

## 데이터마이닝 기법을 활용한 노인장기요양급여 권고모형 개발

한은정<sup>1</sup> · 이정석<sup>2</sup> · 김동건<sup>3</sup> · 강임옥<sup>4</sup>

<sup>1</sup>국민건강보험공단 건강보험정책연구원, <sup>2</sup>국민건강보험공단 건강보험정책연구원  
<sup>3</sup>동덕여자대학교 정보통계학과, <sup>4</sup>국민건강보험공단 건강보험정책연구원

(2009년 2월 접수, 2009년 7월 채택)

### 요약

노인장기요양보험에서 가장 중요한 이슈는 급여대상자의 희망, 건강 및 기능상태에 따라 어떤 급여를 제공할 것인가이다. 이를 해결하고자 노인장기요양보험의 보험자인 국민건강보험공단은 급여대상자에게 ‘표준장기요양이용계획서’를 제공하고 있다. 본 연구에서는 표준장기요양이용계획서 작성의 효율화 방안을 마련하고자 노인장기요양보험 3차 시범사업 표준이용계획 자료를 활용하여 노인장기요양급여 권고모형을 개발하였다. 모형개발에는 데이터마이닝의 의사결정나무모형, 로지스틱회귀모형, 앙상블 모형의 배깅과 부스팅 기법을 사용하였고, 이 중 실무자가 이해하기 쉬운 의사결정나무를 채택하여 권고모형을 설명하였다. 본 연구는 노인장기요양보험제도의 이용계획 수립의 객관성 및 과학성을 확보하고 이용계획 업무를 효율화하는 데에 기여할 것으로 기대된다.

주요용어: 노인장기요양보험, 표준장기요양이용계획서, 데이터마이닝, 의사결정나무.

### 1. 배경

우리나라 65세 이상 노인인구 비율은 2000년에 7.2%로 고령화 사회로 진입한 이후 급속도로 증가하여 2008년에는 10.3%를 돌파하였고 이런 추세라면 2026년에는 20%를 넘는 초고령사회에 진입할 것으로 예상하고 있다 (통계청, 2008). 이와 같은 고령화사회로의 빠른 진입은 노인부양에 대한 사회적 부담을 증가시켜서 노인부양이라는 문제를 사회적 과제로 대두시켰다.

이러한 문제를 해결하고자 정부는 2008년 7월 1일부터 노인장기요양보험을 실시하고 있다. 이 제도는 고령이나 노인성 질병 등으로 일상생활을 혼자 수행하기 어려운 노인 중 장기요양 1~3등급자를 장기요양급여의 대상으로 하고 있다 (보건복지가족부·국민건강보험공단, 2008). 장기요양급여는 크게 시설급여와 재가급여로 구분된다. 시설급여란 노인요양시설 등에 장기간 입소하여 신체활동 지원 및 심신기능의 유지·향상을 위한 교육·훈련 등을 받는 것을 말하며, 재가급여는 대상자가 자택에서 방문요양, 방문목욕, 방문간호, 주야간보호, 단기보호 등 5종의 서비스를 제공받는 것을 말한다. 이와 같은 장기요양급여를 이용하기 위해서는 이용자의 욕구와 선택을 존중한다는 장기요양급여의 기본원칙에 따라 급여대상자는 필요한 급여를 제공하는 기관의 담당자와 직접 계약을 체결해야한다. 하지만 일반적으로 급여대상자는 급여제공기관에 대한 충분한 정보를 갖기 어렵고, 대상자의 건강이나 기능상태도 제각기 다르고, 희망급여도 달라서 그들에게 필요한 급여의 종류나 범위, 강도를 정하기 쉽지 않다. 또한 현재 운영 중인 급여제공기관은 제각기 다른 인적 물적 구조를 가지고 있어서 동일한 수급자라 하더라도 제공되는 급

<sup>4</sup>교신저자: (121-749) 서울시 마포구 염리동 168-9, 국민건강보험공단 건강보험정책연구원, 부연구위원.

E-mail: imok48@nhic.or.kr

여의 범위나 강도가 달라질 수 있다 (강임옥 등, 2008). 이러한 우려로 인해 노인장기요양보험의 사업은 영주체인 국민건강보험공단은 급여대상자가 장기요양급여 이용계획을 수립하고 급여를 원활히 이용할 수 있도록 '표준장기요양이용계획서'(이하 표준이용계획서)를 제공하고 있다.

