

## 표본코호트기반 고지혈증 약제의 저밀도 콜레스테롤 감소량 및 투약순응도 분석<sup>†</sup>

김규진<sup>1</sup> · 전치혁<sup>2</sup> · 이혜선<sup>3</sup> · 김현성<sup>4</sup>

<sup>1,2,3</sup>포항공과대학교 산업경영공학과 · <sup>4</sup>가톨릭대학교 의과대학 의료정보학과

접수 2017년 9월 6일, 수정 2017년 9월 19일, 게재확정 2017년 9월 21일

### 요약

고지혈증은 혈액 중에 지방 성분이 필요이상으로 많은 상태를 의미하며, 특히 저밀도 콜레스테롤이 혈관벽에 달라붙어 심근경색증, 뇌경색 등의 다양한 심혈관계의 질병을 발생시킬 수 있다. 스타틴은 대표적인 고지혈증 치방제로서 치방 받는 환자의 개별적인 특성 및 건강관리행태, 투약순응도 등에 따라 그 효과가 달라진다. 스타틴의 주요 효과는 저밀도 콜레스테롤 수치를 낮추는 것인데, 이는 National cholesterol education program-adult treatment panel (NCEP-ATP III) 가이드라인에서 환자의 조건에 따라 정한 목표 수치에 도달해야 한다. 본 연구에서는 저밀도 콜레스테롤 수치의 감소량과 환자 별 목표 수치 도달여부를 각각 스타틴의 효과로 상정하고, 국민건강보험공단에서 구축한 표본코호트 DB를 이용하여 건강검진기록의 개별특성 (나이, 성별, 흡연, 운동 및 혈액검사결과)과 치방전 기록으로부터 투약순응도를 통합해서 저밀도 콜레스테롤 감소량 및 목표도달률에 미치는 영향을 분석하고자 한다.

주요용어: 고지혈증, 스타틴, 저밀도 콜레스테롤, 투약순응도, 표본코호트.

### 1. 서론

혈액 속에 콜레스테롤이나 중성지방등과 같은 지질이 많을 때 고지혈증이라고 하며 혈중 콜레스테롤이 240mg/dL 이상일때 고콜레스테롤혈증, 혈중 중성지방이 200mg/dL 이상일때 고 중성지방혈증이라고 한다. Boyd 등 (1989)과 Han (2009)에 의하면 고지혈증은 동맥경화증의 가장 흔한 원인 중 하나인데, 콜레스테롤 중 1.019 ~ 1.063 kg/L 분획 사이에 존재하는 저밀도 콜레스테롤은 혈중 농도가 높아지면 산화된 저밀도 콜레스테롤이 혈관벽을 자극하여 동맥경화증과 같은 심혈관 질환을 유발할 수 있고, 그와 반대로 고밀도 콜레스테롤은 혈중농도가 1mg/dL 높아질 때마다 심혈관질환을 2 ~ 3% 정도 줄일 수 있는 것으로 알려져 있다. 고지혈증에는 일반적으로 스타틴을 치방하는데, 스타틴은 HMG CoA 환원제 억제제로서 고콜레스테롤혈증이 있는 사람에게 1차 약물로 널리 선택되어 사용되고 있으며 고지혈증 치료의 1차목표인 저밀도 콜레스테롤의 혈중 농도를 18 ~ 55%까지 낮출 수 있는 것으로 알려져 있

<sup>†</sup> 본 연구는 보건복지부의 재원으로 한국보건산업진흥원의 보건의료기술연구개발사업 지원에 의하여 이루어진 것임(HC15C1362).

<sup>1</sup> (37673) 경북 포항시 남구 청암로 77, 포항공과대학교 산업경영공학과, 석사과정.

<sup>2</sup> (37673) 경북 포항시 남구 청암로 77, 포항공과대학교 산업경영공학과, 교수.

<sup>3</sup> 교신저자: (37673) 경북 포항시 남구 청암로 77, 포항공과대학교 산업경영공학과, 연구교수.

E-mail : [hyelee@postech.ac.kr](mailto:hyelee@postech.ac.kr)

<sup>4</sup> 공동교신저자: (06591) 서울특별시 서초구 반포동 505, 가톨릭대학교 의과대학 의료정보학과 연구조교수.

다 (NCEP ATPIII). 현재까지 다양한 종류의 스타틴이 처방되고 있으며, 이에 대한 경제성 및 복용 효과에 대해 많은 연구가 진행되어 왔다. Kim 등 (2016), Jeong (2017), Jeong 등 (2004) 참조. 하지만 아직까지 스타틴 약물에 대한 환자의 투약순응도가 저밀도 콜레스테롤 감소량에 미치는 영향에 대해서는 연구가 이루어진 바 없다. 이번 연구에서는 건강보험공단에서 제공하는 코호트 DB를 이용하였고, 투약 순응도와 스타틴 처방으로 인한 저밀도 콜레스테롤의 감소량 간의 상관관계를 밝히는 점에서 의의가 있다.

## 2. 연구 방법

### 2.1. 표본코호트 DB 구조

본 연구에 사용된 표본코호트DB는 국민건강보험공단의 ‘국민건강정보 DB’에 근간하여 2002년부터 2013년까지 12년간 건강보험 및 의료급여권자 전체에 대한 진료명세서와 진료내역, 상병내역, 처방전내역 등을 포함하고 있고 청구일 중심으로 수집되었던 원자료를 진료개시일 중심으로 변환한 것이다. 크게는 자격 DB, 진료 DB 및 검진 DB 등으로 구성되어 있고 전체 인구를 대상으로 비례할당 충화하여 추출률 2.2%의 무작위 표본 추출로 구축하였다. 코호트는 그 특성 상 시간이 지남에 따라 사망 등 자격 상실로 인해 코호트 크기가 줄어들게 되지만, 연도별 신생아의 표본을 새로운 코호트로 포함하기 때문에 코호트의 크기가 급격히 감소하지는 않는다. 각 DB는 개인식별 아이디 (Person\_id)를 공통키 값으로 각 DB를 연계하여 분석용 데이터를 구성할 수 있다. 자격DB는 사람의 성별, 나이, 소득분위 등 개인정보를 포함한다. 진료DB는 개인의 질병 (주상병/부상병)과 같은 진료기록과 약의 코드명이나 1일투여량과 같은 처방항목을 포함하며, 마지막으로 건강검진 DB는 개인의 신장, LDL콜레스테롤수치, 혈압과 같은 건강검진 항목을 포함한다. 건강검진 DB의 경우엔 2009년부터 레이아웃이 바뀌는데, 콜레스테롤 항목에 저밀도 콜레스테롤과 고밀도 콜레스테롤을 추가하는 등 기존의 37개 항목에서 41개 항목으로 증가한다. Figure 2.1은 각 DB의 정보와 연결을 나타낸다.

