

데이터마이닝 기법을 활용한 국민건강보험 상해상병 관리모형 개발

박일수¹ · 한준태² · 손혜숙³ · 강석복⁴

¹국민건강보험공단 건강보험정책연구원 · ²국가보훈처 정보화팀 ·

³인제대학교 의과대학 예방의학교실 · ⁴영남대학교 통계학과

접수 2011년 4월 9일, 수정 2011년 5월 17일, 게재확정 2011년 5월 21일

요약

우리나라의 건강보험제도권 내 해당되지 않은 상해상병 진료건 중 국민건강보험으로 부당·잘못 청구되는 진료건을 적발하여, 환수조치하기 위해서는 정확한 상해상병 조사대상자 선정이 필요하다. 그러나, 국민건강보험공단의 한정된 인력으로 증가하는 상해조사관련 업무량을 보다 효율적으로 대처하고, 수행하기 위해서는 상해요인조사 업무 효율화 및 환수 결정을 제고를 위한 조사대상자 발췌기준의 고도화 방안을 마련해야 한다. 이에 본 연구에서는 상해상병 유형에 대해 일정금액 이상 진료건의 발췌 등과 같은 과거의 발췌기준에서 데이터마이닝 기법과 같은 통계적 모형과 업무규칙을 함께 적용한 하이브리드 모형으로서 상해상병 조사대상자 선정기준을 제시하고자하였다.

주요용어: 데이터마이닝, 상해상병, 하이브리드 모형.

1. 서론

우리나라 전체 요양기관에서 연간 국민건강보험공단으로 청구되는 진료비 건수는 2009년 기준으로 12억 건 이상이고, 이에 대해 요양기관으로 지급되는 급여비는 29조1천억 원 이상이다 (국민건강보험공단, 2010). 이는 우리나라 전체 국민이 질병 및 손상 등으로 인해 12억 건 넘게 요양기관에서 진료한 내용을 청구하고 있다는 것이다. 그러나 실질적으로 이 청구내용에는 건강보험이 제도적으로 인정하는 급여가 있는 반면, 건강보험제도권 내에서 건강보험가입자 (피부양자 포함)의 진료내용이 건강보험급여기준에 해당되지 않음에도 불구하고, 고의 또는 중과실로 인한 진료건을 요양기관에서 건강보험으로 청구하는 경우가 존재한다. 이러한 경우 국민건강보험법 제48조에 따라 국민건강보험공단은 보험급여를 받을 수 있는 자가 고의 또는 중대한 과실로 인한 범죄행위에 기인하거나, 고의로 사고를 발생시켰을 때에는 건강보험급여를 제한한다. 또한 보험급여를 받을 수 있는 자가 업무상 또는 공무상 질병·부상·재해로 인하여 다른 법령에 의한 보험급여나 보상 또는 보상을 받게 되는 때에는 보험급여를 하지 아니한다 (국민건강보험, 2003; 국민건강보험, 2006). 이 같이 보험급여제한에 대한 업무특성상 법적인 기준은 마련되어있으나, 그 유형이 너무나 다양하고 복잡하므로 관련 청구건에 대해 실제 조사를 통해 개별적으로 판단해야 하는 경우가 대다수이다.

¹ (121-749) 서울 마포구 염리동 168-9, 국민건강보험공단 건강보험정책연구원, 부연구위원.

² 교신저자: (150-874) 서울 영등포구 방승길 13, 국가보훈처 정보화팀, 사무관. E-mail : hanjt@korea.kr

³ (614-735) 부산 부산진구 개금동 633-165, 인제대학교 의과대학 예방의학교실, 교수.

⁴ (712-749) 경북 경산시 대동 214-1, 영남대학교 통계학과, 교수.

국민건강보험공단 내부 자료에 의하면, 2001년 주상병 기준 상해상병과 상해외인상병으로 진료 받은 건은 1천 8백만건 이상이며, 총진료비는 9천억원 이상이던 것이, 2006년에는 3천 6백만건 이상으로 2배 이상 증가하였으며, 총진료비 또한 2배 이상 증가한 1조9천억원으로 나타났다. 이처럼 상해상병 및 상해외인상병이 증가하고 있는 양상을 띠고 있는 바, 지속적으로 상해요인관련 조사물량은 늘어날 것으로 추측된다. 그러나 늘어나는 조사물량에 비해, 현재 국민건강보험공단에서 상해요인관련 업무를 수행하는 담당자들은 한정되어 있으며, 국민건강보험공단의 2005년 조직진단 및 업무재설계 보고서 기준에 따르면 업무량이 이미 현재 인력으로도 포화상태임을 제시하고 있다 (국민건강보험공단, 2005).

