

부분적 청구경력 고지 간소화에 따른 건강보험 요율 차등화에 관한 연구

A Study on Premium Rate Differentiation of Health Insurance by Simplified Partial Disclosure of Claim History

심 현 우*·임 형 기**·최 양 호***

Hyunoo Shim·Hyungi Im·Yangho Choi

우리나라 민영건강보험의 청구경력 고지의무 간소화는 경증 만성질환이나 치료 이력이 있어서 가입사 거절로 보험 혜택을 받기 어려운 유병력자의 보험 가입을 위한 시장 확대 목적으로 발전되어 왔다. 이러한 변화에 기인한 간편심사 보험은 건강보험의 정책형으로 개발된 상품에 국한되어 왔으며, 실손보상의 경우 유병력자 실손의료보험 상품으로 출시되었다. 건강보험의 주요 요율 산출 요소로 성, 연령의 구분 외에 고지의무 간소화에 따른 효과를 반영할 수 있게 되었으나, 청구경력 고지 유무만 요율에 반영되었을 뿐 고지기간에 따른 세밀한 요율 차등화는 반영되어 있지 않았다.

본 연구는 이러한 보험계약상 청구경력 고지기간의 간소화 정도에 따라 가입자의 의료이용이 증가하는지에 대하여 2002년부터 2015년까지의 국민건강보험 표본 코호트 데이터베이스를 이용하여 로지스틱 회귀분석을 통한 실증분석을 진행하였다. 분석결과, 고지의무의 간소화 정도가 높을수록 의료이용의 확률이 통계적으로 유의하게 높은 것으로 나타났으며, 입원에 비해 외래의 고지기간 차등화 효과는 낮게 나타나 차이를 보였다. 추가적으로 고지기간 간소화의 정도에 따라 건강보험의 연납 보험료를 차등화하여 산출하였으며, 보너스-말러스 시스템으로 고지기간 차등화 정도에 따른 할인·할증 수준을 제시하였다.

국문 색인어: 청구경력 고지, 간편심사, 보험료차등화, 건강보험, 보너스-말러스 시스템

한국연구재단 분류 연구분야 코드: B051603, C030805, E090201

* 한양대학교 보험계리학과 부교수(hyunooshim@hanyang.ac.kr), 제1저자

** 한양대학교 금융보험학과 박사, 보험개발원 선임담당역(limhk@kidi.or.kr), 공동저자

*** 한양대학교 보험계리학과 부교수(ychoi@hanyang.ac.kr), 교신저자

**** 본 연구는 보험개발원의 공식적인 의견과 다를 수 있음을 밝힙니다.

논문 투고일: 2020. 7. 21, 논문 최종 수정일: 2021. 1. 8, 논문 게재 확정일: 2021. 5. 20

I. 서론

현재 우리나라 보험사의 경우, 건강보험 위험률 산출의 핵심적 위험요인은 성과 연령이다. 일반적으로 건강보험의 경우 보험사는 피보험자의 성, 연령 이외에 추가적인 위험요인 파악을 위해 청약서로부터 피보험자의 건강상태를 수집한다. 이를 바탕으로 피보험자의 위험도가 표준체에 비해 좋지 않은 경우 특별조건부로 위험보험료를 할증하여 계약을 체결하고 있다. 이러한 보험료 차등화는 청약서로부터 수집된 건강정보를 토대로 핵심적 위험요인을 추가로 파악할 수 있었기 때문에 가능한 것이다.

하지만 상기 건강보험의 표준화된 청약서를 통해 가입하려는 유병력자는 과거 병력 이력으로 인해 언더라이팅 심사에서 가입이 거절되는 경우가 많았다. 최근에는 표준화된 청약서의 고지사항을 간소화한 정액형 간편심사 보험상품(simplified issue product)이 활성화되고 있다. 실손보상의 경우에는 2018년 4월에 간편심사 실손의료보험 상품으로 개발되었다.¹⁾ Insurance Barometer Report(2019)에 따르면, 미국에서도 2019년 보험산업에서 가장 인기 있는 주제 중의 하나로 간편심사(simplified underwriting)가 선정되었을 정도로 보험업계의 주목을 받고 있다.

그동안 민영건강보험에 대한 연구로는 민영의료보험과 국민건강보험 간의 역할 정립에 대한 연구, 민영의료보험의 가입요인 및 가입자 특성에 대한 연구, 민영의료보험 수요와 보험금 수령 및 의료이용량에 관한 연구, 건강보험공단 코호트 DB를 활용한 개인별 건강 위험에 관한 연구, 이러한 개인별 건강위험을 보험의 언더라이팅에 반영한 연구, 실손의료보험의 역선택에 관한 연구, 마지막으로 실손의료보험에서 요율 산정 시 보험료 차등화에 대한 연구 등이 진행되어 왔다.

오창수·유동완(2018)은 패널프로빗 모형을 이용하여 민영건강보험의 가입에 영향을 미치는 결정요인에 대해 실증 분석하였으며, 입원 및 외래 경험에 있는 경우 민영건강보험의 가입에 정(+)의 영향이 있는 것으로 나타났다. 강성호·김대환·이순재(2019)는 민영건강보험의 가입에 영향을 미치는 주요 결정요인으로 소득과 연령을 꼽았으며, 보험 수요는 소득

1) 금융위원회 보도자료(2018. 3. 30), “4월부터 유병력자 실손의료보험이 출시되고, 실손의료보험 끼워팔기가 금지됩니다.”

이 증가할수록 증가하고 연령이 고령화될수록 감소하는 것으로 확인하였다. 김송희·김윤희(2019)는 보험 가입 여부에 따른 당뇨 유병자의 건강관리 효과에 대해 분석하였으며, 민영건강보험 가입자는 비가입자보다 걷기 및 근력운동을 할 확률이 더 높고 건강관리에 더 높은 관심이 있음을 보여주었다.

이현복·황연희(2014)는 민영건강보험의 보험금 수령에 대한 결정요인 중 연령, 성별, 질환 보유, 입원 등 의료서비스 이용 경험 등이 통계적으로 유의한 영향 요인임을 한국의료패널 자료를 통해 보여주었다.

건강보험공단 코호트 DB를 활용한 연구 중 강용곤(2018)은 생체나이와 주민등록상 나이 간 차이에 따른 사망 및 주요 질병에 대한 위험도(hazard ratio)를 건강보험공단 코호트 DB를 활용하여 실증분석하였다. 2002년부터 2013년의 12년간 116만 명의 데이터를 이용하여 콕스-비례위험모형을 통해 위험도를 분석하였으며, 생체나이가 주민등록상 나이보다 1살 증가할수록 사망위험도는 1.6% 증가, 고혈압 발생위험도는 2.5% 증가한다고 추정하였다. 권태연·임자영·박유성(2017)은 동 코호트 DB를 통해 전체 기대수명에서 특정 질병에 걸린 상태로 삶을 유지하는 건강하지 않은 기간의 비중을 추정하여 기대수명과 대비되는 건강기대수명을 도출하였다.

언더라이팅에 대한 연구 중 박상준(2003)은 보험계약 인수 시 비표준 위험의 언더라이팅 개선방안으로 의적 언더라이팅(medical underwriting)에 의한 해외사례 현황과 선진 기법을 소개하였다. 이찬희·전희주·김석영(2015)은 국민 암 발생률과 보험사의 경험 암 발생률 간 모비율의 차이 검정을 통해, 암보험에 대한 언더라이팅 효과를 실증분석하였다.

김대한·이봉주(2013)는 실손의료보험에 역선택이 존재하는지 여부에 대해 실증적으로 분석하였으며, 입원 및 외래방문 경험이 있을 경우 실손의료보험 가입수요는 증가하는 것으로 나타났다. 김대한(2014)은 실손의료보험의 가입 여부가 의료수요에 미치는 영향을 분석하였으며, 단기적으로는 실손의료보험에 가입한 경우 의료수요를 증가시키는 것으로 나타났지만 1년이 지난 이후에는 의료수요 증가 효과가 나타나지 않는 것으로 확인하였다.

실손의료보험의 보험료 차등화를 위한 연구로 이경아·이항석(2016)은 신뢰도 이론을 적용하여 계약자 경험자료 반영의 타당성을 제시하였으며, 이항석·이가는·이경아(2017)는 현행 요율 변수인 성·연령에 대한 GLM 분석 결과보다 과거 계약자별 보험금 지급실적을 포

함한 GLM 분석 결과가 보험금 지급 건수에 대해 더 높은 설명력을 지닌다는 점을 확인함으로써 실손의료보험 보험료 차등화를 위한 계약자 실적변수 도입의 필요성을 강조하였다.

간편심사 상품은 일반적인 인수심사 기준을 완화하여 판매하므로 장기적인 운용을 위해서는 인수집단의 특성을 고려한 보험료 산정이 선행되어야 한다. 청구경력 정보를 간소화하여 인수심사 기준을 완화하는 경우에는 해당 집단의 효율과 일반 집단의 효율 차이를 반영할 필요가 있다. 따라서 본 연구는 건강보험공단의 표본 코호트 DB를 이용하여 보험계약상 고지의무 간소화가 의료이용량에 어떠한 영향을 미치는지 실증적으로 분석하고자 한다.

