

데이터마이닝을 이용한 건강보험 상해요인 조사 대상 선정 모형 개발 -건강보험 지역가입자 상해상병 진료건을 중심으로-

박일수*, 박소정**, 한준태***, 강성홍**
위덕대학교 보건학과*, 인제대학교 보건행정학과**, 한국장학재단***

Development of the Fraud Detection Model for Injury in National Health Insurance using Data Mining Focusing on Injury Claims of Self-employed Insured of National Health Insurance

Il-Su Park*, So-Jeong Park**, Jun-Tae Han***, Sung-Hong Kang**
Dept. of Health, The Uiduk University*, Dept. of Health Policy and Management, The Inje University**, Korea Student Aid Foundation***

요약 상해상병으로 청구되는 건수가 증가함에 따라 조사 대상을 보다 정교하게 선정하여 상해요인 조사 대상을 줄이면서 환수를 및 환수금액을 올릴 수 있는 방안을 마련할 필요가 있다. 이를 위해서 2006~2011년까지의 상해요인 조사자료를 수집하여 의사결정나무 모형을 활용하여 지역가입자 상해상병 진료건에 대한 부당환수 조사대상 선정 모형을 개발하였다. 최종 개발된 모형결과에 따르면, 조사대상 유형은 18개로 분류되었고, 이러한 분류결과는 실제 조사가 시행될 시, 모형을 적용하지 않았을 때 보다 최고 12.8배 높은 부당환수결정율을 나타낼 수 있을 것으로 분석되었다. 또한, 본 연구에서 개발된 조사 대상자 선정 모형을 실제 업무에 적용하기 위해서는 조사물량 대비 국민건강보험공단의 조사인력 및 운영 계획을 보다 면밀히 검토해야만 모형 적용의 효과성이 극대화 될 수 있을 것으로 판단된다.

주제어 : 상해상병, 국민건강보험, 지역가입자, 데이터 마이닝, 의사결정나무

Abstract According to increasing number of injury claims, the challenge is reducing investigation of cases of injuries by selecting them more delicately, while also increasing the redemption rates and the amount of restitution. In this regards, we developed the fraud detection model for injury claims of self-employed insured by using decision tree after collecting medical claim data from 2006 to 2011 of the National Health Insurance in Korea. As a result of this model, subject types were classified into 18 types. If applying these types to the actual survey compared with if not applying, the redumption collecting rate will be increasing by 12.8%. Also, the effectiveness of this model will be maximize when the number of claims handlers considering their survey volume and management plans are examined thoroughly.

Key Words : Injury, National health insurance, Self-employed insured, Data mining, Decision tree

* 본 논문은 2012 년 "상해요인 조사대상 발취기준 고도화" 연구보고서(국민건강보험공단 발주)에 수록된 내용의 일부를 수정 및 보완한 것임.

Received 22 August 2013, Revised 25 September 2013

Accepted 20 October 2013

Corresponding Author: Sung-Hong Kang(The Inje University)

Email: hcmkang@inje.ac.kr

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

국민건강보험공단의 자료에 따르면 2007년부터 2011년까지 상해질환(S00~T98 : 손상, 중독 및 외인에 의한 특정 기타 결과)으로 병원을 찾는 환자의 수는 매년 약 50만 명씩 증가 하고 있으며 이에 따라 진료비도 해마다 약 1,900억 원씩 증가하여 2011년 한 해 동안에는 상해 환자의 급여비는 2조 2천억 원 이상 지급되었다[1]. 상해 환자의 보험적용은 상해사고의 내용에 따라 자동차보험, 산재보험, 건강보험 등에서 부담을 한다. 즉, 자동차 사고로 인한 상해는 자동차 보험에서 부담하고, 산재사고로 인한 상해는 산재보험에서 부담하고, 이 밖에 건강보험의 급여범위에 해당되는 것은 건강보험에서 부담을 한다. 그러나 건강보험의 적용 대상이 아닌 상해 사고에 대해서 환자나 회사, 자동차 보험회사 등의 경제적 이익 때문에 건강보험을 적용하여 건강보험 재정을 악화시키고 있다. 2006년 한해 건강보험 이용자를 기초로 한 연구에 따르면 건강보험을 이용한 환자 1,238만 800명 중 22.5%(278만 5000명)가 직장에서 재해를 입은 것으로 나타났다[2]. 이 같은 규모는 같은 해 고용노동부가 발표한 산업재해자 수 8만 9900명의 30배가 넘는다. 의료기관에서 자동차 보험 및 건강보험에 이중 청구하는 금액이 연간 366억원, 자동차 보험 대상 환자가 보험회사와 합의 종료 후 자동차 사고로 인한 후유증 질환에 대한 치료비를 건강보험으로 청구하는 금액이 연간 917~3,007억 원 정도 되는 것으로 나타났다[3].

2000년에 4억 건이던 건강보험 청구건수는 2010년 12억 6천 건으로, 2000년 대비 3.06배 증가, 연평균 증가율 약 11.8% 증가하고 있고, 진료비는 2000년 9조원이던 것이 2010년 33조 8천억 원으로, 2000년 대비 3.73배 증가, 연평균 증가율 14.1% 증가함에 따라 국민건강보험공단의 급여비 대상 업무가 크게 증가하고 있으나 인력은 증가하지 않음에 따라 국민건강보험공단의 급여 관리 업무를 효율적으로 수행하기 어려운 실정이다. 특히, 상해 상병인 S,T상병으로 인한 진료 실 인원이 2006년에는 1,093만 명이던 것이 2010년에는 1,332만 명으로 크게 증가 하였으나 이를 담당할 상해요인 업무담당자는 증가하지 않음에 따라 건강보험 공단의 상해요인 업무의 효율적인 방안을 개발할 필요가 있다[4,5] 특히, 상해요인 조사 대상으로 선정된 건수가 2004년 351,598건이던 것이 2010

년에 911,487건으로 증가하였으나 조사인력은 2004년 360명 이던 것이 2010년에는 280명으로 도리어 감소하여 2004년 1인당 월별 평균처리건수가 81.4건이던 것이 2010년에는 276.2건으로 크게 증가함에 따라 상해요인 업무가 효율적으로 이루어지지 못하고 있다[6, 7]. 이를 개선하기 위해서는 상해요인 업무 담당자를 크게 늘여야 하나 현실적으로 단기간 내에 이것이 이루어지기 어렵기 때문에 상해요인 조사 대상 선정을 보다 정교하게 선정하여 상해요인 조사 대상을 줄이면서 환수율 및 환수금액을 올릴 수 있는 방안을 마련할 필요가 있다. 이를 위해서는 상해요인 조사대상자의 발체기준의 정교화 및 데이터마이닝 기법을 활용한 Fraud Detection 방법론을 활용하여 발체기준을 선정할 필요가 있다. 이미 이러한 방법은 미국 국세청의 Triage for Tax AuditTM, 미국 텍사주의 세금탈루 방지 시스템인 ADS(Advanced Detection System), 미국 저소득층 진료기관 부당청구방지를 위한 MFASD(Medicaid Abuse Detection System)에 적용되어 그 효과성이 입증되었다[8-12].

이에 본 연구의 목적은 국민건강보험법 제 48조에 의한 건강보험 급여 제한을 하기 위해 실시하는 국민건강보험공단의 상해요인조사 사후관리업무¹⁾ 중 “상해상병 진료건”에 대해 보다 효과적이고 정교한 조사대상자 발체 기준을 마련하여, 건강보험공단의 상해요인 조사 업무의 효율성을 도모하는 것이다.