이처럼 한정된 인력을 통해서 증가하는 업무량을 보다 효율적으로 수행하기 위해서는 상해요인조사 업무의 효율화 및 환수 결정을 제고를 위한 조사대상자 발체기준의 고도화 방안을 마련해야 한다. 기존 국민건강보험공단의 자료를 통해 맞춤형 고혈압 사후관리 모형 연구나 위암 예측모형 연구는 업무의 효율성을 개선할 수 있을 것으로 제시했으며 (박일수 등; 2008, 2010), 한은정 등 (2009)은 표준장기요양 이용계획 작성의 효율화 방안을 마련하고자 노인장기요양급여 권고모형을 개발하였다. 최근에는 데이터마이닝을 활용한 국민연금의 부정 수급자를 관리하고자 예측모형을 개발한 연구가 발표되었다 (차경엽, 2010).

이에 본 연구에서도 상해상병 유형에 대해 일정금액 이상의 발체 등과 같은 과거의 발체기준에서 통계적 모형을 활용한 고도화된 형태로의 조사대상자 선정기준을 제시하고자 한다.

2. 분석자료 및 모형 구축

2.1. 분석자료 구축

국민건강보험공단의 상해요인 관리업무 중 가장 효율적인 조사대상자 선정이 필요한 상해상병 조사대상을 색출을 하기 위해 2001년~2006년까지의 상해상병 (손상 증독 및 외인에 의한 특정 기타 결과, 국제질병분류 (international classification of disease-10th edition, S00~T98) 및 상해외인상병 (질병이환 및 사망의 원인, V01~Y98)으로 청구된 진료 건 중 상해요인 관리 대상으로 선정된 자들의 기본적인 정보와 이들의 상해환수결정 정보를 연결하여 분석데이터베이스를 구축하였다.

분석을 위해 기본적으로 구축되어진 조사대상자 자료는 표 2.1과 같다.

표 2.1 상해상병 연구대상자 분석자료

구 분	항 목
자적 및 보험료 자료 (2005~2006)	대상자ID, 산정건강보험료, 직역구분, 자격변동, 거주지역, 도시규모
개인요양급여자료 (2005~2006)	대상자ID, 성별, 연령, 요양기관정보, 총진료비 (청구, 심사결정), 공단부담금 (청구, 심사결정), 진료형태, 진료개시일, 진료과목, 주·부상병 (S, T 및 V, W, X, Y 질병코드), 입내원일수, 투약일수
상해환수결정자료 (2001~2006)	대상자ID, 결정일자, 사고발생일자, 처리구분, 처리사유, 결정금액, 상병발생경위, 결정근거 및 사유
파생변수	대상자 ID, 과거 부당결정 환수경험 유무, 과거 1년간 건강보험증 변동횟수, BIDM (Barell Injury Diagnosis Matrix)*

* BIDM은 1986년 이스라엘 보건복지부 내 상해예방연구부와 관련 임상가들에 의해 고안되었고, 상해통계를 위한 국제협력모임에서 Barell 등 (2002)에 의해 개발이 시작된 것으로, 외상 (injury)을 기술하고 있는 모든 질병코드를 포괄하고 있는 2차원 (17개 신체부위 그룹, 16개의 상해 특성)의 매트릭스임.
 - 신체부위 그룹: 머리 및 목 (1. 외상성 손상 2. 다른 머리 3. 목 4. 다른 머리와 목 부위), 척추 및 상부 등 (5. 척추신경 6. 척추),
 몸통 (7. 흉부 8. 배 9. 골반 및 하부 등 10. 배, 하부 등 골반 11. 다른 몸통 부위), 사지 (12. 상부사지 13. 엉덩이 14. 다른 하부 사지),
 분류되지 않은 타 부위 (15. 광범위 신체 손상 16. 여러 기관), 기타 (17. 기타)
 - 상해 특성: 1. 골절 2. 탈구 3. 내부기관 손상 4. 상처 5. 절단 6. 혈관 7. 표피 손상 및 타박상 8. 압박손상 9. 화상 10. 관통상 11. 이물질에 의한 다른 손상
 12. 증독 13. 작용 14. 광범위 상해 15. 다른 특정 상병 16. 기타

2.2. 모형구축

상해상병 관리모형은 상해상병 조사대상자에 대한 조사업무량이 매년 증가됨에 따라, 조사대상 건수를 줄이면서 조사건 단위의 부당결정 환수 결정을 및 환수 결정금액을 높이기 위해 하이브리드 모형

(hybrid model)을 활용하였다. 본 연구에서 적용한 하이브리드 모형은 SAS E-Miner 4.1을 사용한 데이터마이닝 기법으로 개발된 예측모형 (predictive model)과 업무관점의 규칙 (business rule model)이 결합된 모형으로, 두 가지 모델을 활용하여 조건에 만족하는 대상자를 선정하기 위한 모형이다.

상해상병 관리모형의 개발은 건강보험가입자의 직장가입자만 선정하여 개발하였다. 본 연구의 상해상병 관리모형 구축에 대한 보다 자세한 프로세스는 다음과 같다.

Step1. 예측모형 개발·평가를 위해 상해상병 유형에 대한 분석데이터베이스를 모형개발 및 모형내적 평가용 (2005년)과 모형외적평가용 (2006년)으로 구축하였다.

사용된 독립변수는 표 2.1에서 제시된 조사대상자의 기본적인 정보 및 이들로 생성된 파생변수를 활용하였고, 종속변수는 해당년도의 환수결정유무변수로 활용하였다.