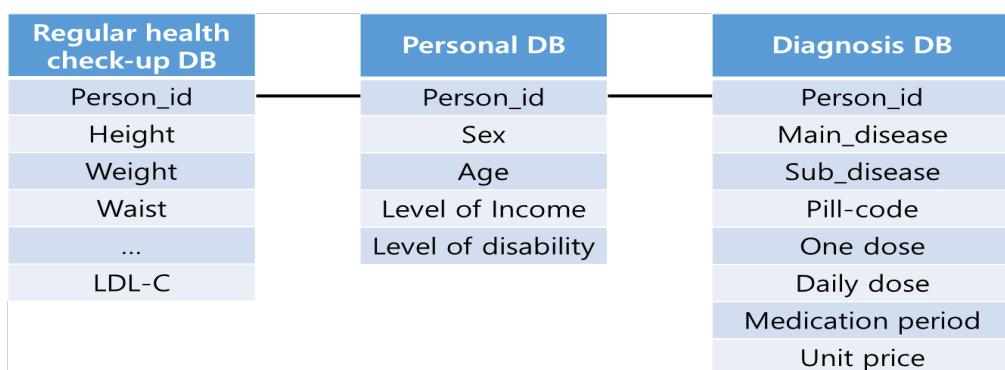


Figure 2.1 Relations of DB

### 2.2. 기술적 통계분석

건강검진 인구의 비율은 11%에서 25%에 근접하게 늘어났다. 건강관리에 대한 사람들의 관심이 높아지고 있음을 나타낸다 (Figure 2.2 참고). 또한 연도별로 어떤 약물이든 처방받은 사람을 봤을 때, 자격

인구 대비 처방자의 비율은 매년 소량씩 증가하여 85%를 넘어감을 볼 수 있다 (Figure 2.3 참고). 특히 스타틴 처방을 고려한다면, 약물 처방자 대비 스타틴 처방자의 비율은 2002년도 기준 1%에서 2013년도 엔 8.5% 까지 약 8배가량 증가함을 알 수 있다 (Figure 2.4 참고). 이와 같이 약물 처방자와 스타틴 처방자의 수는 모두 증가하고 있지만 스타틴이 더 큰 증가율을 보이는 것으로부터 스타틴에 대한 관심이 매우 증가하고 있음을 알 수 있다. 일반인의 전체 콜레스테롤 평균값은 큰 변화없이 유지되는 데에 비해 스타틴 처방자의 콜레스테롤 평균값은 점차 감소하는 추세임을 볼 수 있다. 이는 스타틴의 효과가 작용하고 있음을 보여준다 (Figure 2.5 참고). 고지혈증 환자 수는 매년 꾸준히 증가하며 (Figure 2.6 참고), 고지혈증 환자 중 스타틴을 처방받는 비율 또한 05년도 이후부터 꾸준히 증가하여 90%에 육박한다 (Figure 2.7 참고).

