVIF에 대하여 $(GVIF^{1/(2df)})^2$ 값을 일반적인 VIF와 비교할 수 있음을 제시하였다.

Ⅲ. 실증분석

1. 사전분석

본 연구는 <표 1>에서 기술한 설명변수들을 기반으로 암 발생 여부 및 암 치료 발생 여부를 예측하는 로지스틱회귀모형을 제시하고자 한다. 전체 분석대상에 대해 연령군 및 성별 분포, 발생자수 및 치료자수의 분포는 <표 14>와 같다.

<표 14> 분석대상의 연령군 및 성별 대상자수, 발생자수 및 치료자수

연령군	대상자수		발생자수		치료자수	
	남자	여자	남자	여자	남자	여자
10세 미만	5,307	5,050	1	2	—	1
10~19세	59,138	54,505	43	56	22	19
20~29세	73,578	65,449	88	258	43	101
30~39세	78,324	73,332	260	639	160	346
40~49세	88,298	84,129	572	1,290	362	819
50~59세	81,709	78,458	1,260	1,344	797	879
60~69세	48,048	51,342	1,694	1,018	1,134	675
70~79세	24,248	34,259	1,475	978	979	670
80~89세	7,275	16,910	571	522	336	313
90세~	659	2,628	35	75	20	32
계	466,584	466,062	5,999	6,182	3,853	3,855

간편고지보험 고지항목의 경우, 아래의 세 가지 이유로 인해 고지항목 간의 상관관계가 있음을 예상할 수 있다. 첫째, 과거에 중대한 질병으로 진단받은 경험이 있다면, 비슷한 시점에 수술, 입원, 치료 또는 투약을 받았을 가능성이 높다. 둘째, 연령에 따라서 중대질병 발병 가능성이 높아질 것이기 때문에 연령과는 강한 상관관계가 예상된다. 마지막으로, 가입자의 건강상태라는 잠재변수(latent variable)의 존재로 인해 각 중대질병 사이에 양의 상관관계가 나타날 것으로 예상된다. 높은 수준의 상관관계($r > 0.6$) 또는 높은 GVIF값($GVIF > 10$)을 가지는 설명변수가 있으면 다중공선성을 의심할 근거가 된다고 판단하여(Chan, 2003; Montgomery and Peck, 1992) 예측모형 구축에 앞서 설명변수 간 상관관계분석을 시행하고, GVIF를 계산하여 다중공선성의 존재여부를 확인하였다.

〈표 15〉와 〈표 16〉의 상관관계분석 결과, 남성에서는 높은 수준의 상관관계($r > 0.6$)를 가지는 설명변수가 존재하지 않았다. 상관관계의 크기가 가장 큰 두 설명변수는 고혈압과 치료 여부($r = 0.540$)이었으며, 예상한 대로 모든 설명변수 사이에서 양의 상관관계가 관측되었다. 여성에서는 높은 수준의 상관관계($r > 0.6$)를 가지는 설명변수가 존재하지 않았다. 상관관계의 크기가 가장 큰 두 설명변수는 입원과 수술여부($r = 0.508$)였으며, 예상한 대로 모든 설명변수 사이에서 양의 상관관계가 관측되었다. 또한 남성과 여성 모두 모든 상관관계에 대한 상관관계분석을 시행한 결과 p-값 < 0.0001 로 측정되었는데, 대규모 데이터를 활용하여 관측치가 커졌기 때문에 나타난 결과로 해석된다(남성 $n = 461,277$, 여성 $n = 461,012$).

〈표 17〉~〈표 20〉의 VIF분석 결과, 높은 다중공선성($GVIF > 10$)을 보이는 설명변수는 존재하지 않았다. 또한, 4개의 모델 모두 설명변수 간 $GVIF^{1/(2df)}$ 값을 비교하였을 때 이상치는 발생하지 않았다.

〈표 15〉 설명변수의 상관관계행렬(남성)

남성	SI2	SI3	SI4	SI5	SI6	SI7	SI8	SI9	PAT	SUR	TREAT	MED
SI1	0.250	0.122	0.348	0.101	0.066	0.022	0.284	0.195	0.176	0.111	0.234	0.137
SI2	1	0.052	0.313	0.038	0.022	0.005	0.112	0.087	0.100	0.063	0.098	0.065
SI3		1	0.075	0.443	0.035	0.024	0.225	0.157	0.167	0.070	0.182	0.162
SI4			1	0.066	0.049	0.010	0.186	0.136	0.105	0.067	0.140	0.075
SI5				1	0.024	0.016	0.193	0.134	0.107	0.051	0.136	0.106
SI6					1	0.008	0.059	0.028	0.047	0.027	0.054	0.042
SI7						1	0.053	0.074	0.070	0.041	0.088	0.067
SI8							1	0.390	0.191	0.146	0.540	0.215
SI9								1	0.176	0.129	0.387	0.196
PAT									1	0.449	0.277	0.485
SUR										1	0.223	0.337
TREAT											1	0.361
MED												1

〈표 16〉 설명변수의 상관관계행렬(여성)