Step2. 의사결정나무기법을 이용하여 상해상병 조사대상 예측모형을 개발·평가한다.

Step3. (Step2)에서 개발된 예측모형을 이용하여, 효율적 (적정한 조사건)이고 효과적 (환수 결정금액 상승)인 선정기준결정을 위해 업무규칙인 전체 조사대상건과 진료건당 공단부담금 변화에 따른 시물레이션을 한다.

Step4. 환수 결정금액을 고려해 최종 조사물량을 선정할 수 있는 업무관점의 규칙 모델의 선정기준을 결정한다.

3. 연구결과

3.1. 상해상병 연구대상자의 일반적 특성

2005년 “상해상병 진료건” 유형으로 조사된 52,755건의 상해상병 조사대상자 특성을 파악하기 위해 표 3.1에서와 같이, 수진자의 성, 연령, 건강보험가입자의 직역, 거주지역 및 도시규모, 과거 1년간 건강보험증 변동 횟수에 대하여 살펴보았다. 먼저 성별로 보면, 남성이 여성보다 조사대상건이 상대적으로 많았으며, 실제 환수된 비율분포가 남성의 경우는 환수건 기준 9.6%이고 환수 결정금액 기준은 9.4%로 여성 (3.2%)보다 약 6%p 높은 환수 결정을 분포를 보였다. 연령대에서는 20~40대가 전체 환수 결정 실적의 33.2%를 차지하였고, 다음 50대가 8.2%의 환수분포를 나타냈다. 건강보험 가입자의 직역에 따른 환수분포는 공무원 및 사립학교 교직원 가입자가 직장가입자보다 약 3%p 높은 7.0~7.4%의 환수분포를 보였고, 농어촌에 거주하는 자들이 도시 (대도시, 중소도시) 거주자보다 상해상병으로 인한 진료건은 적었으나, 실제 환수된 실적은 대도시가 상대적으로 높았다. 한편 연구대상자의 건강보험증 변동 횟수가 증가할수록 환수 결정 실적이 높아지는 것으로 나타났으며, 16개 도시직장별로의 환수 결정 실적은 대구, 제주가 가장 높았고 다음 경북, 서울, 경기 등의 순으로 나타났다.

3.2. 상해상병 관리 모형 개발

3.2.1. 상해상병 조사대상자 예측모형 개발

모형은 2005년 당시 조사대상건인 52,755건을 기준으로 국민건강보험공단의 상해요인 업무처리 요령에 따라 환수 결정건 (3,504건)과 결정되지 않는 건 (49,251건)으로 구성된 데이터를 활용하여, 데이터마이닝 기법 중에 하나인 의사결정나무기법 (decision tree)을 이용하여 개발하였다.

개발된 상해상병 조사대상자 예측모형의 안정성 및 정확성을 확보하기 위해 2005년 분석데이터의 60%만을 활용하여 모형을 개발하고 (모형개발용 : training dataset), 나머지 40%를 이용하여 모형의 내적 타당도 (internal validation)를 평가하였다. 또한, 일반화된 모형개발을 위해 모형개발시점의 2005년 데이터와 같은 자료구조의 2006년 데이터에 예측모형을 적용시켜 외적타당도 (external validation)를 평가하였다.

표 3.1 상해상병 연구대상자들의 일반적 특성에 따른 환수 결정을 (직장 : 2005년)

(단위 : 건, 천원, %)

일반적인 현황		대상		환수		환수 결정율	
		건	금액	건	금액	건	금액
성별	남	28,317	43,248,473	2,732	4,079,791	9.6	9.4
	여	24,438	39,527,707	772	1,018,251	3.2	2.6
연령	10세 이하	4,904	5,771,561	367	526,194	7.5	9.1
	20 29세	7,590	10,172,110	951	1,381,653	12.5	13.6
	30 39세	8,492	11,485,465	915	1,251,027	10.8	10.9
	40 49세	5,125	7,003,601	509	760,197	9.9	10.9
	50 59세	3,628	5,128,126	299	420,2440	8.2	8.2
	60세 이상	23,016	43,215,317	463	758,727	2.0	1.8
가입자 직역	일반직장	10,774	17,618,160	386	543,296	3.6	3.1
	공교 (공무원 및 사립학교 교직원)	41,981	65,158,021	3,118	4,554,746	7.4	7.0
거주지역	강원	1,748	2,723,426	119	130,541	6.8	4.8
	경기	9,792	15,807,564	654	1,020,010	6.7	6.5
	경남	4,127	5,850,095	240	311,909	5.8	5.3
	경북	3,911	6,220,603	310	491,037	7.9	7.9
	광주	1,411	2,070,924	91	119,349	6.4	5.8
	대구	2,313	3,523,478	231	302,758	10.0	8.6
	대전	1,500	2,393,010	69	93,053	4.6	3.9
	부산	3,857	5,663,129	191	260,534	5.0	4.6
	서울	8,437	14,547,682	677	1,030,708	8.0	7.1
	울산	1,451	2,122,742	77	90,704	5.3	4.3
	인천	2,570	4,074,110	151	217,169	5.9	5.3
	전남	2,726	3,980,829	116	152,366	4.3	3.8
	전북	2,833	3,976,783	145	198,815	5.1	5.0
	제주	649	1,015,754	61	87,118	9.4	8.6
	충남	2,482	3,838,472	158	206,972	6.4	5.4
	충북	1,943	3,201,806	113	189,021	5.8	5.9
	결측	1,005	1,765,773	101	195,979	10.0	11.1
도시규모	농어촌지역	7,099	11,294,401	452	635,848	6.4	5.6
	대도시	21,453	34,277,898	1,473	2,099,371	6.9	6.1
	중소도시	23,198	35,438,110	1,478	2,166,844	6.4	6.1
	결측	1,005	1,765,773	101	195,979	10.0	11.1
과거 1년간 건강보험증 획득	1회	51,838	81,163,502	3,269	4,670,842	6.3	5.8
	2회	876	1,535,020	220	401,335	25.1	26.1
	3회	37	72,760	12	22,274	32.4	30.6
	4회	3	3,869	2	2,564	66.7	66.2
	5회	1	1,0280	1	1,028	100.0	100.0
전체		52,755	82,776,180	3,504	5,098,042	6.6	6.2