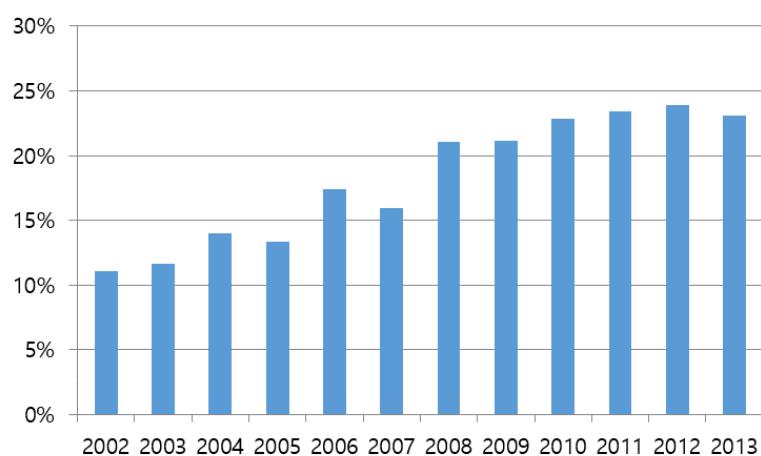


Figure 2.2 Ratio of regular check-up

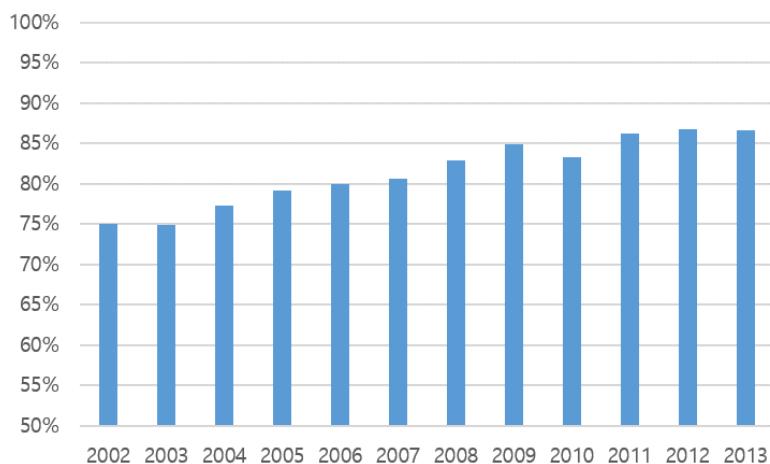
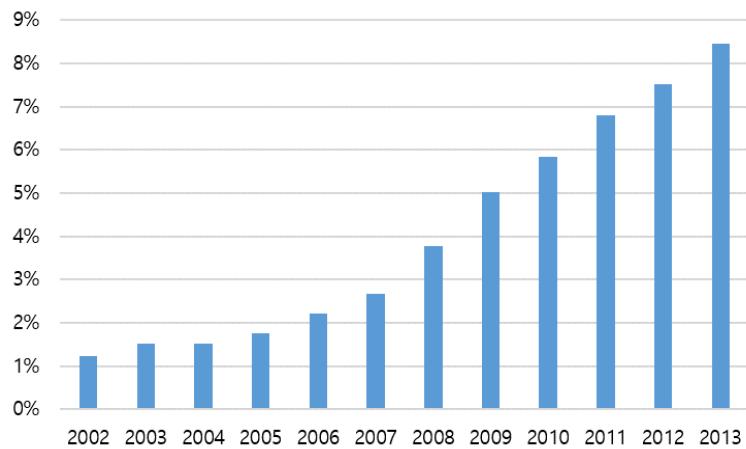
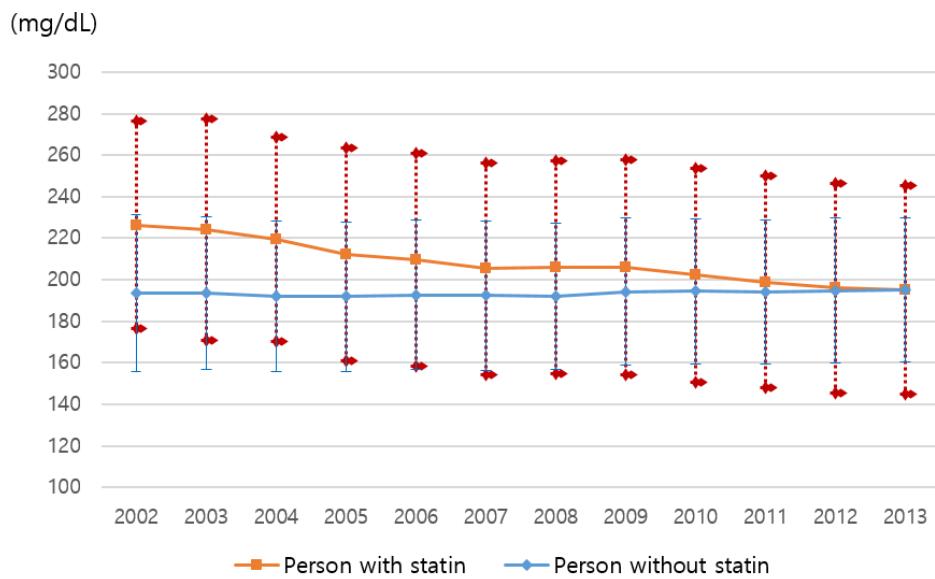


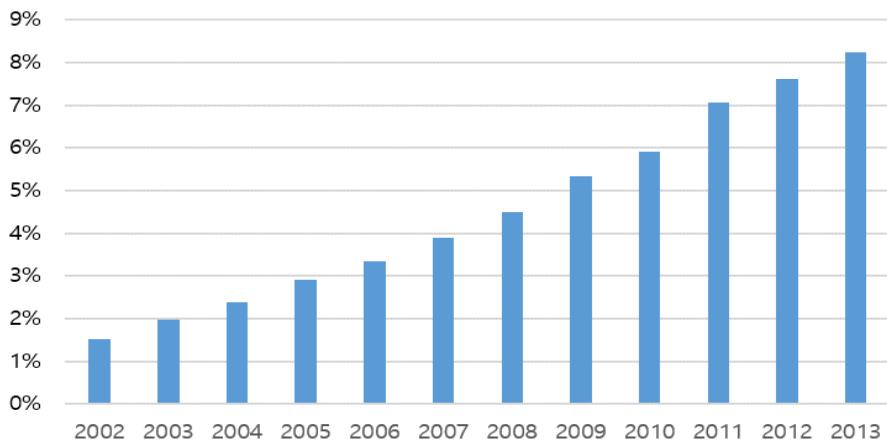
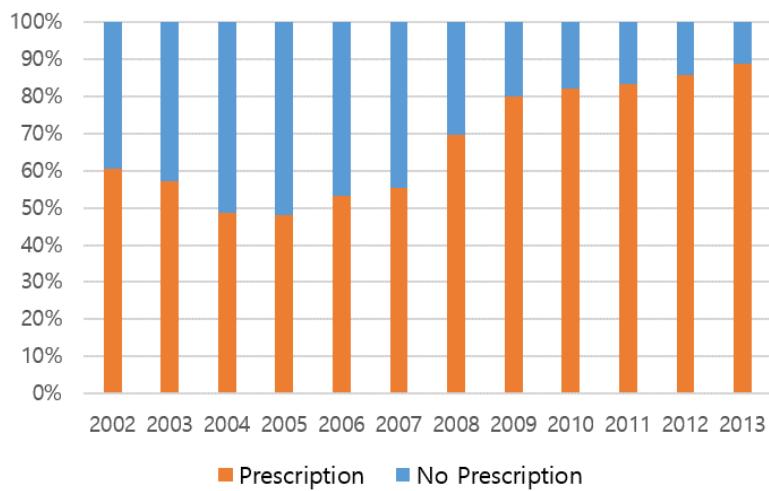
Figure 2.3 Ratio of patients who got medicine prescription



**Figure 2.4** Ratio of patients who got statin prescription out of medicine prescription



**Figure 2.5** Average of total cholesterol

**Figure 2.6** Ratio of hyperlipidemia**Figure 2.7** Percentage of prescription of hyperlipidemia

### 2.3. 분석용 데이터 구성

#### 2.3.1. 분석 대상 조건 필터링 및 이상치 보완

본 연구는 환자가 스타틴을 복용하여 얻는 지질개선효과를 바탕으로 한다. 그러나 Hong 등 (2003)에 의하면 스타틴은 지질개선효과 외에도 다른 효과들이 입증되었기 때문에, 고지혈증이 아니더라도 스타틴을 복용할 수 있다. 따라서 본 연구는 스타틴 복용자 중에 지질개선이 목적인 고지혈증 환자를 대상으로 한다. 고지혈증 환자는 진료DB의 주상병 (main\_disease) 혹은 부상병 (sub\_disease) 항목에 한국표준질병사인분류(Korean standard classification of diseases; KCD)에 의거한 고지혈증 상병코드

