

## 데이터 마이닝을 이용한 당뇨병환자의 관리요인에 관한 연구

김유미<sup>1</sup>, 장동민<sup>2</sup>, 김성수<sup>3</sup>, 박일수<sup>4</sup>, 강성홍<sup>2\*</sup>

<sup>1</sup>한국보건산업진흥원, <sup>2</sup>인제대학교 보건행정학부, <sup>3</sup>서울아산병원, <sup>4</sup>국민건강보험공단

### A Study on Factors of Management of Diabetes Mellitus using Data Mining

Yoo-Mi Kim<sup>1</sup>, Dong-Min Chang<sup>2</sup>, Sung-Soo Kim<sup>3</sup>, Il-Su Park<sup>4</sup> and Sung-Hong Kang<sup>2\*</sup>

<sup>1</sup>Korea Health Industry Development Institute,

<sup>2</sup>Dept. of Health Policy & Management, InJe University,

<sup>3</sup>Dept. of Medical Record and Informatics, Asan Medical Center,

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

**요약** 본 연구의 목적은 당뇨병환자 관리와 관련된 요인을 규명하는데 있다. 2005년 국민건강 영양조사에 참여한 20세 이상의 성인 당뇨병환자를 대상으로 하였다. 데이터마이닝 기법을 이용하여 로지스틱 회귀모형, 의사결정나무, 신경망 모형으로 당뇨병환자관리모형을 개발한 결과 의사결정나무가 가장 설명력이 뛰어났다. 당뇨병인지율과 관련된 요인으로는 연령, 거주지 및 직업이었고 그 중 연령이 가장 중요한 요인으로 나타났다. 당뇨병치료율과 관련된 요인으로는 당뇨병인지여부, 거주지 및 직업이었고 그 중 당뇨병인지여부가 가장 중요한 변수로 나타났다. 당뇨병환자의 관리프로그램은 당뇨병환자의 특성별 군집으로 분류하고 그에 따라 관리해야 한다.

**Abstract** The Objectives: The purpose of this study is to identify the factors related to management of DM in Korea. Methods: The subjects selected by using data of National Health and Nutrition Survey(NHANS) in 2005 were 415 adults, aged 20 and older, and diagnosed with DM. This study used data mining algorithms. This study validated the predictive power of data mining algorithms by comparing the performance of logistic regression, decision tree, and Neural Network on the basic of validation, it was found that the model performance of decision tree was the best among the above three techniques. Result: First, awareness of DM was positively associated with age, residential area, and job. The most important factor of DM awareness is age. Awareness rate of DM with 52 age over is 76.1%. Among the  $\geq 52$  age group, an important factor is family history. Among patients who are 52 years or over with family history of DM, an important factor is job. The awareness rate of patients who are 52 age over, family, history of DM, and professionals is 95.0%. Second, treatment of DM was also positively associated with awareness, region, and job. The most important factor of DM treatment is DM awareness. Treatment rate of patients who are aware of DM is 84.8%. Among patients who have awareness of DM, an important factor is region. The awareness rate of patients who are aware of DM in rural area is 10.4%. Conclusion: Finally, the result of analysis suggest that DM management programs should consider group characteristic of DM patients.

**Key Words** : Management of DM, Data mining, Decision Tree

---

본 논문은 2007년도 인제대학교 학술 연구조성비 보조에 의한 것임.

\*교신저자 : 강성홍(hcmkang@inje.ac.kr)

접수일 09년 03월 23일

수정일 09년 05월 01일

게재확정일 09년 05월 27일

## 1. 서론

### 1.1 연구의 배경 및 필요성

식생활 패턴의 서구화, 운동량 감소, 사회의 복잡성 등 생활양식의 변화로 우리나라도 선진국처럼 당뇨병자가 증가하여 당뇨병자의 유병률이 1970년에 1% 미만으로 추정되던 것이 1980년대 말에는 약 3%로 증가 하였으며, 2005년에는 유병률이 7.3%가 될 정도로 크게 증가하였다[1]. 또한 당뇨병은 국가의 사망원인의 4위로 부상 될 만큼 주요 사망원인이 됨에 따라 당뇨병은 국가차원에서 관심을 가지고 대처해야 할 보건문제가 되었다[2].

당뇨병은 당뇨 자체로 인한 문제와 더불어 합병증 발생으로 인한 건강문제를 야기할 수 있다. 당뇨병의 증가는 망막합병증으로 인한 실명, 당뇨병성 신증으로 인한 말기신부전증으로의 진행, 신경합병증으로 인한 하지 절단 등을 야기하며 특히 심혈관 질환으로 인한 사망의 증가 및 의료비 상승과 직결되고 있다[3-6]. 이와 같이 당뇨로 인한 합병증 문제가 심각함에 따라 당뇨는 예방 못지 않게 당뇨병자 관리 사업의 중요성도 매우 크다.

당뇨는 다른 질병과 달리 자가 관리가 매우 중요한 병으로서 당뇨병자들이 당뇨약을 규칙적으로 복용하고 일생동안 생활습관을 개선하면 합병증 예방 등 질병 결과의 조절이 가능한 질환이다. 당뇨병자 관리를 위한 국가차원의 노력과 국민들의 건강에 대한 관심의 증대에 힘입어 최근에 당뇨병자들의 당뇨에 대한 관리수준이 크게 향상 되고 있다. 1998년에 당뇨병자 인지율이 35.5%였는데 2005년에 67.6%로 크게 증가하였다. 당뇨병에 대한 진단은 1998년에 21.6%이던 것이 2005년에는 67.6%로 크게 상승하였다. 당뇨병의 약물치료율은 1998년에 25.6%이던 것이 2005년에 47.7%로 크게 개선되고 있는 것으로 나타났다[1]. 이와 같이 당뇨병에 대한 관리수준이 크게 개선됨에 따라 당뇨병자에 대한 관리사업의 내용도 변화되어야 할 것 이다.