본 연구의 상해상병 조사대상자 예측모형은 의사결정나무분석 중 ‘지니지수 (Gini index)’를 활용한 마디분리기준을 활용하여 개발되었으며, 그 결과 예측정확도를 나타내는 root ASE (average squared error) 및 오분류율이 각각 0.2340 (모형평가용 : 0.2374) 및 0.0620 (모형평가용 : 0.0653)로 나타나, 안정성과 정확성이 확보된 모형이 도출되었다 (표 3.2 참조).

표 3.2 상해상병 조사대상자 예측모형의 내적 타당도 평가 (2005년)

	모형생성용 (training dataset : 60%)		모형평가용 (validation dataset : 40%)	
	Root ASE	오분류율	Root ASE	오분류율
의사결정나무모형	0.2340	0.0620	0.2374	0.0653

그림 3.1, 표 3.3의 결과를 살펴보면, 예측모형의 적중률이 높을수록 모형의 향상도 (lift)가 높음을 확인할 수 있으며, 그 결과 본 연구에서 개발된 모형의 향상도는 그룹1이 9.14배 (조사건 기준), 9.44배 (환수 결정금액 기준) 정도 향상됨을 알 수 있다. 또한 집단 13까지의 누적 검출력은 조사건을 중심으로 봤을 때는 69.2%, 68.9%인 것으로 나타나, 전체 조사환수결정건 4,209건 중 2,914건 (금액기준 : 46억 원)을 차지하였다.

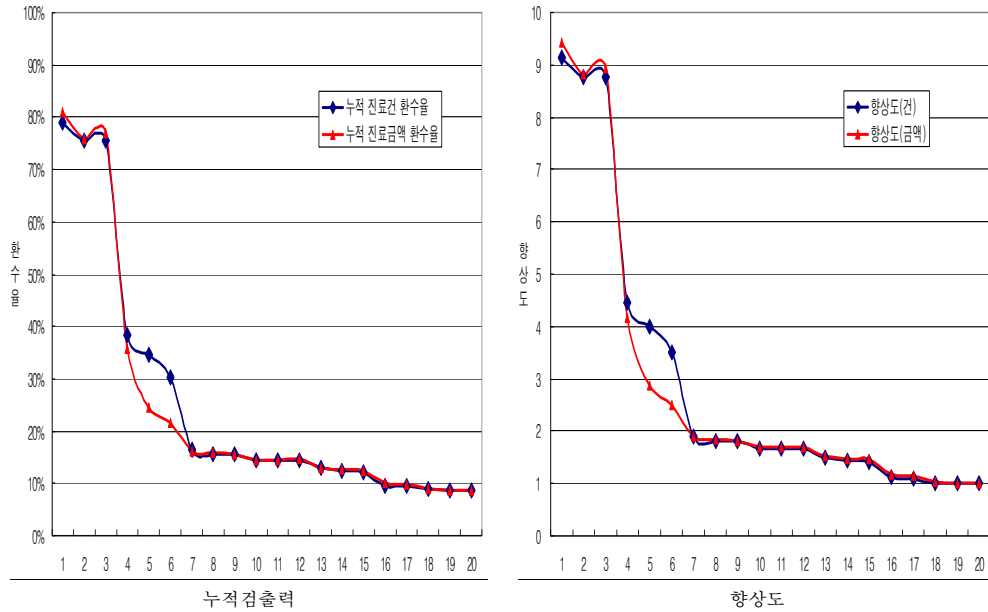


그림 3.1 직장 상해상병 조사대상 예측모형의 환수 결정율, 검출력 및 향상도

표 3.3 상해상병 조사대상자 예측모형의 외적 타당도 평가 (2006년)