케어플랜의 일종이라고 할 수 있는 표준이용계획서는 급여대상자의 건강 및 기능상태, 권고 급여종류 및 양에 대한 정보를 포함하고 있다. 3차례에 걸친 노인장기요양보험 시범사업을 통해 표준이용계획서의 문제점이 제기되었는데, 가장 큰 문제는 이용계획서를 작성하는 자의 특성에 따라 이용계획의 내용이나 범위에 편차가 있다는 것이다 (선우덕 등, 2007). 표준이용계획서는 인정조사결과와 의사소견서 등의 기존 자료를 근거로 수급자 개인의 기능상태, 욕구 및 특성 등을 고려하여 작성되어야 한다. 표준이용계획서의 역할은 수급자에게 적정한 급여의 종류와 양에 대한 정보를 제공하는 것뿐만 아니라 재정 및 자원을 적절히 배분하는 것이다. 이런 역할을 수행하기 위해서는 수급자의 기능 및 건강상태, 희망 급여에 대한 정확한 파악과 급여종류와 양을 결정하기까지의 객관적이고 과학적인 절차가 필요하다 (이정석 등, 2008).

이용계획을 작성하는 다른 나라들의 방식을 살펴보면, 호주는 노인요양보호평가팀(the Aged Care Assessment Team; ACAT)을 운영하여 노인의 기능상태에 맞게 시설 및 재가급여를 통제하고 있고, 일본과 영국은 지역차원에서 케어매니저(care manager)를 고용하여 이들에 의해 장기요양보호가 필요한 사람들의 급여를 조율함으로써 장기요양보호가 통합적으로 제공될 수 있는 환경을 제공하고 있다 (OECD, 2005). 미국과 캐나다의 경우는 대상자의 기능 및 건강상태에 대한 장기요양급여 이용 결정시스템을 구축하여 케어매니지먼트 역할에 활용하고 있는데, 대표적인 것이 미국의 미시간 주에서 사용하는 MI-Choice (Fries 등, 2002)와 캐나다의 MAPLe이라는 시스템 (Hireds 등, 2008)이다. 기존에 다른 나라에서 사용되고 있는 장기요양서비스 이용계획 작성방식들은 각기 그 나라의 장기요양제도의 틀에 맞도록 구축되어 있어 우리나라에 직접적으로 적용하기에는 한계가 있다. 우리나라 노인장기요양보험제도에서도 대상자의 욕구를 최대한 반영하면서 장기요양보험 자원을 효율적으로 배분하기 위해 우리나라 상황에 맞게 객관적이고 과학적인 방식으로 급여의 종류와 양을 결정하는 장기요양서비스 이용계획 시스템을 구축할 필요가 있다.

이에 본 연구는 노인장기요양보험 3차 시범사업 자료를 근거로 권고급여종류의 영향요인 및 결정요인을 분석하고, 데이터마ining 기법을 이용하여 노인장기요양급여 권고모형을 개발 및 평가하고자 하였으며, 개발된 모형을 근거로 대상자에게 필요한 급여를 권고하는 과학적 기초 알고리즘을 도출하고자 하였다.

## 2. 연구방법

### 2.1. 연구대상 및 자료수집

본 연구의 대상자는 2007년 7월 1일부터 12월 31일까지 13개 시·군·구(강릉, 경북북부, 광주동부, 부산북부, 부여, 수원, 완도, 제주, 인천부평, 대구남부, 청주, 전북익산, 하동남해)에서 65세 이상 지역사회에 거주하는 노인을 대상으로 실시한 3차 시범사업 대상자를 포함하였다. 이들 대상자가 장기요양급여를 이용하기 위해서는 일련의 등급판정과정을 거쳐야 하는데, 그 과정은 그림 2.1과 같다.

3차 시범사업신청자는 총 16,157명이었으며 이 중 1~3등급자(이하 등급인정자)는 11,870명이었다. 이 대상자의 건강보험자격DB, 인정조사DB 그리고 표준이용계획서DB를 서로 연계하여 노인장기요양급여 권고모형 개발을 위한 DB를 구축하였다. 등급인정자 11,870명 중 표준이용계획서 작성 누락자 247명, 표준이용계획서에 시설급여와 재가급여를 동시에 권고받은 자 4명, 인구사회학적 변수 결측자 875명(인구사회학적 변수 중 현 주거형태 326명, 1순위 희망급여 549명이 결측), 건강 및 기능상태에 대한 변수가 결측인 14명을 제외한 10,730명을 최종 분석대상자로 정하였다.