당뇨환자 관리사업이 보다 효율적으로 수행되기 위해서는 당뇨병자의 관리요인을 규명하고 이를 기반으로 하여 대상자에 특성에 맞는 맞춤형 당뇨관리사업을 실시하여야 할 것이다. 그러나 기존의 당뇨병자의 관리요인에 대한 연구는 일부지역이나 특정 집단을 대상으로 한 것임에 따라 전국적으로 대표성 있는 자료를 이용하여 이의 요인을 규명하는 연구가 필요하다. 또한 최신 정보기술인 데이터마이닝 기법을 이용하여 당뇨병자의 관리요인을 체계적으로 규명할 필요가 있다. 왜냐하면 데이터마이닝이란 광범위한 데이터에서 의미 있는 정보를 손쉽게 추출하여 만성질환관리 사업을 효율적으로 수행할 수 있는 방안을 마련하는데 도움을 줄 수 있기 때문이다[7].

### 1.2 연구목적

본 연구는 당뇨의 관리 수준을 향상시키는데 적절하게 활용할 수 있는 기초자료를 제공하는 것이다. 이를 달성하기 위한 구체적인 연구목적은 다음과 같다.

첫째, 당뇨 환자의 관리 실태를 파악한다.

둘째, 데이터마이닝 기법을 이용하여 당뇨병자의 관리요인을 규명한다.

## 2. 연구방법

### 2.1 자료수집

2005년도 국민건강 영양조사를 통해 검진에 참여한 20세 이상의 성인을 대상으로 하였다. 검진조사의 최종 대상자는 10,816명이고 이 중 7,597명이 검진조사를 실시하였다. 이 중 20세 이상인 5,502명 중에서 당뇨병자 415명의 자료를 최종 분석 대상으로 하였다[1]. 수집된 자료의 주요 내용은 조사대상자의 성, 연령, 소득, 학력 등의 일반적 특성, 주관적 건강인식, 고혈압유무, 고지혈증 유무 등의 건강수준과 흡연, 음주, 고위험음주, 운동, 건강검진과 같은 건강행태에 관한 내용이다.

### 2.2 자료분석

#### 2.2.1 변수정의

본 연구에서는 2005년도 국민건강 영양조사의 변수정의의를 기반으로 본 연구의 목적에 맞게 표 1과 같이 정의하였다. 당뇨관리의 지표는 당뇨 인지율과 당뇨 치료율을 선정하였다. 이 두 가지 지표를 선정한 이유는 당뇨관리에 있어서 당뇨를 조기에 인지하고, 적절한 치료를 하게 하는 것이 보건사업에서 수행해야할 가장 기본적인 당뇨관리 사업이라 판단하였기 때문이다[1].

[표 1] 변수정의

항 목	정 의
당뇨환자	공복혈당 수치가 126mg/dl 이상 또는 경구용 혈당강하제 복용 또는 인슐린 투여 또는 의사진단을 받은 경우
흡연자	현재 흡연을 하고 있는 자
음주자	현재 음주를 하고 있는 자
고위험음주자	한 번에 소주 5잔 이상에 해당하는 양을 일주일에 3번 이상 마시는 자
스트레스인지	주관적 스트레스 경험으로 평상시 생활중에 '대단히 많이' 혹은 '많이'에 응답한 사람
운동여부	주3회 이상 규칙적으로 운동하는지 여부
건강검진여부	지난 2년간의 건강검진 여부
당뇨인지여부	당뇨병 유병자중 평생 당뇨를 앓은 적이 있다(본인인지)고 응답한 경우
당뇨치료	당뇨병 유병자중 현재 경구용 혈당강하제 복용 또는 인슐린 투여하는 경우

### 2.2.2 자료분석

당뇨환자의 일반적 특성, 건강행태, 당뇨관리 실태에 대해서는 빈도분석을 실시하였다. 일반적 특성과 당뇨관리 실태에 대한 관련성에 대해서는 교차분석을 실시하여 변수들간의 관련성을 보았다. 당뇨관리에 영향을 끼치는 요인분석은 종속변수로는 당뇨인지율, 당뇨치료율 변수를 사용하였으며, 독립변수는 성, 연령, 소득, 학력 등의 일반적 특성, 주관적 건강인식, 고혈압유무, 고지혈증 유무 등의 건강수준과 흡연, 음주, 고위험음주, 운동, 건강검진과 같은 건강행태에 관한 변수를 사용하였다. 데이터 마이닝 분석은 로지스틱 회귀분석, 신경망분석, 의사결정 분석을 이용하여 분석하였다. 본 연구에서 활용된 로지스틱 회귀분석은 단계적 추출방법(Stepwise)을 이용하였고, 의사결정나무(Decision Tree) 분석은 CHAID (Chi-Squared Automatic Interaction Detection)방법 중 카이제곱 통계량(Chi-Squared Statistics)의 p값을 활용하여 최적분리를 하고자 하였고, 신경망 분석은 MLP(Multi Layer Perceptron)를 이용하였다.

세 가지 분석방법 중 최적모형의 선택을 위한 평가지표는 평균오차의 제곱근(Average Squared Error, ASE)를 활용하였다. 최종 선정된 의사결정나무는 의사결정규칙(decision rule)을 도표화하여 관심대상이 되는 집단을 몇 개의 소집단으로 분류(classification)하거나 예측(prediction)을 수행하는 분석방법이다. 의사결정분석은 분석과정이 나무구조에 의하여 표현되기 때문에 판별분석(Discriminant Analysis), 회귀분석(Regression Analysis), 신경망(Neural Networks) 등과 같은 방법들에 비해 연구자가 분석과정을 쉽게 이해하고 설명할 수 있는 장점을 가지고 있다[8].