본 연구는 다음과 같이 구성되어 있다. 2장에서는 분석방법을 살펴보고, 3장에서는 실증분석 자료와 변수선택 결과 및 모형에 따른 분석결과를 설명하였고, 마지막 4장에서는 결론을 제시하였다.

끝으로 본 연구는 한양대학교 기관 생명 윤리위원회(IRB: Institutional Review Board)로부터 인간 대상 및 인체 유래물 연구에 대한 심의면제 승인(HYU-2019-05-017)을 받은 후 연구를 진행하였다. 연구자료는 국민건강보험공단의 표본 코호트 DB(연구관리번호 NHIS-2020-2-061)를 활용한 것으로, 연구의 결과는 국민건강보험공단과 관련이 없음을 밝힌다. 본 논문은 임형기(2021)의 박사학위논문의 일부에 동반하여 작성한 것이다.

II. 분석모형

1. 이항 로지스틱 회귀모형

본 연구에서는 고지의무 간소화 정도에 따른 의료이용량 분석을 위해 국민 건강에 가장 큰 영향을 미치는 3대 질환(암, 뇌혈관질환, 허혈성 심장질환) 발병 여부를 응답변수로 선정하였다. 분석방법으로 이항 로지스틱 회귀모형(binomial logistic regression)을 적용하였다. 회귀분석 시 변수선택 방법은 BIC 기준을 적용하였으며, 이항 로지스틱 회귀모형의 결과를 이용하여 3대 질환에 대한 평균 의료비 차이를 살펴보았다.

일반적으로 설명변수와 종속변수 간 관계를 파악하는 방법으로 사용되는 선형회귀모형

은 종속변수가 연속형이어야 하며, 설명변수와 선형관계를 띄는 경우 사용할 수 있는 모형이다. 이항 로지스틱 회귀모형은 종속변수가 이분형 변수(binary variable)인 경우 종속변수와 설명변수 간의 관계를 분석하는 데에 사용할 수 있는 통계기법이다.

종속변수를 이분형 확률변수 Y 로 나타내자. Y 는 0(미발생) 또는 1(발생)의 값을 가질 수 있다.

$$\begin{cases} \Pr(Y=1) = \pi \\ \Pr(Y=0) = 1 - \pi \end{cases}, \text{ 여기서 } \pi \text{는 사건(발병)의 발생 확률이다. } \pi = E(Y).$$

즉, 확률변수로 변환하면서 추정대상은 S자 곡선 형태인 확률의 누적분포함수(CDF)로 대체된다. 이때 로지스틱 함수의 누적분포함수로 적합하는 것을 로지스틱 회귀분석이라 한다.

특정 사건이 일어날 확률($\pi = \Pr(Y=1)$)을 다음과 같이 로짓변환하여 보자)로 대체된 종속변수를 로지스틱 누적분포함수로 가정한 로지스틱 회귀모형은 다음과 같다.

$$\pi = \frac{\exp(Z)}{1 + \exp(Z)}.$$

위 식을 정리하면, 다음과 같이 π 를 로짓변환하고 이를 Z 로 나타내자.

$$\begin{aligned} \Rightarrow \exp(Z) &= \frac{\pi}{1 - \pi}, \\ \Rightarrow \ln\left(\frac{\pi}{1 - \pi}\right) &= Z. \end{aligned}$$

Z 를 설명변수(X_k)의 선형결합으로 예측하는 모형이 로지스틱 회귀모형이다.

$$\ln\left(\frac{\pi}{1 - \pi}\right) = Z = \beta_0 + \sum_{k=1}^p \beta_k X_k \quad (p: \text{독립변수의 개수}).$$

위의 식 $\frac{\pi}{1 - \pi}$ 를 오즈(Odds)라 하고, 특정 사건이 발생하지 않을 확률 대비 특정 사건이 발생할 확률의 비율을 뜻한다.

2. 간편심사 건강보험 요율 차등화를 위한 보험요율 산출 모형

건강보험공단의 표본 코호트 DB를 활용하여 건강상태 및 무사고기간에 따른 3대 질환별 발생 여부에 미치는 영향을 분석하기 위해 이항 로지스틱 회귀모형을 실시하였다.

$$\ln\left(\frac{\pi}{1-\pi}\right) = \ln(Odds) = \beta_0 + \sum_{i=1}^p \beta_i X_i,$$

$\pi = \Pr(Y=1)$ = 3대 질환의 (입원·외래)발생 확률, p : 독립변수의 개수,

β_0 : 절편, β_i : 이항로지스틱 회귀계수, X_i : 독립변수.

종속변수(Y)는 진료 DB의 주 상병코드를 기준으로 구분된 3대 질환(암, 뇌혈관질환, 허혈성 심장질환)²⁾의 발생 여부(입원·외래)로 구성하였다. 김동욱(2017)은 표본 코호트 DB의 8종류의 암에 대한 조작적 정의로 주 상병과 부 상병, 입원 여부 등에 따른 발병률을 비교하여 적절한 기준을 제시하였으며 주 상병 코드만을 사용한 경우가 부상병 코드를 함께 사용한 경우보다 과대추정오류를 줄일 수 있다는 점을 보여주었다. 이에 따라 발생 여부에 대한 조작적 정의는 각 주 상병별로 최소 두 번 이상의 입원 또는 외래가 있다면 발병한 것으로 정의하였다. 암과 뇌혈관질환, 허혈성 심장질환은 중대 질환으로써 초진 이후에도 지속적인 외래와 입원이 동반될 것으로 보여지기 때문이다.

주요 설명변수의 후보 변수들을 선택하기 위해 아래의 연구를 참고하였다.

김혜련(2006)은 한국인의 주요 질병 발생으로 인한 의료이용에 영향을 미치는 건강 위험요인 중 대사증후군 여부(콜레스테롤, 공복혈당 변수), 고혈압 여부(수축기혈압, 이완기혈압 변수), 혈중 지질 수준(총콜레스테롤, 혈중 중성지방, HDL 콜레스테롤, LDL 콜레스테롤 변수), 혈당(공복혈당 변수), 비만(체질량지수 변수), 간기능검사(AST, ALT 변수), 질병 기왕력, 질병의 가족력 등을 제시하였다. 오영수(2003)는 우리나라 건강보험 가입 시 실제 위험심사에 반영하는 항목으로 콜레스테롤, 혈압수치, 체격(신장/몸무게), 흡연습관, 가족력 등을 포함하고 있음을 조사하였다. 이항석·이민하·백혜연(2018)은 실손의료보험의 할인할증제도 도입을 위해 일반화 혼합 선형모형(GLMM)을 이용하여 무사고 누적연수

2) 암, 뇌혈관질환, 허혈성 심장질환에 해당하는 한국표준질병사인분류 코드는 통계청고시 제 2015-309호에 따라 <Appendix Table 3>에 나타내었다(통계청, 2015).

별 상대도를 제시하였다. 해당 논문은 무사고기간이 누적이므로 독립적이지 못하다는 한계를 가진다. 본 연구에서는 이러한 단점을 보완하고자 중첩되지 않은 무사고기간들을 설명변수로 설정하여 설명변수의 독립성을 높였다. 또한, 단순히 무사고기간 1년, 2년뿐만 아니라 여러 더미 변수의 조합을 통해 직전 1년 전, 직전 2년 전~3년 전 등 여러 실제 사례를 자연스럽게 적용할 수 있다. 김경훈(2010)은 Charlson Comorbidity Index(CCI)를 ICD-10으로 전환한 여러 알고리즘의 예측력을 비교하였고, 모든 상황에서 Quan 등이 제시한 알고리즘이 다른 알고리즘에 비해 결과 예측력이 높았으나 유의미한 차이는 없었음을 제시하였다.

이에 따른 분석모델에서 고려되는 설명변수를 <Table 1>에 나타내었다. 설명변수(X)는 인적 현황에 대한 변수로 성별, 연령구간, 지역, 소득분위를 포함하며, 건강상태 변수로는 동반질환지수(CCI), BMI, 수축기혈압, 이완기혈압, 공복혈당, 중성지방, Non-HDL 콜레스테롤, HDL 콜레스테롤, AST, ALT, 감마GTP를 포함하였다. 본 연구의 중요 변수로 무사고기간(No-claim Period), 주요 질환에 대한 기왕력을 포함하였다. 또한, 흡연 여부, 주당 운동 횟수와 같이 자의적으로 응답할 수 있는 설문변수는 설명변수에서 제외하였다. 무사고기간에 대한 변수 정의는 보험계약 시 청약서상 고지항목으로 최근 연도별 입원 여부를 조사하는 것에 기인한다. 개별 의료수급권자의 과거 입원 진료 여부를 추적하여 직전 연도 t년부터 t+1년까지 사이에 입원 Claim이 있다면 무사고기간 t년으로 정의한다. CCI 변수는 많은 진단 코드가 고려된 Quan(2005)에 따라 부여하였다. 개인별 청구 Claim의 상병코드별로 대응(mapping)된 CCI 최댓값을 연간 합산한 뒤, 이를 다시 0점, 1점, 2점, 3점 이상으로 범주화하였다.

