

A Target Selection Model for the Counseling Services in Long-Term Care Insurance

Eun-Jeong Han^a · Dong-Geon Kim^{b,1}

^aHealth Insurance Policy Research Institute, National Health Insurance Service

^bDepartment of Statistics and Information Science, Dongduk Women's University

(Received July 13, 2015; Revised August 1, 2015; Accepted October 15, 2015)

Abstract

In the long-term care insurance (LTCI) system, National Health Insurance Service (NHIS) provide counseling services for beneficiaries and their family caregivers, which help them use LTC services appropriately. The purpose of this study was to develop a Target Selection Model for the Counseling Services based on needs of beneficiaries and their family caregivers. To develop models, we used data set of total 2,000 beneficiaries and family caregivers who have used the long-term care services in their home in March 2013 and completed questionnaires. The Target Selection Model was established through various data-mining models such as logistic regression, gradient boosting, Lasso, decision-tree model, Ensemble, and Neural network. Lasso model was selected as the final model because of the stability, high performance and availability. Our results might improve the satisfaction and the efficiency for the NHIS counseling services.

Keywords: long-term care insurance, NHIS counseling services, data-mining, Lasso

1. 연구배경

그동안 가족의 영역에 맡겨져 왔던 노인 간병과 장기요양에 대한 문제를 사회연대 원리에 따라 국가와 사회가 분담하고자 2008년 7월에 도입된 노인장기요양보험제도가 올해로 8년차를 맞이하였다. 지난 8년 동안 본 제도를 운영해오면서 등급판정체계 개편, 장기요양 수가체계 개선 등 노인장기요양보험제도를 개선해 왔다. 최근 제도개선의 주요 관심은 재정의 안정성, 재가급여 활성화, 장기요양서비스 질 수준의 제고에 집중되고 있으며 (Sunwoo 등, 2014), 이를 추진하기 위한 핵심 전략으로 공적 영역에서의 케어매니지먼트가 강조되고 있다 (Jegal, 2011; Han 등, 2012; OECD, 2011).

노인장기요양보험제도에서 케어매니지먼트는 ‘이용지원’이라는 제도의 틀에서 이루어지고 있다. 이용지원은 노인장기요양보험법 제 48조에 따라 노인장기요양보험의 관리운영기관인 국민건강보험공단(이하 공단)이 수급자가 심신기능 상태와 욕구에 따라 불이익이나 불편함 없이 적정하고 원활하게 급여를 이용할 수 있도록 정보제공·안내·상담 등을 제공하는 것을 목적으로 한다 (National Health Insurance Service, 2014a). 이용지원체계는 표준장기요양이용계획서(이하 표준이용계획서) 작성, 이용지원 상담, 지역협의회 운영, 등급외자 연계사업 등으로 이루어지는데, 이 중 급여이용 계획인 표준이용계획서에 대

¹Corresponding author: Department of Statistics and Information Science, Dongduk Women's University, 23-1 Wolgok-Dong, Sungbuk-Gu, Seoul 01369, Korea. E-mail: dongg@dongduk.ac.kr

해 설명하고, 장기요양 등급변경이 필요한 수급자에게는 등급을 변경할 수 있도록 안내하는 등의 이용지원 상담 업무가 이용지원 업무의 대부분을 차지한다 (Han 등, 2012). 최근 정부가 수급자의 삶의 질 및 장기요양 재정의 안정화를 위해 재가급여 활성화 정책을 추진하면서, 재가급여 수급자와 가족부양자가 보다 비용효과적이고 적절한 재가급여를 이용할 수 있도록 상담하는 이용지원 상담에 대한 중요성이 대두되고 있다.

재가급여 수급자와 가족부양자를 대상으로 이루어지는 이용지원 상담은 최초상담, 정기상담, 수시상담으로 나뉘는데, 이용지원 상담업무의 대부분은 정기상담을 통해 이루어진다. 정기상담은 수급자와 가족부양자의 특성에 따라 다섯 개의 군으로 나뉘며, 이 군에 따라 각기 다른 주기로 상담이 이루어진다. 1군은 집중 상담이 필요한 대상자(기준: ① 표준이용계획서와 다르게 급여이용, ② (급성기) 의료적 문제, ③ 주수발자 없음)로 상담은 1개월마다 이루어지며, 2군은 보호관찰이 필요한 자(기준: ④ 생애최초 인정자, ⑤ 독거, ⑥ 기능상태변화(악화, 호전), ⑦ 욕창 있음, ⑧ 인지증)로 3개월 마다, 3군은 정기 확인이 필요한 자(기준: ⑨ 환경문제(불량), ⑩ 위생상태(불량), ⑪ 급여이용 만족도(하), ⑫ 외상도(준외상이상))로 6개월 마다, 4군과 5군은 연 1회 확인이 필요한 자로 연 1회 상담이 이루어진다. 이용지원의 상담은 주로 1개월과 3개월의 주기를 갖는 1군과 2군을 대상으로 이루어진다. 재가급여 수급자의 상담군을 결정하는 요인은 장기요양 등급판정시 활용되는 인정조사표와 상담기록표를 기반으로 공단의 상담제공자들의 논의를 거쳐 선정되었고, 이를 근거로 매월 각 군에 해당하는 대상자가 전산시스템에 의해 발체된다 (National Health Insurance Service, 2014b). 그러나 상담군에 대한 정의 및 이를 결정하는 요인이 공단의 상담담당자에 의해 보험자 및 전문가적 관점에서 판단된 것이어서, 현장에서 상담이 필요하다고 판단되는 기준과 맞지 않고, 실제 상담을 원하는 대상자와 공단 상담담당자가 상담이 필요하다고 선정된 자 간의 차이가 존재하며, 이 때문에 수급자와 가족부양자의 입장에서 볼 때 이용지원 상담이 일방적으로 이루어진다고 느껴 만족도와 효과가 낮다고 보고되었다 (Jegal, 2011; Han 등, 2012; Han 등, 2013). Han 등 (2012)이 이용지원 상담 담당자, 수급자와 그 가족부양자를 대상으로 이용지원 상담에 대한 의견을 조사한 결과, 이용지원 상담 담당자는 공단의 일방향적인 상담대상자 선정시스템으로 인해 상담 대상자와의 라포형성이 어렵고 상담대상자가 상담을 감시와 감독으로 여긴다고 하였다. 수급자와 가족부양자는 상담대상자로 왜 선정되었는지 알지 못하며, 상담내용 또한 본인의 욕구와 맞지 않아 불만족스럽다고 하였다. 따라서 이용지원 상담 대상자를 선정함에 있어 상담의 대상이 되는 재가급여 수급자와 가족부양자의 이용지원 상담에 대한 욕구를 면밀히 조사하고, 이를 반영하여 이용지원 상담대상자를 선정할 필요가 있다고 제언하였다.