(‘E78’, ‘E780’, ‘E781’, ‘E782’, ‘E783’, ‘E784’, ‘E785’, ‘E786’, ‘E787’, ‘E788’, ‘E789’) 중 하나 이상 포함되는 경우로 정의한다. 기간은 ‘스타틴 복용 전 → 스타틴 복용 → 복용 후 효과측정’과 같이 세 기간으로 구분한다. Kim (2010)에 의하면 스타틴은 꾸준히 복용해야 하는 약물이고 복용을 중단하고 3개월만 지나도 저밀도 콜레스테롤 수치가 40% 증가하는 것으로 알려져 있다. 따라서 본 연구에서는 스타틴 복용년도를 기준으로 최근 6개월동안 복용하지 않았으면 미복용으로 간주한다. 스타틴 복용년도는 스타틴의 복용효과로서 저밀도 콜레스테롤 변화량을 알 수 있는 연도이어야 한다. 저밀도 콜레스테롤 수치는 행정상 2008년부터 측정하기 시작했지만, 결측치가 많았고, 데이터가 제대로 쌓이기 시작한 연도는 2010년도이므로, 2011년도를 스타틴 복용년도로 지정한다. 그 외에 저밀도 콜레스테롤이 500mg/dL 이상인 경우, 이상치로 간주하고 제거한다 (Figure 2.8 참고).

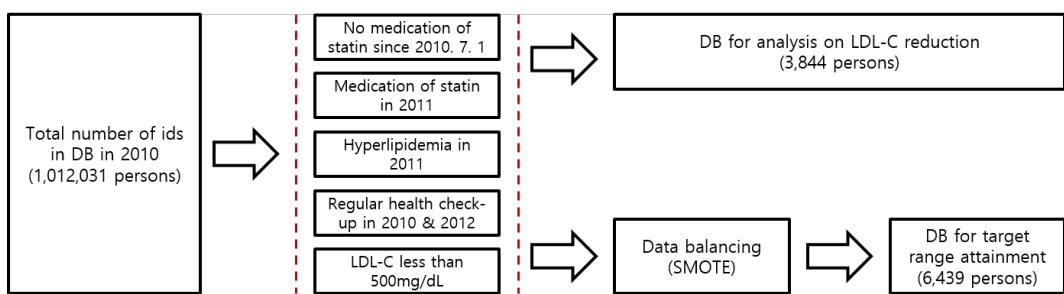


Figure 2.8 Data preprocessing

### 2.3.2. 독립변수와 종속변수의 정의

본 연구에서 투약순응도와 위험군, 그리고 저밀도 콜레스테롤 감소량과 목표도달여부을 계산하여 새로운 변수로서 도입한다. 투약순응도는 환자가 약을 꾸준히 복용했는지를 가늠할 수 있는 변수인데, 진료DB(Diagnosis DB)에 있는 처방전의 총 투여일수를 합산하여 분자로 놓고 투약권장일 수를 분모로 나누어 계산한다. 투약권장일 수는 환자가 투약해야하는 기간인데, 복용년도 (2011년도) 이전년도부터 고지혈증을 이어왔다면 365일로 간주하고 복용년도에 새롭게 고지혈증을 판정을 받았다면 당해 연도에 스타틴을 처음 처방받은 날부터 연말까지의 기간을 계산하여 투약권장일수로 간주한다. US department of health and human services (2001)에 의하면 위험군은 NCEP-ATP III 가이드라인에서 지정한 기준을 따르면 고위험군, 중위험군, 저위험군 총 세 가지로 구분된다. 스타틴의 효과는 총 두 가지로 평가하는데, 저밀도 콜레스테롤 수치의 감소량과 위험군에 따른 목표수치 도달여부이다. 우선 저밀도 콜레스테롤 감소량은 연속값을 가지고 약을 복용하기 전과 후의 감소량을 측정한다. 복용 후에 증가했을 경우엔 저밀도 콜레스테롤 감소량은 음수 값을 가진다. 목표수치 도달여부는 이진수값 (0, 1)으로, NCEP-ATP III 가이드라인에 의해 고위험군의 목표는 복용 후 저밀도 콜레스테롤수치 100mg/dL 미만, 중위험군의 목표는 복용 후 저밀도 콜레스테롤 수치 130mg/dL 미만, 마지막으로 저위험군은 복용 후 저밀도 콜레스테롤 수치가 160mg/dL 미만인 경우 목표에 도달한 것으로 판단하였다.

위에서 계산한 독립변수와 종속변수를 걸러진 환자 데이터에 붙여 분석에 필요한 DB를 생성한다. 독립변수로는 복용 전 연도의 건강검진 데이터, 자격 데이터 및 투약순응도를 계산하여 설정하고, 종속변수로는 위에서 설명한 저밀도 콜레스테롤 감소량과 목표도달여부로 설정한다. 독립변수는 분석 방법을 적용하기 전에 모두 표준화 (normalization) 작업을 거친다.

### 3. 분석 방법

다음의 설명은 본 연구에서 사용한 분석 방법들이다. 분석 도구는 R3.4.0의 lars package, glmnet package, smotefamily package 및 covTest package를 사용하였다.

#### 3.1. SMOTE 알고리즘

데이터를 분류 모형에 사용할 때에 레이블이 한 가지로 지나치게 편중될 경우 모델을 학습시키면 예측 결과가 한 쪽으로 쏠리는 등 신뢰하기 어려운 결과가 나올 수 있다. 데이터를 벨런싱해주는 작업이 필요한데 그 방법으로는 SMOTE (synthetic minority over-sampling technique), ADASYN (adaptive synthetic sampling) 등이 있다. He 와 Garcia (2009) 참고. 본 연구에서는 SMOTE 알고리즘을 사용하는데 간략히 설명하면 다음과 같다. 더 적은 label 집합을 이라고 했을 때, 의 한 샘플을 생각한다. 에서 kNN(k-nearest neighbors)인 안의 샘플들 중 임의로 한 샘플을 선택하여 라고 했을 때, 새로운 샘플을 다음과 같이 계산한다(Figure 3.1 참고). 이때 는 0과 1사이의 무작위 수이다. 벨런싱을 하기 전의 데이터는 목표수치 도달여부 기준으로 (성공/실패) = (3038/806)이고, 벨런싱 후에는 (3224/3215)로 교정되었다. 벨런싱 이전의 데이터는 회귀모형을, 벨런싱 이후의 데이터는 분류모형을 학습시킨다.

