

데이터마이닝 기법을 활용한 맞춤형 고혈압 사후관리 모형 개발

박일수¹ · 용왕식² · 김유미³ · 강성홍⁴ · 한준태⁵

¹국민건강보험공단; ²국민건강보험공단; ³한국보건산업진흥원 의료기관평가지원센터;
⁴인제대학교 보건행정학과; ⁵국민건강보험공단

(2008년 3월 접수, 2008년 4월 채택)

요 약

본 연구는 국민건강보험공단의 건강검진데이터, 자격 및 보험료 그리고 진료비 데이터를 활용하여 고혈압 관리를 위한 맞춤형 고혈압 사후관리모형(고혈압 진료예측모형 및 고혈압 진료순응도세분화모형)을 개발하고자 하였다. 모형 개발에는 데이터마이닝의 로지스틱 회귀모형, 의사결정나무 그리고 앙상블 모형을 활용하였다. 고혈압 진료예측모형에서는 3가지 모형 중 로지스틱 회귀모형이 가장 우수한 모형으로 채택되었으며, 고혈압 진료순응도세분화모형은 의사결정나무모형을 통해 개발되었다. 본 연구는 전국 규모의 수년간 축적된 자료를 데이터마이닝을 활용함으로써 고혈압의 진료 및 진료순응도에 이르는 고혈압 사후관리 프로세스 전반에 걸친 결과를 도출함으로써 우리나라 고혈압 사후관리체계 구축에 기여할 것으로 사료된다.

주요용어: 고혈압 사후관리모형, 데이터마이닝, 로지스틱 회귀모형, 의사결정나무.

1. 배경

많은 국가들에서 날로 증가하고 있는 고혈압등의 만성질환은 국민건강과 건강보험재정을 위협하는 주요 원인이 되고 있다. 김영식 등 (2001)의 코호트 연구에서는 고혈압의 조발생률이 남자에서 1,000명당 17.1명, 여자는 12.9명이었다. 그러나 이 중 의료기관에서 실제로 진단을 받은 사람은 절반에도 미치지 못하며, 진단을 받은 후 치료를 받은 사람도 역시 절반에 미치지 못한다고 추정되고 있다. 이 결과는 우리나라의 '2005년 국민건강·영양조사'에서도 비슷한 양상을 보이고 있는 바, 고혈압의 유병율이 27.9%였으며, 그 중 남자가 30.2%, 여자가 25.6%였다. 그러나 고혈압 유병자 중 정기적으로 혹은 필요시 약물을 복용하는 대상자 비율을 나타내는 치료율은 49.0%(남자 39.2%, 여자 60.0%)로 나타나 병이 발생했으나 그 중 치료를 받는 사람의 비율은 적은 것으로 나타났다. 또한, 2006년도 건강보험의 질

¹(121-749) 서울시 마포구 염리동 168-9, 국민건강보험공단 건강보험연구원, 주임연구원.

E-mail: leosu@nhic.or.kr

²교신저자: (135-934) 서울시 강남구 강남대로 501 미진프라자빌딩, 국민건강보험공단 강남서부지사, 지사장.

E-mail: yongws1306@hanmail.net

³(156-050) 서울시 동작구 노량진동 311-27 대성빌딩, 한국보건산업진흥원 의료기관평가지원센터, 연구원.

E-mail: vega1111@hanmail.net

⁴(621-749) 경남 김해시 어방동 607, 인제대학교 보건행정학과, 교수. E-mail: hcmkang@hanmail.net

⁵(121-749) 서울시 마포구 염리동 168-9, 국민건강보험공단 건강보험연구원, 주임연구원.

E-mail: maru@nhic.or.kr

병 소분류별 다빈도 상병 급여실적(입원, 외래진료비 및 약국 처방조제 진료비 포함(비급여진료비 제외))을 보면 본태성 고혈압으로 진료 받은 사람의 총진료금액이 1조 1천억원에 달했으며, 이는 우리나라의 비급여진료비를 제외한 총진료금액(21조 4천억원)의 5.2%에 해당하는 금액이다 (국민건강보험공단·건강보험심사평가원, 2007). 따라서 고혈압 관리는 개인과 사회, 국가차원에서 중요한 보건문제라고 할 수 있다. 고혈압에 대한 관리는 고혈압 진단과 진료에 대한 예측을 통해 개인의 건강행태를 조기에 변화시키는 것이 중요하며, 효율성 또한 높다.