1) 국민건강보험공단의 보험급여업무의 큰 흐름은 요양기관으로부터 수진자가 진료행위를 받고 난 이후에 진료비의 일부인 공단부담금을 지급하는 것임. 이에 따라 국민건강보험공단은 급여비를 지급하고 난 이후에 지급된 보험급여의 정당성에 대한 급여사후관리업무를 수행함. 국민건강보험공단의 사후관리업무에는 부당수급, 상해요인, 공무상 요양급여, 요양급여비용의 환수, 본인부담환급금 지급, 본인부담액 보상금의 지급, 본인부담액상한제 사후 환급금 지급 등이 해당되며, 그 중 상해요인 관리업무는 사전관리 및 사후관리로 구분됨

- 상해요인 사전관리: 요양기관이 국민건강보험공단(지사)로 조회한 “급여제한여부조회서”를 접수받아 이를 근거로 상병발생원인을 조사하며, 상해발생원인이 “고의 및 범죄행위”, “제 3자에 의한 가해”, “업무상 재해” 등을 확인될 경우, 수진자 및 요양기관에게 보험급여제한/적용, 선급여 후 사후관리 등을 실시함
- 상해요인 사후관리: 사전관리 업무 중 “선급여 후 사후관리” 업무의 연장으로, 현재 국민건강보험공단에서는 총 15가지 유형으로 구분하고 있으며, 그 유형 중 하나가 “상해상병 진료건”에 해당됨

2. 연구방법

2.1 자료수집

상해상병 진료건 조사 대상자 선정 모형 개발을 위해 본 연구에서 활용한 주요 국민건강보험공단 데이터셋은 급여사후대상으로 발체된 자료의 결정내역인 상해요인 결정자료(2006~2011년), 그들의 진료내역을 파악할 수 있는 개인현물급여 자료(2006~2011년) 및 진료내역 상세자료(2010~2011년)이다.

2.2 상해상병 진료건 조사대상자 선정 모형 개발과정

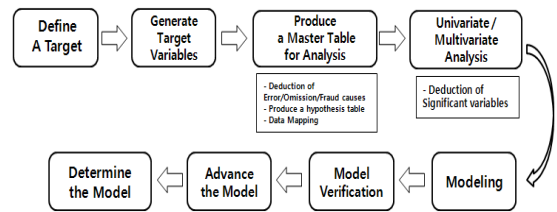
분석 및 모형 구축을 위한 자료구성은 아래 제시된 단계에 따라 생성하였다.

전체 상해요인 사후처리 현황은 전체 가입자를 대상으로 파악함으로써 본 연구에 대한 필요성을 재검토하였고, 조사대상자 선정을 위한 모형개발은 대상자의 특성을 반영하기 위해 지역가입자, 직장가입자, 직장 피부양자로 자격을 분류하여 각각 모형을 개발하였으나, 본 논문에서는 지역가입자에 대한 모형결과만을 제시하고자 한다.

1단계 : 급여사후 결정내역 자료 중 본부 발체건을 중심으로, 발체사유가 상해상병부호 진료발체(질병분류 코드 : S01~T98)이고, 급여사후결정사유가 구상급, 부당이득급이며, 급여 사후처리구분이 부당(완전적용 또는 일부적용), 정당급여인 건만 색출하였다. 단, 국민건강보험공단의 급여사후 결정내역 자료는 진료시점(2010년/2011년)의 형태로 구성함으로써, 2010년 자료는 모형개발 및 내적 모형평가(Internal Validation)용 자료, 2011년 자료는 외적 모형평가(External Validation)용 자료로 활용하였다.

2단계 : 단계 1에서 생성된 데이터셋을 기준으로, 진료내역 자료 및 진료내역 상세 자료 등을 이용하여 최종 분석 데이터셋을 구축하였다. 단, 현행 국민건강보험공단의 조사대상자 발체기준에서 약국 청구건을 제외하고 있음에 따라, 진료내역 자료 중 약국자료는 제외하였다.

위 단계에 따라 개발된 모형의 개발 프로세스는 <그림 1>과 같다.



[Fig. 1] Model Development Process

2.3 변수 정의

최종 분석 및 모형 개발을 위한 데이터셋을 구성하는 주요 변수들이 가지는 특성은 아래와 같다. 단, 아래 제시된 주요 변수는 최적의 모형 구축을 위해 활용된 변수이며, 최종 모형에 사용된 변수는 아님을 밝혀둔다²⁾.

- 인구사회학적 특성 : 수진시점의 연령, 실거주지(시군구명), 지역특성(대도시, 중소도시, 농어촌), 수진 당시 산정보험료, 과거 1년 간 건강보험증 변동횟수(2009년)
- 진료 이용 특성 : 수진시점의 주말여부, 수진당시 물리치료 횟수, 수진당시 수술유무, 거주지와 동일지역 진료 유무(실거주지와 요양기관 주소 비교), 심사결정공단부담금(급여비)
- 과거 상해 부당 결정 특성
 - 대상자 개인 기준 : 과거 1년 전(2009년) 개인별 총 부당 건수 및 금액, 과거 1년 전 개인별 자동차보험 연계 부당 건수, 과거 1년 전 개인별 산재보험 연계 부당 건수
 - 대상자 동일세대(건강보험증 단위) 기준 : 과거 1년 전(2009년) 세대별 총 부당 건수 및 금액, 과거 1년 전 세대별 자동차보험 연계 부당 건수, 과거 1년 전 세대별 산재보험 연계 부당 건수
 - 상해상병별 환수결정율 : 과거 1년전(2009년) 상해상병 조사건 기준의 상해상병별 환수결정율
- 수진당시 요양기관 정보 : 지역특성(대도시, 중소도시, 농어촌), 요양기관 종별
- 상해상병 양상, 상해상병 부위로서 생성된 Barell Matrix : Barell Matrix는 상해상병의 양상을 12가지로 분류하고, 상해상병이 발생하는 부위를 대그

2) 제시된 각 요인변수를 모형 구축 시 모두 활용하였으나, 변수의 중요도에 따라 상대적으로 중요하지 않은 변수는 최종 모형에서 제거됨

롭 5가지, 중분류 9가지, 소분류 36가지로 분류한 것으로, 현재 미국 질병관리본부에서는 Barell Matrix를 ICD-9CM 기준으로 제공하고 있음[13]. 그러나 이러한 질병분류체계는 우리나라 의료기관의 그것과는 상이한 점, 또 본 연구가 의무기록 조사자료가 아닌 청구자료인 점 등을 감안하여, 현재 우리나라 질병관리본부 손상감시사업을 위한 통계 산출 시 이용하는 Barell Matrix와 동일한 기준을 적용하여 ICD-10-CM[14]을 기준으로 새롭게 상해상병을 상해상병 양상별 14가지로 분류하고, 상해상병이 발생하는 부위를 그룹I 8가지, 그룹II 21가지로 분류하여 본 연구에 활용함

- 환수결정유무 : 상해진료건에 대한 부당환수결정 여부. 단, 일부 환수와 전체 환수는 모두 부당환수로 간주함

(Table 1) Top 100 Redemption rates according to ICD-10-CM codes for injuries

No.	S,T codes	redemption rates	No.	S,T codes	redemption rates
1	S26	54.5%	51	S12	12.1%
2	S47	50.0%	52	T90	12.0%
3	T59	45.5%	53	T75	12.0%
4	S78	40.0%	54	T06	11.8%
5	T20	34.3%	55	S06	11.7%
6	S11	32.3%	56	T14	11.3%
7	S21	30.8%	57	S51	11.2%
8	T01	29.2%	58	T26	11.1%
9	S09	27.1%	59	S95	11.1%
10	S57	26.8%	60	S55	11.1%
11	T04	25.0%	61	S65	10.7%
12	S19	25.0%	62	S89	10.5%
13	T27	24.5%	63	S72	10.3%
14	S25	23.5%	64	S30	10.1%
15	T07	22.8%	65	S98	10.0%
16	S75	22.7%	66	S20	10.0%
17	T15	22.5%	67	S69	9.9%
18	S79	22.2%	68	T31	9.6%
19	S97	20.7%	69	T03	9.5%
20	S36	20.6%	70	S90	9.4%
21	S35	20.5%	71	S68	9.1%
22	S05	19.8%	72	T09	8.7%
23	T00	19.5%	73	S54	8.5%
24	S31	19.5%	74	S14	8.4%