## 3. 연구결과

### 3.1 분석대상자의 특성 및 건강행태

당뇨환자의 성별분포는 남자가 214명(51.6%)으로 여자 201명(48.4%)보다 조금 많았으며, 연령은 60대가 123명(29.6%)으로 가장 많고 50대가 115명(27.7%), 40대 85명(20.5%), 70세 이상은 71명(17.2%) 순이었으며 30대 이하는 21명(5%)으로 당뇨환자는 50대 이상에서 많다는 것을 확인할 수 있었다. 지역별로 보면 대도시가 199명(48.0%)으로 가장 많았다. 독신 110명(26.7%)보다 유배우가 304명(73.3%)으로 많았다. 교육수준은 초등학교가 129명(31.1%)으로 가장 많았으며 고등학교 110명(26.5%), 중학교 67명(16.1%), 무학도 55명(13.3%)이나 되었다. 대학졸업은 54명(13.0%)로 가장 낮았다. 가구소득은 100만원 이하가 152명(36.7%), 100~200만원

(28.0%), 201~300만원(18.6%) 순으로 나타났다. 당뇨에 대한 가족력이 있는 응답자는 145명(34.5%) 이었다[표2].

당뇨와 같은 만성질환은 개개인의 건강행태와도 밀접한 관계가 있다. 본 연구에서 이들 특성에 대하여 빈도분석을 실시하였다[표 3].

[표 2] 연구대상자의 인구사회학적 특성

	변 수	N	%
성별	남자	214	51.6
	여자	201	48.4
연령	20~29세	6	1.4
	30~39세	15	3.6
	40~49세	85	20.5
	50~59세	115	27.7
	60~69세	123	29.6
	70세 이상	71	17.2
지역	대도시	199	48.0
	중소도시	141	34.0
	농어촌	75	18.0
결혼상태	독신	111	26.7
	유배우	304	73.3
교육수준	무학	55	13.3
	초등학교	129	31.1
	중학교	67	16.1
	고등학교	110	26.5
	대학교	54	13.0
직업	전문기술관리직	24	5.8
	사무직	13	3.1
	판매서비스업	43	10.4
	농어업	32	7.7
	기능단순노무직	80	19.3
	주부	88	21.2
	무직/군인/학생	135	32.5
의료보장	의료보험	384	92.5
	의료보호	30	7.3
	미가입	1	0.2
가구소득	100만원 이하	152	36.7
	100~200만원	116	28.0
	201~300만원	77	18.6
	301~400만원	26	6.3
	401만원 이상	43	10.4
당뇨 가족력	있음	145	34.9
	없음	270	65.1
전 체		415	100.0

[표 3] 연구대상자의 건강행태

변수	단위 : %	
	당뇨환자	전체 <sup>1)</sup>
현재 흡연율	28.3	28.9
고위험 음주율	41.9	61.2
스트레스 인지율	41.0	35.2
규칙적 운동실천율	45.6	32.2

<sup>1)</sup> 2005년 국민건강영양조사 성인대상

당뇨환자 중 현재 흡연중인 응답자는 28.3%였으며, 19세 이상 전체 응답자는 28.9%로 당뇨병환자가 조금 낮았다. 당뇨병환자의 고위험 음주율은 41.9%이다. 2005년 국민건강영양조사의 전체 자료에 의하면 현 음주자의 고위험 음주율은 61.2%로 본 연구에 사용된 당뇨병환자의 고위험 음주비율이 19.3% 낮았다. 스트레스 인지율에 있어서는 당뇨병환자(41.0%)가 19세 이상 전체 응답자(35.2%)보다 높게 나타났다. 당뇨병의 진단을 받으면 환자는 규칙적이고 지속적인 자기관리행동을 함으로써 당뇨병성 합병증을 예방해야 하는데, 이런 과정 등이 심한 당뇨스트레스를 야기할 수 있다[9]. 따라서 당뇨를 인지하고 있는 환자는 그렇지 않은 환자보다 스트레스를 더 많이 받게 된다. 규칙적 운동실천율은 당뇨병환자가 45.6%로 19세 이상 응답자 32.2%보다 높아 당뇨병환자의 운동실천율이 높음을 알 수 있었다. 당뇨병환자의 건강행태는 19세 이상 전체 응답자보다 스트레스 인지율을 제외하면 더 좋은 것으로 나타났다. 이처럼 당뇨병환자의 건강행태가 전체 성인에 비하여 더 양호한 것은 당뇨에 대한 인지여부가 건강행태에 영향을 미친 것으로 생각된다.

### 3.2 당뇨병환자의 관리실태

#### 3.2.1 당뇨병환자의 인지실태

연구대상자의 일반적 특성별로 당뇨 인지율을 분석하였다. 표 3에서도 나타나듯이 당뇨병환자의 흡연, 음주, 운동과 같은 건강행태는 일반 당뇨가 없는 사람들 보다 양호하게 나타나는데 이는 당뇨를 인지여부에 따라 다르기 때문이다. 따라서 당뇨에 대한 인지여부는 환자의 건강행태에 영향을 미치고 이는 당뇨의 치료와도 관련이 있어 매우 중요하다고 할 수 있다. 성별에 있어서 당뇨인지율은 여자가 71.6%, 남자는 65.3%로 여자의 당뇨인지율이 높았으나 통계적으로 유의하지 않았다. 연령에 따라서는 60대가 76.4%로 가장 높았으며 70대 이상이 76.1%, 50대가 73.7%로 40대의 51.8%와 30대 이하의 33.3%보다 20~40%이상 높아 연령이 증가하면서 당뇨인지율도 증가하는 것으로 나타났다. 지역별 당뇨인지율은 대도시가 72.2%로 가장 높고 중소도시 65.2%, 농어촌이 64.0%로 도시 규모가 클수록 당뇨인지도가 높았으나 유의한 차이는 없었다.