이에 본 연구는 재가급여 수급자와 가족부양자의 이용지원 상담에 대한 욕구조사 자료를 활용하여 이용지원 상담을 원하는 자의 영향요인을 파악하고, 이를 근거로 이용지원 상담 대상자를 선정하는 모델을 개발하고자 하였다.

2. 연구방법

2.1. 연구대상 및 자료수집

본 연구 수행을 위해 2013년 3월 장기요양 재가급여(방문요양, 방문목욕, 방문간호, 주야간보호, 단기보호)를 이용한 수급자 가운데, 2013년 1월부터 3월 사이에 장기요양 등급을 신규 또는 갱신판정 받은 자를 대상으로 이용지원 상담에 대한 욕구를 조사하였다. 조사 모집단은 2013년 3월 장기요양 재가급여 수급자 163,322명 중에서 2013년 1월부터 3월 사이에 신규 및 갱신판정을 받은 수급자 25,862명으로, 거주지역(서울, 경기, 강원, 충청, 전라, 경상), 장기요양등급(1, 2, 3등급), 의료보장형태(일반, 그 외(의료급여, 기초생활수급, 경감)), 재가급여 종류(방문요양, 그 외)를 층화변수로 하여 층화비례확률

추출방법을 통해 조사 표본을 추출하였다. 조사는 객관성 및 전문성을 확보하기 위해 조사전문기관을 통해 진행하였으며, 연구진이 제공한 구조화된 설문지를 바탕으로 조사원이 대상자의 가정 또는 면접장소를 방문하여 1:1 개별면접방식으로 이루어졌다. 자료수집은 2013년 7월 4일부터 7월 24일까지 전문조사업체 (주)한국능률협회컨설팅을 통해 실시되었다. 조사 목표대상자는 2,000명(신뢰수준 95%, 표준오차의 한계 $\pm 2.19\%$ 포인트)으로 정하였고, 최종적으로 2,000명이 조사 완료되었다. 조사에 대한 응답은 장기요양 수급자 대부분이 와상 또는 준와상, 치매질환자임을 고려하여, 이용지원 상담의 대상이 되는 수급자 및 가족부양자 모두가 가능하도록 하였다. 응답자 2,000명 중 173명이 가족부양자가 없는 독거 수급자가 응답하였고, 나머지 1,827명은 가족부양자가 응답하였다.

2.2. 분석변수

2.2.1. 종속변수 본 연구의 종속변수는 이용지원 상담에 대한 욕구가 있는 지 여부로 정하였다. 이에 “장기요양서비스 이용시 불편하거나 어려운 점을 해결하고 원활한 급여이용 지원을 위해 공단 직원이 제공하는 상담이 필요하다고 생각하십니까?”라는 질문에 ‘예’라고 응답한 자를 상담에 대한 욕구가 있는 대상으로 정의하였다.

2.2.2. 독립변수 독립변수는 크게 수급자 특성, 가족부양자 특성, 재가급여이용 관련 특성으로 나누었다. 첫째, 수급자 특성은 인구사회학적 특성(성, 연령, 장기요양등급, 자격, 학력, 지역), 건강 및 기능상태, 질병 및 증상이며, 건강 및 기능상태와 질병 및 증상에 대한 정보는 노인장기요양보험 등급판정시 수집되는 ‘장기요양인정조사표’를 활용하였다. ‘장기요양인정조사표’는 노인장기요양보험 도입 당시에 개발된 것으로 신체기능, 수단적일상생활기능, 인지기능, 행동변화, 간호처치, 재활, 복지용구, 지원형태, 환경평가, 시력·청력상태, 질병 및 증상 등으로 구성된다. 본 상담욕구모형 개발을 위해서는 인정조사표 항목 중 신체기능 13항목, 수단적일상생활기능 10항목, 인지기능 10항목, 행동변화 22항목, 간호처치 10개 항목, 재활 10개 항목, 시력상태, 청력상태를 활용하였다. 이와 더불어 대상자의 중증도 측정 변수로 활용 (Hyun과 Lee, 2012; Lee 등, 2013)되는 신체기능, 인지기능, 행동변화, 간호처치, 재활 5개 영역에 대해서는 영역별 점수(영역별 문항의 점수 합)를 추가로 활용하였다 (Table 2.1). 질병 및 증상에 대해서는 인정조사 시 3개월간 앓고 있는 만성질환에 대해 조사하는 항목을 활용하였고, 치매, 증풍(뇌졸중), 혈압, 당뇨, 관절염, 요통, 골다공증, 골절, 암이 활용되었다.