그 동안 고혈압의 진단과 진료에 영향을 미치는 요인들에 관한 많은 연구들이 수행되었으나, 대부분이 외국에서 수행되어 우리나라 국민들의 특성을 잘 반영하지 못하고 있으며, 우리나라에서 수행된 연구들은 유병률을 조사하면서 부차적으로 고혈압 환자에서 그 당시에 가지고 있던 요인을 분석한 단면연구가 대부분이어서 시간적인 선행성을 알 수 없어 발병요인으로서의 인과관계를 입증할 수 없다는 한계가 있다 (김영식 등, 2001). 또한 연구자료에 있어서도 일개 의료기관을 이용한 환자들이나 일부 지역사회 주민들을 대상으로 하고 있어 자료의 대표성과 연구결과의 일반화에 한계가 있다고 할 수 있다. 국가 차원의 고혈압 관리를 위해 대표성 있는 자료로부터 도출된 고혈압의 진단과 진료, 순응도에 영향을 미치는 요인들에 관한 결과를 근거로 해야 한다. 따라서 전국적인 규모의 자료를 이용하여 고혈압 진단, 진료, 순응도에 영향을 미치는 요인들을 밝혀내는 것이 매우 필요하다.

이에 본 연구는 수년간 축적된 국민건강보험공단이 실시한 건강검진과 문진자료, 건강보험 자격자료와 요양급여 자료를 이용하여 선행 연구된 용왕식 등 (2006)의 “고혈압 발생예측모형 개발”의 후속 연구로서 고혈압 진단 이후 고혈압 사후관리모형을 고혈압 진료 및 진료순응도 두 단계로 걸쳐 개발·평가하고자 하였으며, 개발된 모형을 근거로 진료 및 순응도에 미치는 요인을 밝혀서, 우리나라 고혈압 관리에 필요한 대표성 있는 기초 자료를 도출하고자 하였다. 또한 자료 도출 방법에 있어서는 다양한 산업분야에서 과학적인 자료분석 및 모형개발방법론으로 활용되고 있는 데이터마이닝 기법을 활용함으로써 연구의 질적 향상을 도모하고자 하였다.

2. 방법

2.1. 자료수집 및 모형

1) 모형개발 및 내적타당도(Internal validation) 평가를 위한 데이터

본 연구의 대상자는 2002년에 국민건강보험공단이 실시하는 건강검진을 통하여 고혈압으로 질환의심을 받은 사람(1998년부터 2000년 건강검진 전까지 고혈압과 관련된 진단 및 치료가 없는 사람 중 2002년 건강검진에서 고혈압 진단을 받은 사람) 166,356명을 대상으로 하였다. 연구대상자에 대해 국민건강보험공단의 건강검진DB, 건강보험급여DB 그리고 건강보험가입자의 인구사회학적 특성을 나타내는 자격DB는 2003년 12월 기준으로 기존 연구에서 밝혀진 고혈압 진료여부에 영향을 미치는 요인들을 토대로 주요 독립변수를 추출하였다. 2002년에 고혈압 진단 이후 실제로 고혈압 진료를 위해 2002년 개인별 건강검진일 이후로부터 2003년 12월 31일까지의 건강보험급여자료에서 고혈압으로 인한 진료유무를 종속변수로 활용하여 고혈압 진료 예측모형을 개발하였다. 또한, 고혈압 진료가 지속적인 약물투여와 추후관리를 필요로 하는 점을 고려하여 의사가 처방한대로 진료를 충실히 받고 있는 정도를 나타내는 진료순응도를 투약일수로부터 산출하여 이에 영향을 미치는 요인을 중심으로 진료순응도세분화모형을 도출하고자 하였다. 이에 2003년 12월 31일까지 고혈압 관련 진료를 한번이라도 받은 자 22,631명을 최종연구대상으로 이들의 인구사회학적 특성, 임상학적 특성, 건강행위 특성을 독립변수로 활용하고, 개인별 2002년 건강검진 시점부터 2003년 12월 31일까지의 진료순응도를 예측·세분화하는 고혈압 진료순응도 모형을 개발하였다.