<Table 1> Candidate Explanatory Variables

Variable	Dummy Variables
Sex	X ₁
Age_group (25~69)	X ₂ ~X ₁₀
Region	X ₁₁ ~X ₁₂
Income Decile Groups	X ₁₃ ~X ₁₄
CCI	X ₁₅ ~X ₁₈
No-Claim Period	X ₁₉ ~X ₂₄

Variable	Dummy Variables
Body Mass Index (BMI)	X ₂₅
Systolic Blood Pressure (SBP)	X ₂₆
Diastolic Blood Pressure (DBP)	X ₂₇
Fasting Blood Sugar	X ₂₈
Triglyceride	X ₂₉
Non-High Density Lipoprotein Cholesterol (Non-HDL)	X ₃₀
High Density Lipoprotein Cholesterol (HDL)	X ₃₁
Aspartate Aminotransferase (AST, SGOT)	X ₃₂
Alanine Aminotransferase (ALT, SGPT)	X ₃₃
Glutamyl Transpeptidase (γ-GTP)	X ₃₄
History of Hypertension	X ₃₅
History of diabetes	X ₃₆
History of Hyperlipidemia	X ₃₇
History of Pulmonary Tuberculosis	X ₃₈
Consecutive No-Claim Period (From the Most Recent Hospitalization to Now)	X ₃₉ ~X ₄₅

본 연구의 모델설정 및 변수선택에서 유념할 점은 본 연구의 목적이 종속변수인 주요질병군에 대한 발병 여부와 설명변수 간의 인과관계를 밝히는 것이 아니라는 점이다. 분석의 주요 목적은 추가적인 설명변수의 선택으로 종속변수인 발병 여부에 대한 설명력을 높이기 위한 모델을 설정하는 데에 있다.

회귀모델의 여러 가지 변수선택 방법(variable selection) 중 AIC(Akaike information criterion)와 BIC(Bayesian information criterion)는 다음과 같다.

$$AIC = -2 \ln(L(\hat{\theta})) + 2k,$$

$$BIC = -2 \ln(L(\hat{\theta})) + k \cdot \ln(n),$$

여기서 $\hat{\theta}$ = 추정 모델의 모수 벡터,
 $L(\hat{\theta})$ = 최대우도로 적합된 모델 하에서 해당 데이터의 우도,
 k = 모수의 개수, n = 데이터 수.

일반적으로 bias-variance trade-off 관계에 따라 변수가 많은 모델일수록, bias는 작아지며 반대로 variance는 증가하게 된다. AIC, BIC 값은 작을수록 좋은 모델이라 볼 수 있다. 우도(likelihood)가 클수록, 변수의 수가 적을수록 AIC, BIC 값이 작아진다. 관측값

의 수가 많으므로 AIC에 비해 BIC는 변수의 개수 증가에 대해 더 큰 페널티를 부과한다.

Ⅲ. 실증분석 자료 및 기술통계량

표본 코호트 DB는 국민건강보험공단의 건강보험 및 의료급여권자 전체 약 4,700만 명에 대해 2002~2013년에 걸쳐 모집단의 여러 특성들에 대한 대표성을 확보하여 추출한 표본으로, 약 100만 명 수준으로 구성되어 있다(이준영 et al. 2014). 추출된 표본은 비례 배분(proportional allocation)에 의한 층화무작위추출법(stratified random sampling)을 통해 구축되어 있으며, 전체 모집단을 성별, 연령, 소득수준, 지역으로 세분화하고 목표 변수를 연간 총 의료비로 하여 표본을 추출하였다. 자격 DB는 건강보험가입자 및 의료급여수급권자(외국인 제외)를 대상으로 하며, 성, 연령대, 지역, 가입자 구분, 소득분위 등 대상자의 사회경제적 변수 및 장애, 사망 관련 총 14개 변수로 구성되어 있다. 진료 DB는 요양기관에 방문하여 진료 등을 받은 내역에 대해 요양기관으로부터 요양급여가 청구된 자료로 명세서, 진료, 상병, 처방 관련 총 57개 변수로 구성되어 있다. 건강검진 DB는 건강검진 주요 결과 및 문진에 의한 생활습관 및 행태 관련 자료로, 신장, 체중, 수축기혈압, 콜레스테롤, 중성지방, 공복혈당, 본인 과거력, 가족력, 흡연상태, 음주량 등 2002년부터 2008년까지는 37개 변수로 구성되어 있고, 2009년에 검진제도 개편으로 2009년부터 2013년까지는 41개 변수로 구성되어 있다.

매년 100만 명에 대한 표본 코호트 DB 데이터 전처리 과정은 다음과 같다. 먼저 진료 DB는 치과 및 한방 진료 DB는 제외하고 의과 진료 DB만을 활용하였다. 진료 DB에는 한 명의 개인이 1년 안에 여러 번 청구한 데이터가 포함되어 있다. 앞서 언급한 3대 질환의 조작적 정의에 따라 개인 ID별 연간 1건의 데이터를 생성한다. 암, 뇌혈관질환 및 허혈성 심장질환별로 최소 두 번 이상의 입원 또는 외래가 발생하면 3대 질환이 발병한 것으로 정의한다. 따라서 3대 질환 발생으로 인한 각 개인의 연간 건강보험 급여청구 전체 건이 분석단위이다. 여러 개의 청구 건을 하나의 관측단위로 하였기 때문에 한 개인의 청구건 간 상관관계가 존재하지 않는다. 건강검진 DB는 2009년 검진제도 개편에 따라, 개인별 콜레

스테를 정보와 본인 과거력에 대한 데이터의 연속성 확보를 위해 2009년부터 2013년까지 5개년 횡단자료를 사용하였다. 이후 자격 DB, 진료 DB, 건강검진 DB를 개인 ID기준으로 병합한다. 무사고기간 변수는 2002~2008년 진료 DB를 활용하여 2009년부터 2013년까지 각 연도별 무사고기간 변수 0년부터 5년까지의 자료를 생성한다.

총 표본 수는 2009년부터 5개년간 매년 약 100만 건을 합산하여 5,032,892건이다. 이 중에서 건강보험공단이 제공하는 건강검진을 받은 비중은 23%로, 그 외 77%는 건강검진 데이터가 없으므로 이를 제외하였다. 추가로 건강상태 변수에 대한 결측치와 3대 질환에 해당하는 본인 과거 병력이 있는 데이터는 분석대상에서 제외하였다. 마지막으로 건강검진 데이터가 충분한 연령구간(25세~69세)을 대상으로 자료를 구성하였다. 최종 분석대상 표본 수는 연도별 약 12만 건으로 총 608,250건이다. 회귀분석에 사용된 연속형 변수와 범주형 변수의 기초통계는 각각 <Table 2>와 <Table 3>과 같다.

<Table 2> Descriptive Statistics for Continuous Variables

(N=608,250)

Variable	Mean	S. D.	p5	p50	p95
BMI	23.83	3.25	18.91	23.63	29.41
Systolic Blood Pressure (SBP)	121.61	14.44	100	120	146
Diastolic Blood Pressure (DBP)	76.22	9.89	60	77	91
Fasting Blood Sugar	97.64	23.35	76	93	131
Triglyceride	132.96	91.59	46	108	302
Non-HDL Cholesterol	139.75	40.18	86	138	203
HDL Cholesterol	55.53	22.16	36	53	80
AST	25.07	15.28	15	22	43
ALT	25.57	21.82	10	20	58
γ -GTP	37.52	48.40	10	24	106

Note: S. D. stands for the standard deviation, p5 the 5th percentile, p50 the median, and p95 the 95th percentile, respectively.

〈Table 3〉 Sample Distribution of Category Variables

(N=608,250)

Variable		Number of Samples	Proportion	
Sex	Male	357,898	58.8%	
	Female	250,352	41.2%	
Age_group	25~29	58,530	9.6%	
	30~34	72,230	11.9%	
	35~39	66,587	10.9%	
	40~44	98,188	16.1%	
	45~49	79,790	13.1%	
	50~54	92,684	15.2%	
	55~59	59,942	9.9%	
	60~64	50,764	8.3%	
	65~69	29,535	4.9%	
Region	Seoul and Six Major Metropolitan Cities	277,731	45.7%	
	Provincial Cities	330,519	54.3%	
Income Decile Groups	Lower-income Group (0~5) (0: Medical Benefit Recipient)	210,233	34.6%	
	Higher-income Group (6~10)	398,017	65.4%	
CCI		0 point	507,396	83.4%
		1 point	79,684	13.1%
		2 point	15,701	2.6%
		3+ point	5,469	0.9%
No-Claim Period (hospitalization)	0-year	Existence	53,822	8.8%
	1-year		40,680	6.7%
	2-year		33,106	5.4%
	3-year		27,510	4.5%
	4-year		23,105	3.8%
	5-year		19,579	3.2%
	Non-existence		410,448	67.5%
History of Hypertension		Non-existence	518,506	85.2%
		Existence	89,744	14.8%
History of Diabetes		Non-existence	577,334	94.9%
		Existence	30,916	5.1%
History of Hyperlipidemia		Non-existence	590,570	97.1%
		Existence	17,680	2.9%
History of Pulmonary Tuberculosis		Non-existence	595,093	97.8%
		Existence	13,157	2.2%

각 설명변수 간 상관관계를 살펴보기 위해 피어슨 상관계수(Pearson correlation coefficient)를 검토한 결과는 Appendix 〈Table 2〉과 같다. 대체로 상관관계가 작으나,

수축기혈압(SBP)과 이완기혈압(DBP), 아스파르테이트 아미노전달효소(AST)와 알라닌 아미노전달효소(ALT) 변수 간 상관관계가 상대적으로 높은 것으로 나타났다.