둘째, 가족부양자 특성은 인구사회학적 특성, 부양부담감으로 구성된다. 가족부양자의 인구사회학적 특성은 성, 연령, 학력, 결혼상태, 신체적 수발기간, 주관적 건강상태, 직업, 월평균 소득, 수급자와의 관계, 가족요양보사여부, 경제활동여부이다. 부양부담감은 가족부양자가 수급자를 부양하면서 느낄 수 있는 부담감으로 Kwon (1994)의 부양부담감 측정도구를 활용하여 조사하였다 (Table 2.2). 부양부담감 측정도구는 노인과 주부양자 관계의 부정적 변화, 사회적 활동제한, 가족관계의 부정적 변화, 심리적 부담감, 재정 및 경제활동상의 부담감, 건강상의 부담감인 6개 영역으로 구성되며, 총 28개 항목에 대해 5점 척도로 측정된다. ‘매우 그렇다’라고 응답할수록 부양부담을 많이 느끼는 것으로 판단한다. 부양부담감 변수는 6개 부양부담 영역별 항목의 점수를 합하여 영역별로 점수를 구하고 점수의 분포에 따라 부양부담 정도를 이산형 변수로 변환하여 활용하였다. 노인과 부양자 관계의 부정적 변화와 사회적 활동제한에 대해서는 제 1사분위수(25%)와 제 3사분위수(75%)를 기준으로 3개 항목으로 구분하였고, 가족관계의 부정적 변화, 심리적 부담감, 재정 및 경제활동상의 부담감, 건강상의 부담감에 대해서는 중위수(50%)를 기준으로 2개 항목으로 구분하여 활용하였다.

셋째, 재가급여이용 관련 특성 변수는 이용급여유형, 급여이용 기간, 장기요양급여 만족도, 표준이용계획서 참고여부, 장기요양 급여 외의 다른 서비스 이용 여부, 급여 이용 시 불편한 점, 본인부담금 부담

Table 2.1. The health status items included in the algorithms

영역	항목	(점수)척도	점수분포
신체기능 (13개 항목)	옷벗고입기/세수하기/양치질하기/목욕하기/식사하기/ 체위변경하기/일어나앉기/ 옮겨앉기/방밖으로 나오기/ 화장실사용하기/대변조절하기/소변조절하기/머리감기	1: 완전자립 2: 부분도움 3: 완전도움	13~39
	집안일하기/식사준비하기/빨래하기/금전관리하기	1: 완전 자립 2: 부분도움 3: 완전도움, 4: 수행하지 않음	
수단적 일상생활기능 (10개 항목)	물건사기/전화사용하기/교통수단이용하기	1: 완전자립, 2: 적은 부분도움 3: 많은 부분도움 4: 완전도움	
	근거리외출하기/몸단장하기/약챙겨먹기	1: 완전자립 2: 부분도움 3: 완전도움	
인지기능 (10개 항목)	단기기억장애/시간불인지/장소불인지/나이·생년월일불 인지/지시 불인지/상황 판단력 감퇴/의사소통·전달장애 /계산능력장애/하루일과이해장애/가족·친척 불인지	1: 예 0: 아니오	0~10
행동변화 (22개 항목)	망상/환각·환청/슬픈상태·울기도 함/불규칙수면·주야혼 돈/도움에 저항/서성거림·안절부절못함/길을 잃음/폭언 ·위협행동/밖으로나가려함/물건망가트리기/의미없거나 부적절한 행동/돈·물건 감추기/부적절한 옷입기/대소변 불결행위/화기관리못함/분리불안/고함침/부적절한 성적 행동/이식행위/지나친간섭/식욕의 변화/따라다니기	1: 예 0: 아니오	0~22
간호처치 (10개 항목)	기관지질개관간호/흡인/산소요법/욕창간호/경관영양/ 암성통증간호/도뇨관리/장루간호/간호투석/당뇨발간호	1: 있다 0: 없다	0~10
재활	상하지 (4개 항목) 우측상지/우측하지/좌측상지/좌측하지	1: 운동장애없음 2: 불완전운동장애 3: 완전운동장애	10~30
	관절 (6개 항목) 팔꿈치관절/손목 및 수지관절/고관절/무릎관절/발목 관절	1: 제한없음 2: 한쪽관절제한 3: 양쪽관절제한	
시력상태	정상	정상	
	1M 떨어진 달력은 읽을 수 있으나 더 먼 거리는 안보임/ 근접한 글씨는 읽을 수 있으나 먼 거리는 안보임/거의 보 이지 않음/보이는지 판단 불능	장애있음	
청력상태	정상	정상	
	보통의 소리를 듣기도 하고, 못 듣기도 함/큰 소리라면 들을 수 있음/거의 들리지 않음/들리는지 판단 불능	장애있음	

정도이다. 이용급여유형은 방문요양만을 이용하는 경우, 방문요양과 방문목욕을 함께 이용하는 경우, 그 외의 급여를 이용하는 경우는 기타로 정의하였고, 급여이용기간은 방문요양, 방문목욕, 방문간호, 주야간보호, 단기보호 5가지 급여이용 기간 중 가장 긴 기간을 사용하였다. 장기요양급여 만족도는 불만족(매우 불만족, 불만족), 보통, 만족(매우 만족, 만족) 3단계로 측정했으며, 표준이용계획서 참고여부는 공단이 제공하는 표준이용계획서를 알고 있고 재가급여 결정 시 표준이용계획서를 참고하는 경우 ‘인지하고 참고’, 표준이용계획서를 알고 있으나 서비스 결정시 참고하지 않는다면 ‘인지하지만 미참고’,