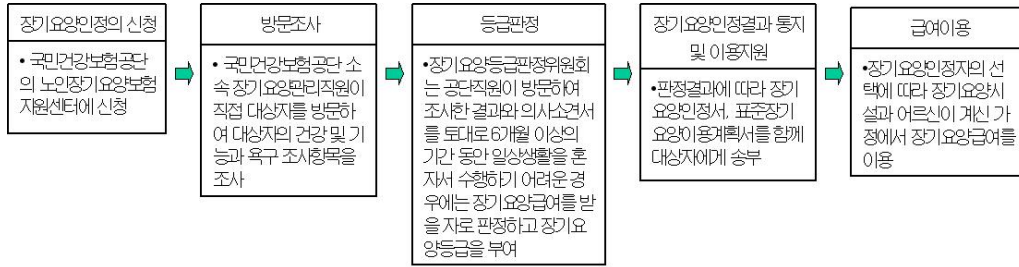


그림 2.1. 장기요양인정 업무 흐름도

## 2.2. 분석변수

### 1) 종속변수

본 연구의 종속변수는 표준이용계획서 상의 권고급여 종류인데, 급여종류를 크게 시설급여와 재가급여로 구분하였다. 재가급여 중 복지용구 급여를 권고 받은 경우는 분석에서 제외하였다.

### 2) 독립변수

본 연구에 사용된 독립변수를 크게 인구사회학적 변수와 건강 및 기능상태 변수로 나누었다. 인구사회학적 변수는 성(남, 녀), 연령(65~74세, 75~84세, 85세 이상), 의료보장형태(건강보험, 의료급여, 기초생활수급자), 지역(대도시, 중소도시, 농어촌), 현 거주형태(주택, 병원, 시설, 기타), 현재 주거환경 상태(양호, 불량), 희망급여(시설, 재가), 장기요양등급(1, 2, 3등급)으로 정하였다. 대상자의 건강 및 기능상태에 대한 변수는 신체기능, 인지기능, 행동변화, 간호처치, 재활 등 다섯 개 영역의 총 59개의 문항으로 정하였다. 신체기능 영역은 13문항으로 구성되어 있고, 각 문항마다 다른 사람의 도움을 받는 정도에 따라 ‘완전자립’, ‘부분도움’, ‘완전도움’으로 평가한다. 인지기능과 행동변화 영역은 각각 10문항, 16문항으로 구성되어 있고, 대상자의 상태가 문항에 해당하는지 여부에 따라 ‘예’ 또는 ‘아니오’로 구분하였다. 간호처치 영역은 10개의 항목에 대한 증상 여부에 따라 ‘있다’ 또는 ‘없다’로 측정하였다. 재활 영역은 운동장애정도와 관절제한정도를 평가하는 10개의 문항으로 구성되어 있는데, 운동장애정도는 ‘운동장애없음’, ‘불완전운동장애’, ‘완전운동장애’로 측정하고, 관절제한 정도는 ‘제한없음’, ‘한쪽관절제한’, ‘양관절 제한’으로 측정하였다. 모든 척도는 값이 클수록 건강이나 기능상태가 나쁜 상태를 나타낸다. 건강 및 기능상태를 측정하는 영역별 세부항목은 표 2.1과 같다.

## 2.3. 분석방법

표준이용계획서 상의 권고급여의 종류(시설/재가)에 따라 인구사회학적 변수의 분포를 파악하였다. 특히, 시설급여 권고군과 재가급여 권고군 간의 그룹 차이에 영향을 미치는 요인을 파악하기 위해 단변량 분석(동질성에 대한 카이제곱검정)을 실시하였다.

노인장기요양급여 권고모형개발을 위한 모형에는 대상자의 인구사회학적 변수를 제외한 건강 및 기능상태 변수만을 포함하였다. 왜냐하면 권고모형은 대상자의 건강 및 기능상태에 근거하여 적절한 급여가 무엇인지를 결정하는 일차적 도구로서의 역할이 강하기 때문에 대상자의 희망 및 주거형태 등을 포함한 인구사회학적 변수는 모형개발에서 제외하고 분석하였다.