2) 외적타당도(External validation) 평가를 위한 데이터

고혈압 진료 및 진료순응도 예측모형의 외적타당도를 평가하기 위해서 모형개발을 위한 데이터와는 시간적으로만 다른 데이터(2003년~2004년)를 수집하였으며, 데이터 수집항목은 모형개발을 위한 데이터와 동일한 항목으로 구축하였다.

2.2. 변수정의

1) 종속변수

고혈압 진료 여부는 건강검진 대상자들이 해당 기간 동안 이용한 의료이용자료 중 고혈압 진단명(I10~I15, I27.0, I67.4, I70.1, G93.2, K76.6)을 포함하고 있는 의료이용이 하나라도 있는지를 가지고 판단하였다. 고혈압 진료를 위해서는 진단을 받은 시점 이후 지속적으로 외래로 방문하여 약물치방을 받아야 한다. 따라서 2002년에 건강검진에서 고혈압 진단을 받은 자들이 진단시점 이후부터 2003년 12월 31일까지 고혈압 약을 복용하는 것을 전제로 하였다. 이러한 전제조건을 기준으로 같은 기간 동안의 외래처방투약일수를 가지고 진료순응도(r)를 산출하였다.

$$r = \frac{1}{D} \sum_t^{2003.12.31} d_t \times 100, \quad (2.1)$$

여기서 d_t =고혈압 처방을 받은 시점(t)에서의 투약일수, D =검진을 통한 고혈압진단 시점부터 2003년 12월 31일까지의 총 일수 이다.

2) 독립변수

활용된 주요 독립변수는 인구사회학적 변수, 건강행위 특성(음주습관, 흡연습관, 운동습관), 임상학적 특성(신장, 체중, 비만도, 혈압수준, 식전혈당(글루코오스)와 콜레스테롤 수준, 고혈압에 대한 가족력), 과거 고혈압 관련 진료이용을 제외한 진료이용 특성이 각각 독립적으로 그리고 상호 교차작용(Interaction)의 조합이 포함되었다.

2.3. 분석방법

1) 모형개발

고혈압 진료예측 모형 및 진료순응도세분화 모형은 데이터마이닝 알고리즘인 CHAID(Chi-Squared Automatic Interaction Detection)방법을 활용한 의사결정나무모형, 단계적 추출방법을 활용한 로지스틱 회귀모형, 앙상블 알고리즘으로 도출한 후 각 모형을 비교·검증하고 그 중 가장 우수한 모형을 채택하였다. 통계분석은 SAS 9.1을 사용하였으며, 모형개발을 위한 데이터마이닝 툴은 SAS사의 Enterprise Miner 4.3을 사용하였다.

2) 모형평가

고혈압 진료 및 진료순응도 세분화모형에 대한 평가는 내적타당도 검증과 외적타당도 검증을 실시하였다.

- (1) 내적타당도(Internal validation): 모형생성을 위해 구축되었던 자료 중 60%는 모형개발용 분석용 자료(Train Data)로 활용하였고, 나머지 40%로서 모형평가 자료(Validation Data)로 활용하여 모형의 내적 타당도 검증을 실시하였다.
- (2) 외적타당도(External Validation): 외적타당도 검증은 모형개발용 데이터를 이용하여 개발된 모형을 외적타당도 평가를 위한 데이터에 적용시켜 고혈압 진료예측확률 및 진료순응도를 예측한