Table 2.2. The family caregiver care-burden scale included in the algorithms

영역	항목
노인과 부양자 관계의 부정적 변화	노인을 부양하는 것에서 벗어나고 싶다/노인을 부양하는 것이 지겨울 때가 있다/노인을 부양하다보면 짜증이 날 때가 있다/노인을 부양할 때 화가 나는 경우가 있다/노인과 같이 있는 것이 즐겁지 않다/앞으로 감당하기 힘든 부양문제가 일어날 것 같아 불안하다/노인을 마주 대하기가 싫을 때가 있다
사회적 활동제한	나 혼자 자유롭게 쓸 수 있는 시간이 없다/친구를 만날 수가 없다/가족과 함께 외출이나 회식을 할 수가 없다/이웃과 만날 수 있는 기회가 없다/바깥일을 볼 수 있는 시간이 없다/충분한 휴식을 취할 수가 없다
가족관계의 부정적 변화	가족 간의 갈등을 일으키는 경우가 많다/가족들이 나의 어려움을 이해해주지 못한다/나는 하느라고 하는데 가족이 알아주지 않으면 화가 난다/다른 가족이 노인을 돌볼 수 있음에도 아무것도 하지 않는다/노인의 부양으로 인하여 가족관계가 나빠졌다
심리적 부담감	죽고 싶다고 느낄 정도로 괴롭다/나의 미래가 어떻게 될 지 불안하다/요즘 아무 것도 하고 싶지 않다/나의 미래는 그렇게 밝을 것 같지 않다
재정 및 경제 활동상의 부담감	노인의 간호와 치료에 드는 비용이 가계에 부담이 된다/노인의 간호와 치료에 드는 비용이 부족하다/노인의 부양으로 인하여 경제활동에 지장을 받는다
건강상의 부담감	피로를 자주 느낀다/충분한 수면을 취하지 못한다/건강이 좋지 않아 불안하다

표준이용계획서를 인지하지 못하는 경우 ‘모름’으로 구분하였다. 장기요양급여 외의 다른 서비스 이용 여부는 노인장기요양보험의 재가급여 이외에 별도로 이용하고 있는 서비스가 있는지에 대한 변수이고, 급여이용 시 불편사항 변수는 불편사항이 없다고 응답한 경우와 그렇지 않은 경우 2가지로 구분하였다. 본인부담금 부담정도는 본인부담금을 납부하지 않는 기초생활수급자와 부담됨(매우 부담된다, 부담된다), 적당함, 부담되지 않음(부담되지 않는다, 전혀 부담되지 않는다)로 구분하였다.

2.2.3. 분석방법 이용지원 상담 대상자 선정모형은 ‘가족지지체계가 있는 수급자 모형’과 ‘가족지지체계가 없는 수급자 모형’으로 구분하여 개발하였다. ‘가족지지체계가 있는 수급자 모형’은 주 상담대상자가 가족부양자로, 이용지원 상담에 대한 욕구가 있는 가족을 선정하기 위한 모형이다. 모형개발을 위해 가족부양자가 응답한 1,827명의 데이터를 활용하였고, 독립변수로 수급자 특성, 가족부양자 특성, 재가급여 이용관련 특성을 사용하였다. ‘가족지지체계가 없는 수급자 모형’의 주 상담대상자는 가족부양자가 없는 독거노인으로, 실제 상담에 대한 욕구가 있는 수급자를 선정하기 위한 모형이다. 모형 개발을 위해서 수급자가 응답한 173명의 데이터를 활용하였고, 독립변수로 수급자 특성, 재가급여이용 관련 특성을 사용하였다.

이용지원 상담 대상자 선정모형을 개발하기 위해 로지스틱 회귀모형, 의사결정 나무모형, Lasso 모형, 자동 신경망모형(auto neural network), 그래디언트 부스팅(gradient boosting), 그리고 고려한 각종 모형의 결과를 다수결로 결합하는 앙상블 모형을 사용하였고, 대화식 방법으로 가지를 조정한 의사결정 나무모형을 구축하였다. 모형 개발에는 SAS Enterprise Miner 13.1에서 제공하는 다양한 분류모형을 적용하였다 (SAS Institute, 2014). 가장 기본적인 방법인 로지스틱 회귀모형에서는 독립변수가 많기 때문에 후방제거법(제거 유의수준 0.05)으로 변수를 선택하였다. 나무모형은 Gini계수를 이용한 이진 분류나무를 적용하였다. 축소방법(shrinkage method)의 하나로 자연스럽게 변수선택이 되는 Lasso 모형에서는 SBC(Schwarz Bayesian Criterion) 기준으로 최적 모형을 선택하였으며, 신경망 모형으로는 Multi-Layer Perceptron 방법을 적용하였다. 약한 분류기법의 성능을 강화시키는 그래디언트 부스팅에서는 개별 모형으로 깊이가 2인 이진 나무를 사용하였다. 전문한 5개의 모형의 결과를 다수결로 결합한 앙상블 모형을 구축하였다. 또한 적합된 나무모형에 일반화하기에 무리가 있는 분지가 있어 대체 규

칙(surrogate rule)으로 변형한 나무모형도 함께 구축하였다.