모형집단	대상건	환수건	대상금액(천원)	환수 결정금액(천원)	환수 결정율(%)		누적 환수 결정율(%)		누적 검출력(%)		향상도	
					건	금액	건	금액	건	금액	건	금액
1	109	86	167,547	135,806	78.9	81.1	78.9	81.1	2.0	2.0	9.14	9.44
2	59	41	96,391	63,876	69.5	66.3	75.6	75.7	3.0	3.0	8.76	8.81
3	74	56	121,946	96,807	75.7	79.4	75.6	76.8	4.3	4.4	8.76	8.95
4	789	213	1,639,394	429,958	27.0	26.2	38.4	35.9	9.4	10.9	4.45	4.18
5	216	34	2,468,698	377,789	15.7	15.3	34.5	24.6	10.2	16.5	3.99	2.86
6	252	24	1,244,046	127,808	9.5	10.3	30.3	21.5	10.8	18.4	3.51	2.50
7	9,019	1,263	12,963,325	1,795,914	14.0	13.9	16.3	16.2	40.8	45.3	1.89	1.89
8	2,205	280	2,632,658	330,577	12.7	12.6	15.7	15.7	47.4	50.3	1.82	1.83
9	263	26	433,438	25,766	9.9	5.9	15.6	15.5	48.1	50.7	1.80	1.81
10	3,821	381	4,790,152	470,407	10.0	9.8	14.3	14.5	57.1	57.7	1.66	1.69
11	120	12	219,920	21,609	10.0	9.8	14.3	14.5	57.4	58.0	1.65	1.69
12	217	44	583,280	106,165	20.3	18.2	14.3	14.6	58.4	59.6	1.66	1.69
13	5,530	454	7,911,653	619,666	8.2	7.8	12.9	13.0	69.2	68.9	1.49	1.52
14	1,911	108	2,529,353	150,667	5.7	6.0	12.3	12.6	71.8	71.2	1.42	1.46
15	561	35	594,348	47,745	6.2	8.0	12.2	12.5	72.6	71.9	1.41	1.46
16	12,844	610	18,993,471	944,772	4.7	5.0	9.7	10.0	87.1	86.0	1.12	1.17
17	1,077	45	1,821,127	62,203	4.2	3.4	9.5	9.8	88.2	86.9	1.10	1.14
18	8,198	447	15,832,829	804,434	5.5	5.1	8.8	8.8	98.8	99.0	1.02	1.03
19	651	39	1,356,874	53,734	6.0	4.0	8.8	8.7	99.7	99.8	1.02	1.02
20	847	11	1,376,741	14,019	1.3	1.0	8.6	8.6	100.0	100.0	1.00	1.00
전체	48,763	4,209	77,777,189	6,679,722	8.6	8.6	8.6	8.6	100.0	100.0	1.00	1.00

3.2.2. 상해상병 조사대상자 예측모형 결과

상해상병으로 진료를 받은 수진자의 특성을 반영한 상해상병 조사대상자 예측모형은 그림 3.2와 같다. 건강보험 직장가입자 중 상해상병 관련 부당환수결정에 영향을 주는 수진자 특성요인은 과거 상해 환수결정 유무, BIDM에서 신체부위 및 상해특성 상병코드, 수진자의 연령대, 직역구분, 심사결정공단 부담금, 과거 1년간 건강보험증 변동 횟수로 나타났으며, 이러한 요인에 의해 20그룹으로 예측·분류되었다. 그림 3.2에서와 같이 분류된 예측결과를 보면, 2005년 당시 진료건 중심의 환수 결정율 6.5%였으나 ‘과거 상해사후관리업무를 통해 환수결정경험이 있는 경우’이고 BIDM의 상해특성별 분류에서 1. 골절, 2. 탈구, 3. 내부기관 손상, 7. 표피손상 및 타박상, 8. 압착손상 그리고 14. 광범위한 상해일 사람 중 연령대가 20대 이하인 사람과 40대인 사람의 환수 결정율이 94.8% (집단 1)로 나타났으며, 동일한 기준에서 연령대만 30대, 50대 이상인 사람들의 환수 결정율은 81.3% (집단 2)였다. 또한 과거 ‘상해사후관리업무를 통해 환수결정경험’이 있고 BIDM의 상해특성별 분류가 4. 상처, 5. 절단, 6. 혈관, 9. 화상, 10. 관통상, 11. 이물질에 의한 손상, 12. 중독 13. 독작용 15. 다른 특정 상해 그리고 16. 기타에 의해 진료 받았던 사람들은 50.0%의 환수 결정율을 (집단 3)을 보였다.