다중공선성(multi-collinearity)을 살펴보기 위해 살펴본 분산팽창요인(VIF: Variance Inflation Factor) 산출결과는 <Table 4>와 같다. Kennedy(1992)에 의하면 일반적으로 VIF가 10을 넘는 경우 다중공선성 문제가 있다고 본다. 다중공선성을 체크한 결과 VIF 값이 모든 변수에서 5 이하로 나타나 다중공선성 문제를 유발하는 변수가 포함될 가능성이 낮다고 판단된다.

<Table 4> Results of VIF

Variable	VIF (set of parameters)
BMI	1.35526
SBP	2.66623
DBP	2.61240
Fasting Blood Sugar	1.09285
Triglyceride	1.51873
Non-HDL Cholesterol	1.26120
HDL Cholesterol	1.22470
AST	2.56639
ALT	2.66629
γ -GTP	1.43681

IV. 분석결과

1. 단순 로지스틱회귀 모델(p=1)

3대 질환에 대한 입원 또는 외래 발생에 영향을 미치는 보험계약 특성 및 건강상태 등의 설명변수 X_i ($i=1, \dots, p$) 중 변수 하나씩으로만 구성된 이항 로지스틱 회귀분석을 실시한 결과는 아래 <Table 5>와 같다. 초기 모델은 무사고기간 변수를 추가하지 않은 모델로 개별 변수가 갖는 모형 설명력을 살펴보기 위한 비교 모델로써 의미를 둔다. 분석결과 모든 변수($i=1, \dots, 33$)에 대하여 높은 통계적 유의성이 존재하였으며, 여성의 경우 남성보다 3대 질환이 발생하지 않을 확률대비 발생할 확률이 $1.243 (= e^{0.2174})$ 배 높은 것을 알 수 있

다. 3대 질환 발병확률은 서울을 포함한 6대 광역시가 지방보다 높았고, 소득 구간별로는 저소득자가 고소득자보다 높았다. 또한, 발병확률은 기저질환에 대응되는 CCI 점수가 높을수록, 연령군(Age_group)이 높을수록, 고혈압(Hypertension)·폐결핵(Pulmonary Tuberculosis)·당뇨(Diabetes)·고지혈증(Hyperlipidemia)에 대한 과거 병력이 있을수록 증가하였다.

〈Table 5〉 Regression Estimates for the Simple Logistic Model

Selected Variable (X _i)		Estimate (β_i) ¹⁾	Standard Error	Wald	AUC ²⁾	BIC ³⁾
Sex - Female (Base Male)		0.2174***	0.0178	149.37	0.527	125,562
Region - Provincial Cities (Base Metropolitan Cities)		-0.0372*	0.0178	4.39	0.505	125,706
Income Decile Groups - Higher Group (Base Lower Group)		-0.0827***	0.0184	20.14	0.509	125,690
CCI (Base 0 point)	1 point	0.8672***	0.0215	1633.53	0.600	122,509
	2 point	1.4252***	0.0341	1745.05		
	3+point	1.8241***	0.0476	1471.50		
Age_group (Base 25~29)	30~34	0.6117***	0.1207	25.68	0.759	114,397
	35~39	1.2037***	0.1131	113.17		
	40~44	1.8044***	0.1055	292.72		
	45~49	2.2024***	0.1047	442.70		
	50~54	2.7729***	0.1026	729.96		
	55~59	3.2207***	0.1027	984.04		
	60~64	3.5974***	0.1023	1235.71		
	65~69	3.8827***	0.1029	1422.80		
BMI		0.1737***	0.00852	416.20	0.558	125,306
SBP		0.2400***	0.00842	813.0579	0.571	124,925
DBP		0.1327***	0.00867	234.07	0.540	125,480
Fasting Blood Sugar		0.2046***	0.00632	1047.22	0.586	124,867
Triglyceride		0.0644***	0.00834	59.66	0.525	125,653
Non-HDL Cholesterol		-0.0845***	0.00902	87.75	0.524	125,622
HDL Cholesterol		-0.2046***	0.00950	464.34	0.558	125,224

AST	0.0755***	0.00473	254.59	0.569	125,522
ALT	0.0576***	0.00577	99.53	0.552	125,636
γ -GTP	0.0790***	0.00586	181.97	0.541	125,572
History of Hypertension	1.3680***	0.0183	5570.92	0.626	120,879
History of Diabetes	1.1136***	0.0263	1790.69	0.543	124,312
History of Hyperlipidemia	0.9766***	0.0350	780.53	0.522	125,100
History of Pulmonary Tuberculosis	1.2549***	0.0359	1225.01	0.524	124,810

Notes: 1) Significance '***' 0.001, '**' 0.01, '*' 0.05

2) Area Under ROC (Receiver Operating Characteristic) Curve

3) Bayesian Information Criterion

2. 최적 변수선택 결과(무사고기간 변수 제외)

다음은 무사고기간 변수를 여전히 제외한 상태에서 BIC 기준에 의한 최적의 변수선택을 실시한 모델(최적 모델 1)이다(〈Table 6〉 참조). BIC(111,201)는 단순 로지스틱회귀 모델보다 개선되었으며, 모형 설명력인 AUC(0.791)도 증가하였다. 변수선택결과 성별(Sex), 지역(Region), 수축기혈압(SBP), 공복혈당(Fasting Blood Sugar), 중성지방(Triglyceride), 아스파르테이트 아미노전달효소(AST), 알라닌 아미노전달효소(ALT) 및 감마글루타밀 전이효소(γ -GTP) 변수가 제외되었다. 설명력이 높은 변수로는 연령군, CCI, 폐결핵 과거력, 고혈압 과거력, 고지혈증 과거력 순으로 나타났다.

〈Table 6〉 Regression Estimates for the Best Model 1
(Excluding the No-Claim Variables of Inpatient)

Deviance		110,897			
Deviance Value/DF		0.1824			
Pr > ChiSq		1.0000			
AUC		0.791			
SC (BIC)		111,201			
Variable (X_i)		Estimate (β_i)	Standard Error	Wald	Odds Ratio (e^{β_i})
Intercept		-6.627***	0.102	4,262.464	
CCI (Base 0 point)	1 point	0.433***	0.022	384.091	1.542
	2 point	0.767***	0.035	473.708	2.153
	3+point	1.013***	0.049	424.749	2.752

Age_group (Base 25~29)	30~34	0.615***	0.121	25.921	1.850
	35~39	1.202***	0.113	112.360	3.326
	40~44	1.774***	0.106	281.581	5.895
	45~49	2.136***	0.105	413.583	8.467
	50~54	2.634***	0.103	651.338	13.935
	55~59	2.996***	0.104	838.300	19.998
	60~64	3.256***	0.103	991.832	25.957
	65~69	3.408***	0.104	1,068.368	30.218
Income Decile Groups - Higher Group (Base Lower Group)		0.115***	0.019	36.910	1.121
BMI		0.035**	0.010	11.963	1.035
Non-HDL Cholesterol		-0.227***	0.010	553.342	0.797
HDL Cholesterol		-0.114***	0.010	142.774	0.892
DBP		-0.045***	0.009	22.866	0.956
History of Hypertension		0.506***	0.021	573.209	1.658
History of Diabetes		0.166***	0.028	35.108	1.180
History of Hyperlipidemia		0.252***	0.036	48.350	1.287
History of Pulmonary Tuberculosis		0.947***	0.037	641.246	2.577

3. 최적 모델 1에 무사고기간(No-Claim Period) 변수를 추가한 결과

다음으로 고지의무 간소화의 정도를 나타내는 무사고기간 변수를 최적 모델 1에 추가하여 로지스틱 회귀분석을 실시한 결과가 <Table 7>이다. 모델 2는 입원에 대한 무사고기간 변수를 추가한 모형으로 모델 1에 비해 통계적으로 더 높은 설명력을 보였다. 모델 2는 모델 1에 비해 BIC값이 개선(111,201→107,771)되었으며, AUC도 더 높은 것(0.791→0.816)으로 확인되었다.

무사고기간 0년(현재부터 1년 전까지 1년간 입원 Claim이 있는 경우)부터 무사고기간 5년(현재부터 5년 전까지 입원 Claim이 없고 5년 전부터 6년 전까지 1년간 입원 Claim이 있는 경우)까지 모든 변수에 대한 회귀계수는 통계적으로 유의하며 부호는 양수로 나타났다. 무사고기간 0년에 따른 β 은 1.279, 오즈비는 $3.592(=e^{1.279})$ 배로, 직전연도에 입원하지 않은 경우 대비 입원한 경우의 3대 질환 발병(입원 또는 외래) 확률이 3.592배 높은 것으로 분석되었다.