과적합을 방지하고 모형의 성능을 객관적으로 측정하기 위하여 데이터를 분할하였는데 ‘가족지지체계가 있는 수급자 모형’은 대상자가 1,827명으로 충분하여 분석용 40%, 평가용 30%, 검증용 30%로 분할하였다. ‘가족지지체계가 없는 수급자 모형’은 독립변수는 많지만 대상자가 173명으로 적어 분석용 70%, 검증용 30%로 분할하였다. 모형의 성능을 비교하기 위하여 오분류율(misclassification rate), ROC 곡선의 면적인 AUC(Area Under Curve)를 이용하였다.

3. 연구결과

3.1. 인구사회학적 특성

분석에 활용된 수급자의 인구사회학적 특성을 살펴보면, 전체 수급자 중에서 ‘여성’(71.4%)이 ‘남성’(28.6%)보다 많았고, ‘75~84세’가 다른 연령대보다 많았다. 장기요양등급별로 살펴보면, ‘3등급’(69.7%)이 가장 많았으며, ‘2등급’(20.3%), ‘1등급’(10.1%) 순이었다. 자격유형별로 살펴보면, ‘일반대상자’(65.8%)가 가장 많았으며, ‘경감대상자’(17.2%), ‘기초수급’(15.4%), ‘의료급여’(1.7%) 순이었다. 교육수준의 경우, 노인이 대부분인 조사대상자의 특성상 ‘중학교졸업 이하’가 대부분(87.0%)을 차지했다. 거주지역은 ‘경기/인천’(25.8%)이 가장 많았으며, ‘서울’(14.9%), ‘전라’(14.9%), ‘경상-부산권’(13.1%), ‘충청’(12.9%) 순이었다 (표로 제시하지 않음).

가족부양자의 인구사회학적 특성을 살펴보면, 전체 가족부양자 중 ‘여성’(68.5%)이 ‘남성’(31.5%)보다 많았고, ‘45~64세’가 다른 연령대보다 많았다. 결혼상태별로는 ‘기혼’(87.4%)이 가장 많았으며, 교육수준은 ‘고졸 이상’이 절반(57.5%) 이상을 차지했다. 신체적 수발기간별로 살펴보면, ‘1년 이상 5년 미만’(44.5%)이 가장 많았고, ‘5년 이상 10년 미만’(27.0%), ‘10년 이상’(24.5%) 순이었다. 대상자의 절반(65.8%) 이상이 경제활동을 하지 않는 것으로 나타났다. 직업별로 살펴보면, ‘전업주부’(43.0%)가 가장 많았고, ‘기타’(26.8%), ‘회사원’(14.7%), ‘자영업’(12.6%) 순으로 나타났다. 가구 월평균 소득의 경우, ‘100만원 이상 300만원 미만’(43.2%)이 가장 많았으며, ‘100만원 미만’(36.6%), ‘300만원 이상 500만원 미만’(17.1%), ‘500만원 이상’(3.1%) 순으로 나타났다 (표로 제시하지 않음).

3.2. 가족지지체계가 있는 수급자 모형

‘가족지지체계가 있는 수급자 모형’의 분석 결과를 살펴보면, 로지스틱 회귀모형이 검증용 데이터에서 오분류율 38.91%로 가장 높은 성능을 보였고, 다음으로 그래디언트 부스팅, Lasso 순으로 나타났다. AUC기준으로는 그래디언트 부스팅, 앙상블모형, 신경망모형, Lasso 모형, 나무모형 순으로 우수한 것으로 나타났다 (Table 3.1). 그러나 그래디언트 부스팅, 앙상블모형, 신경망모형은 예측력은 좋은 반면 로지스틱 회귀모형이나 나무모형과 같이 예측 결과를 쉽게 이해할 수 없다는 단점이 있다. 모형의 결과를 이해하기 쉬운 장기요양 등급판정모형 개발 (Lee 등, 2013), 표준이용계획 산출을 지원하는 표준급여모형 개발 (Han 등, 2014)에 활용된 나무모형의 경우, 분석용 데이터에서의 성능은 오분류율 31.41%로 우수한 편이었지만, 평가용과 검증용 데이터에서 로지스틱 회귀모형에 비해 낮은 성능을 보였다. 로지스틱 회귀모형, 그래디언트 부스팅이나 Lasso 모형 간의 오분류율 차이는 1.09%p로 그렇게 크지 않았고, 나무모형과는 2.36%p 정도의 차이를 보였다.

변수의 중요도를 산출할 수 있는 로지스틱 회귀모형, 그래디언트 부스팅, Lasso 모형, 나무모형을 통해 ‘가족지지체계가 있는 수급자 모형’의 중요 변수를 살펴보았다. 그래디언트 부스팅은 변수 중요도가 높은 변수 상위 10개, Lasso 모형은 최종 모형에서 남은 변수를, 그리고 나무모형에서는 변수 중요도가 높은 변수를 요약하여 정리하였다 (Table 3.2). 로지스틱 회귀모형은 총 15개의 변수를 모형의 중요 변수

Table 3.1. Assessments of the algorithms for beneficiaries with family caregivers

	Training data(40%)		Validation set(30%)		Test set(30%)	
	오분류율	AUC	오분류율	AUC	오분류율	AUC
Logistic	0.3388	0.719	0.3905	0.652	0.3891	0.627
Gradient Boosting	0.3032	0.761	0.3485	0.679	0.4018	0.644
Lasso	0.3868	0.655	0.3686	0.663	0.4018	0.623
Ensemble	0.3100	0.750	0.3504	0.688	0.4055	0.640
Neural network	0.3388	0.727	0.3631	0.667	0.4055	0.628
Tree (surrogate rule)	0.3141	0.724	0.3814	0.643	0.4127	0.596
Tree	0.3278	0.698	0.3686	0.662	0.4200	0.597