표 3.1. 인구사회학적 특성별 고혈압 진료

	특성(2000년)	대상자수(%)	고혈압진료(2002 ~ 2003)	
			진료자수	진료율
연령	60세 이상	41,883(25.2)	6,443	15.4
	50세 ~ 59세	54,283(32.6)	8,432	15.5
	40세 ~ 49세	48,132(28.9)	6,257	13.0
	39세 이하	22,058(13.3)	1,499	6.8
성별	남	127,930(76.9)	16,615	13.0
	여	38,426(23.1)	6,016	15.7
거주지	대도시	72,408(43.5)	9,886	13.2
	중소도시	70,345(42.3)	9,307	13.7
	농어촌	22,319(13.4)	3,257	14.6
	없음	1,284(0.8)	181	14.1
직역	지역	29,565(17.8)	4,752	16.1
	공교	39,695(23.9)	7,016	17.7
	직장	95,812(57.6)	10,682	11.2
	없음	1,284(0.8)	181	14.1
보험료	상위 75% 이상(69,930원 이상)	50,556(30.4)	8,011	15.9
	상위 50% ~ 75%미만(50,820원 ~ 69,930원 미만)	33,440(20.1)	4,391	13.1
	상위 25% ~ 50%미만(31,610원 ~ 50,820원 미만)	33,982(20.4)	4,027	11.9
	상위 25%미만(31,610원 미만)	48,378(29.1)	6,202	12.8
전체		166,356(100)	22,631	13.6

(단위 : 명, %)

후 이들 자료를 2003년의 실제 고혈압 진료를 및 순응도 값과 비교하여 개발된 모형의 외적타당도를 비교·검증하였다.

- (3) 모형평가지표: 개발된 모형의 내적, 외적타당도 평가와 최종 선정된 모형의 효율성 평가는 ASE(Average Squared Error), 예측력(%Response), 모형적용효과(%Captured Response) 및 향상도(Lift)지표를 사용하였다.

3. 연구결과

3.1. 연구대상자의 인구사회학적 특성

2002년 건강검진에서 고혈압 진단을 받은 시점부터 2003년 12월 31일 사이에 요양기관에서 고혈압 진료를 받은 사람들의 인구사회학적 특성에 따른 고혈압 진료율은 (표 3.1)과 같다. 연령 분포에서 진료를 받은 환자수는 50대가 15.5%로서 가장 많으며 타 연령대에 비해 높았다. 성별 진료율은 남자(13.0%)보다 여자(15.7%)가 높게 나타났으며, 경제적 수준을 나타내는 보험료수준이 가장 높은 집단에서 고혈압 진료율이 높았다.

3.2. 고혈압 사후관리 예측모형의 개발 및 평가

1) 모형개발 및 내적타당도(Internal Validation)

2002년에 고혈압으로 진단 받은 자가 진료 당해년도인 2002년부터 2003년까지 고혈압으로 진료 받을

표 3.2. 고혈압 진료예측모형 결과 비교

모형	Root ASE			Misclassification Rate		
	2002 ~ 2003		2003 ~ 2004	2002 ~ 2003		2003 ~ 2004
	Train(60%)	Test I (40%)	Test II	Train(60%)	Test I (40%)	Test II
Logistic Regression	0.3286	0.3321	0.2810	0.1345	0.1383	0.1010
Decision Tree	0.3345	0.3374	0.3013	0.1347	0.1377	0.1010
Ensemble	0.3299	0.3331	0.2859	0.1347	0.1381	0.1010

표 3.3. 고혈압 진료순응도세분화 모형결과 비교

모형	Root ASE		
	2002 ~ 2003		2003 ~ 2004
	Train(60%)	Test I (40%)	Test II
Liner Regression	0.2724	0.2704	0.2939
Decision Tree	0.2702	0.2689	0.2921
Ensemble	0.2702	0.2686	0.2920

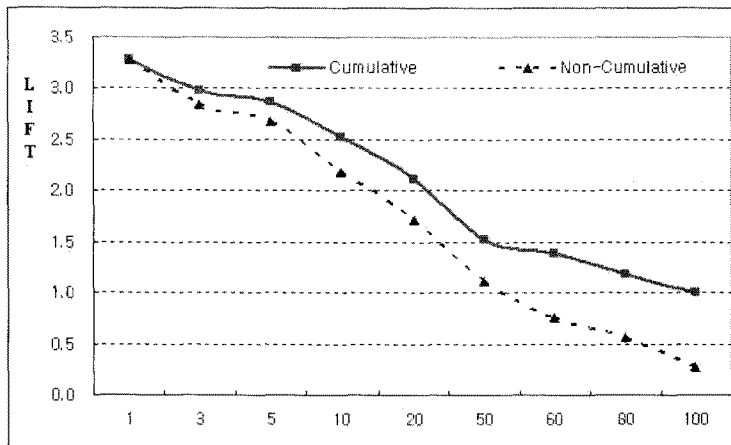


그림 3.1. 고혈압진료예측모형 향상도표(LIFT CHART: 2002~2003년)

확률에 대한 모형을 데이터마이닝 알고리즘인 로지스틱 회귀모형, 의사결정나무, 앙상블로 모형을 구축한 후 모형 비교를 통하여 가장 좋은 모형인 로지스틱 회귀모형을 선택하였다 (표 3.2).