25	S58	18.8%	75	S86	8.4%
26	S08	17.6%	76	S81	8.2%
27	S45	17.6%	77	S92	8.1%
28	S02	17.6%	78	S76	8.1%
29	S29	16.7%	79	T91	8.1%
30	S87	16.5%	80	S70	7.8%
31	S03	16.4%	81	T13	7.7%
32	T21	16.1%	82	S34	7.6%
33	S04	15.9%	83	S99	7.4%
34	S15	15.8%	84	S40	7.4%
35	S10	15.4%	85	T93	7.0%
36	S60	15.4%	86	T79	7.0%
37	S37	15.3%	87	S22	6.8%
38	S01	14.9%	88	T54	6.7%
39	S84	14.8%	89	S32	6.6%
40	T29	14.0%	90	S80	6.4%
41	S71	13.4%	91	S24	6.2%
42	S41	13.3%	92	S50	6.2%
43	S77	13.3%	93	S82	6.0%
44	S39	13.0%	94	S42	5.9%
45	T92	13.0%	95	S74	5.9%
46	S61	12.7%	96	T58	5.9%
47	S44	12.5%	97	T30	5.5%
48	S16	12.5%	98	S91	5.5%
49	S27	12.2%	99	S94	5.4%
50	S07	12.1%	100	T65	5.1%

2.4 모형개발 및 평가

지역가입자 상해상병 진료건 조사 대상자 선정 모형은 2010년 자료를 기반으로 데이터마이닝 기법 중 의사결정나무모형을 활용하였고, 다양한 의사결정나무 분류 기준 중 Entrophy reduction를 활용한 Interactive Method(분석자 관점 및 각 의사결정나무 분류 과정에서의 통계량 등 각 주요변수의 중요도를 고려하여 모델링하는 방법)을 적용하였다. 의사결정나무 모형을 활용한 이유는 다른 데이터마이닝 기법에 비해 분류나 예측규칙을 명확하게 나타내는 장점을 가짐에 따라, 본 연구와 같이 결정의 정확도 및 그 결정요인을 밝혀내는 것이 중요한 분석에 유용하기 때문이다. 또한, 모형의 안정성을 위해, 앞서 제시된 바와 같이, 2010년 데이터셋을 기준으로 60%를 모형개발용으로, 40%는 내적 모형평가용으로 활용하였으며, 2011년 데이터셋을 외적 모형평가용으로 활용하였다.

(Table 2) Barell Matrix classification by Nature of the injury

No.	Nature of the injury	ICD-10-CM codes	
1	Fracture	less than 3 digit	S02, S12, S22, S32, S42, S52, S62, S72, S82, S92, T02, T08, T10, T12
		4 digit	T142, T902, T911, T912, T921, T922, T931, T932
2	Dislocation	3 digit	S03, S13, S23, S33, S43, S53, S63, S73, S93
		4 digit	S030, S031, S032, S033, S130, S131, S132, S133, S230, S231, S232, S330, S331, S332, S333, S334, S430, S431, S432, S433, S530, S531, S630, S631, S632, S730, S830, S831, S930, S931, S933
3	Sprains & Strains	less than 3 digit	S10, S30, S40, S50, S60
		4 digit	S034, S035, S134, S135, S136, S233, S234, S235, S335, S336, S337, S434, S435, S436, S437, S534, S635, S636, S637, S700, S731, S834, S835, S836, S934, S935
4	Internal Organ	3 digit	S24, S26, S39
		less than 3 digit	S06, S36, S37
		4 digit	S140, S141, S142, S240, S241, S260, S270, S271, S272, S273, S274, S275, S276, S278, S279, S340, S341, S343, S396, S397, T065, T093, T905, T913, T914, T915
5	Open Wound	3 digit	T92, T93
		less than 3 digit	S01, S11, S21, S31, S41, S51, S61, S71, S81, S91, T01
		4 digit	S052, S053, S054, S055, S056, S057, S080, S092, T091, T111, T131, T141, T901, T920, T930
6	Amputations	3 digit	S08, S18, S28, S38, S48, S58, S68, S78, S88, S98, T05
		less than 3 digit	S18, S48, S58, S68, S78, S88, S98, T05
		4 digit	S081, S088, S089, S281, S382, S383, T096, T116, T136
7	Blood Vessels	3 digit	S15, S25, S35, S45, S55, S65, S75, S85, S95
		4 digit	S090, T063, T114, T134, T145
8	Contusion/ Superficial	3 digit	S70, T90
		less than 3 digit	S00, S20, S80, S90, T00
		4 digit	S050, S051, S701, S707, S708, S709, T090, T110, T130, T140, T336, T900
9	Crush	3 digit	T04, S28, S38
		less than 3 digit	S07, S17, S47, S57, S67, S77, S87, S97
		4 digit	S280, S380, S381, T040, T041, T042, T043, T044, T047, T048, T049
10	Burns	less than 3 digit	T20, T21, T22, T23, T24, T25, T26, T27, T28, T29, T30, T31, T32, T95
11	Nerves	3 digit	S34
		less than 3 digit	S04, S44, S54, S64, S74, S84, S94
		4 digit	S143, S144, S145, S146, S242, S243, S244, S245, S246, S342, S344, S345, S346, S348, T060, T061, T062, T094, T113, T133, T144, T903, T924, T934
12	Poisoning	less than 3 digit	T36~T65, T96, T97
13	Others	3 digit	S05, S14, S27, S29, S49, S59, S69, S79, S83, S89, S99, T06, T09, T11, T13, T33
		less than 3 digit	S16, S46, S56, S66, S76, S86, S96, T03, T15, T16, T17, T18, T19, T34, T35, T66, T67, T68, T69, T70, T71, T73, T74, T75, T79
		4 digit	S058, S091, S097, S098, S197, S198, S268, S277, S290, S297, S298, S390, S398, S497, S498, S532, S533, S597, S598, S633, S634, S697, S698, S797, S798, S832, S833, S897, S898, S932, S936, S997, S998, T064, T068, T092, T095, T098, T112, T115, T118, T132, T135, T138, T143, T146, T147, T330, T331, T332, T333, T334, T335, T337, T338, T339, T784, T814, T849, T873, T904, T908, T910, T918, T923, T925, T926, T928, T933, T935, T936, T938, T980, T982
14	Unspecified	3 digit	S09, S19, T14, T91, T94, T98
		less than 3 digit	T07
		4 digit	S059, S099, S199, S269, S299, S399, S499, S599, S699, S799, S837, S899, S999, T099, T119, T139, T148, T149, T909, T919, T929, T939, T940, T941, T981