여성	SI2	SI3	SI4	SI5	SI6	SI7	SI8	SI9	PAT	SUR	TREAT	MED
SI1	0.167	0.140	0.277	0.117	0.088	0.018	0.299	0.202	0.154	0.095	0.215	0.158
SI2	1	0.056	0.174	0.035	0.026	0.005	0.091	0.069	0.060	0.038	0.060	0.052
SI3		1	0.093	0.407	0.048	0.016	0.255	0.170	0.150	0.069	0.176	0.158
SI4			1	0.070	0.054	0.006	0.186	0.131	0.078	0.050	0.111	0.080
SI5				1	0.033	0.009	0.224	0.143	0.102	0.052	0.132	0.111
SI6					1	0.008	0.083	0.037	0.052	0.032	0.067	0.051
SI7						1	0.036	0.042	0.033	0.022	0.055	0.037
SI8							1	0.401	0.199	0.133	0.498	0.285
SI9								1	0.176	0.123	0.339	0.229
PAT									1	0.508	0.291	0.485
SUR										1	0.229	0.348
TREAT											1	0.413
MED												1

〈표 17〉 암 발생(남성) 다중공선성 분석 결과

설명변수	GVI	df	GVI ^(1/2df)
MAGE	1.4947	7	1.0291
SI1	1.2519	1	1.1189
SI2	1.1159	1	1.0564
SI3	1.3208	1	1.1493
SI4	1.2045	1	1.0975
SI5	1.2564	1	1.1209
SI6	1.0112	1	1.0056
SI7	1.0260	1	1.0129
SI8	1.4975	1	1.2237
SI9	1.1907	1	1.0912
PAT	1.6401	1	1.2807
SUR	1.3418	1	1.1584
TREAT	1.6710	1	1.2927
MED	1.4978	1	1.2239

〈표 18〉 암 치료(남성) 다중공선성 분석 결과

설명변수	GVI	df	GVI ^(1/2df)
MAGE	1.4621	7	1.0275
SI1	1.2427	1	1.1148
SI2	1.1121	1	1.0546
SI3	1.3063	1	1.1429
SI4	1.1984	1	1.0947
SI5	1.2493	1	1.1177
SI6	1.0100	1	1.0050
SI7	1.0267	1	1.0133
SI8	1.5003	1	1.2249
SI9	1.1930	1	1.0923
PAT	1.6425	1	1.2816
SUR	1.3468	1	1.1605
TREAT	1.6785	1	1.2956
MED	1.5002	1	1.2248

〈표 19〉 암 발생(여성) 다중공선성 분석 결과

설명변수	GVIF	df	GVIF ^(1/2df)
MAGE	1.9792	7	1.0500
SI1	1.1988	1	1.0949
SI2	1.0420	1	1.0208
SI3	1.2758	1	1.1295
SI4	1.1190	1	1.0578
SI5	1.2191	1	1.1041
SI6	1.0135	1	1.0067
SI7	1.0084	1	1.0042
SI8	1.7293	1	1.3150
SI9	1.2544	1	1.1200
PAT	1.6280	1	1.2759
SUR	1.3800	1	1.1747
TREAT	1.6626	1	1.2894
MED	1.5050	1	1.2268

〈표 20〉 암 치료(여성) 다중공선성 분석 결과

설명변수	GVIF	df	GVIF ^(1/2df)
MAGE	2.0785	7	1.0537
SI1	1.2113	1	1.1006
SI2	1.0510	1	1.0252
SI3	1.2914	1	1.1364
SI4	1.1239	1	1.0602
SI5	1.2234	1	1.1061
SI6	1.0145	1	1.0072
SI7	1.0073	1	1.0037
SI8	1.7548	1	1.3247
SI9	1.2589	1	1.1220
PAT	1.6282	1	1.2760
SUR	1.3738	1	1.1721
TREAT	1.6643	1	1.2901
MED	1.5083	1	1.2281

2. 통계적 모형

암 발생 여부 및 암 치료 여부에 적합한 예측모형으로는 로지스틱회귀모형을 적용하였다. 가장 먼저 반응변수 Y1, Y2에 대해 설명변수들을 로지스틱회귀모형에 적합하여 유의수준 0.05에서 유의한 설명변수들을 선택하였고, 구별된 설명변수들을 대상으로 후진제거법(backward elimination method)을 사용하여 AIC가 가장 작은 모형을 최적모형으로 선택하였다. 로지스틱회귀모형의 경우에는 예측치와 실제 관측치의 차이를 보여주는 적합도 통계량인 Hosmer and Lemeshow 카이제곱 통계량이 $p\text{-값} > 0.05$ 인지를 추가로 확인하였다.

(1) 암 발생 여부 예측 모형

5년(2012~2016년) 동안 암 발생 이력이 없는 사람을 분석 대상으로 정하였다.

1) 남성

〈표 21〉은 20~89세 남성의 암 발생 여부를 적합하기 위한 최적 로지스틱회귀예측모형 결과이다. 남성의 경우 협심증(SI1), 간경화증(SI7), 고혈압(SI8), 당뇨병(SI9), 입원(PAT), 수술(SUR), 계속투약(MED), 계속치료(TREAT)가 설명변수로 선택되었으며, 연령군의 증가는 모든 구간에서 암 발생률을 높이는 방향으로 적합되었다.