그리고 2002년에 고혈압으로 진단 받은 자가 진료 당해년도인 2002년부터 2003년까지 진료 받은 사람의 진료순응도에 대하여서는 회귀모형, 의사결정나무, 앙상블 모형 중에서 데이터마이닝의 의사결정나무 알고리즘을 활용하여 진료순응도모형을 개발하였다. 이 세가지 모형을 평가한 결과 앙상블 모형의 Root ASE가 가장 낮게 나타났으나 의사결정나무모형과 거의 차이가 없음을 확인할 수 있다. 이에 본 연구에서는 고혈압 진료순응도세분화 모형을 설명함에 있어서 앙상블 모형보다는 의사결정나무모형이 효과적이라고 판단되어 의사결정나무모형을 선택하였다 (표 3.3).

2) 외적타당도(External Validation)

표 3.4. 고혈압 진료예측모형(LOGISTIC REGRESSION)

특성요인		추정회귀계수	상대진료 예측도
절편		-2.3218***	
인구사회학적 특성	연령		
	40세 미만		1.000
	40대	0.1155***	1.736
	50대 이상	0.3206***	2.131
	상위 75% 이상		1.000
	최종보험료 (경제적 수준)		
상위 50%~75%미만	-0.1302***	1.038	
상위 25%~50%미만	0.0583***	1.253	
상위 25%미만	0.2394***	1.502	
임상학적 특성	혈압 mmHg		
	정상		1.000
	경계역	0.1203***	1.272
	수축기 혈압변화 (2000→2002)		
	변화없음		1.000
	정상A, B→질환의심	0.5845***	2.813
정상A→정상B	-0.1347***	1.370	
건강행위 특성	이완기 혈압변화 (2000→2002)		
	변화없음		1.000
	정상B→질환의심	0.2589***	2.001
	정상A→정상B, 질환의심	0.1759***	1.842
건강행위 특성	흡연기간		
	5년 미만		1.000
	5~9년	-0.0328	1.041
	10~19년	-0.0438	1.029
	20~29년	0.0441	1.124
30년 이상	0.1024***	1.192	

주: 1)혈압: 정상(수축기혈압: 139이하, 이완기혈압: 89이하),
경계역(수축기혈압: 140이상~159이하, 이완기혈압: 90이상~94이하)
2)수축기혈압: 정상A(139이하), 정상B(140이상~159이하), 질환의심(160이상),
이완기혈압: 정상A(89이하), 정상B(90이상~94이하), 질환의심(95이상)
3)**: $p < 0.05$, ***: $p < 0.01$

로지스틱 회귀모형으로 개발된 고혈압진료예측모형을 검증용 자료(Test: 2003~2004년)에 적용하였을 때 진료예측확률이 모형개발을 위한 훈련용 자료(Train: 2002~2003년)와 유사하게 상위 1%에서 임의의 확률보다 3.43배 높게 나타나 이 연구에서 개발된 모형의 적합성 및 우수성이 입증되었다 (그림 3.1).

3.3. 모형의 결과

이 연구에서 개발된 고혈압 진료예측모형은 연구모형에서 제시된 인구사회학적 특성, 임상학적 특성, 건강행위 특성, 진료이용량 및 건강행위의 변화를 고려하여 개발되었다. 그 결과 인구사회학적 특성인 연령과 보험료, 임상학적 특성인 혈압 그리고 건강행위 특성에서 흡연기간 및 건강위험행위의 변화가 가장 유의한 변수로 나타났다. 다른 특성요인들은 고혈압진료에 통계적으로 유의한 영향을 미치지 못함에 따라 주요 요인변수에는 포함되지 않았다. 모형에 포함될 경우 모형의 신뢰도를 떨어뜨리는 이유가 되므로 고혈압 진료에 대해 유의하지 않는 변수는 고혈압 진료모형에 포함하지 않았다.