독신인 응답자의 당뇨인지율이 73.0%로 배우자가 있는 응답자의 66.7%보다 높았으나 통계적으로 유의하지 않았다. 학력별로 살펴본 당뇨인지율에서도 중학교 졸업 학력인 응답자가 74.2%로 가장 높은 반면 대학이상 졸업한 응답자는 53.7%로 가장 낮았으나 통계적으로 유의하지 않았다. 직업별로는 무직/군인/학생이 74.6%로 가장

높았으며 주부 72.7%, 기능/단순노무직 68.8%, 농어업 50.0%, 사무직 30.8% 순이었으며 통계적으로 유의한 차이가 있었다. 의료보험 대상자보다 의료보호 대상자의 당뇨인지율이 19.9% 높았으나 통계적으로 유의하지 않았다. 가구소득에 있어서는 월 301~400만원이 76.9%로 가장 높았고 다음으로 100만원 이하 76.3%, 100~200만원 69.8%, 401만원 이상 58.1%, 201~300만원 52.6% 순으로 나타났다. 이는 통계적으로 유의하였다. 응답자의 당뇨에 대한 가족력 유무에 따라서는 가족력이 있는 응답자의 당뇨인지율이 72.9%로 그렇지 않은 응답자보다 7% 더 높았으나 통계적으로 유의하지 않았다[표 4].

[표 4] 일반적 특성에 따른 당뇨 인지율

변 수	인지		비인지		p-value	
	N	%	N	%		
성별	남자	139	65.3	74	34.7	0.163
	여자	144	71.6	57	28.4	
연령	20~29세	2	33.3	4	66.7	0.000
	30~39세	5	33.3	10	66.7	
	40~49세	44	51.8	41	48.2	
	50~59세	84	73.7	30	26.3	
	60~69세	94	76.4	29	23.6	
	70세 이상	54	76.1	17	23.9	
지역	대도시	143	72.2	55	27.8	0.265
	중소도시	92	65.2	49	34.8	
	농어촌	48	64.0	27	36.0	
결혼 상태	독신	81	73.0	30	27.0	0.222
	유배우	202	66.7	101	33.3	
교육 수준	무학	39	70.9	16	29.1	0.061
	초등학교	95	73.6	34	26.4	
	중학교	49	74.2	17	25.8	
	고등학교	71	64.5	39	35.5	
직업	대학교	29	53.7	25	46.3	0.009
	전문기술관리직	15	62.5	9	37.5	
	사무직	4	30.8	9	69.2	
	판매서비스업	29	67.4	14	32.6	
	농어업	16	50.0	16	50.0	
	기능단순노무직	55	68.8	25	31.3	
	주부	64	72.7	24	27.3	
	무직/군인/학생	100	74.6	34	25.4	
의료 보장	의료보험	256	66.8	127	33.2	0.063
	의료보호	26	86.7	4	13.3	
	미가입	1	100.0	0	0.0	
가구 소득	100만원 이하	116	76.3	36	23.7	0.003
	100~200만원	81	69.8	35	30.2	
	201~300만원	40	52.6	36	47.4	
	301~400만원	20	76.9	6	23.1	
	401만원 이상	25	58.1	18	41.9	
당뇨가 가족력	있음	105	72.9	39	27.1	0.145
	없음	178	65.9	92	34.1	

### 3.2.2 당뇨병환자의 치료실태

당뇨환자에 있어서 현재 자신이 당뇨병인 것을 인식하고 이를 치료하기 위하여 혈당강하제를 복용하거나 인슐린을 투여하는 것과 같은 치료는 매우 중요하다. 따라서 연구대상자의 일반적 특성별로 당뇨병에 대한 치료율을 분석하는 것은 당뇨병환자의 실태를 파악하는데 의미가 있다.

성별에 있어서 당뇨병치료율은 여자가 60.0%로 남자 55.9%에 비하여 높았으나 통계적으로 유의하지는 않았다. 이는 당뇨병인지율과 같은 결과이다. 연령별로는 50대의 당뇨병치료율이 64.0%로 가장 높았다. 다음으로 60대 63.1%, 70대 이상 59.2%, 40대 47.1%, 30대 이하 33.3% 순이었다. 결과적으로 50대 이상에서는 당뇨병 치료율이 50%를 넘는 것으로 나타났다. 이는 통계적으로 유의한 결과이다.

대도시에 거주자들의 당뇨병치료율은 72.1%로 가장 높았고 중소도시 65.2%인 반면 소도시는 75명중 5명만이 치료를 받았다고 응답하여 6.7%의 치료율을 보였으며 지역별 당뇨병치료율은 통계적으로 차이가 있는 것으로 나타났다. 독신이 62.2%로 유배우자의 56.3%에 비하여 다소 높았으나 통계적 유의성은 없었다.

교육수준별로는 주부가 63.2%로 가장 높았으며, 판매서비스업 종사자가 62.8%, 전문기술관리직과 기능단순노무직 종사자가 62.5%로 같았다. 반면 사무직 종사자는 30.8%만이 당뇨병에 대한 치료를 하였고 농어업 종사자는 18.8%의 치료율을 보여 50%에도 미치지 못하는 매우 낮은 치료율을 보였다.