(Table 3) Barell Matrix classification by body region

Body region of the injury				ICD-10-CM codes			
No.	Group I	No.	Group II				
1	Head and Neck	1.1	Traumatic	less than 3 digit	S01, S06, S07		
			Brain Injury	4 digit	S020, S021, S023, S027, S028, S029, S040, S097, S098, S099, T010, T020, T040, T060, T901, T902, T904, T905, T908, T909		
		1.2	Other head, face and neck	3 digit	S04, T90		
				less than 3 digit	S00, S03, S05, S08, T15, T16, T26		
		4 digit			S022, S024, S025, S026, S041, S042, S043, S044, S045, S046, S047, S048, S049, S090, S091, S092, T170, T171, T180, T330, T340, T900		
			1.3	Neck	3 digit	S15, T17	
		less than 3 digit			S10, S11, S16, S17, S18, S19		
		4 digit			S128, S129, S132, S133, S135, S136, S143, S144, S145, S146, S150, S152, S153, S154, S155, S156, S157, S158, S159, T172, T173, T174, T270, T274, T331, T341		
			1.4	Other & unspecified	3 digit	S02, S09, T02, T20, T95	
		4 digit			T000, T030, T200, T201, T202, T203, T204, T205, T206, T207, T280, T285, T352, T903, T950, T980		
		2	Spine and Back	2.2	Spinal cord	4 digit	S140, S141, S240, S241, S340, S341, S343, T093, T913
				2.2	Vertebral column	3 digit	S12, S13, S14, S22, S23, S24, S32
less than 3 digit	T08						
4 digit	S120, S121, S122, S127, S130, S131, S134, S142, S151, S220, S221, S230, S231, S233, S242, S320, S321, S322, S330, S331, S332, S335, S336, S342, S344, T911						
3	Torso	3.1	Chest (Thorax)	less than 3 digit	S20, S21, S25, S26, S27, S28, S29		
				4 digit	S222, S223, S224, S225, S226, S227, S228, S229, S232, S234, S235, S243, S244, S245, S246, T175, T281, T286, T332, T342, T914		
		3.2	Abdomen, Pelvis	3 digit	S33, S32, T18		
				less than 3 digit	S30, S31, S35, S36, S37, S39, T19		
		4 digit			S323, S324, S325, S327, S328, S333, S334, S337, S345, S346, S348, S380, S381, S382, S383, T001, T011, T021, T031, T041, T065, T182, T183, T184, T185, T283, T288, T333, T343, T353, T915		
			3.3	Other & unspecified	3 digit	S34, S38, T09, T21, T27, T28, T91	
4 digit	T090, T091, T092, T094, T095, T096, T098, T099, T178, T179, T181, T188, T189, T210, T211, T212, T213, T214, T215, T216, T217, T272, T273, T276, T277, T282, T287, T912, T951						
4	Upper Extremities	4.1	Sholder & upper arm	less than 3 digit	S40, S41, S42, S43, S44, S45, S46, S47, S48, S49, T22		
				4 digit	T334, T344		
		4.2	Forearm & elbow	less than 3 digit	S50~S59		
		4.3	Wrist, hand & fingers	less than 3 digit	S60, S61, S62, S63, S64, S65, S66, S67, S68, S69, T23		
				4 digit	T050, T335, T345, T922		
		4.4	Other & unspecified	3 digit	T10, T11, T92		
				4 digit	T002, T012, T022, T024, T032, T042, T051, T052, T100, T101, T110, T111, T112, T113, T114, T115, T116, T118, T119, T354, T920, T921, T923, T924, T925, T926, T928, T929, T952		

Body region of the injury				ICD-10-CM codes	
No.	Group I	No.	Group II		
5	Lower Extremities	5.1	Hip	3 digit	S70, S72
				less than 3 digit	S73
				4 digit	S700, S710, S720, S721, S722, S760, S770, S780
		5.2	Thigh & upper leg	3 digit	S74, S75, S79
				4 digit	S701, S707, S708, S709, S711 S717, S718, S723, S724, S727, S728, S729, S761, S762, S763, S764, S767, S771, S772, S781, S789, T336, T346
		5.3	Knee & lower leg	less than 3 digit	S80~S89
				4 digit	T337, T347
		5.4	Ankle & foot	less than 3 digit	S90, S91, S92, S93, S94, S95, S96, S97, S98, S99, T25
				4 digit	T338, T348
		5.5	Other & unspecified	3 digit	S71, S76, S77, S78, T12, T13, T24, T93
				4 digit	T003, T013, T023, T025, T033, T043, T053, T054, T055, T120, T121, T130, T131, T132, T133, T134, T135, T136, T138, T139, T240, T241, T242, T243, T244, T245, T246, T247, T355, T930, T931, T932, T933, T934, T935, T936, T938, T939, T953
		6	Multiple	6.1	Multiple
less than 3 digit	T07				
4 digit	T006, T008, T009, T016, T019, T026, T027, T028, T029, T034, T038, T039, T044, T047, T048, T049, T056, T058, T059, T061, T062, T063, T064, T271, T275, T289, T350, T351, T356, T910, T918, T919, T940				
7	System-wide	7.1	System-wide	less than 3 digit	T36~T75, T29, T79, T96, T97
				4 digit	T982
8	Unclassifiable by site	8.2	Unclassifiable by site	3 digit	T06, T33, T34, T94, T98
				less than 3 digit	T14, T30, T31, T32
				4 digit	T018, T068, T284, T339, T349, T357, T784, T814, T849, T873, T941, T981, T954, T958, T959, T810, T811, T812, T813, T815, T816, T817, T818, T819, T840, T841, T842, T843, T844, T845, T846, T847, T848, T870, T871, T872, T874, T875, T876

3. 연구결과

3.1 상해요인 사후처리 결과 현황

2006년 상해요인 조사 대상 전체 건수는 496,621건 이었으며 이들에 대한 상해요인 부당환수율은 55.5% (275,625건)로 나타났다. 2008년부터 데이터마이닝 기법을 이용하여 새로운 발취기준에 의거하여 734,661건을 조사대상으로 선정하고 조사를 한 결과, 환수율은 72.2%로 증가하였다. 그러나 새로운 기준에 의거하여 2010년에 조사 물량을 선정한 결과 조사 물량은 911,489건 이었으며 환수율은 60.9%로 떨어짐에 따라 기존 모형의 고도화가 필요한 것으로 판명되었다.³⁾

3) 지사 환수율이 본부 환수율 보다 월등히 높은 이유는 본부에서 선정된 상해요인 조사대상자 중 부당으로 판명된 조사건과 연계된 청구건을 조사하기 때문

환수금액 기준으로 보았을 때, 새로운 기준이 적용된 2008년도의 본부 환수율이 25.5%였는데 비하여 2009년은 31.2%, 2010년은 29.9%로 나타났다.

3.2 연구대상자의 특성분석

지역가입자 상해상병 진료건 조사대상 선정모형 개발을 위한 연구대상자의 특성을 인구사회학적 특성, 수진시점의 진료이용 특성 및 요양기관 특성, 과거 상해요인 부당결정 특성, 상해상병 양상 및 부위 특성으로 구분하여 살펴보았다.

3.2.1 인구사회학적 특성, 수진시점의 진료이용 특성 및 요양기관 특성

성, 연령, 거주지역 특성, 수진당시 경제적 수준을 나

<Table 4> Post-processing results of Injuries by year (focusing of number of surveys)

Year	Subject (N)			Fraud case (N)			Redemption rate (%)		
	Total	Headquarters	Branch	Total	Headquarters	Branch	Total	Headquarters	Branch
2006	496,621	333,618	163,003	275,625	122,025	153,600	55.5	36.6	94.2
2007	563,006	372,992	190,014	311,774	132,111	179,663	55.4	35.4	94.6
2008	734,661	290,305	444,356	530,667	106,044	424,623	72.2	36.5	95.6
2009	741,792	347,239	394,553	536,974	177,697	359,277	72.4	51.2	91.1
2010	911,487	546,187	365,300	554,653	124,142	346,389	60.9	22.7	94.8
2006	300,595	278,847	21,749	72,358	55,394	16,964	24.1	19.9	78.0
2007	337,952	311,966	25,986	80,591	60,665	19,926	23.8	19.4	76.7
2008	313,918	267,693	46,225	106,920	68,157	38,763	34.1	25.5	83.9
2009	332,611	266,325	66,286	112,369	83,101	29,268	33.8	31.2	44.2
2010	347,106	317,603	29,503	117,859	95,026	22,833	34.0	29.9	77.4

<Table 5> Post-processing results of Injuries by year (focusing of costs)

Year	Subject (Cost in 1 million)			Fraud case (Cost in 1 million)			Redemption rate (%)		
	Total	Headquarters	Branch	Total	Headquarters	Branch	Total	Headquarters	Branch
2006	300,595	278,847	21,749	72,358	55,394	16,964	24.1	19.9	78.0
2007	337,952	311,966	25,986	80,591	60,665	19,926	23.8	19.4	76.7
2008	313,918	267,693	46,225	106,920	68,157	38,763	34.1	25.5	83.9
2009	332,611	266,325	66,286	112,369	83,101	29,268	33.8	31.2	44.2
2010	347,106	317,603	29,503	117,859	95,026	22,833	34.0	29.9	77.4