Table 3.2. Variable importance lists for beneficiaries with family caregivers

Logistic		Gradient Boosting	
① 수급자(행동)-화기관리 못함		① 가족-이용지원 상담 인지도	
② 가족(부담)-심리적 부담감점수		② 가족-수급자와의 관계	
③ 수급자(수단)-근거리외출하기		③ 가족-급여이용시 표준이용계획서 참고여부	
④ 수급자(인지)-길잃음		④ 가족-재가급여이용시 어려움 여부	
⑤ 가족(부담)-재정 및 경제활동상 부담점수		⑤ 가족(부담항목)-노인과 있는 게 즐겁지 않다	
⑥ 가족-이용지원상담 인지도		⑥ 수급자-재활영역점수	
⑦ 수급자(신체)-양치질하기		⑦ 가족-결혼상태	
⑧ 수급자(신체)-방밖으로 나오기		⑧ 가족(부담)-재정 및 경제활동 상 부담점수	
⑨ 수급자(재활)-우측하지		⑨ 수급자-요통	
⑩ 수급자(수단)-금전관리		⑩ 가족(부담)-사회적 활동제한 부담점수	
⑪ 수급자(재활)-우측상지			
⑫ 가족-재가급여이용시 어려움 여부			
⑬ 가족-성별			
⑭ 가족-급여이용시 표준이용계획서 참고여부			
⑮ 수급자(질병)-요통			
Lasso		Tree (surrogate)	
① 가족-이용지원 상담 인지도		① 가족-이용지원 상담 인지도	
② 가족-재가급여이용시 어려움 여부		② 가족(부담)-재정 및 경제활동상 부담점수	
③ 가족-급여이용시 표준이용계획서 참고여부		③ 가족-재가급여이용시 어려움 여부	
		④ 가족-수급자와의 관계	
		⑤ 가족-주관적 건강상태	
		⑥ 가족(부담항목)-노인과 있는 게 즐겁지 않다	
		⑦ 수급자(신체)-체위변경하기	
		⑧ 가족(부담)-노인을대하기 싫다	
		⑨ 수급자(신체)-목욕하기	
		⑩ 수급자(신체)-머리감기	
		⑪ 수급자(인지)-하루일과 이해장애	

로 선택하였다. 15개의 중요 변수는 수급자의 건강 및 기능상태 변수가 주를 이루었으며, 가족부양자의 특성 중에는 가족부양자 부양부담감과 이용지원 상담에 대한 인지도, 재가급여이용시 어려움 여부, 급여 이용시 표준이용계획서 참고여부 등이 포함되었다. 그래디언트 부스팅은 가족부양자의 특성을 중요 변수로 꼽았고, 가족부양자의 이용지원 상담에 대한 인지도, 수급자와의 관계, 급여이용시 표준이용계획서 참고여부를 가장 중요한 변수로 선택하였다. Lasso 모형의 경우, 이용지원 상담에 대한 인지도, 재가급여 이용시 어려움 여부, 급여이용시 표준이용계획서 참고여부가 최종 모형에 남은 변수였다. 나무모형

Table 3.3. Assessments of the algorithms for beneficiaries without family caregivers

	Training data(70%)		Test set(30%)	
	오분류율	AUC	오분류율	AUC
Ensemble	0.1083	0.982	0.3396	0.566
Lasso	0.3000	0.709	0.3396	0.639
Gradient Boosting	0.1583	0.949	0.3585	0.680
Tree(surrogate rule)	0.1917	0.858	0.3585	0.617
Tree	0.1333	0.927	0.3774	0.618
Logistic	0.2000	0.850	0.5094	0.388
Neural network	0.0083	0.999	0.5660	0.433

Table 3.4. Variable importance lists for beneficiaries without family caregivers

Logistic	Gradient Boosting	Lasso	Tree (surrogate)
① 신체-체위변경하기	① 이용지원 상담 인지도	① 이용지원 상담 인지도	① 신체-체위변경하기
② 간호-욕창간호	② 신체-체위변경하기	② 신체-체위변경하기	② 급여이용시 표준이용 계획서 참고여부
③ 수단-빨래하기	③ 급여이용시 표준이용계획서 참고여부		③ 재활-어깨관절
④ 신체-옷벗고 입기	④ 인지기능점수		④ 수급자 학력
⑤ 수단-식사준비하기	⑤ 신체-화장실 사용하기		⑤ 수단-교통수단이용
⑥ 신체-양치질	⑥ 재가급여 이용기간		⑥ 신체-일어나 앉기
⑦ 수단-몸단장하기	⑦ 이용중인 재가급여 종류		⑦ 신체-식사하기
⑧ 행동-화기관리못함	⑧ 신체기능영역 일상생활점수		⑧ 이용지원 상담 인지도
⑨ 이용중인 재가급여 종류	⑨ 수단-교통수단이용		⑨ 행동변화점수
⑩ 질병-골절	⑩ 재가급여이용시 어려움 여부		
⑪ 질병-혈압			

또한 가족부양자의 변수 중 이용지원 상담에 대한 인지도, 재가급여이용시 어려움 여부, 급여이용시 표준이용계획서 참고여부를 중요 변수로 꼽았고, 수급자의 건강 및 기능상태 관련 변수 중 신체기능상태 변수를 중요 변수로 꼽았다.