의료보장별로 살펴보면 의료보험 환자는 56.8%였으나 의료보험 환자는 당뇨병치료율이 70.0%로 비교적 높았으나 통계적인 유의성은 없었다. 가구소득별 당뇨병인지율은 통계적으로 차이가 있었으나 치료율은 그렇지 않았다. 가구소득이 월 301~400만원인 응답자는 65.4%로 가장 높은 치료율을 나타냈다. 다음은 100만원 이하, 100~200만원, 401만원 이상, 201~300만원 순으로 나타났으며 소득이 증가하여도 당뇨병치료율은 증가하지는 않았다. 당뇨가족력 여부별로는 당뇨가족력이 있는 경우 63.9%, 그렇지 않은 경우 54.6%로 당뇨가족력이 있는 환자의 당뇨병치료율이 높았으나 통계적으로 유의하지는 않았다[표 5].

### 3.3 당뇨병관리 모형개발

#### 3.3.1 당뇨의 인지모형

로지스틱 회귀모형, 의사결정나무, 신경망 분석방법 중 의사결정나무가 평균제곱오차가 가장 낮아서 당뇨병인지 모형의 최종 모형으로 선정하였다[표 6].

[표 5] 일반적 특성에 따른 당뇨 치료율

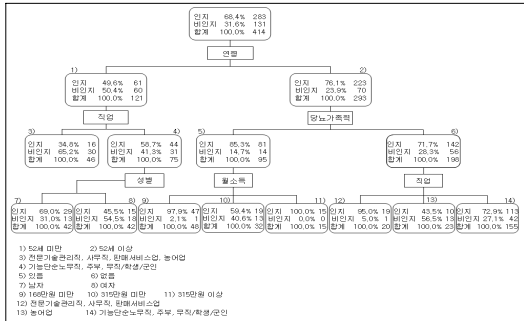
변 수	치료		미치료		p-value	
	N	%	N	%		
성별	남자	119	55.9	94	44.1	0.395
	여자	120	60.0	80	40.0	
연령	20~29세	2	33.3	4	66.7	0.029
	30~39세	5	33.3	10	66.7	
	40~49세	40	47.1	45	52.9	
	50~59세	73	64.0	41	36.0	
	60~69세	77	63.1	45	36.9	
	70세 이상	42	59.2	29	40.8	
지역	대도시	142	72.1	55	27.9	0.000
	중소도시	92	65.2	49	34.8	
	소도시	5	6.7	70	93.3	
결혼 상태	독신	69	62.2	42	37.8	0.284
	유배우	170	56.3	132	43.7	
교육 수준	무학	32	58.2	23	41.8	0.937
	초등학교	76	59.4	52	40.6	
	중학교	40	60.6	26	39.4	
	고등학교	62	56.4	48	43.6	
	대학교	29	53.7	25	46.3	
직업	전문기술관리직	15	62.5	9	37.5	0.000
	사무직	4	30.8	9	69.2	
	판매서비스업	27	62.8	16	37.2	
	농어업	6	18.8	26	81.3	
	기능단순노무직	50	62.5	30	37.5	
	주부	55	63.2	32	36.8	
의료 보장	무직/군인/학생	82	61.2	52	38.8	0.257
	의료보험	217	56.8	165	43.2	
가구 소득	의료보호	21	70.0	9	30.0	0.436
	미가입	1	100.0	0	0.0	
	100만원 이하	91	60.3	60	39.7	
	100~200만원	69	59.5	47	40.5	
당뇨 가족력	201~300만원	37	48.7	39	51.3	0.070
	301~400만원	17	65.4	9	34.6	
	401만원 이상	24	55.8	19	44.2	
당뇨 있음	92	63.9	52	36.1	0.070	
당뇨 없음	147	54.6	122	45.4		

[표 6] 당뇨병인지모형 결과 비교

모 형	평균오차의 제곱근(ASE)
로지스틱 회귀모형	0.4480
의사결정나무	0.4213
신경망	0.4397

의사결정나무 결과를 통한 당뇨병인지와의 관련요인에 대한 분석결과는 총 9개의 집단으로 세분화되었다. 모형을 결정하는 요인은 연령변수가 집단의 세분화 요인 중 가장 큰 요인으로 나타났으며, 직업, 당뇨가족력, 월소득이 집단 세분화에 영향을 미치는 요인으로 작용하였다. 당뇨병인지모형을 통한 집단별 특성을 살펴보면, 최초 연구

대상자의 당뇨인지율은 68.4%였으나, 연령이 52세 이상이고, 당뇨가족력이 있으며, 월소득이 315만원 이상인 집단은 당뇨인지율이 100%였으며, 월소득이 168만원 미만인 사람은 97.9%인 것으로 나타났다. 그러나, 연령이 52세 이상이고, 당뇨가족력이 없으며, 농어촌에 거주하는 사람일 경우는 평균 당뇨인지율보다 낮은 것으로 나타났다[그림 1].



[그림 1] 당뇨인지에 대한 의사결정나무

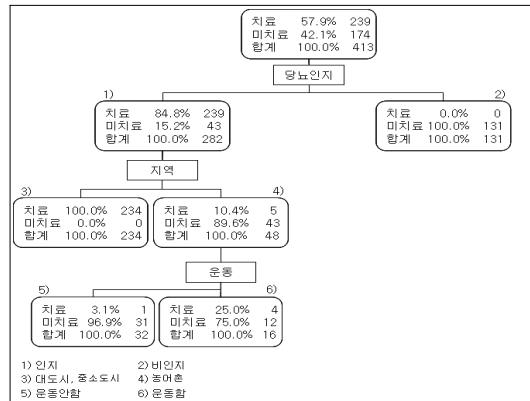
### 3.3.2 당뇨의 치료모형

로지스틱 회귀모형, 의사결정나무, 신경망 분석방법 중 의사결정나무가 평균제곱오차가 가장 낮아서 당뇨인지 모형의 최종 모형으로 선정하였다[표 7].