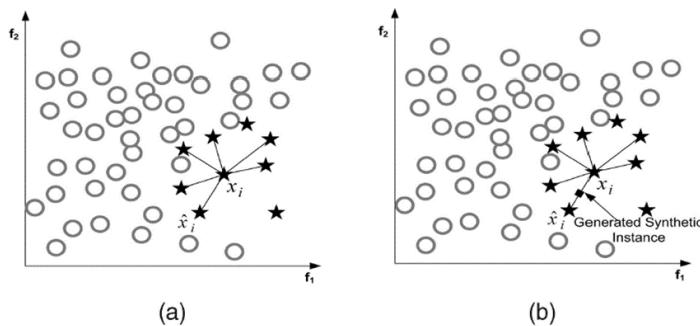


Figure 3.1 Example of SMOTE algorithm (source: He와 Garcia, 2009)

#### 3.2. LASSO 회귀모형 (저밀도 콜레스테롤 감소량 / 목표수치 도달 여부)

최근에 Park (2013)이 주성분 분석 도구로서 사용하고, Kwak 과 Kim (2014)에 의해 분류모델에도 적용된 LASSO (least absolute shrinkage and selection operator)는 Tibshirani (1996)에 의해 개발된 알고리즘으로, 변수선택과 규제화를 진행하여 모형의 성능이나 설명력을 높이는 장점이 있다. 최소화할 대상인 objective function에 계수들의 절대값을 더함으로써 penalty를 주거나, coefficient의 크기에 직접적으로 제약조건을 주는 방식으로 상대적으로 덜 중요한 변수들의 계수 값이 0으로 수렴하도록 유도 한다. 이렇게 희소화된 회귀모형 (sparse model)은 예측성능이 향상되거나 변수들을 해석하기가 용이하다는 이점이 있다.  $y_i$ 를 목표값으로 놓고  $x_i := (x_1, x_2, \dots, x_p)^T$ 를 공변량으로 놓으면 다음과 같은 제약식을 세울 수 있다.

$$\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^N (y_i - \beta_0 - x_i^T \beta)^2 \quad \text{subject to} \sum_{j=1}^P |\beta_j| \leq t. \quad (3.1)$$

### 3.3. 로지스틱 회귀모형

선형회귀와 비슷하게 종속변수와 독립변수간의 관계를 구체적인 함수로 나타내어 예측 모델에 사용한다. King 과 Zeng (2001) 참고. 하지만 단순 선형회귀모델과는 다르게 최종적으로 예측하는 결과가 특정 분류로 나누어지기 때문에 분류기법으로 볼 수 있다. 기본적인 모형은 다음과 같다.

$$p(y = 1|x) = \frac{\exp(\beta^T x)}{1 + \exp(\beta^T x)}. \quad (3.2)$$

본 연구에서는 LASSO 기법에 적용하여 목표수치 도달여부를 예측하는 데에 사용한다.

### 3.4. 모형 평가

LASSO 모형은 제약조건을 주는 하이퍼 파라미터 파라미터를 조정하는 과정이 필요하다. 이 때 적절한 파라미터를 정하는 기준이 필요한데, 저밀도 콜레스테롤 감소량 예측 모형에서는 평균제곱오차를 사용했고 목표수치 도달여부 예측 모형에서는 AUC (area under ROC curve)를 사용했다. 모두 10-fold 교차검증을 통해 평가했다. 선택된 변수에 대한 계수의 유의성 검정은 Lockhart 등 (2014)이 제안한 covTest-package를 사용하여 p-value 0.05를 기준으로 하였다.

## 4. 분석 결과

### 4.1. 분석대상자 기술적 통계분석

위 과정을 거쳐 걸러진 분석대상자 3844명의 주요 변수들에 대한 기술적 통계분석을 나타낸다. 먼저 투약순응도의 분포를 보면 1에서 높은 밀도를 보이고 있어 많은 환자들이 꾸준히 투약을 이행하고 있음을 알 수 있다 (Figure 4.1 참고). 약제 처방비용의 경우, 낮은 가격대에서 높은 밀도를 보이는데, 연간 약 5만원 근처에서 가장 높은 비율을 보이고 있다 (Figure 4.2 참고). 마지막으로 저밀도 콜레스테롤의 경우, 스타틴을 투약하기 전과 후에 평균 약 22mg/dL 정도의 감소효과를 볼 수 있었고, 두 해의 표준편차는 큰 차이를 보이진 않았다 (Figure 4.3 참고). Table 4.1은 각 변수들의 평균과 표준편차를 나타낸 것이다.