모형개발을 위하여 분석DB를 학습용 분석자료(training set) 70%, 평가용분석자료(test set) 30%로 분할하여 모형 구축 및 평가에 활용하였다. 급여 권고모형 도출을 위해 CART방법(분리기준 GINI 계수)

표 2.1. 건강 및 기능상태 변수

영역	항목	척도
신체기능(13개 항목)	옷 벗고 입기, 세수하기, 양치질하기, 목욕하기, 식사하기, 체위변경하기, 일어나 앉기, 옮겨 앉기, 방 밖으로 나오기, 화장실 사용하기, 대변 조절하기, 소변 조절하기, 머리감기	1 = 완전자립 2 = 부분도움 3 = 완전도움
인지기능(10개 항목)	단기 기억장애, 날짜 불인지, 장소 불인지, 나이·생년월일 불인지, 지시 불인지, 상황 판단력 감퇴, 의사소통·전달 장애, 계산 장애, 하루 일과 불인지, 가족·친척 불인지	0 = 아니오 1 = 예
행동변화(16개 항목)	망상, 환각·환청, 슬픈 상태·울기도 함, 불규칙 수면·주야 혼돈, 도움에 저항, 서성거림·안절부절못함, 길을 잃음, 폭언·위협행동, 밖으로 나가려함, 물건 망가트리기, 의미 없거나 부적절한 행동, 돈·물건 감추기, 부적절한 옷 입기, 대소변불결행위, 화기관리 불가, 혼자 있는 것의 두려움	0 = 아니오 1 = 예
간호처치(10개 항목)	간호, 흡인, 산소요법, 욕창간호, 경관 영양, 암성통증간호, 당뇨관리, 장루간호, 투석간호, 당뇨 발간호	0 = 없다 1 = 있다
재활	상하지(4개 항목)	우측상지, 우측하지, 좌측상지, 좌측하지
	관절(6개 항목)	팔꿈치 관절, 손목 및 수지관절, 고관절, 무릎관절, 발목관절
		1 = 운동장애없음 2 = 불완전운동장애 3 = 완전운동장애
		1 = 제한없음 2 = 한쪽관절제한 3 = 양쪽관절제한

을 사용한 의사결정나무분석, 단계적 변수추출방법을 활용한 로지스틱회귀분석, 앙상블 모형인 배깅, 부스팅을 사용하였다. 개발된 모형의 평가를 위해서 정분류율, 모형 향상도(lift), ROC 곡선을 이용하였다. 모형 중 가장 우수하고 실무자가 이해하기 쉬운 모형을 채택하여 급여 권고모형을 설명하였다. 통계분석은 SAS 9.1을 사용하였으며, 모형개발을 위한 데이터마이닝 툴은 SAS사의 Enterprise Miner 4.3을 사용하였다 (SAS Institute, 2003).

### 3. 연구결과

#### 3.1. 인구사회학적 특성

연구대상자 10,730명 중 표준이용계획서에서 시설급여를 권고받은 사람은 3,800명(35.4%)이었고, 재가급여를 권고받은 사람은 6,930명(64.6%)이었다 (표 3.1). 시설급여 권고군과 재가급여 권고군의 분포의 차이가 인구사회학적 특성에 따라 통계적으로 모두 유의하였다( $p < 0.01$ ). 시설급여 권고군의 경우 80세 이상 후기노령인구의 비율이 재가급여 권고군보다 10.0% 더 높았고, 기초생활수급권자의 비율도 18.8% 더 높았다. 시설급여 권고군의 경우 현 거주형태가 시설인 자가 82.8%였고, 재가급여 권고군은 자택거주자가 70.0%였다. 시설급여 권고군의 97.9%는 시설급여를 희망한 것으로 나타났고, 재가급여 권고군의 72.3%는 재가급여를 희망한 것으로 나타났다.

#### 3.2. 예측모형의 개발 및 평가

표 3.2는 학습용 분석자료와 평가용 분석자료에 각 기법을 적용하여 얻은 정분류율 및 민감도와 특이도를 정리한 것이다. 이 결과는 분류기준값(threshold value)을 0.5로 하여 계산된 것이며, 일반적으로는 베이스 규칙(Bayes rule)을 사용하여 오분류율을 최소로 하거나 이익(profit)을 최대로 하는 분류기

표 3.1. 인구사회학적 특성에 따른 권고급여의 분포 (단위: 명, %)