[표 7] 당뇨치료모형 결과 비교

모형	평균오차의 제곱근(ASE)
로지스틱 회귀모형	0.4937
의사결정나무	0.0980
신경망	0.1119

의사결정나무 결과를 통한 당뇨치료와의 관련요인에 대한 분석결과는 총 4개의 집단으로 세분화되었다. 모형을 결정하는 요인은 당뇨인지 여부가 집단의 세분화 요인 중 가장 큰 요인으로 나타났으며, 거주지역의 특성, 운동여부가 집단 세분화에 영향을 미치는 요인으로 작용하였다. 당뇨치료모형을 통한 집단별 특성을 살펴보면, 최초 연구대상자의 당뇨치료율은 57.9%였으나, 당뇨를 인지하고 있고, 중소도시 이상 거주하는 자들은 100%가 당뇨치료를 하고 있는 것으로 나타났다. 그러나, 당뇨를 인지하고 있으나, 농어촌에 거주하는 자들의 당뇨치료율은 3.1%~25.0%로 나타나, 지역적인 차이가 당뇨치료율에도 많은 영향이 있는 것으로 분석결과 나타났다[그림 2].



[그림 2] 당뇨치료에 대한 의사결정나무

## 4. 결론 및 토의

본 연구는 국민건강증진법(제16조)에 근거하여 국민들의 건강수준 및 건강행태를 파악하기 위해 전국적인 규모의 표본을 추출하여 조사한 2005년도 국민건강영양조사자료 중 건강검진이 이루어진 20세 이상 성인 당뇨병환자 415명을 대상으로 하여 국가적 당뇨병환자 관리체계를 개선하는데 필요한 기초자료를 제공하기 위하여 수행되었다.

당뇨병과 관련한 기존의 연구를 살펴본바 기존의 국내 선행연구들은 당뇨에 대한 임상적 치료와 개인적 차원의 관리에 관한 내용이 주를 이루는 것이 많았다. 또한 환자 교육과 검사결과 모니터링을 병행한 연구들이 적지 않았고 이들 연구는 대부분 소수의 의료기관에서 외래방문환자들을 대상으로 하는 연구로 관찰기관도 1년 미만의 단기간에 그치고 있다[10-12].

근래에는 진단과 치료 중심보다는 지역사회 중심, 역학적 관점, 근거기반 의료, 건강결과와 예방과 같은 전 사회적인 차원의 지원시스템이 필요하다는 주장이 제기되고 있다[13-14].

이를 바탕으로 한 최근의 국외 사례를 살펴보면 영국의 Quality and Outcome Framework(2003), 미국의 민간기구 Bridges to Excellence의 Diabetes Care Link(2003), 미국 CMS(Centers for Medicare & Medicaid Services)의 Physician Group Practice 시범사업(2005) 등은 당뇨병 등 주요 만성질환들을 효과적으로 관리하기 위한 전 사회적 프로그램들이라 할 수 있다.

2005년 국민건강영양조사에서 20세 이상 성인 중 당뇨 유병률은 7.5%였다. 이들 당뇨병환자의 건강행태는 흡연율이 28.3%, 고위험음주율 41.9%로 나타났으며, 스트

레스인지율은 41.0%였다. 규칙적 운동실천율은 전체 성인 응답자보다 많은 45.6%였다. 이는 당뇨에 대한 인지가 건강행태에 영향을 미치고 있음을 보여준다.

본 연구보다 앞서 1998년 국민건강영양조사 자료를 분석한 김상아 등(2005)의 연구는 당뇨 유병률과 운동, 흡연, 음주와 같은 건강행태의 실천정도 및 임상수치 그리고 이들에 대한 관리 상태를 단면적으로 분석하였다. 연구결과에 의하면 당뇨 유병률이 남자 9.73%, 여자 7.95%로 본 연구보다는 높았다.

최근 국가차원의 당뇨연구는 이준영 등(2007)의 연구로 인구통계자료와 보험심사청구자료, 그리고 field survey 자료 등을 사용하여, 498명(남자 249명, 여자 249명)에 대하여 만 20세 이상 당뇨병 유병률을 추정하였는데 남자의 당뇨유병률은 8.7%였으며 여자는 이보다 조금 낮은 8.4%였다. 415명의 당뇨환자를 대상으로 한 본 연구에서는 전체적인 유병률이 7.5%로 나타나 다소 낮았다. 이처럼 각 연구마다 다소 차이가 있으나 인구 1000명 당 5~6명의 당뇨환자가 발생하는 추세가 계속된다면 당뇨환자는 400만 명을 넘어 2030년이면 전체 인구의 10%를 넘을 수 있다[15].

2004년 질병관리본부에서 실시한 역학연구 조사에 의하면 50대와 60대의 남자에서 전체 평균이상의 높은 당뇨병 유병률을 보여 주었으며 여자에서도 50대와 60대의 유병률이 높은 것으로 나타났다[16]. 본 연구에서도 당뇨인지에 가장 큰 요인이 연령인 것으로 나타나 선행연구와 같은 맥락임을 보여주었다. 또한 당뇨 발병은 연령과 양의 상관관계를 보이며 이러한 관계성은 선형(linear)보다는 문턱값(threshold)현상으로 40대 이후에 급격한 증가현상을 보인다. 하지만 40대에서는 50대나 60대에 비해 관리 비율이 절반밖에 안 되며, 이를 체계적으로 관리하지 않아 질환을 키워 심각한 예후를 자처하는 의료행태를 보이고 있다고 하였다[17]. 이러한 행태로 인하여, 미국에서 당뇨병 관리를 위한 연간 의료비용을 분석한 결과, 비당뇨병 환자에 비해 당뇨병 환자의 의료비용이 5배 이상 높은 것으로 조사되었다[18].