무사고기간 5년부터 무사고기간 0년까지 고지의무 간소화기간이 짧아질수록 β 는 증가하였다. 무사고기간 5년의 β 는 0.619, 무사고기간 4년의 β 는 0.724, ..., 무사고기간 1년의 β 는 1.100, 무사고기간 0년의 β 는 1.279로 점차 증가하였다. 즉, 오래 전 입원 여부보다 최근 입원 여부에 대해 더 큰 회귀계수가 추정되었음을 알 수 있다. 이를 통해 3대 질환 위험상대도는 보험 가입 시 청약서상 입원에 대한 고지기간의 간소화 정도가 클수록 증가하는 것을 확인할 수 있다.

추가적으로 무사고기간에 대한 변수 정의를 입원(inpatient)이 아닌 외래(outpatient)로 변경하여 동일한 방법으로 분석을 진행하였다. 모델 3은 최적 모델 1에 외래에 대한 무사고기간 변수를 추가한 모델이며, 로지스틱 회귀분석을 실시한 그 결과를 <Table 7>에 포함하였다. 입원에 대한 무사고기간 변수를 추가한 모델과 달리 최적 모델 1과 통계적으로 비슷한 설명력을 보였다. 모델 3은 최적 모델 1 대비 BIC값은 111,201에서 111,001로, AUC는 0.791에서 0.793로 모두 근소하게 개선되었다. 이는 외래에 대한 3년 이전 무사고기간 변수에 대한 β 가 유의하지 않은 결과(유의수준 $p=0.1$)로 나타났기 때문이다. 즉, 3년 이전의 외래 여부는 3대 질환의 발병에 있어 유의미한 영향이 있다고 볼 수 없으므로, 외래에 대한 보험계약상 고지기간의 간소화는 3년 정도면 충분하다고 보인다.

<Table 7> Regression Estimates for the Best Model 2, 3 by BIC Statistics
(Including the No-Claim Variables of Inpatient / Outpatient)

Model		Model 2 (No-Claim Variables of Inpatient)		Model 3 (No-Claim Variables of Outpatient)	
Deviance		107,388		110,617	
Deviance Value/DF		0.1766		0.1819	
Pr > ChiSq		1.0000		1.0000	
AUC		0.816		0.793	
SC (BIC)		107,771		111,001	
Variable(X_i)		Estimate for β_i (standard error)	Wald	Estimate for β_i (standard error)	Wald
Intercept		-6.954*** (0.102)	4,660.401	-8.145*** (0.318)	654.550
CCI (Base 0 point)	1 point	0.371*** (0.022)	278.068	0.412*** (0.022)	346.928
	2 point	0.651*** (0.036)	333.696	0.744*** (0.035)	446.250

	3+point	0.825*** (0.050)	272.934	0.991*** (0.049)	407.766
Age_group (Base 25~29)	30~34	0.521*** (0.121)	18.606	0.615*** (0.121)	25.854
	35~39	1.128*** (0.113)	98.852	1.194*** (0.113)	110.831
	40~44	1.706*** (0.106)	259.975	1.758*** (0.106)	276.245
	45~49	2.051*** (0.105)	380.758	2.115*** (0.105)	405.159
	50~54	2.503*** (0.103)	587.117	2.603*** (0.103)	635.451
	55~59	2.851*** (0.104)	757.905	2.959*** (0.104)	817.288
	60~64	3.090*** (0.104)	890.814	3.217*** (0.103)	967.102
	65~69	3.210*** (0.104)	945.498	3.368*** (0.104)	1,042.737
Income Decile Groups - Higher Group (Base Lower Group)		0.099*** (0.019)	27.288	0.110*** (0.019)	33.848
No-Claim Period	0-year	1.279*** (0.025)	2,716.071	1.606*** (0.303)	28.144
	1-year	1.100*** (0.029)	1,450.836	1.035*** (0.308)	11.266
	2-year	1.011*** (0.033)	969.378	0.671* (0.331)	4.104
	3-year	0.885*** (0.037)	574.586	0.324 (0.393)	0.680
	4-year	0.724*** (0.043)	286.041	0.604 (0.419)	2.078
	5-year	0.619*** (0.048)	165.375	0.233 (0.541)	0.186
BMI		0.011 (0.010)	1.213	0.030*** (0.010)	8.860
Non-HDL Cholesterol		-0.216*** (0.010)	499.896	-0.227*** (0.010)	551.791
HDL Cholesterol		-0.108*** (0.010)	127.696	-0.113*** (0.010)	140.438
DBP		-0.033*** (0.009)	12.548	-0.040*** (0.009)	18.458
History of Hypertension		0.502*** (0.021)	558.161	0.480*** (0.021)	516.249
History of Diabetes		0.083* (0.028)	8.545	0.154*** (0.028)	30.553
History of Hyperlipidemia		0.237*** (0.037)	41.963	0.243*** (0.036)	44.779
History of Pulmonary Tuberculosis		0.885*** (0.038)	547.445	0.944*** (0.037)	637.123

4. 요율 차등화를 위한 평점화 및 의료비 추정

Milliman(2011)에서 소개된 요율 차등화를 위한 요소에는 평균식 보험료, 차등화된 보험료, 위험조정을 위한 위험조정계수가 필요하다. 여기서 위험조정계수란 집단내 계약자들의 리스크 속성을 반영한 수치화된 변수를 말하며, 해당 집단전체의 경험데이터 및 개별 경험데이터를 평점화(scoring)하는 과정이 필요하다. 본 연구에서는 로지스틱 회귀분석 결과의 예측확률을 이용하여 평점화한다. 모델 2로 추정된 예측확률에 다음과 같이 보정(scaling)을 하였다. 오즈(*Odds*)가 2배가 되는 점수 구간(Point to Double Odds, PDO)을 20점으로 설정한다. $\text{logit}(\ln(Odds))$ 의 상한과 하한의 차이에 PDO(20)를 곱하여 base score를 180으로 설정한다. 전체 평균점수를 50점으로 설정하기 위해 base odds를 0.93으로 설정한다.

$$Risk\ Score = offset + factor \times \ln\left(\frac{\pi}{1-\pi}\right), \pi: \text{로지스틱 회귀모형 예측확률.}$$

$$\begin{aligned} offset &= base\ score - factor \times \ln(base\ odds) \\ factor &= PDO(20)/\ln(2) \\ base\ score &= 180 \\ base\ odds &= 0.93 \end{aligned}$$

요율 차등화를 위해 보험계약 시 청약서상 고지항목으로 대변되는 현재 시점 연속무사고기간(Consecutive No-Claim Period)³⁾ 변수, $I(CNC, t\ years)$ 를 아래와 같이 정의한다. 즉, 청약서상 고지기간인 연속무사고기간이 짧을수록 고지기간 간소화 정도가 크다고 할 수 있다.

$$I(CNC, t\ years) = \begin{cases} 1 : \sum_{i=0}^{t-1} I(No-Claim, i^{th}\ year) = 0 \text{ and } I(No-Claim, t^{th}\ year) = 1 \\ 0 : otherwise \end{cases}$$

이제 Risk Score 값을 현재시점 연속무사고기간으로 등급화한 결과를 <Table 8>에 나타내었다. 전체 평균점수 50점 대비 최근 연속 0년간 무사고자(최근 1년간 사고자)의 Risk Score 값은 85점, 최근 5년간 입원 사고이력이 없는 경우 38점으로 큰 차이를 보였다.

건강보험 요율 차등화를 위한 고지기간 간소화 정도에 따른 기대의료비 차이에 대한 분

3) 여기서 현재시점 연속무사고기간은 개인별 현재부터 과거 가장 최근 입원(the most recent hospitalization)일까지의 무사고기간을 의미하며, 이는 앞서 회귀분석모델의 설명변수로 정의된 무사고기간별(현재~1년 전, ..., 5년 전~6년 전) 입원 존재 여부(existence of hospitalization by no-claim period for each 0~5 year period)와 차이가 있다.

석절차는 아래와 같다. 먼저 데이터 그룹을 분할하여 성별, 연령군별, 3대 질환 그룹별로 분할된 평균 급여의료비를 산출한다. 여기에 3대 질환 발생에 대한 이항 로지스틱 회귀분석 결과를 이용하여, 고지기간 간소화 정도에 따른 기대의료비 차이를 분석한다. 연간 기대의료비는 급여부분 법정 본인부담금과 비급여의료비를 합한 것으로 실손의료보험의 연간 위험보험료(annualized risk premium)에 대응된다. 연간 기대의료비 추정을 위해, 국민건강보험공단의 주요 질환에 대한 급여 및 비급여 비율⁴⁾을 활용하였다.