<Table 6> Social demographic, Medical utilization and Medical Institution Characteristics

Categories		Fraud		p	
		No N(%)	Yes N(%)		
Social demographic characteristics	Gender	Male	24,382(90.6)	2,544(9.5)	<0.0001
		Female	15,029(95.7)	677(4.3)	
	Age	20-29	9,828(90.4)	1,046(9.6)	<0.0001
		30-39	7,201(92.1)	614(7.9)	
		40-49	9,789(92.5)	794(7.5)	
		50-59	10,362(94.3)	627(5.7)	
		>=70	2,231(94.1)	140(5.9)	
	Region	Urban area	15,777(92.5)	1,288(7.6)	<0.0001
		Small and Medium Cities	17,514(92.7)	1,343(7.1)	
		Rural area	6,107(91.2)	587(8.8)	
Insurance Premium		63,729±62,960*	61,066±59,684*	0.0154	
A number of changing a health insurance card during the past a year	0	3740(89.8)	393(10.2)	<0.0001	
	1	21,174(91.3)	2,025(8.7)		
	2	11,525(94.9)	616(5.1)		
	3	2,563(94.6)	146(5.4)		
	>=4	679(94.3)	41(5.7)		

Medical utilization characteristics	Weekday / Weekend	Weekday	32,665(92.8)	2,537(7.2)	<0.0001	
		Weekend	6746(90.8)	684(9.2)		
	A number of getting physical therapy	0		18,250(89.5)	2,144(10.5)	<0.0001
		1		1,504(93.2)	110(6.8)	
		2		1,245(93.4)	88(6.6)	
		3		6,604(95.7)	298(4.3)	
		4		7,664(95.6)	349(4.4)	
		5		2,747(95.0)	146(5.1)	
Surgery	No		24,594(94.5)	1,425(5.5)	<0.0001	
	Yes		14,817(89.2)	1,796(10.8)		
Treatment performed in patients residence	No(other area)		20,943(91.3)	1,998(8.7)	<0.0001	
	Yes(Identical area)		18,455(93.8)	1,220(6.2)		
National Health Insurance Fee			1,450,218±1,893,127 *	1,822,515±2,576,858 *	<0.0001	
Medical institution characteristics	Location	Urban area	17,872(92.2)	1521(7.8)	0.0210	
		Small and Medium Cities	17,922(92.8)	1384(7.2)		
		Rural area	3,617(92.0)	316(8.0)		
	Type	Tertiary hospital		4,756(85.1)	833(14.9)	<0.0001
		General hospital		11,146(89.9)	1,256(10.1)	
		Hospital		11,550(93.8)	759(6.2)	
		Long-term hospital		941(91.2)	91(8.8)	
		Clinic		10,968(97.6)	273(2.4)	
Dental Hospital			47(83.9)	9(16.1)		
Public health center		3(100.0)	0(0.0)			
Total			39,411(92.4)	3,221(7.6)		

*Mean±SD

타내는 건강보험료, 과거 1년간 건강보험 자격변동을 나타내는 건강보험증 변동횟수로 구성된 인구사회학적 특성을 살펴보면, 모든 인구사회학적 특성과 부당환수결정여부와 통계적으로 유의한 관계가 나타났으며, 여자보다는 남자, 고연령층보다는 저연령층, 대도시 및 중소도시보다는 농어촌에서 부당환수결정율이 높은 것으로 나타났다. 수진시점의 진료이용 특성 및 요양기관 특성 또한, 부당환수결정여부와 통계적으로 유의한 관계가 나타났으며, 특히, 수진 당시 수술을 한 경우에서 부당환수결정율이 전체 부당환수결정율 7.6%보다 높은 10.8%로 나타났으며, 요양기관 종별에서도 상급종합병원에서의 부당환수결정율이 14.9%로 매우 높게 나타났다.

3.2.2 과거 상해요인 부당결정 특성

연구대상자의 과거 상해요인 부당결정 특성에 대해 살펴보았다. 그 결과, 과거 1년전 개인 또는 동일세대원 중 상해요인과 관련한 부당환수경험이 있는 자들이 현재 시점에서도 부당환수결정율이 높은 것으로 나타났으며 이는 통계적으로 유의하였다. 과거 개인별 부당건수가 없는 자들이 부당환수결정율이 7.1%인데 비해, 과거에 부당환수가 있었던 자들의 현재 조사시점에서의 부당환수결정율은 90~96.6%인 것으로 매우 높게 나타났다. 이는 과거 진료건에 있어 건강보험 제도권에 해당되는 진료료가 아니지만, 다양한 이유로 인해 건강보험으로 부당하게 진료 받은 경험이 있는 자들이 또 다시 부당하게 건강보험 진료를 받는 확률이 높음을 의미한다.

〈Table 7〉 Analysis Results of the characteristics of the subjects' fraud experiences over the past year

Categories			Fraud		p
			No N(%)	Yes N(%)	
By individual	The number of treatment fraud	0	39,392(93.2)	2,974(7.1)	<0.0001
		1-10	9(3.4)	257(96.6)	
		>=11	10(10.0)	90(90.0)	
	Costs of treatment fraud		5,878±386,953*	857,133±4,691,237*	<0.0001
Workers' compensation fraud	honest	39,410(92.5)	3,206(7.5)	<0.0001	
	fraud	1(6.3)	15(93.8)		
Automobile insurance fraud	honest	39,411(92.5)	3,193(7.5)	<0.0001	
	fraud	0(0.0)	28(100.0)		
By Household	The number of treatment fraud	0	39,392(93.0)	2,951(7.0)	<0.0001
		1-10	11(5.1)	203(94.9)	
		>=11	8(10.7)	67(89.3)	
	Costs of treatment fraud		4,037±278,385*	546,191±3,211,736*	<0.0001
Workers' compensation fraud	honest	39,411(92.5)	3,211(7.5)	<0.0001	
	fraud	0(0.0)	10(100.0)		
Automobile insurance fraud	honest	39,411(92.5)	3,204(7.5)	<0.0001	
	fraud	0(0.0)	17(100.0)		
Total			39,411(92.4)	3,221(7.6)	