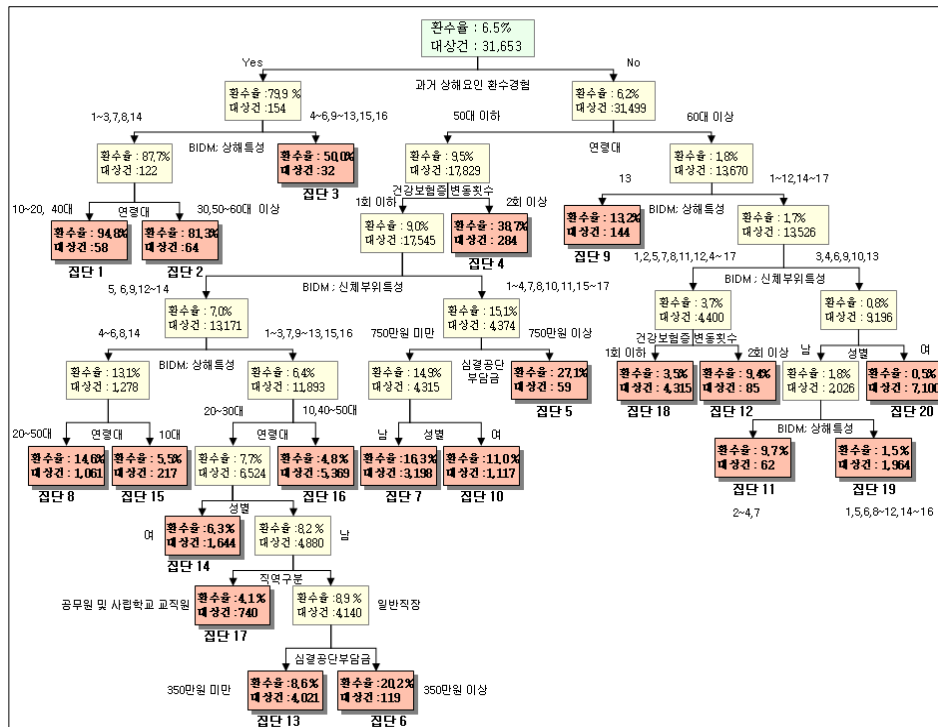


그림 3.2 의사결정나무분석에 의한 상해상병 조사대상 예측모형(2005년)

3.3. 업무규칙 결정을 위한 시뮬레이션

본 연구에서는 환수조사건 단위의 상해상병 조사대상자 예측모형만을 활용하여 최종 상해상병 관리 모형을 구축하는 것이 아닌 예측환수대상건의 진료금액을 업무규칙으로 적용시켜 하이브리드 형태의 모

형으로 개발하였다. 효과적이고 효율적인 업무규칙을 결정하기 위해 그림 3.2의 예측모형에 업무규칙을 적용하여 전체 상해상병 유형 조사대상건 (2006년 상해상병 유형 전체 조사대상건 기준 : 48,763건) 으로부터 조사물량이 50%, 60%, 70%, 80%인 경우 각각에 대하여 환수대상건의 진료금액수준을 최소 67만원부터 최대 300만원까지 10만원 단위로 변화를 주는 시뮬레이션을 실시 (2006년도와 똑같은 전체 조사대상건수로 발췌한 후, 이를 기준으로 50% 80%로 조사물량을 줄일 경우, 가장 효과적이고 효율적인 발췌기준을 재설정하는 것)하였다. 조사건과 조사건의 진료금액을 고려한 시뮬레이션 결과는 그림 3.3과 같으며, 환수건률, 환수금액율, 환수건검출력, 환수금액검출력을 근거로 최종 조사대상 물량을 조절할 수 있는 배분기준을 결정하였다. 그 결과, 하이브리드 모형으로서 직장 상해상병 관리의 최적 모형은 전체 조사대상건의 50% 추출 시에는 상해상병 조사대상 예측모형을 적용시켜 도출된 총 20개의 집단 중 1~13번 집단까지 적용 시키는 동시에 예측환수대상건의 진료금액은 210만원 이상 해당되는 조사건을 발췌하는 것이 가장 적절한 것으로 나타났다. 또한 조사물량을 2006년 전체 직장 상해상병 조사대상건의 60%로 할 경우는 예측모형결과 1~13번 그룹과 190만원을 적용시키고, 70%일 경우는 예측모형결과 1~14번 집단과 130만원 이상 건을 적용시키고, 80%일 경우는 예측모형결과 1~13번 집단과 90만원 이상건을 적용시키는 것이 가장 효율적인 발췌인 것으로 시뮬레이션 결과 나타났다.

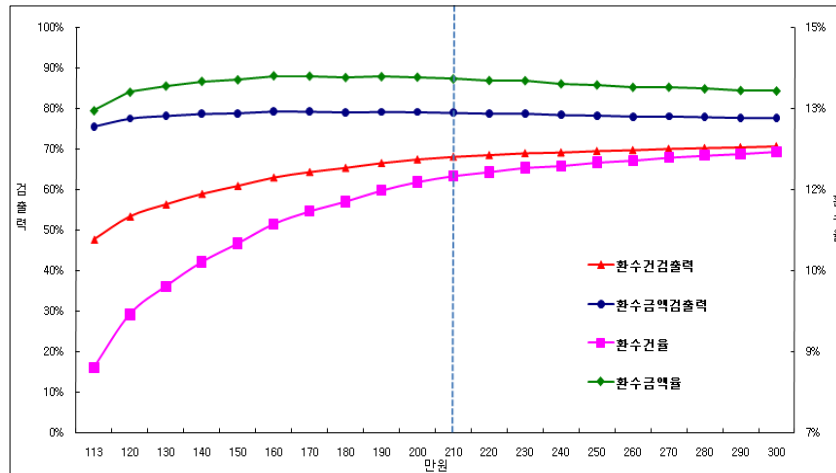


그림 3.3 조사대상건 추출률에 따른 조사대상건 당 진료금액의 변화 시뮬레이션 결과 (전체 조사대상건 중 조사물량 50% 적용 시)