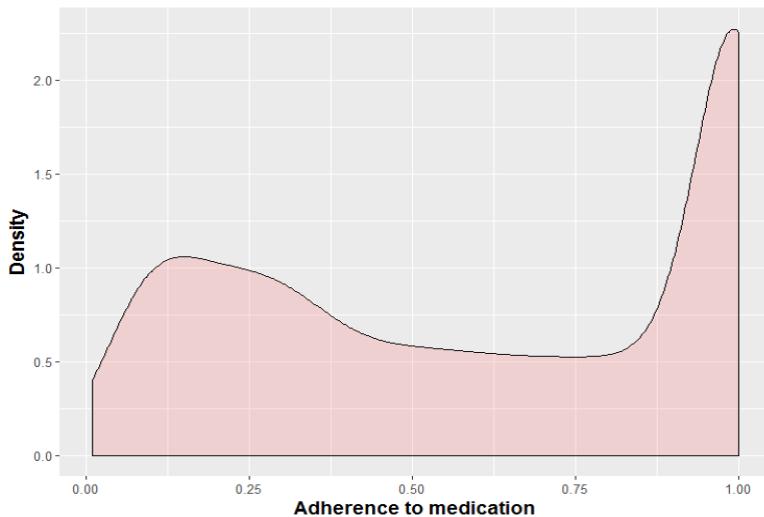


Figure 4.1 Density of adherence to medication

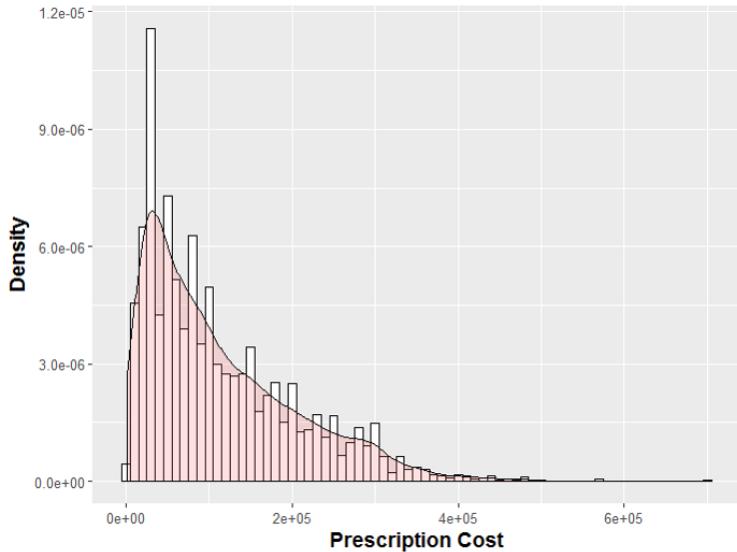
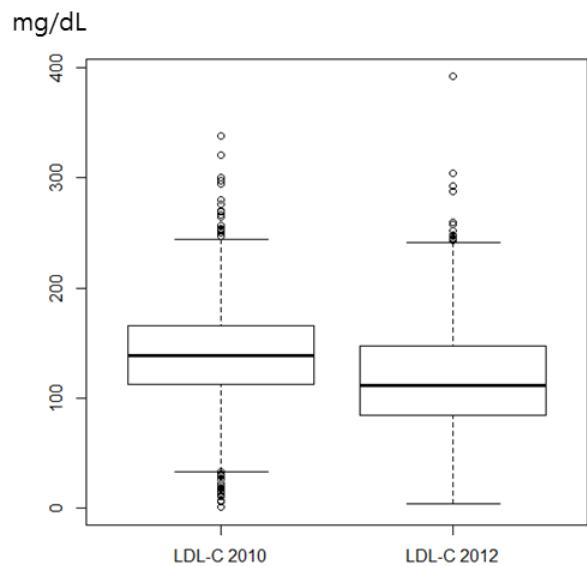


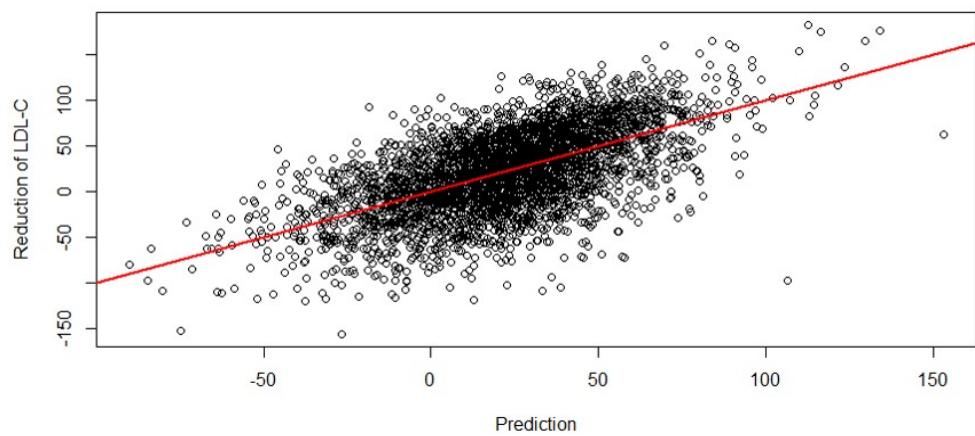
Figure 4.2 Density of statin prescription cost

#### 4.2. LASSO 회귀모형

LASSO 회귀모형을 이용하여 저밀도 콜레스테롤 감소량을 예측하는 경우 p-value와 계수의 크기를 함께 고려하면, 직전년도의 저밀도 콜레스테롤 수치 (26.2), 투약순응도 (8.23), 스타틴 처방비용 (6.88), 연령대 (3.29), 직전년도의 공복혈당 (3.19) 순으로 선택되었다 (Table 4.1 참고). 이는 투약순응도를 0.35 (투약순응도의 표준편차)정도 증가시킬 경우, 저밀도 콜레스테롤 감소량은 8.2mg/dL 정도 증가한다는 것으로 해석할 수 있다. 적절한 하이퍼 파라미터로 학습한 모형은  $R^2 = 0.38$ 로 모형 자체의 설명력은 높지 않다. 주요 변수로 선택된 변수들은 대부분 사전에 예상된 변수들인데, 계수의 크기를 기준으로 보면 투약순응도는 두 번째 변수로서 상당히 중요한 변수임을 알 수 있다. Figure 4.1은 실제 값과



**Figure 4.3** Boxplots of LDL-C of before and after statin prescription



**Figure 4.4** Prediction and reduction of low-density lipoprotein cholesterol

모형의 결과값을 plotting한 것이며 직선은  $y = x$ 를 나타낸다.