*Mean±SD

〈Table 8〉 Analysis Results of the injury characteristics

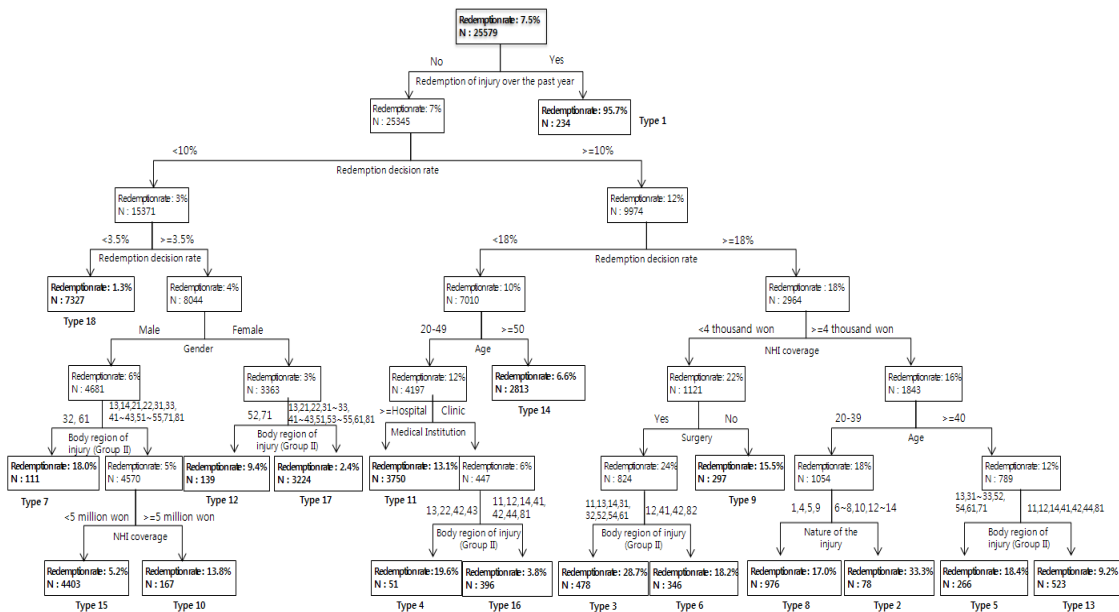
Categories				Fraud		p	
				No N(%)	Yes N(%)		
Nature of the injury	1	Fracture		10,350(89.9)	1,158(10.1)	<0.0001	
	2	Dislocation		212(93.8)	14(6.2)		
	3	Sprains & Strains		1,3147(98.0)	269(2.0)		
	4	Internal Organ		7,886(88.6)	1,011(11.4)		
	5	Open Wound		2,419(87.5)	345(12.5)		
	6	Amputations		93(88.6)	12(11.4)		
	7	Blood Vessels		248(86.4)	39(13.6)		
	8	Contusion/Superficial		773(88.6)	99(11.4)		
	9	Crush		133(86.9)	20(13.1)		
	10	Burns		511(91.4)	48(8.6)		
	11	Nerves		226(90.4)	24(9.6)		
	12	Poisoning		706(95.1)	36(4.9)		
	13	Others		1,891(95.3)	93(4.7)		
	14	Unspecified		816(93.9)	53(6.1)		
Body region of the injury	1	Head and Neck	1.1	Traumatic Brain Injury	8,878(87.2)	1,304(12.8)	<0.0001
			1.2	Other head, face and neck	1,571(84.1)	296(15.9)	
			1.3	Neck	410(88.9)	51(11.1)	
			1.4	Other & unspecified	21(77.8)	6(22.2)	

2	Spine and Back	2.2	Spinal cord	800(92.6)	64(7.4)
		2.2	Vertebral column	10,317(97.8)	234(2.2)
3	Torso	3.1	Chest (Thorax)	1,413(91.6)	130(8.4)
		3.2	Abdomen, Pelvis	1418(86)	230(14)
		3.3	Other & unspecified	49(84.5)	9(15.5)
4	Upper Extremities	4.1	Sholder & upper arm	2,052(95.4)	100(4.6)
		4.2	Forearm & elbow	333(90.2)	36(9.8)
		4.3	Wrist, hand & fingers	841(90.1)	92(9.9)
		4.4	Other & unspecified	49(92.5)	4(7.5)
5	Lower Extremities	5.1	Hip	739(94.4)	44(5.6)
		5.2	Thigh & upper leg	486(86.6)	75(13.4)
		5.3	Knee & lower leg	5,795(95.4)	282(4.6)
		5.4	Ankle & foot	1,476(96.8)	49(3.2)
		5.5	Other & unspecified	65(95.6)	3(4.4)
6	Multiple	6.1	Multiple	329(81.6)	74(18.4)
7	System-wide	7.1	System-wide	1,012(94.6)	58(5.4)
8	Unclassifiable by site	8.2	Unclassifiable by site	1,357(94.4)	80(5.6)
Total				39,411(92.4)	3,221(7.6)

3.2.3 상해상병 양상 및 부위 특성

상해상병으로 진료 받은 연구대상자들의 질병 특성을 Barel Matrix에 의해 14개의 상해상병 양상과 대그룹 8개와 세부 그룹 21개의 그룹으로 분류된 상해상병 부위에 따른 부당환수결정율을 살펴보았다. 그 결과, 상해상병 양상 및 부위별과 부당환수결정유무 간에는 통계적으

로 유의한 차이를 보였으며, 특히, 상해상병 양상에서는 혈관손상, 압착손상 등의 경우에는 전체 부당환수결정율보다는 높은 부당환수결정율을 나타냈으며, 상해상병 부위에서는 기타 머리, 머리 및 목 기타, 몸통 기타, 엉덩관절 외 넓적 다리 그리고 다발성에서 매우 높은 부당환수결정율이 나타났다.



[Fig. 2] Result of decision tree

<Table 9> Decision tree classification rules

Type	Rule	Response Rate of Redemption decision
1	Redemption of injury over the past year(+)	95.7%
2	Redemption of injury over the past year(-) & Redemption decision rate($\geq 10\%$) & Redemption decision rate($\geq 18\%$) & Rated insurance fee(≥ 4 thousand won) & Age(20-39) & Nature of the injury(6~8,10,12~14)	33.3%
3	Redemption of injury over the past year(-) & Redemption decision rate($\geq 10\%$) & Redemption decision rate($\geq 18\%$) & Rated insurance fee(< 4 thousand won) & Surgery(+)& Body region of the injury(Group II : 11,13,14,31,32,52,54,61)	28.7%
4	Redemption of injury over the past year(-) & Redemption decision rate($\geq 10\%$) & Redemption decision rate($< 18\%$) & Age(20-49) & Medical institution(Clinic) & Body region of the injury(Group II : 13,22,42,43)	19.6%
5	Redemption of injury over the past year(-) & Redemption decision rate($\geq 10\%$) & Redemption decision rate($\geq 18\%$) & Rated insurance fee(≥ 4 thousand won) & Age(≥ 40) & Body region of the injury(Group II : 13,31~33,52,54,61,71)	18.4%
6	Redemption of injury over the past year(-) & Redemption decision rate($\geq 10\%$) & Redemption decision rate($\geq 18\%$) & Rated insurance fee(< 4 thousand won) & Surgery(+)& Body region of the injury(Group II : 12,41,42,82)	18.2%
7	Redemption of injury over the past year(-) & Redemption decision rate($< 10\%$) & Redemption decision rate($\geq 3.5\%$) & Gender(M) & Body region of the injury(Group II : 32,61)	18.0%
8	Redemption of injury over the past year(-) & Redemption decision rate($\geq 10\%$) & Redemption decision rate($\geq 18\%$) & Rated insurance fee(≥ 4 thousand won) & Age(20-39) & Nature of the injury(1,4,5,9)	17.0%
9	Redemption of injury over the past year(-) & Redemption decision rate($\geq 10\%$) & Redemption decision rate($\geq 18\%$) & Rated insurance fee(< 4 thousand won) & Surgery(-)	15.5%
10	Redemption of injury over the past year(-) & Redemption decision rate($< 10\%$) & Redemption decision rate($\geq 3.5\%$) & Gender(M) & Body region of the injury(Group II : 13,14,21,22,31,33,41~43,51~55,71,81) & NHI coverage(≥ 5 million won)	13.8%
11	Redemption of injury over the past year(-) & Redemption decision rate($\geq 10\%$) & Redemption decision rate($< 18\%$) & Age(20-49) & Medical institution(\geq Hospital 이상)	13.1%
12	Redemption of injury over the past year(-) & Redemption decision rate($< 10\%$) & Redemption decision rate($\geq 3.5\%$) & Gender(F) & Body region of the injury(Group II : 52,71)	9.4%
13	Redemption of injury over the past year(-) & Redemption decision rate($\geq 10\%$) & Redemption decision rate($\geq 18\%$) & Rated insurance fee(≥ 4 thousand won) & Age(≥ 40) & Body region of the injury(Group II : 11,12,14,41,42,44,81)	9.2%
14	Redemption of injury over the past year(-) & Redemption decision rate($\geq 10\%$) & Redemption decision rate($< 18\%$) & Age(≥ 50)	6.6%
15	Redemption of injury over the past year(-) & Redemption decision rate($< 10\%$) & Redemption decision rate($\geq 3.5\%$) & Gender(M) & Body region of the injury(Group II : 13,14,21,22,31,33,41~43,51~55,71,81) & NHI coverage(≥ 5 million won)	5.2%
16	Redemption of injury over the past year(-) & Redemption decision rate($\geq 10\%$) & Redemption decision rate($< 18\%$) & Age(20-49) & Medical institution(Clinic) & Body region of the injury(Group II : 11,12,14,41,42,44,81)	3.8%
17	Redemption of injury over the past year(-) & Redemption decision rate($< 10\%$) & Redemption decision rate($\geq 3.5\%$) & Gender(F) & Body region of the injury(Group II : 13,21,22,31~33,41~43,51,53~55,61,81)	2.4%
18	Redemption of injury over the past year(-) & Redemption decision rate($< 10\%$) & Redemption decision rate($< 3.5\%$)	1.3%
	Base	7.5%