3.4. 상해상병 관리 모형 결과

하이브리드 모형으로 개발된 상해상병 관리모형을 이용한 최종 조사대상건의 발췌기준은 전체 조사대상건이 2006년 상해상병 전체 조사대상건 대비 50%일 경우 표 3.4와 같다. 즉, 조사건수가 확정되면, 예측모형 그림 3.2에서 집단1~13로 분류된 대상건은 우선적으로 선정하고, 다음 선정되지 않은 집단14~20에서는 업무규칙을 적용하여 조사건당 진료비가 210만원 이상인 건은 전수를 조사대상에 포함하는 것이다. 그 결과 전체 조사대상건 기준의 환수율은 11.7% (조사금액기준 환수율 13.5%), 환수건 기준에서의 환수건검출력은 68.0% (환수금액검출력 78.8%)로 나타났다.

표 3.4 상해상병 조사대상건 발체 규칙 (전체 조사대상건 중 조사물량 50% 적용 시)

모형	집단	상해상병 관리 대상자 특성
예측모형	1	과거상해환수경험 있음 & 상해특성 (1~3, 7, 8, 14) & 연령대 (10~20, 40대)
	2	과거상해환수경험 있음 & 상해특성 (1~3, 7, 8, 14) & 연령대 (30, 50~60대 이상)
	3	과거상해환수경험 있음 & 상해특성 (4~6, 9~13, 15, 16)
	4	과거상해환수경험 없음 & 연령대 (50대 이하) & 건강보험증 변동횟수(2회 이상)
	5	과거상해환수경험 없음 & 연령대 (50대 이하) & 건강보험증 변동횟수(1회 이하) & 신체부위특성 (1~4, 7, 8, 10, 11, 15~17) 심결공단부담금 (750만원 이상)
	6	과거상해환수경험 없음 & 연령대 (50대 이하) & 건강보험증 변동횟수(1회 이하) & 신체부위특성 (5, 6, 9, 12~14), 상해특성 (1~3, 7, 9~13, 15, 16) & 연령대 (20~30대) & 성별 (남) & 직역구분 (직장) & 심결공단부담금 (350만원 이상)
	7	과거상해환수경험 없음 & 연령대 (50대 이하) & 건강보험증 변동횟수(1회 이하) & 신체부위특성 (1~4, 7, 8, 10, 11, 15~17) 심결공단부담금 (750만원 미만) & 성별 (남)
	8	과거상해환수경험 없음 & 연령대 (50대 이하) & 건강보험증 변동횟수(1회 이하) & 신체부위특성 (5, 6, 9, 12~14), 상해특성 (4~6, 8, 14), 연령대 (20~50대)
	9	과거상해환수경험 없음 & 연령대 (60대 이상) & 상해특성 (13)
	10	과거상해환수경험 없음 & 연령대 (50대 이하) & 건강보험증 변동횟수(1회 이하) & 신체부위특성 (1~4, 7, 8, 10, 11, 15~17) 심결공단부담금 (750만원 미만) & 성별 (여)
	11	과거상해환수경험 없음 & 연령대 (60대 이상) & 상해특성 (1~12, 14~17) & 신체부위특성 (3, 4, 6, 9, 10, 13) & 성별 (남) 상해특성(2~4, 7)
	12	과거상해환수경험 없음 & 연령대 (60대 이상) & 상해특성 (1~12, 14~17) & 신체부위특성 (1, 2, 5, 7, 8, 11, 12, 14~17) 건강보험증 변동횟수 (2회 이상)
	13	과거상해환수경험 없음 & 연령대 (50대 이하) & 건강보험증 변동횟수(1회 이하) & 신체부위특성 (5, 6, 9, 12~14), 상해특성 (1~3, 7, 9~13, 15, 16) & 연령대 (20~30대) & 성별 (남) & 직역구분 (직장) & 심결공단부담금 (350만원 미만)
업무규칙	14~20	진료금액 210만원 이상
전체		환수율 (조사대상건: 11.7%, 금액: 13.5%), 검출력 (조사대상건: 68.0%, 금액: 78.8%)