변수	전체		시설		재가		
	N	%	N	%	N	%	
전체	10,730	100.0	3,800	35.4	6,930	64.6	
성별	남	2,926	27.3	831	21.9	2,095	30.2
	녀	7,804	72.7	2,969	78.1	4,835	69.8
연령(세)	65~69	1,206	11.2	325	8.6	881	12.7
	70~74	1,960	18.3	592	15.6	1,368	19.7
	75~79	2,408	22.4	812	21.4	1,596	23.0
	80~84	2,397	22.3	904	23.8	1,493	21.5
	85~89	1,672	15.6	708	18.6	964	13.9
	90 이상	1,087	10.1	459	12.1	628	9.1
사회보장형태	건강보험	7,529	70.2	2,258	59.4	5,271	76.1
	의료급여	332	3.1	64	1.7	268	3.9
	기초수급	2,869	26.7	1,478	38.9	1,391	20.1
지역	대도시	2,869	26.7	847	22.3	2,022	29.2
	중소도시	6,420	59.8	2,562	67.4	3,858	55.7
	농어촌	1,441	13.4	391	10.3	1,050	15.2
현거주형태	주택	5,248	48.9	399	10.5	4,849	70.0
	병원	2,040	19.0	455	12.0	1,585	22.9
	시설	3,186	29.7	2,896	76.2	290	4.2
	기타	256	2.4	50	1.3	206	3.0
주거환경평가	양호	9,527	88.8	3,730	98.2	5,797	83.7
	불량	1,203	11.2	70	1.8	1,133	16.3
1순위 희망급여	재가	5,101	47.5	89	2.3	5,012	72.3
	시설	5,629	52.5	3,711	97.7	1,918	27.7

표 3.2. 세부 모형에 대한 평가표 (단위: 명, %)

분류 모형	학습용 데이터			평가용 데이터		
	정분류율	민감도	특이도	정분류율	민감도	특이도
의사결정나무분석(Decision Tree)	69.29	37.010	87.292	66.95	33.27	84.11
로지스틱회귀분석(Logistic Regression)	70.00	40.890	86.020	68.44	38.20	84.93
배깅(Bagging)	70.00	33.970	88.917	66.95	30.11	87.04
부스팅(Boosting)	69.37	31.947	89.940	66.88	28.52	87.81

준 값을 정할 수 있다 (한상태 등, 2004). 의사결정나무모형에서 사전 정지기준을 나무의 깊이 6으로 정하였고 사후 가지치기는 교차타당성을 통하여 시행하였다. 앙상블모형에서 재표본 크기는 10으로 정하였고, 모형 구축에는 의사결정나무를 사용하였다. 표 3.2를 살펴보면 학습용 데이터에 대한 정분류율은 로지스틱회귀모형과 배깅이 70.0%, 부스팅 69.37%, 의사결정나무모형 69.29%로 나타났고, 권고 급여를 시설로 예측하는 성능은 로지스틱이 68.44%로 나타났고, 의사결정나무모형, 배깅 그리고 부스팅은 67%를 상회하였다.

평가용 데이터의 결과를 이용하여 모형의 정확도를 분석한 결과는 그림 3.1과 같다. Lift 도표를 살펴보면 상위 10%에 대해서 시설급여를 권고할 경우 기대 반응률(expected response rate)이 각각 1.79%, 1.98%, 1.97%, 1.93%로 나타났으며, 그 중 로지스틱 모형의 Lift 값이 가장 높게 나타났다. ROC곡선에서도 로지스틱회귀모형이 다른 모형에 비해 조금 위쪽에 위치하고 있어 정확도가 높게 나타났지만 다