위험 집단 k 의 연간 기대의료비 (E_k)

$$= \sum_{i=1}^{N_k} \frac{p_i^k}{N_k} \times E(C_i) \times RC_i$$

여기서, i = 의료수급권자

k = (0년연속무사고, 1년연속무사고, ..., 5년연속무사고, 5년이상연속무사고) 그룹

N_k = 연속무사고기간 그룹(k)별 개체수

p_i^k = k 집단에 속한 의료수급권자 i 의 3대질환 발병여부에 대한 예측확률

C_i = i 의 급여의료비 (연도_y/성_s/연령군_a/3대질환_m 별)

$E(C_i)$ = i 의 기대급여의료비

RC_i = (급여대비) 비급여포함 본인부담비율 (연도_y/3대질환_m 별)

연도 = 2009, 2010, ..., 2013

성 = 1(남자), 2(여자)

연령군 = 25 ~ 29, 30 ~ 34, ..., 65 ~ 69

3대질환 = 암, 뇌혈관질환, 허혈성심장질환

분석결과(〈Table 8〉 참조), 무사고기간별 3대 질환 발병 예측확률(p)의 평균은 고지기간 간소화 정도가 클수록 커지며, 이는 앞서 분석모델에서 추정된 무사고기간 변수의 회귀

4) 표본코호트DB는 급여부분 공단부담금과 법정 본인부담금에 대한 자료만을 포함한다. 실손의료보험의 보장범위는 급여부분 법정 본인부담금과 비급여 본인부담금에 해당된다. 이에 따라 비급여 본인부담금은 국민건강보험공단 건강보험환자 진료비실태조사 보고서의 급여 및 비급여비율을 활용하여 추정하였다. 급여의료비 대비 본인부담금의 비율

($= \frac{\text{법정 본인부담률} + \text{비급여 본인부담률}}{\text{건강보험 보장률} + \text{법정 본인부담률}}$)을 〈Appendix Table 4〉로부터 산출하였다.

계수(β)의 크기순서와 동일한 것을 알 수 있다. 3대 질환에 대한 연간 기대의료비(E_k)는 고지기간이 짧을수록(k 가 작을수록) 증가하였으며, 최근 5년 동안 사고이력이 없는 경우에 비해 직전연도에 사고이력이 있는 경우의 연간 위험보험료(=연간 기대의료비) 차이는 92,859원(=117,753원-24,894원)으로 나타났다.

〈Table 8〉 Summary Statistics for Estimated Probability (p)
and Annualized Risk Premium (E_k)

(N=608,250)

Category		Mean ²⁾	p25 ²⁾	p75 ²⁾	Risk Score	E_k (KRW) ³⁾
Consecutive No-Claim Period (K) (From the Most Recent Hospitalization to Now)	0-year	0.0559	0.0152	0.0775	85	117,753
	1-year	0.0442	0.0116	0.0609	78	93,323
	2-year	0.0395	0.0105	0.0540	75	83,613
	3-year	0.0343	0.0092	0.0463	71	72,483
	4-year	0.0286	0.0079	0.0381	66	60,652
	5-year	0.0258	0.0074	0.0344	63	54,746
	Non-existence ¹⁾	0.0116	0.0028	0.0151	38	24,894
Total		0.0214	0.0041	0.0246	50	45,355

Notes: 1) Non-existence: There is no hospitalization history for five years.

2) Mean is the average of p , p25 its 25th percentile, and p75 its 75th percentile.

3) Expected medical expenses for the three major diseases (=risk premium)

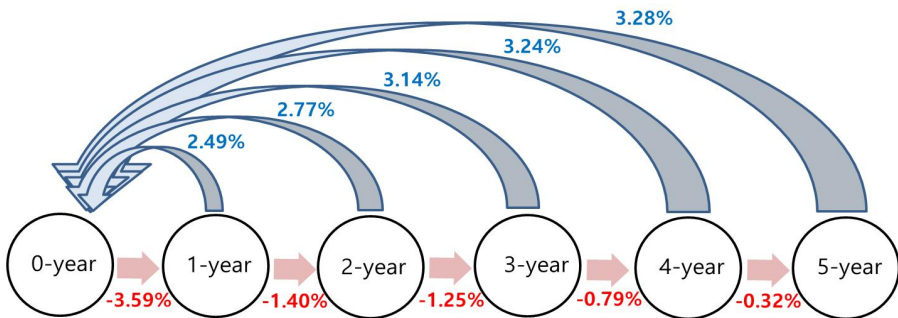
5. 할인할증제도(Bonus-Malus System) 요소

할인할증제도를 적용하기 위해 앞서 연속무사고기간 변수를 추가한 최적 모델 결과를 활용하여 예측 모델링을 실시하였다. 할인할증을 위한 예측 모델링의 과정은 다음과 같다. 첫째, 현재 시점 각 피보험자의 연속무사고기간 변수의 1년 뒤 전이행렬(transition matrix) 값을 구한다. 둘째, 각 전이행렬에 의해 변환된 연속무사고기간 변수를 새로운 입력변수로 설정한다. 셋째, 현재 시점 결정된 최적 모델로 1년 후 시점 연속무사고기간별 3대 질환 발병 예측확률(p)을 산출한다. 넷째, 각 전이행렬별 할인할증 요소로 피보험자별 예측확률(p)의 평균을 비교한다.

분석결과를 〈Figure 1〉에 나타내었다. 0년 연속무사고 구간의 경우 1년간 사고가 발생하지 않으면 할인율은 3.59%로 가장 컸으며, 4년 연속무사고 구간의 경우 1년간 사고가

발생하지 않으면 5년 연속무사고 구간에 적용되어 0.32%의 할인율을 적용받을 수 있다. 반대로 5년 연속무사고 구간의 1년간 사고 발생에 대한 할증률은 3.28%이며, 1년 연속무사고 구간의 1년간 사고 발생에 대한 할증률은 2.49%를 적용받게 된다. 한편 할인율은 매년 누적되어 5년 연속무사고 구간까지 단계별로 할인되지만, 할증률은 한 번 사고로 인해 0년 연속무사고 구간에 적용되므로 할인 폭보다 할증 폭이 더 크다고 볼 수 있다.

〈Figure 1〉 Markov-Chain Rate Change for Consecutive No-Claim Periods



본 연구는 보험계약자의 부분적 정보제공 조건을 고려하여 연속무사고기간을 상태변수로 정의한 후 할인할증 제도를 예시적으로 분석하였다. 이 경우 0년 연속무사고자, 1년 연속무사고자, ..., 5년 연속무사고자들의 각 집단이 전체에서 차지하는 비중과 예측확률을 고려하였기 때문에 전 계약집단의 보험료는 공평하다고 할 수 있으나, 연속무사고기간 이전의 사고정보 미비로 인하여 개별계약의 보험료는 공평하지 않을 수 있다. 반면에, 보험계약자의 완전한 정보제공 조건이 충족될 때에는 과거 사고기록(최근 5년) 중 각 해의 사고 유무들을 모두 모아 상태변수로 정의한 후 계약단위로 공평한 할인할증 제도를 설계할 수도 있다. 예를 들어, 0~1년 사이 0건, 1~2년 사이 1건, 2~3년 사이 1건, 3~4년 사이 0건, 4~5년 사이 0건인 경우 '(0 1 1 0 0)' 이런 식으로 상태를 정의하고 전이행렬과 할인할증 요율을 결정하는 것이다. 이 경우 연속무사고기간뿐만 아니라 과거의 모든 사고경험을 반영하기 때문에 개별계약별로도 보험료가 공평하다.

V. 결론

본 연구는 건강보험공단의 표본코호트DB를 활용하여 각종 건강정보, 기왕력 및 무사고 기간에 따른 3대질환 발병확률을 로지스틱 회귀모형으로 분석하였다. 또한, 고지기간 축소에 따른 기대의료비의 차이를 추정하였고, 할인할증 언더라이팅 시스템하에서 연속무사고기간 경험에 따라 보험료의 할인할증이 어떻게 달라지는지를 분석하였다.

본 연구는 보험계약상 고지의무 간소화에 대해 다음과 같은 분석결과를 도출하였다. 첫째, 건강보험 요율 차등화 시 보험계약상 고지의무 간소화 정도를 뜻하는 기왕력과 무사고기간 변수에 대한 필요성을 확인하였다.

둘째, 3대 질환(암, 뇌혈관질환, 허혈성 심장질환) 발병 확률은 기왕력(폐결핵, 고혈압, 고지혈증, 당뇨)과 입원 또는 외래에 대한 무사고기간에 대한 간소화 정도가 클수록 증가하였다. 기왕력에 대한 간소화 정도는 폐결핵, 고혈압, 고지혈증, 당뇨 순으로 해당 변수를 간소화 시 3대 질환 발병확률이 증가하는 것으로 나타났다. 입원에 대한 고지기간별 간소화는 그 정도가 클수록(무사고기간 5년→0년) 3대 질환 발병확률이 크게 높아지는 것으로 추정된다. 또한, 3년 이상 외래에 대한 고지항목 간소화는 3대 질환 발병 여부에 통계적으로 유의미한 영향이 있다고 보기 어렵다.

셋째, 3대 질환에 대한 연간 기대의료비는 고지기간이 축소될수록 크게 증가하는 것으로 나타났으며, 예측모델링을 활용한 할인할증률은 현재 0년 연속무사고 구간에서 할인폭이 가장 큰 것으로 나타났다.

또한, 요율 차등화는 고지기간별 위험집단에 대한 위험 계량화를 통해 저위험군은 더 저렴한 보험 가입이 가능하며, 인수 거절되었던 고위험군은 보장 혜택을 누릴 수 있다. 또한, 보험사는 효율적인 언더라이팅이 가능하며, 보험계약자가 스스로 리스크를 감소하는 유인책이 되어 사회 전체적인 효용도 증대시킬 수 있을 것이다. 나아가 향후 연구에서는 포괄적인 건강보험의 고지의무 간소화를 위한 다양한 요율 차등화 분석이 이루어지기를 기대해본다.