설명변수 개별적으로 살펴보면, 모든 설명변수에서 무사고기간이 증가할수록 암 발생 가능성이 적은 것으로 예측되었다. 유의한 영향을 끼친 중대질병에 대하여 협심증(SI1), 간경화증(SI7), 고혈압(SI8), 당뇨병(SI9) 무사고기간이 한 단위(1년) 증가할 때마다 암 발생 오르는 각각 0.9921배, 0.8828배, 0.9921배, 0.9900배로 감소하는 것으로 나타났다. 이와 함께 질병이력 중에서 간경화증 무사고 기간이 $p < 0.0001$ 로 가장 유의한 변수로 나타났으며, 협심증 무사고 기간의 경우 $\alpha = 0.05$ 하에서는 유의하지 않지만 $\alpha = 0.1$ 하에서는 유의한 변수임이 관측되었다.

질병 이력 외에 입원, 수술, 계속투약, 계속치료 무사고기간의 경우 모두 한

단위 증가할 때마다 암 발생 오즈를 낮추는 것으로 관측되었으며, 특히 투약, 입원, 치료이력이 수술이력 대비 더 큰 영향을 미치는 것으로 나타난다.

또한, Hosmer and Lemeshow 카이제곱 통계량 분석 결과 p-값이 0.1166으로 나타나, 모형적합도 검증 기준을 통과한 것을 확인할 수 있었다.

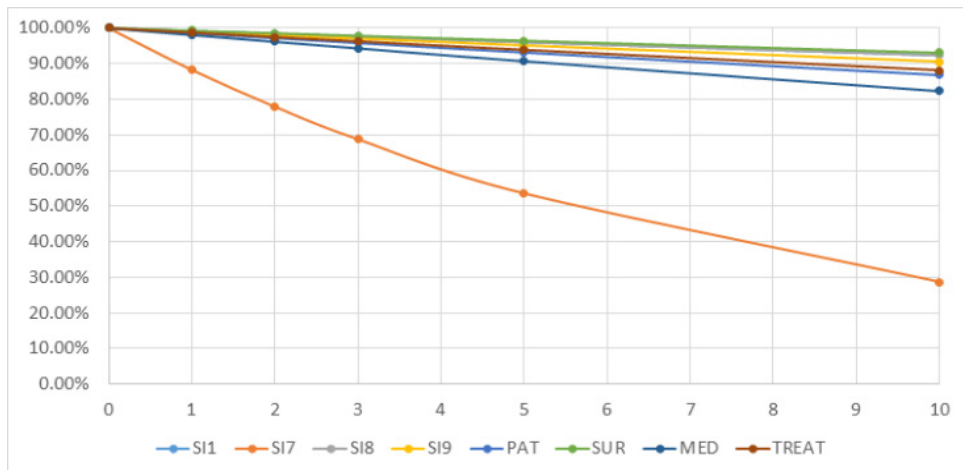
〈표 21〉 암 발생(남성) 로지스틱회귀모형

설명변수		df	β	s.e.	p-value	$\exp(\beta)$
Intercept		1	-0.9809	0.0962	<0.0001	-
연령(AGE)	20~29	1	-4.3558	0.1613	<0.0001	0.0128
	30~39	1	-3.8495	0.1190	<0.0001	0.0213
	40~49	1	-2.8658	0.0802	<0.0001	0.0569
	50~59	1	-2.2661	0.0635	<0.0001	0.1037
	60~69	1	-1.4945	0.0532	<0.0001	0.2244
	70~79	1	-0.7440	0.0500	<0.0001	0.4752
	80~89	1	-0.2325	0.0504	<0.0001	0.7925
	90+	0	0	0	<0.0001	1.0000
협심증(SI1)		1	-0.0079	0.0047	0.0884	0.9921
간경화증(SI7)		1	-0.1247	0.0079	<0.0001	0.8828
계속투약(MED)		1	-0.0193	0.0045	<0.0001	0.9809
입원(PAT)		1	-0.0141	0.0041	0.0005	0.9860
계속치료(TREAT)		1	-0.0126	0.0039	0.0013	0.9875
당뇨병(SI9)		1	-0.0100	0.0034	0.0029	0.9900
고혈압(SI8)		1	-0.0079	0.0034	0.0190	0.9921
수술(SUR)		1	-0.0071	0.0040	0.0758	0.9929
통계량		자유도	값	p-value		
Hosmer and Lemeshow 카이제곱		8	12.865	0.1166		

무사고기간 0년을 기준(100%)으로 했을 때, 위 모델을 바탕으로 기대되는 고지항목 기간별 암 발생 상대위험도는 〈표 22〉과 〈그림 2〉와 같다.

〈표 22〉 무사고기간별 기대 암 발생 상대위험도(%) (남성)

설명변수	1년	2년	3년	5년	10년
간경화증(SI7)	88.28	77.93	68.79	53.61	28.74
계속투약(MED)	98.09	96.21	94.37	90.79	82.43
입원(PAT)	98.60	97.22	95.86	93.19	86.85
계속치료(TREAT)	98.75	97.51	96.29	93.89	88.16
당뇨병(SI9)	99.00	98.02	97.04	95.11	90.46
고혈압(SI8)	99.21	98.42	97.64	96.10	92.36
협심증(SI1)	99.21	98.43	97.65	96.12	92.38
수술(SUR)	99.29	98.58	97.88	96.50	93.12



〈그림 2〉 무사고기간별 기대 암 발생 상대위험도(남성)

2) 여성

〈표 23〉는 20~89세 여성의 암 발생 여부를 적합하기 위한 최적 로지스틱회귀예측모형 결과이다. 여성의 경우 간경화증(SI7), 고혈압(SI8), 당뇨병(SI9), 수술(SUR), 계속투약(MED), 계속치료(TREAT)이 설명변수로 선택되었으며, 연령군의 증가는 모든 구간에서 암 발생률을 높이는 방향으로 적합되었다.