3.3. 가족지지체계가 없는 수급자 모형

‘가족지지체계가 없는 수급자 모형’의 분석 결과를 살펴보면, 앙상블모형과 Lasso 모형이 검증용 데이터에서 오분류율 33.96%로 가장 높은 성능을 보였고, 다음으로 그래디언트 부스팅, 나무모형 순으로 나타났다. AUC 기준으로는 그래디언트 부스팅, Lasso 모형, 나무모형 순으로 성능이 좋게 나타났다 (Table 3.3).

로지스틱 회귀모형, 그래디언트 부스팅, Lasso 모형, 나무모형을 통해 ‘가족지지체계가 없는 수급자 모형’의 중요 변수를 살펴보면 Table 3.4와 같다. 로지스틱 회귀모형은 총 11개의 변수를 모형의 중요 변수로 선택하였고, 수급자의 건강 및 기능상태 변수가 주를 이루었으며, 주로 신체기능 및 수단적 일상생활 기능 변수였다. 그래디언트 부스팅은 로지스틱 회귀모형과 달리 주로 이용지원 상담에 대한 인지도, 급여이용시 표준이용계획서 참고여부, 재가급여 이용시 어려움 여부 등을 중요 변수로 선택하였다. Lasso 모형은 이용지원 상담에 대한 인지도와 신체기능 영역의 체위변경하기를 중요 변수로 선정하였다. 마지막으로 나무모형은 다른 모형과 마찬가지로 신체기능 변수를 중요 변수로 선정하였고, 이와 더불어 급여이용시 표준이용계획서 참고여부, 이용지원 상담 인지도를 중요 변수로 선정하였다.

Table 4.1. Comparison Table for the current counselling target group and the results predicted by Lasso model

Lasso 모형 추정 결과			현행 이용지원 상담군							
			1군(1개월)		2군(3개월)		3군(6개월)		4군(12개월)	
구분	N	%	N	%	N	%	N	%	N	%
상담필요	883	50.6	30	41.7	430	50.5	350	51.3	73	52.1
상담불필요	863	49.4	42	58.3	422	49.5	332	48.7	67	47.9

4. 결론

본 연구는 통계적 모형을 적용하여 이용지원 상담 요구가 있는 대상자를 선정함으로써 이용지원 상담 대상자 선정 시 대상자 욕구 미반영에 따른 문제점을 해결하고 이용지원 상담의 효과를 높이고자 수행되었다. 본 연구는 이용지원 상담 대상자 선정모형으로 ‘가족지지체계가 있는 수급자모형’과 ‘가족지지체계가 없는 수급자모형’을 제안하였다.

연구결과, 재가급여 수급자 중 가족부양자가 있어 주 상담대상인 상담욕구가 있는 가족부양자를 선정하는 모형인 ‘가족지지체계가 있는 수급자 모형’의 경우 분석에 활용된 모형 중 15개의 변수로 수정된 로지스틱 회귀모형의 성능이 가장 우수하였으나, 모형의 간결성을 고려하면 Lasso 모형이 단 3개의 변수로 로지스틱 회귀모형에 가까운 성능을 보였다. 이에 Lasso 모형의 결과를 ‘가족지지체계가 있는 수급자 모형’의 최종 모형으로 선정하였다. 한편, 부양가족이 없는 재가 수급자 중 이용지원 상담 욕구가 있는 대상자를 선정하는 모형인 ‘가족지지체계가 없는 수급자모형’은 신경망모형이 분석용 데이터에서 최고의 성능을 보였으나 검증용 데이터에서 낮은 성능을 보여 과적합되었고, 5개 모형의 결과를 결합한 앙상블 모형 또한 과적합의 양상을 보였다. 데이터의 숫자가 173개로 적고 독립변수는 상대적으로 많기 때문에 로지스틱회귀모형이나 신경망모형의 성능이 매우 불안정하였다. 반면 Lasso 모형은 단 2개의 변수로 검증용 데이터에서 가장 낮은 오분류율과 높은 AUC를 보였는데 이는 Lasso 모형이 회귀계수가 커지는 것을 벌점으로 제약하여 궁극적으로 중요하지 않은 변수를 제거하는 축소방법의 하나로 데이터는 적고 변수는 많은 고차원 데이터(high-dimensional data) 분석에 적절하기 때문으로 보인다 (Tibshirani, 1996; Hastie 등, 2009). 따라서 ‘가족지지체계가 없는 수급자모형’의 최종 모형으로 데이터 수가 변수의 수에 비해 상대적으로 적은 고차원 데이터상황에서 가장 간결하고 안정적인 성능을 보인 Lasso 모형을 선택하였다.

본 연구에서 제안한 이용지원 상담 대상자 선정모형은 이용지원 상담의 대상이 되는 재가급여 수급자 및 가족부양자의 상담욕구를 반영하여 통계적 알고리즘을 통해 개발하였다는 데에 의의가 있다. Lasso 모형은 이용지원 상담에 대한 욕구가 있는 대상자를 선정하는 중요 변수로 ‘이용지원 상담 인지도’, ‘재가급여이용시 어려움 여부’, ‘급여이용시 표준이용계획서 참고 여부’와 수급자의 신체기능 상태 중 ‘체위변경하기’를 채택하였다. 본 연구에서 제안한 Lasso 모형에서 중요 변수로 채택한 변수를 통해 분석대상자 2,000명의 상담대상자를 추정하였고, 이 결과를 현행 상담기준에 따른 상담군 결과와 비교하였다 (Table 4.1). 전체 2,000명 중 재가급여에서 시설급여로 변환(8.3%)하거나 급여를 미이용(4.4%)하여 현행 재가급여 상담군에서 제외된 자를 제외한 1,746명에 대해 다음 회귀에 진행될 상담군이 할당되었다. 이 1,746명을 Lasso 모형에 적용한 결과, 50.6%가 상담필요 대상으로 선정되었고 49.4%가 상담불필요군, 즉, 요구가 없는 자로 선정되었다. 현행 상담군별 Lasso 모형의 추정 결과를 비교해 보면, 각군의 절반가량이 상담불필요 군으로 선정되었다. 이는 모든 재가급여 수급자와 가족부양자가 현행 상담선정기준에 따라 이용지원 상담 대상으로 선정되어 주기적으로 상담을 제공받지만 대상자의 절반가량이 공단의 상담을 불필요하다고 느낀다고 해석할 수 있다. 따라서 향후 이용지원 상담 대상자 선정시 Lasso 모형에서 채택한 변수를 인적조사표 또는 상담기록표에 포함하여 상담 필요도가 있는 대상자를