Appendix

〈Appendix Table 1〉 Description of Explanatory Variables¹⁾

Variable	Dummy Variable	Unit	Category of Dummy Variables				
Sex	X ₁		1 (Female), 2 (Male)				
Age_Group	X ₂ ~X ₁₀		25~29 (X ₂), 30~34 (X ₃), 35~39 (X ₄), 40~44 (X ₅), 45~49 (X ₆), 50~54 (X ₇), 55~59 (X ₈), 60~64 (X ₉), 65~69 (X ₁₀)				
Region	X ₁₁ ~X ₁₂		Seoul and Six Major Metropolitan Cities (X ₁₁), Provincial Cities (X ₁₂)				
Income Decile Groups	X ₁₃ ~X ₁₄		Lower (0~5) Income Group (0: Medical Benefit Recipient) (X ₁₃), Higher (6~10) Income Group (X ₁₄)				
CCI	X ₁₅ ~X ₁₈		0 point (X ₁₅), 1 point (X ₁₆), 2 point (X ₁₇), more than 3 point (X ₁₈)				
No-Claim Period	X ₁₉ ~X ₂₄		No-Claim Period of Hospitalization ²⁾ 0-year (X ₁₉), 1-year (X ₂₀), 2-year (X ₂₁), 3-year (X ₂₂), 4-year (X ₂₃), 5-year (X ₂₄) Value: 0 (Non-existence), 1 (Existence)				
Body Mass Index (BMI)	X ₂₅	Kg/m ²	Weight/(Height/100) ²				
Systolic Blood Pressure (SBP)	X ₂₆	mm Hg		Optimal	PreHypertension	Hypertension	
			value	<120	120~139	≥140	
Diastolic Blood Pressure (DBP)	X ₂₇	mm Hg		Optimal	PreHypertension	Hypertension	
			value	<80	80~89	≥90	
Fasting Blood Sugar	X ₂₈	mg/dL		Optimal	Prediabetes	diabetes	
			value	<100	100~125	≥126	
Triglyceride	X ₂₉	mg/dL		Optimal	Intermediate	High	Very High
			value	<150	150~199	200~499	≥500
Non-High Density Lipoprotein Cholesterol (Non-HDL)	X ₃₀	mg/dL		Optimal	Intermediate	High	Very High
			value	<130	130~159	160~189	≥190
High Density Lipoprotein Cholesterol (HDL)	X ₃₁	mg/dL		Optimal	Intermediate	High	
			value	≥60	40~59	<40	

Variable	Dummy Variable	Unit	Category of Dummy Variables				
Aspartate Aminotransferase (AST, SGOT)	X ₃₂	U/L		Optimal	Inter-mediate	High	
			value	≤ 40	41~50	>50	
Alanine Aminotransferase (ALT, SGPT)	X ₃₃	U/L		Optimal	Inter-mediate	High	
			value	≤ 35	36~45	>45	
Glutamyl Transpeptidas (γ-GTP)	X ₃₄	U/L			Optimal	Inter-mediate	High
			value	M	11~63	64~77	≥ 78
				W	8~35	36~45	≥ 46
History of Hypertension	X ₃₅		Value: 0 (Non-Existence), 1 (Existence)				
History of Diabetes	X ₃₆						
History of Hyperlipidemia	X ₃₇						
History of Pulmonary Tuberculosis	X ₃₈						
Consecutive No-Claim Period	X ₃₉ ~X ₄₅		Consecutive No-Claim Period from the Most Recent Hospitalization to Now 0-year (X ₃₁), 1-year (X ₃₂), 2-year (X ₃₃), 3-year (X ₃₄), 4-year (X ₃₅), 5-year (X ₃₆) Value: 0 (Non-existence), 1 (Existence)				

Notes: 1) KMI Korea Medical Institute and Ahn(2016); National Health Insurance Service (2020).

- 2) If at least one hospitalization is observed in 2008, no-claim period 0-year is set to '1' for 2009. If at least one hospitalization is observed in 2007, no-claim period 1-year is set to '1' for 2009.

〈Appendix Table 2〉 Pearson Correlation Coefficient

	BMI	SBP	DBP	Fasting Blood Sugar	Triglyceride
BMI	1.00000	0.31763	0.29263	0.17802	0.32048
SBP		1.00000	0.78240	0.19408	0.22121
DBP			1.00000	0.16013	0.22401
Fasting Blood Sugar				1.00000	0.21073
Triglyceride					1.00000

	Non-HDL	HDL	AST	ALT	γ -GTP
BMI	0.29094	-0.29067	0.15249	0.29699	0.18511
SBP	0.16485	-0.10047	0.13126	0.15465	0.18602
DBP	0.16613	-0.09352	0.12568	0.15629	0.18872
Fasting Blood Sugar	0.12861	-0.09911	0.10663	0.13946	0.18486
Triglyceride	0.41432	-0.36571	0.17041	0.25191	0.32036
Non-HDL	1.00000	-0.22947	0.08717	0.17488	0.13052
HDL		1.00000	-0.04474	-0.15320	-0.05603
AST			1.00000	0.76014	0.46709
ALT				1.00000	0.44336
γ -GTP					1.00000

〈Appendix Table 3〉 The Three Major Diseases-related Covered Diseases

Disease Classification	KCD-code
Cancer-related Covered Diseases	C00~C14, C15~C26, C30~C39, C40~C41, C43~C44, C45~C49, C50, C51~C58, C60~C63, C64~C68, C69~C72, C73~C75, C76~C80, C81~C96, C97, D45, D46, D47.1, D47.3, D47.4, D47.5
Cerebrovascular-related Covered Diseases	I60~I69
Ischaemic Heart-related Covered Diseases	I20~I25

Note: The diseases are classified according to Statistics Korea(2015).

〈Appendix Table 4〉 Coverage Rate & Co-payment Rate for the Three Major Diseases¹⁾

The Three Major Diseases		Benefit Coverage Rate by National Health Insurance	Co-payment Rate by Patients' Share	Co-payment Rate for Coverage Exclusion
Cancer Related Diseases ²⁾	2009	67.9%	5.4%	26.7%
	2010	72.4%	6.7%	20.9%
	2011	72.6%	5.8%	21.6%
	2012	74.1%	6.2%	19.7%
	2013	72.7%	6.7%	20.6%
Cerebrovascular Related Diseases ³⁾	2009	60.9%	12.6%	26.6%
	2010	72.2%	7.1%	20.8%
	2011	70.2%	7.1%	22.7%
	2012	73.8%	5.7%	20.5%
	2013	74.4%	6.3%	19.3%
Heart Related Diseases ⁴⁾	2009	63.8%	9.2%	27.0%
	2010	73.0%	6.8%	20.2%
	2011	73.5%	5.3%	21.1%
	2012	74.8%	5.1%	20.1%
	2013	78.0%	5.6%	16.4%

Notes: 1) NHIS(National Health Insurance Service) Health Insurance Policy Research Institute, Survey on The Benefit Coverage Rate of National Health Insurance (2009; 2013; 2015).

2) Cancer related diseases: C00~C97, D00~D09, D32~D33, D37~D48

3) Cerebrovascular related diseases: I60~I67, I720, I77, Q280~Q283, S06

4) Heart related diseases: D151, I01, I05~I09, I20~I26, I28, I30~I51, I700, I71, I790~I791, M314, Q20~Q25, Q260~Q264, Q268, Q269, S25~S26

참고문헌

- 강성호·김대환·이순재 (2019), “고령층의 민영건강보험 수요결정요인과 잠재적 수요 분석”, **보험금융연구**, 제30권 제1호, pp. 107-140.
- (Translated in English) Kang, S., D., Kim and S., Lee (2019). “Estimation of Determinants and Potential Demand for Private Health Insurance for the Elderly in Korea”, *Journal of Insurance and Finance*, 30(1):107-140.
- 국민건강보험공단 (2020), **건강검진 실시기준 및 운영세칙**, 건강관리기관포털 (Retrieved from <https://sis.nhis.or.kr>).
- (Translated in English) National Health Insurance Service (2020). *Physical Examination Implementation Standards and Operating Rules*, Health Care Organization Portal (Retrieved from <https://sis.nhis.or.kr>).
- 국민건강보험공단 건강보험정책연구원 (2009; 2013; 2015), **건강보험환자 진료비 실태조사 보고서**.
- (Translated in English) NHIS (National Health Insurance Service) Health Insurance Policy Research Institute (2009; 2013; 2015). *Survey on The Benefit Coverage Rate of National Health Insurance*.
- 권태연·임자영·박유성 (2017), “국민건강보험 표본코호트DB를 이용한 한국인의 건강기대수명 연구”, **응용통계연구**, 제30권 제3호, 한국통계학회, pp. 475-486.
- (Translated in English) Kwon, T., J., Lim and Y., Park (2017). “Health Life Expectancy in Korea based on Sample Cohort Database of National Health Insurance Services”, *The Korean Journal of Applied Statistics*, 30(3):475-486.
- 김경훈 (2010), “Charlson 동반질환의 ICD-10 알고리즘 예측력 비교연구”, **예방의학회지**, 제43권 제1호, pp. 42-49.
- (Translated in English) Kim, K. (2010). “Comparative Study on Three Algorithms of the ICD-10 Charlson Comorbidity Index with

Myocardial Infarction Patients”, *Journal of Preventive Medicine and Public Health*, 43(1):42-49.