당뇨사업에서 저소득 취약계층이 표적대상으로 선정되어 실시한 당뇨병 프로그램이 효과적이라는 연구[19]는 소득 및 직업을 당뇨의 인지와 치료에 영향요인으로 보고 있으며 Anderson 등(2003)의 연구도 지역사회 보건소들을 중심으로 저소득 취약계층의 당뇨병 환자들을 대상으로 한 자기관리 프로그램을 실시하여 효과성을 입증하였다. 본 연구결과에서도 소득, 직업, 지역이 당뇨인지와 치료에 유의한 요인으로 밝혀졌으며 이들 요인이 국가당뇨사업에서 가장 중요한 취약계층이고 주요 대상임이 밝혀졌다.

당뇨에 대한 전체적인 인지율은 68.4%였다. 전체적으로 연령이 증가할수록 당뇨 인지도율이 증가하였으며 60대가 가장 높은 것으로 나타났다. 또한 가구소득에 따라서도 당뇨 인지도율의 차이를 보였는데 월소득 100만원 이하인 응답자가 76.3%로 가장 높았다.

최근 여성의 고혈압, 당뇨의 인지와 치료에 관련한 요인을 분석한 한 국내 연구결과에서는 연령과 결혼상태 순으로 영향을 미치고 있다고 보고하고 있다[20].

당뇨와 교육 및 사회경제적 수준에 대한 국외 연구를 살펴보면 이들 요소가 건강에 대한 관심을 높여 개인 건강생활습관에 영향을 미치므로 당뇨발생이 낮아진다고 하였다[21]. 또한 미국에서 여성 당뇨환자는 결혼, 고용상태와 같은 생활수준과는 독립적으로 사회경제적 수준이 낮을 가능성이 큰 것으로 나타났다[22].

종합하여 보면, 생활습관의 급속한 변화로 환경적 요인이 중요하게 여겨지고 있으며 이러한 생활습관은 교육과 사회경제적수준에 기인하며 당뇨병뿐 아니라 각종 만성질환에도 기여하는 바가 크다고 하겠다. 본 연구의 결과를 살펴보아도 당뇨에 대한 인지에 영향을 미치는 주요 요인이 연령, 직업, 소득, 성별로 나타났고 치료모형에서는 당뇨인지, 지역이 주요 요인으로 작용하는 것으로 나타나, 선행연구의 결과를 뒷받침하고 있었다. 선행연구에서 지역과 당뇨에 대한 연구를 접할 수 없었으나 우리나라처럼 서울과 수도권에 많은 인구가 몰려있고 국가경제에 상당부분을 차지하는 것으로 보아 지역 또한 사회경제적 요인과 관련이 있다고 하겠다.

연령, 지역, 직업은 당뇨치료와 통계적으로 유의한 관련이 있는 것으로 나타났다. 특히, 농어촌 지역의 당뇨치료율이 현저히 떨어지는 것으로 나타났다. 이는 직업과도 관련이 있는 것으로 농어업 종사자의 당뇨치료율이 가장 낮았다. 건강행태별 당뇨치료율은 현재 비흡연자, 고위험 음주를 하지 않는 자, 규칙적 운동을 수행하는 응답자가 높게 나타났다. 스트레스는 많이 느끼는 응답자가 당뇨치료율이 높았으나 통계적으로 의미 있는 결과는 아니었다.

본 연구의 궁극적인 목적인 당뇨환자의 관리요인 규명을 위해 여러가지 데이터마이닝 알고리즘 모형을 구축한 후 모형비교를 통하여 의사결정나무를 최적의 모형으로 선택하였다. 먼저 당뇨인지에 대한 모형에서는 연령변수가 가장 중요한 변수이었으며 다음으로 직업, 당뇨가족력 및 성별과 소득으로 구분되었다. 당뇨인지에 대하여 가장 취약한 집단으로는 연령이 52세 이상이고 당뇨가족력이 없으며, 직업이 농업이나 어업에 종사하는 집단으로 나타났다. 당뇨치료에 대한 모형을 구축한 결과 당뇨인지가 가장 중요한 변수였으며, 다음으로 지역, 운동 순으로 나타났다. 당뇨를 인지하고 있음에도 치료율이 가장 낮은

집단은 농어촌에 거주하고 규칙적인 운동을 하지 않은 것으로 나타났다. 이를 볼 때 인지율을 높이기 위한 당뇨 조기검진 사업은 53세 이상으로 당뇨 가족력이 없으며, 직업이 농어민인 집단에 대해 중점적으로 실시하여야 한다는 것을 알 수 있었다. 또한 당뇨 치료 사업의 경우, 농어촌지역이 대도시나 중소도시에 비해 당뇨 치료율이 낮으므로 농어촌 지역을 당뇨 치료사업의 주요 대상으로 하는 것이 합리적이라 판단된다.

본 연구는 전국적인 대표성을 가진 표본자료임에도 불구하고 2차 자료를 이용하였으므로 당뇨와 관련된 기타 요인을 분석에 포함할 수 없었으나 정기적이고 국가적인 조사로서 향후 심도 있는 관련 요인을 추가하여 분석에 포함함으로써 효과적인 당뇨관리 사업에 중요한 자료가 될 수 있을 것이다.

국가 보건사업에서 불특정 다수를 대상으로 사업효과가 높기 위해서는 정확한 사업대상의 선별이 중요하다. 본 연구결과에 나타난 주요 요인을 중심으로 대상환자를 선별하여 질환을 조기에 발견하고 적절한 치료를 제공하는 국가적 당뇨병환자의 관리 사업이 절실 하다고 하겠다.