**Table 4.1** Mean and standard deviation of variables

Variable	Mean	Standard deviation
Statin prescription cost	112985.2	92312.92
Total cholesterol of last year	228.2815	39.9316
High-density lipoprotein cholesterol	54.85874	27.30092
Low-density lipoprotein cholesterol of last year	138.8583	39.60738
Triglyceride of last year	178.1067	120.3593
Low-density lipoprotein cholesterol of next year	116.9054	42.88155
Height	161.5096	9.532401
Weight	65.19693	11.54322
Waist	83.38137	8.553773
Systolic blood pressure of last year	127.4909	15.68975
Diastolic blood pressure of last year	79.13163	10.54203
Fasting blood sugar of last year	104.5442	29.67049
Hemoglobin of last year	13.95392	1.528662
Proteinuria	1.102237	0.48629
Serum creatinine	1.054773	1.176685
Serum glutamlc oxaloacetic transaminase	27.30073	18.57229
Serum glutamlc pyruvic transaminase	28.49298	20.45223
Gamma-GTP	42.79995	49.37221
Smoking period	4.646982	10.66191
amount of smoking per day	3.288241	7.479952
Drinking days per week	1.950312	1.524499
Amount of drinking per time	2.596774	4.191767
More than 20 minutes of intense exercise per week	2.101197	1.778645
More than 30 minutes of mid-intense exercise per week	2.283559	1.886096
More than 30 minutes of walking per week	3.633975	2.495586
Difference of LDL-C	21.95293	45.56312
Attainment to target range	0.7903226	0.4071313
Adherence to medication	0.6057656	0.3515576

#### 4.3. 로지스틱 회귀모형

LASSO에 로지스틱 회귀모형을 적용하여 목표수치 도달여부를 예측하는 경우 p-value와 계수의 크기를 함께 고려하면, 주요변수로는 직전년도 저밀도 콜레스테롤 (-0.643), 처방비용 (0.411), 위험군 (-0.410), 투약순응도 (0.312), 공복혈당 (0.157)등의 순서로 되었다 (Table 4.2 참고). 이는 투약순응도를 0.35 (투약순응도의 표준편차)정도 증가시킬 경우, 목표수치에 도달할 확률의 승산비가 배로 증가한다. 10-fold 교차검증 결과 모형의 예측 정확도는 0.722 (표준편차=0.025)이고 AUC값은 0.79 (표준편차=0.0071)이다. 학습한 모형을 전체 데이터에 적용한 결과 confusion matrix은 Table 4.3과 같다. 선택된 변수들을 보면 저밀도 콜레스테롤이 가장 주요한 변수로 선택되고 처방비용과 위험군이 그 다음을 차지하고 있다. 계수의 부호에 의해 처방비용과 공복혈당, 투약순응도가 높을수록, 그리고 직전년도 저밀도 콜레스테롤 수치와 위험군의 수준이 낮을수록 목표 수치에 도달할 확률이 높아짐을 알 수 있다. 저밀도 콜레스테롤 감소량을 예측할 때와는 달리 투약순응도가 다른 변수들보다 영향력 순위가 비교적 낮은 것으로 나왔는데, 이는 위험군에 따라 목표수치가 달라지므로 그 과정이 단순히 저밀도 콜레스테롤 수치를 낮추는 것과 차이가 있기 때문인 것으로 보인다.

**Table 4.2** Coefficients and p-value of variables of LASSO regression

Variable	Coefficient of LASSO regression	p-value
Low-density lipoprotein cholesterol of last year	<b>26.271</b>	0
Adherence to medication	<b>8.2316</b>	0
Statin prescription cost	<b>6.8897</b>	0
Proteinuria	3.3105	0.4618
Age	<b>3.2937</b>	0.0069
Fasting blood sugar of last year	<b>3.1969</b>	0.0369
amount of smoking per day	1.67464	0.9556
Diastolic blood pressure of last year	1.46428	0.9994
Systolic blood pressure of last year	1.34958	0.4464
Hemoglobin of last year	1.30092	0.7733
Smoking period	-1.16583	0.2396
Serum glutamyl oxaloacetic transaminase	1.15154	0.0419
Gamma-GTP	1.12793	0.6866
Triglyceride of last year	-1.02664	0.2878
Waist	-0.8097	0.8948
Level of disability	0.76335	0.6421
Serum creatinine	-0.68459	0.6973
Serum glutamyl pyruvic transaminase	0.51804	0.9168
Amount of drinking per time	0.48956	0.9673
More than 30 minutes of mid-intense exercise per week	0.16961	0.9964
More than 30 minutes of walking per week	-0.04966	0.9735
Level of income	0.04579	0.9896
risk-level	0.04059	0.9999
Drinking days per week	0.00382	0.9891
High-density lipoprotein cholesterol	0	1
SEX	0	0.9751
Height	0	1
Weight	0	1
More than 20 minutes of intense exercise per week	0	1

## 5. 결론 및 제언

본 연구에서는 건강보험공단이 구축한 표본코호트 DB를 이용하여 고지혈증환자가 스타틴을 복용했을 때, 복용 효과를 예측하는 모형을 만들고, 결과에 영향을 미치는 요인들을 분석하였다. 저밀도 콜레스테롤 수치를 낮추는 저밀도 콜레스테롤 감소량의 경우엔 처방에 들어가는 비용보다도 환자의 투약순응도가 더 영향을 갖는데, 이는 복용 효과를 높이기 위해 환자에게 꾸준한 복용을 권장하는 근거가 될 수 있다. 두 번째로 NCEP-ATP III 가이드라인에 의한 환자의 위험군에 따라 목표수치 도달여부를 성공과 실패로 나눈 경우엔 다른 변수들보다 처방 비용이 가장 큰 영향을 미치는 변수로 나왔는데, 환자의 위험군에 따라 높은 강도의 스타틴을 처방하고 이는 처방 비용에 직결됐던 것으로 여겨진다. 전반적으로 모든 경우에 처방 비용은 결과에 유의미하게 영향을 미친 것으로 보이며 보다 세부적인 연구로 경제성을 연구할 바탕이 될 수 있을 것이다.

스타틴은 저밀도 콜레스테롤 수치를 낮추는 효과가 있지만 간수치 증가, 근염, 혈당 상승 등의 다양한 부작용이 있는 것으로 알려져 있다. 이런 부작용을 함께 고려하면 변수들의 중요도나 경제성에 영향을 미쳐 기존 연구와는 다른 결론을 도출할 수 있을 것으로 예상된다.