설명변수 개별적으로 살펴보면, 모든 설명변수에서 무사고기간이 증가할수록 암 발생 가능성이 적은 것으로 예측되었다. 유의한 영향을 끼친 중대질병에

대하여 간경화증(SI7), 고혈압(SI8), 당뇨병(SI9) 무사고기간이 한 단위(1년) 증가할 때마다 암 발생 오즈는 각각 0.9107배, 0.9920배, 0.9835배로 감소하는 것으로 나타났다. 이와 함께 질병이력 중에서 간경화증, 당뇨병 무사고기간이 $p < 0.0001$ 로 매우 유의한 변수로 나타났으며, 고혈압 무사고기간 역시 $\alpha = 0.05$ 하에서 유의한 변수임이 관측되었다.

질병 이력 외에 수술, 계속투약, 계속치료 무사고기간의 경우 모두 한 단위 증가할 때마다 암 발생 오즈를 낮추는 것으로 관측되었으며, 특히 수술과 계속투약 무사고기간이 계속치료 무사고기간 대비 더 유의한 변수로 나타났다. 또한, Hosmer and Lemeshow 카이제곱 통계량 분석 결과 p-값이 0.8763으로 나타나, 모형적합도 검증 기준을 통과한 것을 확인할 수 있었다.

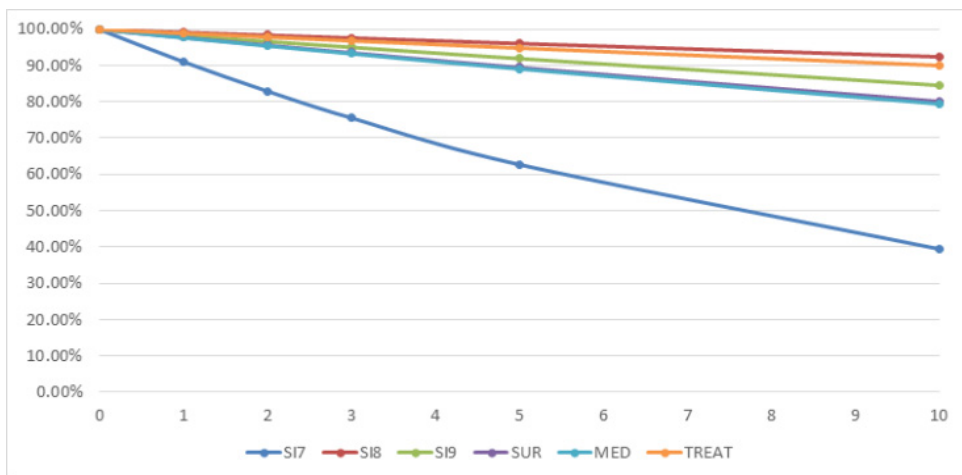
〈표 23〉 암 발생(여성) 로지스틱회귀모형

설명변수		df	β	s.e.	p-value	$\exp(\beta)$
Intercept		1	-2.2740	0.1382	<0.0001	-
연령(AGE)	20~29	1	-3.0064	0.1446	<0.0001	0.0495
	30~39	1	-1.6974	0.0825	<0.0001	0.1832
	40~49	1	-0.9987	0.0652	<0.0001	0.3684
	50~59	1	-0.4279	0.0571	<0.0001	0.6519
	60~69	1	-0.4277	0.0533	<0.0001	0.6520
	70~79	1	-0.3784	0.0534	<0.0001	0.6850
	80~89	1	-0.0847	0.0529	0.1095	0.9188
	90+	0	0	0	<0.0001	1.0000
간경화증(SI7)		1	-0.0935	0.0131	<0.0001	0.9107
계속투약(MED)		1	-0.0233	0.0041	<0.0001	0.9770
수술(SUR)		1	-0.0220	0.0036	<0.0001	0.9782
당뇨병(SI9)		1	-0.0167	0.0038	<0.0001	0.9834
계속치료(TREAT)		1	-0.0106	0.0037	0.0040	0.9895
고혈압(SI8)		1	-0.0081	0.0036	0.0271	0.9919
통계량		자유도	값	p-value		
Hosmer and Lemeshow 카이제곱		8	3.7818	0.8763		

무사고기간 0년을 기준(100%)으로 했을 때, 위 모델을 바탕으로 기대되는 고지항목 기간별 암 발생 상대위험도는 〈표 24〉과 〈그림 3〉과 같다.