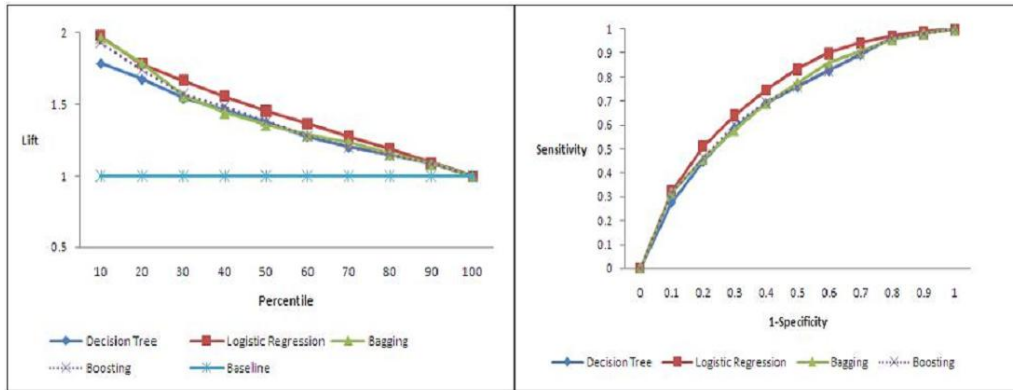


그림 3.1. 세부 모형별 ROC 곡선과 Lift 도표

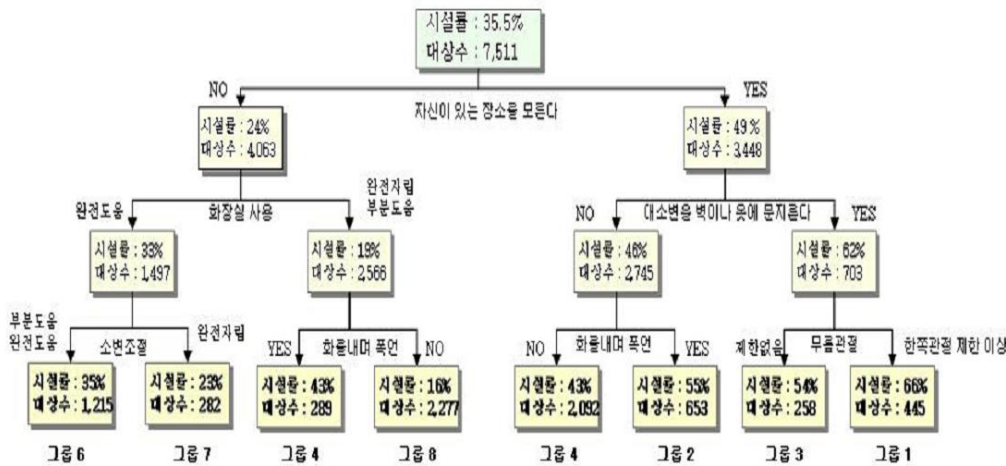


그림 3.2. 노인장기요양급여 권고모형 의사결정나무 분석 결과

큰 모형들과의 차이는 매우 적은 것으로 보인다.

이상의 결과를 통해 살펴본 바, 네 가지의 모형 중 로지스틱 모형의 성능이 가장 높은 것으로 나타났다. 그러나 의사결정나무모형 역시 로지스틱 모형과 비교할 때 성능의 차이가 미미하고 장기요양급여 권고 모형을 설명함에 있어서 실무자들이 이해하기 쉽고 영향변수와 목표변수의 영향력을 설명하기에 효과적이라는 장점이 있어 노인장기요양급여 권고모형에 의사결정나무모형을 선택하였다.

의사결정나무분석 결과를 살펴보면, 장기요양급여 권고그룹은 총 8개로 세분됨을 볼 수 있다 (그림 3.2). 인지기능 영역의 ‘자신이 있는 장소를 잘 모른다’가 가장 중요한 요인으로 나타났으며, 화장실 사용하기, 대소변을 벽이나 옷에 묻지른다, 소변조절, 화를 내며 폭언을 한다, 무릎관절제한 여부가 집단 세분화에 영향을 미치는 요인으로 나타났다. 세분화 모형을 통한 집단별 특성을 살펴보면, 자신이 있는 장소를 모르고, 대소변을 벽이나 옷에 묻지르고, 한쪽 또는 양쪽의 무릎에 관절제한이 있는 경우 시 설급여를 권고받을 확률이 66%였다. 한편 자신이 있는 장소를 모르고 대소변을 벽이나 옷에 묻지르지 않지만 화를 내며 폭언을 하는 경우 시 설급여를 권고받을 확률이 55%였다. 그와 반대로 자신이 있는 장

표 3.3. 노인장기요양급여 권고모형 알고리즘

---

<b>시설급여 권고 자격기준</b>	
● 그룹 1: 권고확률 66%	if 자신이 있는 장소를 모름 and 대소변을 벽이나 옷에 묻지름 and 한쪽 또는 양쪽의 무릎에 관절제한이 있음
● 그룹 2: 권고확률 55%	if 자신이 있는 장소를 모름 and 대소변을 벽이나 옷에 묻지르지 않음 and 화를 내며 폭언을 함
<b>재가급여 권고 자격기준</b>	
● 그룹 1: 권고확률 84%	if 자신이 있는 장소를 알고 있음 and 화장실 사용이 부분도움이거나 완전자립 and 화를 내며 폭언을 하지 않음
● 그룹 2: 권고확률 77%	if 자신이 있는 장소를 알고 있음 and 화장실 사용이 완전도움 and 소변조절이 완전자립

---

소를 알고, 화장실 사용이 부분도움이거나 완전자립이고 화를 내며 폭언을 하지 않는 경우 시설을 권고 받을 확률이 16%로 이런 특성을 갖는 대부분이 재가급여를 권고받는 것으로 나타났다. 이를 바탕으로 작성한 노인장기요양급여 권고모형 알고리즘은 표 3.3과 같다.