**Table 4.3** Coefficients and p-value of variables of LASSO with logistic regression

Variable	Coefficient of LASSO regression	p-value
Low-density lipoprotein cholesterol of last year	<b>-0.643</b>	0
Statin prescription cost	<b>0.4110</b>	0
risk-level	<b>-0.410</b>	0
Adherence to medication	<b>0.3120</b>	0
Fasting blood sugar of last year	<b>0.1578</b>	0.0108
Systolic blood pressure of last year	<b>0.1557</b>	0.0146
Age	<b>0.1024</b>	0
Serum glutamlc oxaloacetic transaminase	0.0604	0.0384
More than 20 minutes of intense exercise per week	0.0463	0.3054
High-density lipoprotein cholesterol	0.0141	0.5342
Proteinuria	0.0125	0.8581
Level of disability	0.0081	0.7795
Gamma-GTP	0.0037	0.7257
Amount of drinking per time	0.0009	0.4602
SEX	0	0.7651
Level of income	0	0.9833
Triglyceride of last year	0	0.8484
Height	0	0.9542
Weight	0	0.9347
Waist	0	0.0938
Diastolic blood pressure of last year	0	1
Hemoglobin of last year	0	0.9987
Serum creatinine	0	0.8891
Serum glutamlc pyruvic transamlnase	0	0.8711
Smoking period	0	0.6287
amount of smoking per day	0	0.0058
Drinking days per week	0	0.0678
More than 30 minutes of mid-intense exercise per week	0	0.9164
More than 30 minutes of walking per week	0	0.6793

**Table 4.4** Confusion matrix

Fitted value	Real value	
	Success (=1)	Fail (=0)
Success (=1)	2380	937
Fail (=0)	844	2278

## References

- Jeong, Y. J., Ko, S. K., Park, H. S. and Yang, B. M. (2004). Meta-analysis and economic evaluation for the effect of the statins. *Korean Society of Lipidology and Atherosclerosis*, **14**, 120-130.
- Han, K. H.. (2009). Control of dyslipidemia. *Journal of the Korean Medical Association*, **52**, 299-311.
- Hong, Y. J., Jeong, M. H., Lim, J. H., Park, H. W., Kim, H. G., Park, O. Y., Kim, J. H., Kim, W., Ahn, Y. K., Cho, J. G., Park, J. C., and Kang, J. C. (2003). The prognostic significance of statin therapy according to the level of C-reactiveprotein in acute myocardial infarction patients who underwent percutaneouscoronary intervention. *Korean Circulation Journal*, **33**, 891-900.
- Boyd, H. C., Gown, A. M., Wolfbauer, G. and Chait, A. (1989). Direct evidence for a protein recognized by a monoclonal antibody against oxidatively modified LDL in atherosclerotic lesions from a Watanabe heritable hyperlipidemic rabbit. *The American Journal of Pathology*, **135**, 815.
- He, H. and Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and*

- Data Engineering*, **21**, 1263-1284.
- Jeong, Y. J., Kim, H., Baik, S. J., Kim, T. M., Yang, S. J., Lee, S. H., Cho, J. H., Lee, H., Yim, H. W., Choi I. Y., Yoon, K. H. and Kim, H. S. (2017). Analysis and comparison of the cost-effectiveness of statins according to the baseline low-density lipoprotein cholesterol level in Korea. *Journal of Clinical Pharmacy and Therapeutics*, **42**, 292-300.
- Kim, H. S., Kim, H., Lee, H., Park, B., Park, S., Lee, S. H., Cho, J. H., Song, H., Kim, J. H., Yoon, K. H. and Choi, I. Y. (2016). Analysis and comparison of statin prescription patterns and outcomes according to clinical department. *Journal of Clinical Pharmacy and Therapeutics*, **41**, 70-77.
- Kim, S. H. (2010). *Pharmacological properties of statins. Evidence in statin clinical trials*, edited by S. H. Baek, Korean Society of Lipid and Atherosclerosis, Korea, 13-21.
- King, G. and Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, **9**, 137-163.
- Kwak, S. and Kim, H. (2014). Comparison of ensemble pruning methods using Lasso-bagging and WAVE-bagging. *Journal of the Korean Data & Information Science Society*, **25**, 1371-1383.
- Lockhart, R., Taylor, J., Tibshirani, R. J. and Tibshirani, R. (2014). A significance test for the lasso. *Annals of Statistics*, **42**, 413.
- National cholesterol education program (NCEP) Expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III). (2001). *Third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report*, US Department of Health and Human Services, US.
- Park, C. (2013). Simple principal component analysis using Lasso. *Journal of the Korean Data & Information Science Society*, **24**, 533-541.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**, 267-288.

# Analyzing association between low-density lipoprotein reduction by statin and adherence to medication using national health insurance service-national sample cohort (NHIS-NSC)<sup>†</sup>

Kyu-Jin Kim<sup>1</sup> · Chi-Hyuck Jun<sup>2</sup> · Hyeseon Lee<sup>3</sup> · Hun-Sung Kim<sup>4</sup>

<sup>1,2,3</sup>Department of Industrial and Management Engineering, Pohang University of Science and Technology

<sup>4</sup>Department of Medical Informatics, College of Medicine, The Catholic University of Korea

Received 6 September 2017, revised 19 September 2017, accepted 21 September 2017

## Abstract

Hyperlipidemia, the status of blood with high level of low-density lipoprotein cholesterol (LDL-C), is known as a main cause of coronary artery diseases such as myocardial infarction or brain infarct. Statin is the representative prescription to hyperlipidemia and the effects of it depend on the patient's individual conditions such as health-caring habits or adherence to medication. The main effect of statin is reducing LDL-C, which should reach the target range based on National Cholesterol Education Program-Adult Treatment Panel III (NCEP-ATP III) guideline. In this research, the reduction of LDL-C and attainment to patient's target range are considered effects of statin. The association between factors - individual conditions and adherence to medication of patients - and the effects of statin is analyzed with National Health Insurance Service-National Sample Cohort (NHIS-NSC).

**Keywords:** Adherence to medication, hyperlipidemia, low-density lipoprotein cholesterol, national sample cohort, statin.

<sup>†</sup> This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (HC15C1362).

<sup>1</sup> Graduate student, Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang 37673, Korea.

<sup>2</sup> Professor, Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang 37673, Korea.

<sup>3</sup> Corresponding author: Professor, Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang 37673, Korea. E-mail : [hyelee@postech.ac.kr](mailto:hyelee@postech.ac.kr)

<sup>4</sup> Co- Corresponding author: Professor, Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul 06591, Korea.