〈표 24〉 무사고기간별 기대 암 발생 상대위험도(%) (여성)

설명변수	1년	2년	3년	5년	10년
간경화증(SI7)	91.07	82.94	75.53	62.65	39.24
계속투약(MED)	97.70	95.45	93.25	89.01	79.23
수술(SUR)	97.82	95.69	93.60	89.57	80.22
당뇨병(SI9)	98.35	96.72	95.12	92.01	84.65
계속치료(TREAT)	98.95	97.91	96.88	94.85	89.96
고혈압(SI8)	99.20	98.40	97.61	96.05	92.25



〈그림 3〉 무사고기간별 기대 암 발생 상대위험도(여성)

(2) 암치료여부 예측 모형

1) 남성

〈표 25〉는 20~89세 남성의 암 치료 여부를 적합하기 위한 최적 로지스틱회귀예측모형 결과이다. 남성의 경우 간경화증(SI7), 고혈압(SI8), 당뇨병(SI9), 계속투약(MED), 계속치료(TREAT)이 설명변수로 선택되었으며, 연령군 증가는 모든 구간에서 암 치료율을 높이는 방향으로 적합되었다.

설명변수 개별적으로 살펴보면, 모든 설명변수에서 무사고기간이 증가할수록 암 치료 가능성이 적은 것으로 예측되었다. 유의한 영향을 끼친 중대질병에

대하여 간경화증(SI7), 고혈압(SI8), 당뇨병(SI9) 무사고기간이 한 단위(1년) 증가할 때마다 암 치료 오즈는 각각 0.8846배, 0.9923배, 0.9892배로 감소하는 것으로 나타났다. 더하여 질병이력 중에서 간경화증 무사고기간이 $p < 0.0001$ 로 가장 유의한 변수로 나타났으며, 고혈압의 경우 $\alpha=0.05$ 하에서는 유의하지 않지만 $\alpha=0.1$ 하에서는 유의한 변수임이 관측되었다.

질병 이력 외에 계속투약, 계속치료 무사고기간의 경우 모두 한 단위 증가할 때마다 암 치료 오즈를 낮추는 것으로 관측되었으며, 계속투약, 계속치료 무사고기간이 한 단위(1년) 증가할 때마다 암 치료 오즈는 각각 0.9894배, 0.9848배로 감소하는 것으로 나타났다.

또한, Hosmer and Lemeshow 카이제곱 통계량 분석 결과 p-값이 0.6191로 나타나, 모형적합도 검증 기준을 통과한 것을 확인할 수 있었다.

〈표 25〉 암 치료(남성) 로지스틱회귀모형

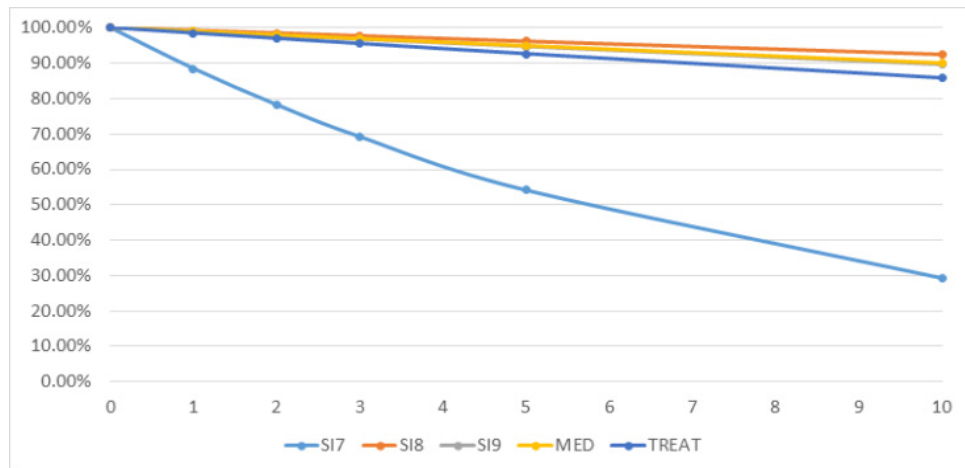
설명변수		df	β	s.e.	p-value	$\exp(\beta)$
Intercept		1	-1.7005	0.1105	<0.0001	-
연령(AGE)	20~29	1	-4.5945	0.2232	<0.0001	0.0101
	30~39	1	-4.1327	0.1663	<0.0001	0.0160
	40~49	1	-2.9022	0.1018	<0.0001	0.0549
	50~59	1	-2.2540	0.0801	<0.0001	0.1050
	60~69	1	-1.4537	0.0674	<0.0001	0.2337
	70~79	1	-0.6229	0.0629	<0.0001	0.5364
	80~89	1	-0.1090	0.0635	0.0858	0.8967
	90+	0	0	0	<0.0001	1.0000
간경화증(SI7)		1	-0.1226	0.0096	<0.0001	0.8846
계속치료(TREAT)		1	-0.0153	0.0044	0.0005	0.9848
당뇨병(SI9)		1	-0.0109	0.0042	0.0088	0.9892
계속투약(MED)		1	-0.0106	0.0055	0.0517	0.9895
고혈압(SI8)		1	-0.0078	0.0042	0.0626	0.9922
통계량			자유도	값	p-value	
Hosmer and Lemeshow 카이제곱			8	6.2514	0.6191	

무사고기간 0년을 기준(100%)으로 했을 때, 위 모델을 바탕으로 기대되는

고지항목 기간별 암 치료 상대위험도는 <표 26>와 <그림 4>와 같다.