#### 4. 결론

본 연구는 노인장기요양보험의 수급자에게 가장 적절한 장기요양 급여종류를 결정하기 위한 알고리즘을 도출하기 위한 것이다. 의사결정나무에 의해 도출된 장기요양급여 종류를 시설로 결정짓는 주요 요인들은 자신이 있는 장소를 잘 모른다, 화장실 사용하기, 대소변을 벽이나 옷에 묻지름다, 소변조절, 화를 내며 폭언을 한다, 무릎관절제한 여부였다. 이는 MI-choice에서 대상자를 너싱홈 서비스로 결정하는 영향 요인 중 의사결정능력에 심한 손상이 있는 경우, 정신적 기능문제가 변화하였거나 갑자기 또는 새롭게 나타난 경우, 다른 사람에게 위협이나 욕설을 퍼붓는지 여부, 휠체어를 사용하거나 문밖으로 나가는 데에 어려움이 있는 경우 등 (Fries 등, 2002)과 대부분 일치하는 것이다. 그러나 대상자의 건강 및 기능상태 변수만을 포함한 본 연구모형은 66.9%로 다소 낮은 정확도를 나타내었다. 반면 현 거주형태와 1순위 희망급여변수 등 인구사회학적 변수를 모형에 포함하여 분석을 실시했을 때 모형의 정확도가 89%로 나타났다. 이러한 결과를 미루어 볼 때, 본 연구모형의 정확도가 다소 낮게 나타난 것은 권고급여 종류와 가장 강력하게 연관되어 있는 대상자의 현 거주형태와 1순위 희망급여 변수를 모형에서 제외하여 나타난 결과로 해석할 수 있다. 대상자에게 적절한 서비스를 결정하는 모형을 개발하는 데 있어 대상자의 인구사회학적 변수와 희망급여를 반영해야 하는 지는 논란의 여지가 있다. 우리나라의 노인장기요양보험은 이용자 본인의 욕구와 선택을 존중한다는 것을 장기요양급여의 기본원칙으로 하고 있기 때문에 희망급여와 같은 변수가 표준이용계획서의 권고 급여종류를 결정하는 데에 강하게 영향을 미쳤을 가능성이 있다. 또한 사람들은 스스로 자신의 건강 및 기능상태, 사회적 지지정도 등을 고려하여 특정한 급여

를 희망하기 때문에, 그런 의미에서 본인의 희망급여가 강력한 설명요인으로 작용하는 것은 당연한 현상일 수 있다. 하지만 일반적으로 대상자에게 어떤 서비스가 필요한 지를 판단하고자 할 때에는 일차적으로 대상자의 건강상태나 기능상태를 근거로 판단하고, 그 이후 대상자의 거주형태나 독거여부와 같은 상황적 요소를 파악하여 최종적으로 대상자의 급여종류를 판단해야 한다. MI-choice의 경우도 대상자의 희망, 인구사회학적 변수, 가족으로부터 제공받는 케어 등의 변수는 대상자 간에 변이가 너무 커서 모형에 적용하지 않고 있다 (Fries 등, 2002). 순수하게 연구대상자의 기능 및 건강상태에 의거하여 급여종류를 결정하고자 한다면 미국의 MI-Choice에서와 같이 잘 훈련된 평가자(trained assessor)를 구성하여 지역의 시설인프라 상황, 대상자의 현 거주형태 및 희망욕구를 배제한 상태에서 수급자의 건강 및 기능상태만을 근거로 이용계획을 작성하고 이 자료를 이용하여 노인장기요양급여 권고알고리즘을 구축하는 작업이 추가로 필요하다.