3.3 지역가입자 상해상병 진료건 조사대상 선 정모형 개발

앞서 살펴본 연구대상자들의 특성을 바탕으로 개발된 지역가입자 상해상병 진료건 조사대상 선정모형 결과를 살펴보면, 건강보험 지역가입자들의 상해상병 진료 부담환수결정에 영향을 주는 주요 특성요인은 과거 1년전 개인별 상해요인 환수유무, 상해상병별 환수결정율, Borell Matix를 이용한 상해상병 부위/특성, 수진당시 보험료, 수진 요양기관 종별, 수진당시 수술유무, 환자 연령, 심사결정공단부담금 등으로 나타났으며, 이러한 요인에 의해 지역가입자 조사대상자는 모두 18개 유형으로 분류되었다.

<그림 2>에서와 같이 2010년 당시 조사건 중심의 부담환수결정율이 7.5%였으나, 유형 1에 해당하는 과거 1년 전 개인별 상해 요인 환수 유인 경우로 건수 기준으로 환수될 확률은 95.7%에 달했다. 유형 2는 과거 1년 전 개인별 상해요인 환수 무 & 상해상병별 환수율 10% 이상 & 상해상병별 환수결정률 18% 이상 & 수진당시 산정보험료 4만원 이상 & 연령 20~30대 & 상해상병 양상(6~8, 10, 12~14)인 경우로 33.3%로 나타났다. 유형 3의 부담환수결정율은 28.7%, 유형 4의 부담환수결정율은 19.6%, 유형 5의 환수결정율은 18.4%, 유형 6의 부담환수결정율은 18%, 유형 7의 부담환수결정율은 18.0%, 유형 8의 부담환수결정율은 17% 등으로 나타났다. <그림 2>는 지역가입자 고도화 모형을 도식화한 것이며, <표 9>는 <그림 2>의 의사결정나무 분류를 각 유형별로 그 규칙을 나타낸 표이다.

3.4 지역가입자 상해상병 진료건 조사대상 선 정모형 평가

지역가입자 조사대상 선정모형의 성능에 대해서 향상도⁴⁾를 기준으로 하여 부담환수결정건수와 부담환수결정 금액에 대해서 평가를 하였다. 개발모형 기준으로 부담환수결정건수에 대해서 평가한 결과, 유형 1을 조사대상으로 했을 경우의 향상도는 12.8로 나타났다. 이는 2008년부터 적용하는 기존 모형에 비해서 환수건수 기준으로 부담환수결정율은 12.8배됨을 나타낸다. 유형 1과 유형 2

를 조사대상으로 하는 경우는 향상도가 10.7, 유형1/유형2/유형3을 조사대상으로 하는 경우에 향상도는 6.5로 나타났다.

개발된 모형을 기준으로 부담환수결정금액에 대해서 평가한 결과, 유형 1을 조사 대상으로 할 경우의 향상도는 10.0, 유형1과 유형 2를 조사 대상으로 할 경우의 향상도는 8.9, 유형1/유형2/유형3을 조사 대상으로 하는 경우에 향상도는 5.3으로 나타남으로서 본 연구에서 개발된 모형의 향상도는 기존의 그것보다 우수함을 판단할 수 있다.

4. 고찰 및 결론

국민건강보험공단의 자료에 따르면 2007년부터 2009년까지 100가지 상해질환으로 병원을 찾는 환자의 수는 매년 50만 명씩 증가하고 있으며 이에 따라 진료비도 해마다 약 3,000억 원씩 증가하여 2009년 한 해 동안에는 상해 환자의 급여비는 2조 4천억 원 정도 지급되었다[1]. 상해환자의 보험적용은 상해사고의 내용에 따라 자동차보험, 산재보험, 건강보험 등에서 부담을 한다. 즉, 자동차 사고로 인한 상해는 자동차 보험에서 부담하고, 산재 사고로 인한 상해는 산재보험에서 부담하고, 이 밖에 건강보험의 급여범위에 해당되는 것은 건강보험에서 부담을 한다. 그러나 건강보험의 적용 대상이 아닌 상해 사고에 대해서 환자나 회사, 자동차 보험회사 등의 경제적 이익 때문에 건강보험을 적용하여 건강보험 재정을 악화시키고 있다. 국민건강보험공단에서 이러한 부담 급여로 인한 보험급여비의 누수를 막기 위해 상해요인 업무를 수행하고 있다. 상해요인업무란 상해요인으로 인한 상해요인자가 건강보험으로 치료를 받았을 경우 국민건강보험공단은 건강보험의 급여(치료비)를 ‘부담 이익금’으로 간주하여 당사자에게 구상권을 행사, 지급한 치료비를 돌려받는 업무를 말한다. 이에 해당되는 경우를 국민건강보험법 제 53조에서 제시하고 있으며 주로 폭행에 의한 사고나 교통사고 및 산업재해이다[15, 16].

상해요인 조사 대상으로 선정된 건수가 2004년 351,598건이던 것이 2010년에 911,487건으로 증가하였으나 조사인력은 2004년 360명 이던 것이 2010년에는 280명으로 도리어 감소하여 상해요인 업무가 체계적으로 이

4)
$$\text{Lift} = \frac{\text{Cr}(\%)}{\text{Base Cr}(\%)}$$
, Cr(%)=예측모형에 의한 적중률,
BaseCr(%)=현재 사용되고 있는 모형에서의 적중률

〈Table 10〉 Gain Table for Model Assessment

Type	Total (N)	Responses (N)	Total injury costs (₩)	Redemption decision costs (₩)	Response Rate of Redemption decision		Cumulative Response Rate of redemption decision		Lift value	
					Case	Cost	Case	Cost	Case	Cost
1	234	224	383,488,220	360,633,730	95.7%	94.0%	95.7%	94.0%	12.8	10.0
2	78	26	111,640,570	53,503,780	33.3%	47.9%	80.1%	83.6%	10.7	8.9
3	478	137	931,994,590	290,025,520	28.7%	31.1%	49.0%	49.3%	6.5	5.3
4	51	10	38,241,020	6,442,310	19.6%	16.8%	47.2%	48.5%	6.3	5.2
5	266	49	615,242,100	159,169,280	18.4%	25.9%	40.3%	41.8%	5.4	4.4
6	346	63	589,704,820	140,145,740	18.2%	23.8%	35.0%	37.8%	4.7	4.0
7	111	20	179,213,300	34,311,750	18.0%	19.1%	33.8%	36.6%	4.5	3.9
8	976	166	1,339,333,230	217,692,680	17.0%	16.3%	27.4%	30.1%	3.6	3.2
9	297	46	291,405,430	45,717,340	15.5%	15.7%	26.1%	29.2%	3.5	3.1
10	167	23	1,503,368,450	183,785,270	13.8%	12.2%	25.4%	24.9%	3.4	2.7
11	3,750	490	6,976,892,880	953,758,280	13.1%	13.7%	18.6%	18.9%	2.5	2.0
12	139	13	188,660,690	21,490,510	9.4%	11.4%	18.4%	18.8%	2.5	2.0
13	523	48	764,850,870	60,321,380	9.2%	7.9%	17.7%	18.2%	2.4	1.9
14	2,813	186	6,489,103,230	467,006,530	6.6%	7.2%	14.7%	14.7%	2.0	1.6
15	4,403	228	5,687,042,920	298,533,820	5.2%	5.2%	11.8%	12.6%	1.6	1.3
16	396	15	291,485,670	12,017,440	3.8%	4.1%	11.6%	12.5%	1.5	1.3
17	3,224	77	4,006,321,350	94,721,150	2.4%	2.4%	10.0%	11.2%	1.3	1.2
18	7,327	97	7,186,803,230	131,950,310	1.3%	1.8%	7.5%	9.4%	1.0	1.0
Total	25,579	1,918	37,574,792,570	3,531,226,820	7.5%	9.4%	-	-	1.0	1.0