로지스틱 회귀모형을 통해 나타난 고혈압진료예측모형 결과에 의하면 연령이 증가할수록 고혈압진료 확률이 높았다. 또한, 흡연기간이 길어질수록 고혈압으로 진료 받을 확률이 높아짐을 알 수 있었다 (표 3.4).

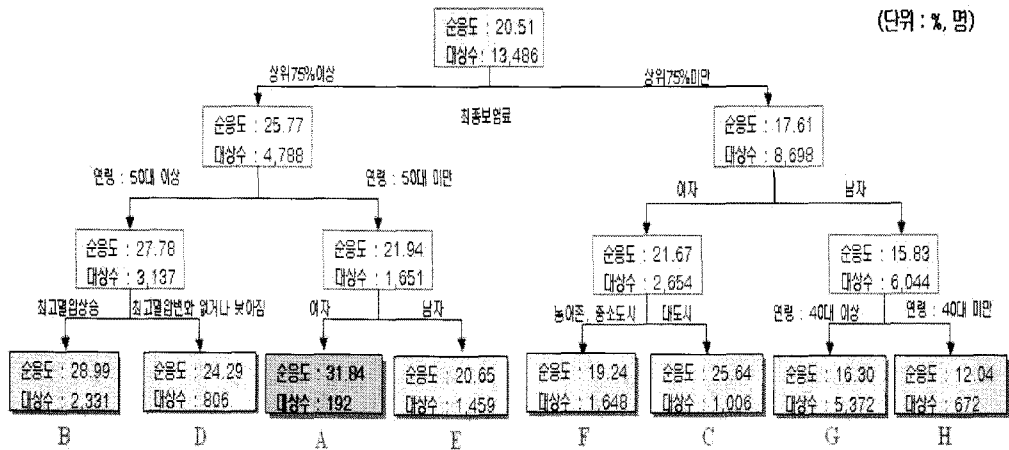


그림 3.2. 고혈압진료순응도 의사결정나무 분석 결과

의사결정나무분석 결과를 통한 고혈압진료순응도의 관련요인에 대한 분석 결과는 총 8개의 집단으로 세분화되었다. 소득수준을 나타내는 보험료가 집단의 세분화 요인 중 가장 큰 요인으로 나타났으며, 50대 이상 여부, 성별, 수축기혈압의 상승여부, 성별, 거주지, 40대 이상여부가 집단 세분화에 영향을 미치는 요인으로 나타났다. 고혈압 진료순응도 세분화 모형을 통한 집단별 특성을 살펴보면, 보험료 수준이 상위 75% 이상이면서 연령이 50대 미만인 여성으로 구성된 A집단은 전체 집단의 평균 순응도보다 훨씬 높은 31.84%의 순응도를 보였다. 즉, 소득이 높은 중년층 여성은 고혈압과 관련된 치료를 잘 받고 있음을 알 수 있다. 그러나 집단 F, G, H에서 고혈압에 대한 진료순응도가 평균 이하로 나타났다. 특히 보험료가 상위 75% 미만이면서 40대 미만인 남성인 경우 순응도가 12.04%로 가장 낮았다. 따라서 F, G, H의 특성집단에 대한 진료순응도를 높일 수 있는 접근방법이 필요하다 (그림 3.2).

4. 결론

본 연구는 국민건강보험공단에서 실시하고 있는 가입자들에 대한 건강검진 자료와 요양급여 자료를 이용하여, 국가 차원에서 주요 관리대상 질환으로 주목받고 있는 고혈압의 진단 이후 고혈압 사후관리 모형을 진료예측 및 진료순응도세분화 모형 두 단계로 구분하여 도출함으로써 향후 우리나라 고혈압관리 체계의 기초자료를 제공하기 위해 수행되었다. 데이터마이닝 프로세스에 의하여 로지스틱 회귀모형으로 만들어진 고혈압 진료예측모형은 인구사회학적 특성에서는 연령, 경제적 수준이, 임상학적 특성에서는 고혈압 진단전의 혈압수준 및 혈압변화가 유의하였다. 건강행위 특성에서는 흡연기간이 고혈압 진료에 가장 영향을 주는 요인으로 나타났다.