4. 고찰 및 결론

건강보험제도권 내에서 건강보험가입자의 진료내용이 건강보험급여기준에 해당되지 않는 고의 또는 중대한 과실로 인한 범죄행위에 기인하거나, 고의로 사고를 발생시켰을 때에는 당해 보험급여제한을 할 수 있다. 보험급여제한의 한 유형인 건강보험 상해요인과 관련하여 국민건강보험공단에서는 일정한 기준에 의해 발체된 15가지 유형에 대해 각 지사가 전수 사실·확인조사를 실시하고 있다. 그러나 해마다 상해상병관련 진료건의 증가, 담당인력수의 감소, 업무유형의 복잡성과 다양성 등으로 인하여 현재 상해상병 조사 대상자 발체기준보다 고도화된 발체기준을 개발하여 업무의 효율화를 도모하기 위한 방안이 절실히 요구되는 상황이다. 이에 본 연구에서는 건강보험 상해요인조사 업무의 효율화 방안을 도출하고자 상해상병조사 대상자 선정기준을 마련하기 위한 관리모형을 개발하였다. 모형개발 방법론은 조사대상건의 효율성과 조사환수 결정금액의 효과성을 고려하여, 데이터마이닝 기법 중 의사결정나무분석 알고리즘을 적용시킨 조사대상 예측모형과 업무규칙을 혼합한 하이브리드 모형 형태로 개발하였다.

그 결과 데이터마이닝 기법 중 의사결정나무모형으로 개발된 상해상병 조사대상자 예측모형에서는 모형 적용 전보다 그 효과성 (향상도 기준)이 최상위 (집단 1) 조사건 기준 9.14배, 조사금액 기준 9.44배 증가하였으며, 본 모형에 진료건 당 금액 (업무규칙) 변화를 시뮬레이션 시켜봄으로서 전체 조사대상자 건수가 확정되었을 경우, 조사대상자 대비 환수건 및 금액에 대해 최대의 효율성과 효과성을 거둘 수 있는 하이브리드 모형을 제시하였다. 그 예로 전체 조사대상건 대비 조사물량을 50%로 할 경우, 전체 조사대상건 중 환수건율, 전체 조사금액 중 환수금액율, 환수건 검출력 그리고 환수금액검출력을 고려하여, 상해상병 조사대상자 예측모형에서 나타난 13개 그룹 및 조사환수대상 진료건 당 금액을 210만원 이상으로 적용하는 것이 가장 효과적인 것으로 분석되었다.

참고문헌

국민건강보험공단 (2003). <건강보험 상해요인 업무총람>, 국민건강보험공단, 서울.

- 국민건강보험공단 (2005). <2005년 조직진단 및 업무재설계(BPR)보고서>, 국민건강보험공단, 서울.
- 국민건강보험공단 (2006). <업무처리요령>, 국민건강보험공단, 서울.
- 국민건강보험공단 (2010). <2009 건강보험통계연보>, 국민건강보험공단, 서울.
- 박일수, 용왕식, 김유미, 강성홍, 한준태 (2008). 데이터마이닝 기법을 활용한 맞춤형 고혈압 사후관리 모형 개발. <응용통계연구>, **21**, 639-647.
- 박일수, 한준태, 강석복, 지재훈 (2010). 데이터마이닝을 이용한 위암 예측모형개발과 활용. <한국데이터정보과학회지>, **21**, 1253-1260.
- 차경엽 (2010). 데이터마이닝을 이용한 국민연금 부정수급 예측모형 개발. <한국통계학회논문집>, **17**, 1-8.
- 한은정, 이정석, 김동건, 강임옥 (2009). 데이터마이닝 기법을 활용한 노인장기요양급여 권고모형 개발. <응용통계연구>, **22**, 1229-1237.
- Barell, V., Aharonson-Daniel, L., Fingerhut, L. A., Mackenzie, E. J., Ziv, A., Boyko, V., Abargel, A., Avitzour, M. and Heruti, R. (2002). An introduction to the Barell body region by nature of injury diagnosis matrix. *Injury Prevention*, **8**, 91-96.

Developing the administrative model using the data mining technique for injury in National Health Insurance

Il-Su Park¹ · Jun-Tae Han² · Hae-Sook Sohn³ · Suk-Bok Kang⁴

¹Health Insurance Policy Research Institute, National Health Insurance Corporation

²Informatization Team, Ministry of Patriots and Veterans Affairs

³Department of Preventive Medicine, Inje University School of Medicine

⁴Department of Statistic, Yeungnam University

Received 9 April 2011, revised 17 May 2011, accepted 21 May 2011

Abstract

We developed the hybrid model coupled with predictive model and business rule model for administration of injury by utilizing medical data of the National Health Insurance in Korea. We performed decision tree analysis using data mining methodology and used SAS Enterprise Miner 4.1. We also investigated under several business rule for benefits (expense paid by insurer) and claims of injury in National Health Insurance Corporation. We can see that the proposed hybrid model provides a quite efficient plausible results.

Keywords: Data mining, hybrid model, injury.

¹ Senior Researcher, National Health Insurance Policy Research Institute, National Health Insurance Corporation, Seoul 121-749, Korea.

² Corresponding author: Deputy Director, Ministry of Patriots and Veterans Affairs, Seoul 150-874, Korea. E-mail: hanjt@korea.kr

³ Professor, Department of Preventive Medicine, Inje University School of Medicine, Busan 614-735, Korea.

⁴ Professor, Department of Statistics, Yeungnam University, Gyeongsan 712-749, Korea.