루어지지 못하고 있다. 이를 개선하기 위해서는 상해요인 업무 담당자를 크게 늘여야 하나 현실적으로 단기간 내에 이것이 이루어지기 어렵기 때문에 상해요인 조사 대상 선정을 보다 정교하게 선정하여 상해요인 조사 대상을 줄이면서 환수율 및 환수금액을 올릴 수 있는 방안이 절실히 요구되는 상황이다. 이에 본 연구에서는 상해조사 업무의 효율화 방안을 도출하고자 상해상병조사 대상자 선정기준을 마련하기 위한 모형을 개발하였다.

그 결과 데이터마이닝 기법 중 의사결정나무모형으로 개발된 상해상병 조사대상자 선정모형에서는 모형 적용 전보다 그 효과성(향상도 기준)이 최상위(유형 1) 조사건 기준 12.8배, 환수금액 기준 10.7배 증가하였다.

상해상병 선정 모형을 적용하기 위해서는 우선적으로 국민건강보험공단이 적정 조사물량을 선정하고 이를 최적화하는데 고도화 모형을 적용해야만 업무의 효율화를 볼 수 있으며, 산업재해 승인자료 연계건에서 발췌되는 상병 C(악성신생물), I10(본태성 고혈압)은 대부분 자연

질병으로써 부당결정율이 아주 낮은 것으로 확인된 바, 발췌기준에서 제외하는 것이 업무효율성이 높을 것으로 보인다. 그러나, 본 연구에서 개발된 모형을 다양한 현실적 문제로 인해 실제 상해조사업무에 직접 활용하여 평가된 결과를 제시하지 못한 것이 본 연구의 한계로 판단된다. 이러한 모형에 대한 실제 검증 과정과 건강보험가입자의 부당환수특성을 보다 다양한 측면에서 고려한 후속연구가 진행된다면 보다 정확한 건강보험 상해요인 조사 대상 선정모형 개발이 이루어질 것으로 사료된다.

ACKNOWLEDGMENTS

This work was supported by the research fund of NHIS(National Health Insurance Service).

REFERENCES

- [1] National Health Insurance Corporation, Injury patients and charge, National Health Insurance Corporation, 2012.
- [2] Kun-Hee Park, Sang-Jun Eun, Eun-Jeong Lee, Chae-Eun Lee, Doo-Yong Park, Kyoung-Hun. Han, Yoon Kim, Jin-Seok Lee, The incidence and patterns of unintentional injuries in daily life in Korea: A nationwide study, *J Prev Med Public Health*, Vol. 41, No. 4, pp. 256-271, 2008.
- [3] Research Affairs of Seoul National University-Korea Insurance Research Institute, A study on improvement of cooperation between national and private insurance for preventing insurance leakage, *Health Insurance Review & Assessment Service · Financial Supervisory Service*, 2011.
- [4] National Health Insurance Corporation, 2005 Organization diagnosis and Business process Reengineering(BPR) Report, National Health Insurance Corporation, 2005.
- [5] National Health Insurance Corporation, the National Health Insurance Statistical Yearbook for 2009, National Health Insurance Corporation, 2010
- [6] Il-Su Park, Ae-Kyoung Lee, Jun-Tae Han, Geun-Woo Ryu, Jong-Uk Won, A study on the measures for efficiency of post-processing management of injuries in National Health Insurance, National Health Insurance Corporation, 2007.
- [7] Il-Su Park, Jun-Tae Han, Hae-Sook Sohn, Suk-Bok Kang, Developing the administrative model using the data mining technique for injury in National Health Insurance, *Journal of the Korean Data & Information Science Society*, Vol 22, No 3, pp. 467 - 476, 2011.
- [8] [Http://the-modeling-agency.com/triage-for-tax-auditstm](http://the-modeling-agency.com/triage-for-tax-auditstm)
- [9] D. Micci-Barreca, Satheesh. R, E. Analytics. Improving tax administration with data mining, <http://www.spss.com> Last accessed April 26, 2006.
- [10] K. Giannangelo, Mining medicare and medicaid data to detect fraud, *Journal of AHIMA* Vol, 78, No.7 pp.66-67, 2007.
- [11] D. W. Abbott, I. P. Matkovsky, J. F. Elder, An evaluation of high-end data mining tools for fraud detection, *IEEE International Conference*, Vol.3, pp.2836-2841, 1998.
- [12] H. C. Koh, G. Tan, Data mining application in healthcare, *Journal of Healthcare Information Management*, Vol. 19, No. 2, pp. 64-72, 2005.
- [13] V. Barell, L. Aharonson-Daniel, L. A. Fingerhut, E. J. Mackenzie, A. Ziv, V. Boyko, A. Abargel, M. Avitzour, and R. Heruti, An introduction to the Barell body region by nature of injury diagnosis matrix, *Injury Prevention*, Vol. 8, pp. 91-96, 2002.
- [14] WHO, *International Statistical Classification of Diseases and Related Health Problems*, 10th Revision (ICD-10) Version, WHO, 2010.
- [15] Division of Healthcare for the Poor, Ministry of Health and Welfare, *Job manual for medicare injury factors*, Ministry of Health and Welfare 2013.
- [16] National Health Insurance Corporation, from : <http://www.nhic.or.kr>

박 일 수(Park, Il Su)



- 2001년 2월 : 인제대학교 보건관리학과(보건학사)
- 2003년 2월 : 인제대학교 일반대학원 데이터정보학과(이학석사)
- 2009년 8월 : 인제대학교 일반대학원 보건학과(보건학박사)
- 2003년 6월 ~ 2012년 2월 : 국민건강보험공단 건강보험정책연구원 부연구위원
- 2012년 3월 ~ 현재 : 위탁대학교 보건학과 조교수
- 관심분야 : 의료정보, 데이터마이닝, 건강보험, 보건통계
- E-Mail : ispark@uu.ac.kr

박 소 정(Park, So Jeong)



- 2010년 2월 : 인제대학교 보건행정학과 (보건행정학사)
- 2013년 8월 : 인제대학교 일반대학원 보건행정학과 (보건행정학석사)
- 관심분야 : 의료정보, 의무기록, 의료 의질향상, 글로벌헬스
- E-Mail : peachpsj@hanmail.net

한 준 태(Han, Jun Tae)



- 2005년 2월 : 영남대학교 일반대학원 통계학과 (이학석사)
- 2009년 2월 : 영남대학교 일반대학원 통계학과 (이학박사)
- 2012년 1월 ~ 현재 : 한국장학재단 조사연구팀 과장
- 관심분야 : 중도표본절단, 모수 추정, 데이터마이닝

· E-Mail : hanjt1003@gmail.com

강 성 홍(Kang, Sung Hong)



- 1990년 2월 : 서울대학교 보건대학원 보건관리학과(보건학석사)
- 1997년 2월 : 인제대학교 일반대학원 보건학과(보건학박사)
- 1998년 3월 ~ 현재 : 인제대학교 보건행정학과 교수
- 관심분야 : 보건정보, 의무기록, 데이터마이닝, 건강증진

· E-Mail : hcmkang@inje.ac.kr