<표 26> 무사고기간별 기대 암 치료 상대위험도(%)(남성)

설명변수	1년	2년	3년	5년	10년
간경화증(SI7)	88.46	78.25	69.22	54.16	29.33
계속치료(TREAT)	98.48	96.99	95.52	92.64	85.82
당뇨병(SI9)	98.92	97.84	96.78	94.70	89.67
계속투약(MED)	98.94	97.90	96.86	94.82	89.91
고혈압(SI8)	99.23	98.46	97.70	96.19	92.53



<그림 4> 무사고기간별 기대 암 치료 상대위험도(남성)

2) 여성

<표 27>은 20~89세 여성의 암 치료여부를 적합하기 위한 최적 로지스틱회귀예측모형 결과이다. 여성의 경우 간경화증(SI7), 고혈압(SI8), 당뇨병(SI9), 수술(SUR), 계속투약(MED), 계속치료(TREAT)가 설명변수로 선택되었으며, 연령군 증가는 20~79세에서 암 치료율을 높이는 방향으로 적합되었다.

설명변수 개별적으로 살펴보면, 군단연령 80~89세를 제외한 모든 설명변수에서 무사고기간이 증가할수록 암 치료 가능성이 적은 것으로 예측되었다. 유의한 영향을 끼친 중대질병에 대하여 간경화증(SI7), 고혈압(SI8), 당뇨병(SI9)

무사고기간이 한 단위(1년) 증가할 때마다 암 치료 오즈는 각각 0.8967배, 0.9914배, 0.9830배로 감소하는 것으로 나타났다. 이와 함께 질병이력 중에서 간경화증 무사고 기간이 $p < 0.0001$ 로 가장 유의한 변수로 나타났으며, 고혈압의 경우 $\alpha = 0.05$ 하에서는 유의하지 않지만 $\alpha = 0.1$ 하에서는 유의한 변수임이 관측되었다.

질병 이력 외에 수술, 계속투약, 계속치료 무사고기간의 경우 모두 한 단위 증가할 때마다 암 치료 오즈를 낮추는 것으로 관측되었으며, 수술, 계속투약, 계속치료 무사고기간이 한 단위(1년) 증가할 때마다 암 치료 오즈는 각각 0.9892배, 0.9903배, 0.9861배로 감소하는 것으로 나타났다.

또한, Hosmer and Lemeshow 카이제곱 통계량 분석 결과 p -값이 0.7646으로 나타나, 모형적합도 검증 기준을 통과한 것을 확인할 수 있었다.

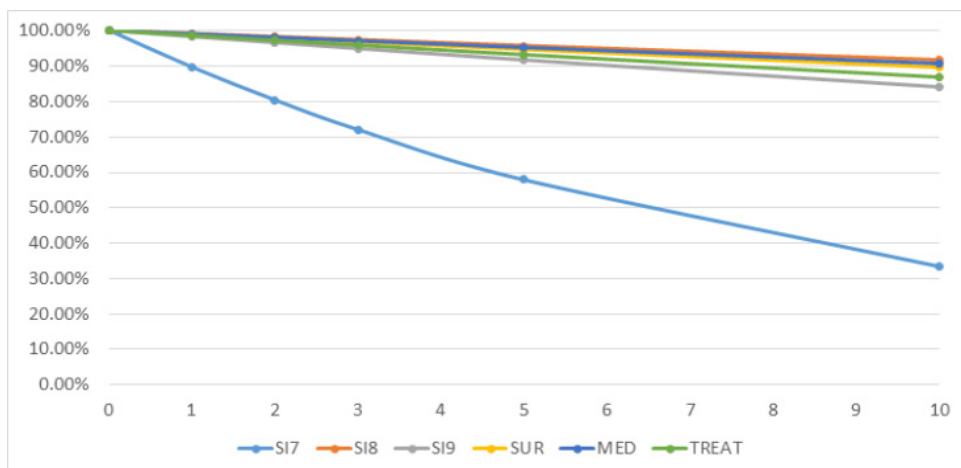
〈표 27〉 암 치료(여성) 로지스틱회귀모형

설명변수		df	β	s.e.	p-value	$\exp(\beta)$
Intercept		1	-2.7214	0.1618	<0.0001	-
연령(AGE)	20~29	1	-3.6347	0.2400	<0.0001	0.0263
	30~39	1	-2.1632	0.1214	<0.0001	0.1150
	40~49	1	-1.1073	0.0856	<0.0001	0.3305
	50~59	1	-0.3785	0.0731	<0.0001	0.6849
	60~69	1	-0.3179	0.0683	<0.0001	0.7277
	70~79	1	-0.2373	0.0681	0.0005	0.7888
	80~89	1	0.0960	0.0671	0.1522	1.1008
	90+	0	0	0	<0.0001	1.0000
간경화증(SI7)		1	-0.1091	0.0152	<0.0001	0.8966
고혈압(SI8)		1	-0.0086	0.0045	0.0591	0.9914
당뇨병(SI9)		1	-0.0172	0.0047	0.0003	0.9829
수술(SUR)		1	-0.0109	0.0046	0.0171	0.9892
계속투약(MED)		1	-0.0097	0.0052	0.0601	0.9903
계속치료(TREAT)		1	-0.0140	0.0046	0.0027	0.9861
통계량			자유도	값	p-value	
Hosmer and Lemeshow 카이제곱			8	4.9337	0.7646	

무사고기간 0년을 기준(100%)으로 했을 때, 위 모델을 바탕으로 기대되는 고지항목 기간별 암 치료 상대위험도는 <표 28>과 <그림 5>와 같다.