김대환 (2014), “실손의료보험이 의료수요에 미치는 영향”, **보험학회지**, 제98집, 한국보험학회, pp. 61-90.

(Translated in English) Kim, D. (2014). “Effects of Fee-for-service Health Insurance on Medical Consumption”, *Korean Insurance Journal*, 98:61-90.

김대환·이봉주 (2013), “실손의료보험 역선택 분석”, **보험학회지**, 제96집, 한국보험학회, pp. 25-50.

(Translated in English) Kim, D., and B., Lee (2013). “An Analysis on Adverse Selection in Fee-For Service Health Insurance”, *Korean Insurance Journal*, 96:25-50.

김동욱·김동욱·이선미·임현선·최정규·박해용·육태미·강민진·홍정화·한규태·배세진 (2017), **건강보험 청구자료에 근거한 질병의 조작적 정의에 관한 연구**, 국민건강보험 일산병원 연구소.

(Translated in English) Kim, D., et al. (2017). *A Study on the Operational Definition of Diseases Based on Health Insurance Claims Data*, NHIS(National Health Insurance Service) Ilsan Hospital Institute of Health Insurance & Clinical Research.

김석영·변혜원(2014), **간편심사 보험상품 활성화 방안**, 보험연구원.

(Translated in English) Kim, S., and H., Byun (2014). *Improving and Enhancing the Simplified Issue Insurance Product*, The Korea Insurance Research Institute.

김송희·김윤희 (2019), “민영보험 가입 여부에 따른 당뇨 유병자의 건강관리 차이 분석”, **보험금융연구**, 제30권 제1호, pp. 39-69.

(Translated in English) Kim, S., and E., Kim (2019). “Analysis on Health Behavior Difference in Diabetic Patients by Private Medical

Insurance”, *Journal of Insurance and Finance*, 30(1):39-69.

김혜련·강영호·박은자·최정수·이연희·김영삼 (2006), **한국인의 사망과 질병 및 의료이용의 요인분석과 정책과제**, 한국보건사회연구원, pp. 47-56; 73-77.

(Translated in English) Kim, H., et al. (2006). *Socioeconomic, Behavioral, Nutritional, and Biological Determinants of Morbidity, Mortality, and Medical Utilization in South Korea*, Korea Institute for Health and Social Affairs. pp. 47-56; 73-77.

박상준 (2003), “언더라이팅 기능 강화를 위한 신기법 도입방안”, **보험의학회지**, 제22권, 보험의학회, pp. 99-120.

(Translated in English) Park, S. (2003). “A Plan for Introduction of New Performance-Enhancing Underwriting Technique”, *Journal of Korean Life Insurance Medical Association*, 22:99-120.

오영수·이경희 (2003), “민영건강보험의 언더라이팅 선진화 방안”, **보험연구원 연구보고서**, 보험연구원.

(Translated in English) Oh, Y., and K., Lee (2003). “Improvement Plan of Private Health Insurance Underwriting”, *KIRI Research Paper*.

오창수·유동완 (2018), “민영건강보험의 가입 결정요인에 관한 연구”, **보험학회지**, 제114집, 한국보험학회, pp. 35-56.

(Translated in English) Ouh, C., and D., You (2018). “A Study on the Determinants of Private Health Insurance”, *Korean Insurance Journal*, 114:35-56.

이경아·이항석 (2016), “실손의료보험의 역선택과 보험료 차등화”, **리스크관리연구**, 제27권 제3호, 리스크관리학회, pp. 93-128.

(Translated in English) Lee, K., and H., Lee (2016). “Adverse Selection and Rates Differentiation in the Korean Private Health Insurance”, *Journal of Risk Management*. 27(3):93-128.

이준영·김기환·이지성 (2014), **국민건강정보 데이터베이스를 이용한 표본 코호트 DB**

구축, 국민건강보험공단.

(Translated in English) Lee, J., K., Kim, and J., Lee (2014). *Construction of an Appropriate Sampling Design and a Sample Database using the National Health Information Database*, National Health Insurance Corporation.

이찬희·전희주·김석영 (2015), “암보험 언더라이팅 효과 실증분석 - 국민 암발생률과 보험회사 암발생률 비교 -”, **보험학회지**, 제102집, 한국보험학회, pp. 59-85.

(Translated in English) Lee, C., H., Chun and S., Kim (2015). “An Empirical Study on the Underwriting Effect on Cancer Coverage - Comparing National Cancer Incidence Rate with Cancer Incidence Rate of Insurer -”, *Korean Insurance Journal*, 102:59-85.

이항석·이가은·이경아 (2017), “실손의료보험에서 요율산정과 일반화선형모형(GLM)의 활용”, **리스크관리연구**, 제28권 제4호, 리스크관리학회, pp. 37-71.

(Translated in English) Lee, H., G., Lee and K., Lee (2017). “A Ratemaking of the Korean Private Health Insurance using GLM”, *Journal of Risk Management*. 28(4):37-71.

이항석·이민하·백혜연 (2018), “실손의료보험 할인할증제도의 실증분석”, **보험학회지**, 제116집, 한국보험학회, pp. 33-66.

(Translated in English) Lee, H., M., Lee and H., Baek (2018). “An Empirical Analysis of Bonus-Malus System in Private Health Insurance”, *Korean Insurance Journal*, 116:33-66.

이현복·정홍주 (2016), “실손의료보험 보장범위 축소에 따른 가입자 의료이용 변화에 관한 연구”, **보험학회지**, 제106집, 한국보험학회, pp. 67-91.

(Translated in English) Lee, H., and H., Jung (2016). “An Empirical Study of Healthcare Utilization Under Indemnity Health Insurance Coverage Changes”, *Korean Insurance Journal*, 106:67-91.

이현복·황연희 (2014), “민영건강보험 가입자의 보험금 수령 결정요인에 대한 연구”, **사**

회보장연구, 제30권, 한국사회보장학회, pp. 171-193.

(Translated in English) Lee, H., and Y., Hwang (2014). “A Study Identifying the Factors Determining Private Health Insurance Claims Paid”, *Korean Social Security Studies*, 30:171-193.

임형기 (2021), “청구경력 고지 간소화에 따른 건강보험 요율 산정에 관한 연구 - 상대위험도 비교를 중심으로 -”, 박사학위논문, 한양대학교 대학원.

(Translated in English) Im, H. (2021). “A Study on the Ratemaking of Health Insurance by Simplified Disclosure of Claim History - Focused on the Comparison of Relative Risk -”, Ph.D. Dissertation, Graduate School of Hanyang University.

통계청 (2015), **제7차 한국표준질병사인분류(KCD-7)**, 제2015-309호, 2016. 1. 1.

(Translated in English) Statistics Korea (2015). The 7th Korean Standard *Classification of Diseases (KCD-7)*, No. 2015-309, 2016. 1. 1.

KMI한국의학연구소·안지현 (2016), **건강검진 사용 설명서**, 사이프레스.

(Translated in English) KMI Korea Medical Institute and J., Ahn (2016). *Manual for Physical Examination*, Cypress.

Akaike H. (1974). “A New Look at the Statistical Model Identification”, *I.E.E.E. Transactions on Automatic Control*, 19(6):716-723.

Donnelly, P. (2011). “Enhanced Simplified Underwriting”, *Munich Re*.

Ho, C. (2013). “U.S. Individual Life Guaranteed and Simplified Issue Persistency”, *LIMLA*.

Kang, Y. et al. (2018). “Biological Age as a Health Index for Mortality and Major Age-related Disease Incidence in Koreans: National Health Insurance Service - Health Screening 11-year Follow-up Study”, *Clinical Interventions in Aging*, 13:429-436.

Kennedy, P. (1992). *A Guide to Econometrics*. Oxford: Blackwell.

LIFE HAPPENS and LIMRA (2019). *2019 Insurance Barometer Report*.

Milliman (2011). "Application of Credibility Theory to Group Life Pricing",
Research Report.

Schwarz, G. (1978). "Estimating the Dimension of a Model", *Annals of Statistics*, 6:461-464.

Quan H. et al. (2005). "Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data", *MedCare*, 43(11):1130-1139.

Abstract

The simplification of the obligation to disclose the claim history of private health insurance in Korea has been developed for the purpose of expanding the market for individuals whose coverage denied by health insurance underwriting due to a mild chronic illness or treatment history. Following policy change, the launch of simplified issue insurance has been limited to fixed benefit products and supplementary private health insurances for those with pre-existing conditions in the private sector. The simplified disclosure obligation is a major factor in determining rates aside from gender and age; however, it does not reflect subtle differences in rate over the period of the disclosed claim history.

This study conducted an empirical analysis through logistic regression using the National Health Insurance Sample Cohort Database to investigate change in healthcare utilization according to the degree of simplification for disclosure of claim history in insurance contract. The results suggest that a higher degree of simplification of disclosure increases the probability of healthcare utilization at a statistically significant level, and inpatients are relatively more affected than outpatients. In addition, the annual premium rate for private health insurance and the level of bonus-malus were derived by differentiating the degree of simplification of disclosure of claim history.

※ **Key words:** Disclosure of Claim, Simplified Issue, Premium Rate Differentiation, Health Insurance, Bonus-Malus System