과약하고 상담을 제공한다면, 기존 상담대상자 선정이 공단의 일방향적 판단에 의해 이루어져 이용지원 상담에 대한 효율성과 만족도가 낮았던 부분을 해소할 수 있을 것으로 기대된다.

References

- Han, E. J., Kwon, J. H., Lee, J. M., Lee, J. S., Choi, J. K., Park, J. D. (2013). *Improvement of LTC Service Counseling System for the Community-Dwelling Elderly*, National Health Insurance Service, Seoul.
- Han, E. J., Lee, J. M., Jo, J. W. and Kim, D. H. (2012). *Improvement of LTC Service Management System*, National Health Insurance Service, Seoul.
- Han, E. J., Lee, J. S., Kim, D. G. and Kwon, J. H. (2014). A decision-support system for care plan in Long-term care insurance, *The Korean Journal of Applied Statistics*, **27**, 667–679.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Second Ed, Springer Verlag, New York.
- Hyun, K. R. and Lee, S. M. (2012). Effects on the functional status changes of LTC(Long-Term-Care) services, *Journal of the Korean Gerontological Society*, **32**, 593–609.
- Jegal, H. S. (2011). Improvement of Long-term care insurance service quality: development of integrated community system for the elderly health and welfare and introduce necessary of care management, Public policy institute for people, Seoul.
- Kang, H. C., Han, S. T., Choi, J. H., Lee, S. G. and Kim, E. S. (2014). *Data Mining Methodology for Big Data Analysis*, Freedom Academy, Seoul.
- Kwon, J. D. (1994). A study on the assessment of caregiver burden in caring for the demented elderly in Korea, Doctoral thesis of Yonsei university, Seoul.
- Lee, Y. K., Jung, K. H., Kim, J. S., Kim, C. W., Park, G. W., Jung, M. Y., Kim, S. J. and Nam, H. J. (2013). Reform LTC Level-Decision system and development of dementia management model, National Health Insurance Service-Korea Institute for Health and Social Affairs, Seoul.
- National Health Insurance Service (2014a). *Act on Long-Term Care Insurance for Senior Citizens*, National Health Insurance Service, Seoul.
- National Health Insurance Service (2014b). *A Source Book for the Work Processing of Long-Term Care Insurance*, National Health Insurance Service, Seoul.
- OECD (2011). *Help Wanted?*, Providing and paying for long-term care.
- SAS Institute Inc (2014). SAS Enterprise Miner 13.1 Reference Help. Cary, NC.
- Sunwoo, D., Lee, T. H., Seo, D. M., Chung, S. D. and Kim, S. J. (2014). *Improvement for Advanced Long-Term Care Insurance System*, National Health Insurance Service, Seoul.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso, *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.

노인장기요양보험 이용지원 상담 대상자 선정모형 개발

한은정^a · 김동건^{b,1}

^a국민건강보험공단 건강보험정책연구원, ^b동덕여자대학교 정보통계학과

(2015년 7월 13일 접수, 2015년 8월 1일 수정, 2015년 10월 15일 채택)

요약

우리나라 노인장기요양보험에서는 수급자와 그 가족부양자가 수급자의 심신기능 상태와 욕구에 따라 불이익이나 불편함이 없이 비용-효과적으로 장기요양 급여를 이용할 수 있도록 지원하고자 이용지원 상담을 제공하고 있다. 본 연구는 재가급여 이용자의 이용지원 정기상담 대상자 선정시 상담 대상자의 욕구를 반영하지 않아 이용지원 상담의 만족도와 효율성이 낮은 문제를 통계학적 모형을 활용하여 해결하고자 수행되었다. 모형 개발을 위해 2013년 3월 장기요양 재가급여를 이용한 수급자와 가족부양자를 대상으로 이용지원 상담에 대한 욕구와 관련 변수를 조사하였으며, 2,000명이 조사를 완료하였다. 조사 자료를 바탕으로 이용지원 상담 대상자 선정모형을 다양한 데이터마이닝 기법(로지스틱 회귀모형, 의사결정 나무모형, Lasso 모형, 자동 신경망모형, 그래디언트 부스팅, 앙상블 모형)을 통해 개발하였고, 이 중 가장 안정적이고 현장 적용이 쉽고 성능이 좋은 Lasso 모형 결과를 최종모형으로 선정하였다. 본 연구가 이용지원 상담의 만족도를 높이고 업무를 효율화 하는데 기여할 것으로 기대된다.

주요용어: 노인장기요양보험, 이용지원 상담, 데이터마이닝 기법, Lasso 모형

¹교신저자: (01369) 서울 도봉구 삼양로 144길 33, 동덕여자대학교 정보통계학과.

E-mail: dongg@dongduk.ac.kr