<표 28> 무사고기간별 기대 암 치료 상대위험도(%)(여성)

설명변수	1년	2년	3년	5년	10년
간경화증(SI7)	89.67	80.40	72.09	57.97	33.60
당뇨병(SI9)	98.30	96.62	94.98	91.77	84.22
계속치료(TREAT)	98.61	97.25	95.90	93.26	86.97
수술(SUR)	98.92	97.84	96.78	94.70	89.68
계속투약(MED)	99.03	98.07	97.12	95.25	90.72
고혈압(SI8)	99.14	98.30	97.46	95.80	91.77



<그림 5> 무사고기간별 기대 암 치료 상대위험도(여성)

IV. 결론 및 시사점

간편고지보험의 판매량이 증가하고 간편고지 시장의 경쟁이 심화됨에 따라 고지항목에 변화를 주어 다양한 유병자 집단을 목표로 하는 트렌드가 이어지

고 있다. 이를 위해 수술 무사고기간, 중대질병의 정의 등을 조금씩 변화시키며 신규 간편고지 상품이 출시되고 있으며, 각 고지항목의 요소들을 통합하여 질병발생의 상대위험도를 측정하는 모형에 대한 필요성이 증가하고 있다.

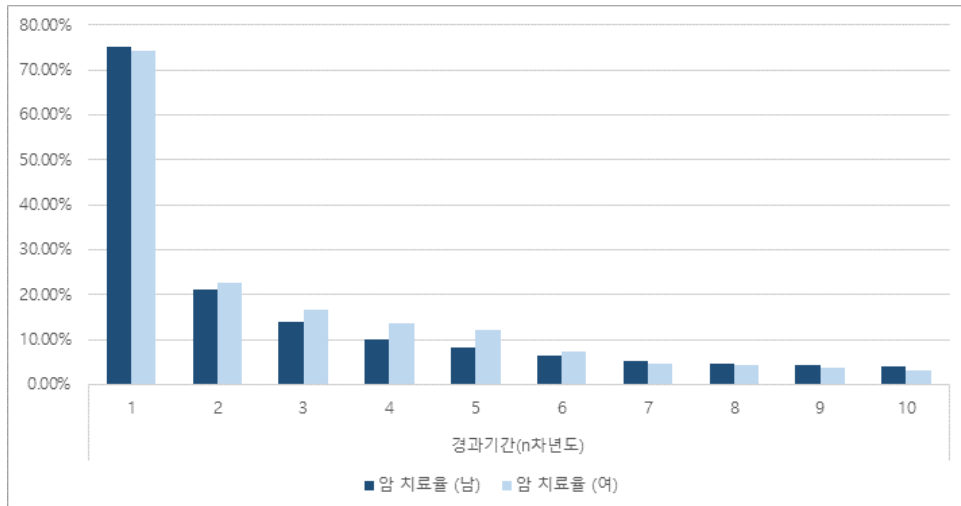
이에 본 연구는 국민건강보험공단의 표본코호트2.0DB를 활용하여 각 간편고지 고지항목을 설명변수로 하고, 암 발생률 및 치료율을 관측변수로 하는 로지스틱회귀모형을 구성하였다. 그 결과 암 발생률 및 치료율에 유의한 영향을 끼치는 고지항목을 판별하였고, 과거 무사고기간이 끼치는 영향의 심도를 객관적으로 측정하였다.

연구 결과를 종합하면, 중대질병 1~9 및 입원, 수술, 계속투약, 계속치료와 관련하여 무사고기간이 증가함에 따라 남녀 모두에서 암 발생 및 치료 위험 상대도가 감소한다는 점을 확인할 수 있었다. 특히, 중대질병 중에서도 협심증(SI1), 간경화증(SI7), 고혈압(SI8), 당뇨병(SI9)이 유의한 변수로 나타났다. 이 중에서 간경화증, 고혈압, 당뇨병은 4개의 모형에서 모두 유의하게 나타났다. 특히 간경화증(SI7) 무사고기간의 증가는 다른 중대질병에 비해 암 발생 및 치료 위험 상대도를 더 크게 감소시키는 것으로 확인되었다. 이는 간경변증이 간세포성 암종의 직접적인 위험요인으로 작용하기 때문으로 해석할 수 있다(Pinter, 2016). 그 외 설명변수인 입원과 수술 무사고기간은 일부 모형에서, 계속투약 및 계속치료 무사고기간은 모든 모형에서 무사고기간과 관련하여 유의한 변수로 확인되었다.

본 연구가 시사하는 바는 다음과 같다. 먼저, 설명변수를 이른바 ‘325보험’, ‘335보험’이 아닌 각 고지항목으로 세분화함으로써 특정 고지항목의 무사고기간이 변화함에 따른 상대위험도를 개별적으로 분석하였다. 이를 통해 전체 고지항목 조합의 경우의 수를 포괄하는 일반화된 모형을 수립했다는 데에 의의가 있다.