본 연구에 사용된 자료가 알고리즘 개발을 목적으로 훈련된 평가자에 의해 축적된 것이 아니어서 모형 결과를 일반화하기 어렵다는 한계점을 가지고 있다. 그러나 우리나라에서 처음 시행된 노인장기요양보험에서 케어플랜 역할을 담당하는 이용계획 업무를 데이터마이닝을 활용하여 객관화 및 과학화 시킬 수 있는 단초를 마련하였다는 데 그 의의가 있다고 하겠다. 일반적으로 이용계획을 결정하는 평가자는 이용계획 수립시 대상자의 희망 서비스, 생활환경, 지역의 인프라 등 여러 상황을 고려하여야 하기 때문에 대상자마다 이용계획의 차이가 존재할 수밖에 없다. 따라서 장기요양급여 권고모형은 평가자가 좀 더 객관적이고 적절한 이용계획을 수립하는데 있어 일차적인 방향성을 제시해주는데 많은 도움을 줄 수 있을 것이다. 이런 의미에서 이 연구의 결과가 노인장기요양보험제도의 이용계획 수립에 관한 업무를 효율화 하는 데에 기여할 수 있을 것으로 기대된다.

## 참고문헌

- 강임옥, 이정석, 한은정, 박일수, 광민정 (2008). 노인장기요양보험 이용계획 작성 및 운영방안, <국민건강보험공단 건강보험연구원>.
- 보건복지가족부-국민건강보험공단 (2008). 노인장기요양보험관련법령집.
- 선우덕, 김찬우, 최정수 등 (2007). 노인장기요양보험제도 시범사업 평가연구 (2차) - 장기요양수요조사, 장기요양 인정체계 및 장기요양서비스이용지원체계 부문 -, <보건복지부한국보건사회연구원>.
- 이정석, 한은정, 권진희, 강임옥 (2008). 표준장기요양이용계획서와 장기요양급여비용과의 상관성 분석, <국민건강보험공단 건강보험연구원>
- 통계청 (2008). 2008 고령자 통계, 서울
- 한상태, 강현철, 이성건, 정요전 (2004). 교차판매(Cross-Sell) 스코어링 모형 개발, <응용통계연구>, **17**, 229-238.
- Fries, B. E., Shugarman, L. R., Morris, J. N., Simon, S. E. and James, M. (2002). A screening system for Michigan's home- and community-based long-term care programs, *The Gerontologist*, **42**, 462-474.
- Hirdes, J. P., Poss, J. W. and Curtin-Telegdi, N. (2008). The method for assigning priority levels(MAPLe): A new decision-support system for allocating home care resources, *BMC Medicine*, **6**.
- OECD 아시아 사회정책센터 (2005). <노인 장기요양보호>, 원본: Long-term care for older people, OECD.
- SAS Institute (2003). *Data Mining Using SAS Enterprise Miner: A Case Study Approach*, SAS Institute INC.



# A Recommending System for Care Plan(Res-CP) in Long-Term Care Insurance System

Eun-Jeong Han<sup>1</sup> · Jung-Suk Lee<sup>2</sup> · Dong-Geon Kim<sup>3</sup> · Im-Ok Kang<sup>4</sup>

<sup>1</sup>National Health Insurance Corporation; <sup>2</sup>National Health Insurance Corporation

<sup>3</sup>Department of Data & Information Science, Dongduk Women's University

<sup>4</sup>National Health Insurance Corporation

(Received February 2009; accepted July 2009)

---

## Abstract

In the long-term care insurance(LTCI) system, the question of how to provide the most appropriate care has become a major issue for the elderly, their family, and for policy makers. To help beneficiaries use LTC services appropriately to their needs of care, National Health Insurance Corporation(NHIC) provide them with the individualized care plan, named the Long-term Care User Guide. It includes recommendations for beneficiaries' most appropriate type of care. The purpose of this study is to develop a recommending system for care plan(Res-CP) in LTCI system. We used data set for Long-term Care User Guide in the 3rd long-term care insurance pilot programs. To develop the model, we tested four models, including a decision-tree model in data-mining, a logistic regression model, and a boosting and boosting techniques in an ensemble model. A decision-tree model was selected to describe the Res-CP, because it may be easy to explain the algorithm of Res-CP to the working groups. Res-CP might be useful in an evidence-based care planning in LTCI system and may contribute to support use of LTC services efficiently.

Keywords: Long-term care insurance, long-term care user guide, data mining, decision-tree analysis.

---

<sup>4</sup>Corresponding author: National Health Insurance Corporation, 168-9 Yeomri-Dong, Mapo-Gu, Seoul 121-749, Korea. E-mail: imok48@nhic.or.kr