고혈압진료순응도 모형에서는 최종보험료 수준, 연령, 성별, 최고 혈압수치의 변화 그리고 거주지역이 고혈압 진료의 순응도 세분화를 결정하는 주요한 요인으로 나타났다.

이 연구는 자료의 특성상 연구 모형에 포함된 변수들이 이론적 틀을 통해 도출되지 않고 이미 구축된 자료에 포함되어 있는 변수들만을 이용하였다는 제한점을 가지나, 연구 결과가 기존의 연구결과들과 유사하며, 현실적으로 이용 가능한 자료를 가지고 설명력 있는 모형을 도출하였다는 점에서 의의를 가진다고 판단된다. 또한 기존의 연구들이 소규모의 일개 의료기관이나 지역에 국한된 자료를 사용함으로써 연구

결과의 대표성과 일반화에 한계를 가졌던 점 그리고 기본적인 통계분석을 통한 연구결과의 도출이라는 점과 비교할 때 이 연구는 우리나라 전체의 수년간 축적된 자료를 데이터마이닝을 활용함으로써 고혈압의 진료 및 진료순응도에 이르는 고혈압 사후관리 프로세스 전반에 걸친 결과를 도출함으로써 우리나라 고혈압 사후관리체계 구축에 기여할 것으로 사료된다.

참고문헌

- 국민건강보험공단·건강보험심사평가원 (2007). <2006 건강보험통계연보>.
- 김영식 (2001). <한국인 고혈압과 당뇨병의 발병요인 규명을 위한 코호트 연구>, 보건복지부 보건의료기술 연구 개발사업 보고서.
- 용왕식, 박일수, 강성홍, 김원중, 김공현, 김광기, 박노례 (2006). 고혈압 발생 예측 모형 개발, <보건교육·건강증진학회지>, **23**, 13-28.

A Development of a Tailored Follow up Management Model Using the Data Mining Technique on Hypertension

Il-Su Park¹ · Wang-Sik Yong² · Yu-Mi Kim³ · Sung-Hong Kang⁴ · Jun-Tae Han⁵

¹National Health Insurance Corporation;

²National Health Insurance Corporation, Daehan University;

³Korea Health Industry Development Institute;

⁴Dept. of Health Policy & Management, INJE University;

⁵National Health Insurance Corporation

(Received March 2008; accepted April 2008)

Abstract

This study used the characteristics of the knowledge discovery and data mining algorithms to develop tailored hypertension follow up management model - hypertension care predictive model and hypertension care compliance segmentation model - for hypertension management using the Korea National Health Insurance Corporation database(the insureds' screening and health care benefit data). This study validated the predictive power of data mining algorithms by comparing the performance of logistic regression, decision tree, and ensemble technique. On the basis of internal and external validation, it was found that the model performance of logistic regression method was the best among the above three techniques on hypertension care predictive model and hypertension care compliance segmentation model was developed by Decision tree analysis. This study produced several factors affecting the outbreak of hypertension using screening. It is considered to be a contributing factor towards the nation's building of a Hypertension follow up Management System in the near future by bringing forth representative results on the rise and care of hypertension.

Keywords: Hypertension follow up management model, data mining, logistic regression, decision tree analysis.

¹Researcher, National Health Insurance Corporation, 168-9 Yeomri-Dong, Mapo-Gu, Seoul 121-749, Korea.
E-mail: leosu@nhic.or.kr

²Corresponding author: Gangnam Western Branch office President, National Health Insurance, Corporation, MijinPlaza Bldg. 501 Gangnamdaero Gangnam-Gu, Seoul 135-934, Korea.
E-mail: yongws1306@hanmail.net

³Researcher, Korea Health Industry Development Institute, Daesung Bldg Noryangjin-Dong, Dongjak-Gu, Seoul 156-050, Korea. E-mail: vegal1111@hanmail.net

⁴Professor, Dept. of Health Policy & Management, INJE University, Obang-Dong, Gimhae, Gyeongnam 621-749, Korea. E-mail: hcmkang@hanmail.net

⁵Researcher, National Health Insurance Corporation, 168-9 Yeomri-Dong, Mapo-Gu, Seoul 121-749, Korea.
E-mail: maru@nhic.or.kr