또한, 암 발생뿐만 아니라 발생 후 치료에 대한 모형을 수립함으로써 고지항목 변경에 따른 상대위험도 변화의 분석범위를 확장했다. 본 연구에서는 반응변수의 기간을 고려하여 발생 후 1년 이내 치료 여부를 집중적으로 분석하였으며, <그림 6>과 같이 2차 연도 이후 치료율의 변화 트렌드를 고려한다면 발생 후 1년 이내 치료뿐만 아니라 장기 치료에 대해서도 연구의 결과를 확장하여

적용할 수 있을 것으로 기대한다.



〈그림 6〉 경과기간별 암 치료율

마지막으로, 국민건강보험 표본코호트2.0DB를 활용하여 분석의 정확성을 확보하였다. 대상 전수와 최대한 유사한 자료를 사용함으로써 각 고지항목이 암 발생 및 치료 상대위험도에 끼치는 영향을 더욱 정확하게 측정할 수 있도록 했다. 모수를 충분히 확보하여 특정 이상치가 결과에 끼치는 영향을 최소화하였고, 이를 통해 결과에서 제시된 모형의 안정성과 범용성을 확보할 수 있었다.

본 연구의 한계점으로는 암 발생은 의사소견인 상병코드만을 가지고 암 발생을 정의하여 실제 암 발생보다 과대 추정될 가능성이 존재한다. 또한 본 연구에서는 암 발생과 암 치료의 정확한 발생률을 구하기 위해 분석의 편의를 위해 간편고지 항목인 무사고 기간만을 사용했지만 암 발생과 암 치료 사고에 영향을 줄 수 있는 내생성 문제도 존재할 것이다.

참 고 문 헌

- 권용진, 박미혜, “건강보험 청구자료를 활용한 약물관련 약골괴사의 국내 현황 및 비용 분석,” *약학회지* 64(4) (2020): 312-318.
- 권혁성, 오피재, 강민용, 우경석, “건강보험통계 기반 건강 수준 평점화 모형을 활용한 간편고지 상품 유형별 보험요율 차등화에 관한 연구,” *리스크관리연구* 32(4) (2021): 99-147.
- 백혜연, 손지훈, 신지민, “유병자 보험의 보장성 확대를 위한 유병자들의 중증질환 발생률 비교,” *보건정보통계학회지* 43(4) (2018): 318-328
- 심현우, 임형기, 최양호, “부분적 청구경력 고지 간소화에 따른 건강보험 요율 차등화에 관한 연구,” *보험금융연구* 32(2) (2021): 43-75.
- 전희주, 인태교, “국민건강보험 표본코호트2.0DB를 활용한 건강상태에 따른 암발생과 암수술건수 상대위험도 연구,” *리스크관리연구* 33(4) (2022): 53-83.
- 전희주, 최경진, “생체연령을 고려한 퇴직연금 프로그램 인출방식에 관한 연구” *리스크관리연구* 32(2) (2021): 101-134.
- 전희주, 최경진, “국민건강보험 빅데이터를 활용한 치아보험 요율 차등화에 관한 연구: 치아우식증, 치주질환을 중심으로,” *보험금융연구* 33(3) (2022): 119-147.
- 전희주, 최경진, “주택연금 가입자의 신탁방식 전환 의향 분석 및 시사점,” *금융감독연구* 9(2), (2022): 171-200.
- 전희주, 최경진, “국민건강보험 건강검진코호트DB와 표본코호트DB를 활용한 크라운, 치수치료 담보 치아보험 상품의 상대위험도 연구,” *금융감독연구* 10(1), (2023): 137-160.
- 통계청 사회통계기획과, 2022 통계청 고령자 통계, 2022.09.29.
- 현경래, 건강보장정책 수립을 위한 주요 질병의 사회경제적 비용 분석. 강원도: 국민건강보험공단 건강보험정책연구원, 2017.
- 황지연, 이상엽, 주석훈, “건강보험통계를 활용한 간편고지 상품의 담보별 최적 언더라이팅 기준 산출,” *보험학회지* 124, (2020): 1-34.
- Chan, Y. H. “Biostatistics 104: Correlational analysis,” *Singapore Medical Journal*, 44(12), (2003): 614-619.
- David, M., “Auto insurance premium calculation using generalized linear models,” *Procedia Economics and Finance* 20 (2015): 147-156.

- de Jong, P., and Heller, G., *Generalized Linear Models for Insurance Data*. Cambridge, NY: Cambridge University Press, 2008.
- Fox, J., and Monette, G. (1992). "Generalized Collinearity Diagnostics," *Journal of the American Statistical Association* 87(417) (1992): 178–183.
- Park, H. G., Youn, D., Baik, J. M., & Hwang, J. H., "Epidemiology of Achilles Tendon Rupture in South Korea: Claims Data of the National Health Insurance Service from 2009 to 2017," *Clinics in Orthopedic Surgery* 13(4) (2021): 539–548.
- Pinter, M., Trauner, M., Peck-Radosavljevic, M., & Sieghart, W. (2016). "Cancer and Liver Cirrhosis: Implications on Prognosis and Management," *ESMO Open* 1(2) (2016): e000042.
- Montgomery, Douglas C., and Elizabeth A. Peck. *Introduction to Linear Regression Analysis*. New York, NY: John Wiley, 1992.
- Naufal, N., Devila, S., and Lestari, D., "Generalized Linear Model (GLM) to Determine Life Insurance Premiums," *AIP Conference Proceedings* 2168(1) (2019): 020036.