### 참고문헌

- [1] 질병관리본부. 국민건강영양조사 제3기 조사결과심층 분석 연구 보고서: 검진부문, 질병관리본부, 2007
- [2] 통계청. 사망원인통계연보(2005). 통계청, 2006.
- [3] Chodosh J, Morton SC, Mojica W, Maglione M, Suttrop MJ, Hilton L, Rhodes S, Shekelle P: Meta-analysis: Chronic disease self-management programs for older adults. *Ann Intern Med* 2005;143:427-38.
- [4] The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 1993; 329:977-86.
- [5] UK Prospective Diabetes Study (UKPDS) Group: Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998;352:837-53.
- [6] Vijan S, Stevens DI, Herman WH, Funnell MN, Stanford CJ: Screening, preventing, counseling, and treatment for the complications of type II diabetes mellitus: putting evidence into practice. *J Gen Intern Med* 1997; 12:567-80.
- [7] 박일수 외 4명. 데이터마이닝 기법을 활용한 맞춤형 고혈압 사후관리 모형 개발. *응용통계연구*, 2008; 21(4)
- [8] 최중후, 서두성. 데이터마이닝 의사결정나무의 응용. *통계분석연구* 1999;4(1):61 -83.
- [9] 전진수, 신강현, 김완석, 권정혜. 당뇨병스트레스가 우울에 미치는 영향:개인차 변수들의 조절효과. *한국심리학회지 : 건강* 2007;12(4):851-868.
- [10] 박봉숙 외 7명. 청장년 당뇨병환자의 자기관리와 삶의 질. *당뇨병* 2005;29:254 -261.
- [11] 송민선 외 10명. 체계적인 당뇨병 교육이 당화혈색소가 높은 제2형 당뇨병환자에게 미치는 장기효과:4년간 추적조사 연구. *당뇨병* 2005;29:140-150.
- [12] 장경순, 이관, 임현술. 일개 의원에서 당뇨병 교육을 통한 혈당조절과 건강상태. *당뇨병* 2006;30:73-81.
- [13] Desai J et al. Public health surveillance of diabetes in United States. *Journal of Public Health Management and Practice* 2003;Suppl:S44-51.
- [14] Ibrahim MA, Saviz LA, Carey TS, Wagner EH. Population-based health principles in medical and public health practice. *Journal of Public Health Management and Practice* 2001;7:75-81.
- [15] 이지혜. 20년 뒤엔 국민 10%가 당뇨병환자 조선일보 2007년 11월 9일.
- [16] 질병관리본부 국립보건연구원 유전체 연구부. 한국인 유전체 역학 조사 사업 기초성과보고서 2004 제1기 지역사회 코호트 연구(안성안산). 서울:질병관리본부, 2004.
- [17] 조남한. 우리나라 당뇨병의 유병률과 관리 상태. *대한내과학회지* 2005;68(1):10 -17.
- [18] Hogan P, Dall T, Nikolov P. Economic Costs of Diabetes in the U.S. in 2002. *Diabetes Care* 2003;26:917- 932.
- [19] McCall, D. T., J. E. Reusch, A. Sauaia, P. Barton, and R. F. Hamman. "Are low-income elderly patients at risk for poor diabetes care?" *Diabetes Care* 2004;27:1060-1065.
- [20] 최정수. 고혈압·당뇨관리와 흡연 및 음주행태의 관련성 분석연구. *보건사회연구* 2007;27(1):103-130.
- [21] Ko GT, Chan JC, Yeung VT, Chow CC, Tsang LW, Cockram CS: A low socio-economic status is an additional risk factor for glucose intolerance in high risk Hong Kong Chinese. *Eur J of Epidemiology* 2001;17:289-95.
- [22] Nathan DM, Davidson MB, DeFronzo RA, Heine RJ, Henry RR, Pratley R, Zinman B: Impaired fasting glucose and impaired glucose tolerance : implications for care. *Diabetes care* 2007;30:753-9.



김 유 미(Yoo-Mi Kim)

[정회원]



- 2000년 2월 : 인제대학교 보건대학원 보건학과(보건학석사)
- 2008년 2월 : 인제대학교 일반대학원 보건학과(보건학박사)
- 1995년 3월 ~ 2004년 10월 : 인제대학교부산백병원 의무기록실 직원/실장
- 2004년 11월 ~ 현재 : 한국보건산업진흥원 연구원

<관심분야>

보건정보, 의무기록, 의료질관리

박 일 수(Il-Su Park)

[정회원]



- 2003년 2월 : 인제대학교 일반대학원 데이터정보학(데이터정보학석사)
- 2009년 2월 : 인제대학교 일반대학원 보건학과 (보건학 박사 수료)
- 2003년 6월 ~ 현재 : 국민건강보험공단 건강보험연구원 주임 연구원

<관심분야>

보건학, 데이터마케팅

장 동 민(Dong-Min Jang)

[정회원]



- 1996년 2월 : 서울대학교 보건대학원 보건학과(보건학박사)
- 1996년 3월 ~ 현재 : 인제대학교 보건행정학부 부교수

<관심분야>

보건행정, 병원행정

강 성 흥(Sung-Hong Kang)

[정회원]



- 1990년 2월: 서울대학교 보건대학원 보건관리학과 (보건학석사)
- 1997년 2월: 인제대학교 일반대학원 보건학과 (보건학박사)
- 1998년 3월 ~ 현재: 인제대학교 보건행정학부 부교수

<관심분야>

보건정보, 의무기록, 데이터마케팅, 건강증진

김 성 수(Sung-Su Kim)

[정회원]



- 2001년 8월 : 인제대학교 보건대학원 보건학과 (보건학석사)
- 2007년 3월 ~ 현재 : 인제대학교 대학원 보건학과(박사과정)
- 2001년 7월 ~ 현재 : 서울아산병원 의무기록팀

<관심분야>

보건통계, 의무기록