

당뇨병 유병률의 지역 간 변이 요인

김 유 미
(상지대학교)

조 대 곤
(포항공과대학교)

강 성 흥*
(인제대학교)

본 연구는 당뇨병 유병률의 지역 간 변이 요인을 규명하여 지역별 특성에 맞는 당뇨병 관리 사업을 지원하기 위한 기초자료를 제공하기 위해 수행되었다. 이를 위해 질병관리 본부의 2011년도 시군구 지역사회건강조사 230건 자료를 지역별 사회경제학 지표 등과 연계하여 생태학적 연구를 위한 자료를 구축한 후 단계적 회귀분석, 의사결정나무 등의 기법으로 분석하였다. 단계적 회귀분석 결과 인지된 고혈압 유병률과 경제활동 비율이 높고, 비만율이 낮을수록 당뇨병 유병률이 높아지는 것으로 나타났다. 지역 간에 상이한 변이 요인을 보다 구체적으로 알아보고자 의사결정나무 모형을 이용하여 지역 간 변이 요인을 규명한 결과 인지된 고혈압 유병률, 비만율, 고위험 음주율, 유배우자 비율, 인구 밀도 등이 당뇨병 유병률의 주요 변이 요인으로 나타났다. 당뇨병 유병률의 지역 간 변이 요인은 세부 지역별로 다양한 양상으로 나타났는데, 본 연구 결과는 지역별 맞춤형 당뇨병 관리 사업 계획을 수립하는데 기초자료로 활용될 수 있겠다.

주요용어: 당뇨병, 지역 간 변이, 지역사회건강조사, 생태학적 방법론, 데이터마ining

* 교신저자: 강성흥, 인제대학교(hcmkang@inje.ac.kr)

■ 투고일: 2014.7.22 ■ 수정일: 2014.9.7 ■ 게재확정일: 2014.9.16

I. 서론

우리나라 당뇨병 유병률은 1998년 남자 9.5%, 여자 7.9%이던 것이 2012년에는 남자 10.1%, 여자 8.0%로 증가하였다. 2012년 당뇨병 미인지율은 29.3%, 비치료율은 37.1%로 당뇨병이 제대로 관리되지 못하고 있다(보건복지부, 2013). 또한 당뇨병 환자 중 혈당조절목표(HbA1c<6.5%)에 도달한 환자는 30% 미만이다. 급속한 고령화 추세에 따라 당뇨병 환자는 지속적으로 증가할 것으로 예상된다. 2010년에 320만 명 수준이던 당뇨병 환자는 2050년에는 약 600만 명으로 크게 증가할 것으로 전망된다(대한당뇨병학회, 2012). 당뇨병 유병률의 증가는 심각한 합병증을 유발하여 삶의 질을 저하시키고, 더 나아가 사망에 이르게 할 수 있다. 또한 사회적으로는 의료비 증대 등 여러 가지 문제를 야기할 수 있으므로 국가차원에서 당뇨병을 효과적으로 관리하는 것은 필수적이라 할 수 있다(김신곤 외, 2009). 일반적으로 국가차원의 당뇨병 관리사업은 지역사회 중심으로 이루어지는 것이 바람직하다. 지역주민에 대한 접근성과 지역주민들의 특성과 상황에 맞는 맞춤형 사업을 수행하는 것이 당뇨병을 관리하는데 효과적이기 때문이다. 이러한 필요성에 따라 각 지역 보건소에서는 당뇨병 환자 등록사업, 국민건강보험공단의 각 지사에서는 당뇨병 환자에 대한 사례관리 사업이 이루어지고 있다(신은영 등, 2003; 박경민 등, 2010). 그러나 우리나라의 현행 당뇨병관리 사업은 지역적 특성이 사업계획과 수행에 체계적으로 반영되지 못하고 있다. 이를 개선하기 위해서는 당뇨병 유병률의 지역별 변이와 이의 요인을 체계적으로 분석할 필요가 있다. 우리나라의 시도별 당뇨병 유병률은 전북 7.5%, 서울 9.7%, 인천 11.0%로 차이가 있는데 이와 같은 지역 간 변이 요인을 규명하는 연구는 거의 없는 실정이다(대한당뇨병학회, 2012). 특히 지역사회의 당뇨병 관리사업이 사군구 단위를 중심으로 이루어진다고 볼 때, 사군구 기준의 당뇨병 유병률의 변이와 이의 요인을 규명하는 연구가 필요하다.

당뇨병 유병률의 지역별 변이요인을 규명하기 위해서는 성, 연령, 소득수준, 거주 지역의 도시화정도 등의 환경 요인과 당뇨병의 위험요인인 유전적 소인, 운동, 영양, 비만, 고혈압, 이상지질혈증 등과 같은 개인들의 건강정보가 체계적으로 구축되어 있어야 한다. 특히 정확한 분석을 위해서는 전국적으로 통일된 기준 하에 각 지역의 대표성을 보장할 수 있는 표본규모를 유지해야 하는데 이와 같은 자료들이 체계적으로 축적되고 분석에 활용되는 것은 미흡한 실정이었다. 미국과 독일 등 선진국에서도 최근에 이르러

서야 당뇨병에 대한 지역적 특성과 분포를 파악하는 자료를 구축하고 방법론을 개발하여 연구를 수행하고 있다(Barker et al., 2011; Schipf et al., 2012). 이러한 여건 하에서 당뇨병 유병률의 변이 요인을 규명하기 위해서는 지역별로 당뇨병 유병률에 영향을 끼칠 수 있는 다양한 지표를 수집하여 변이 요인을 규명하는 생태학적 연구 방법론을 도입하여 활용할 필요가 있다. 물론 생태학적 방법에 의해 지역별 당뇨병 유병률의 변이 요인을 규명하는 것은 실제 주민들의 개인자료를 수집하여 분석하는 전통적인 역학적 방법론에 비해 신뢰도에 문제가 있다는 지적이 있다. 그러나 개인의 특성과 집단간 생태학적 특성을 일치시키는 것이 오류일 수 있고 질병관리 정책을 수립하는데 기초자료로 활용할 수 있다는 측면에서는 집단의 건강수준과 건강행태의 영향요인은 생태학적 접근에 따라 분석하는 것이 의미가 있다 할 수 있다(Kasl, 1979; Schwarz et al., 1994; Macintyre et al., 2000; Oggioni et al., 2014).

생태학적 방법으로 당뇨병 유병률의 지역 간 변이 요인을 규명하는 연구에 있어서 전통적인 다중회귀분석을 이용하는 것은 많은 제한점이 있다. 지역별 변이 요인 규명의 주목적이 지역 특성에 맞는 당뇨병 관리 사업을 유도하기 위한 것이라고 볼 때 다중회귀 분석은 국가 전체의 위험요인 및 중요도만을 알려주기 때문에 지역차원의 당뇨병 관리 사업에 활용하는 데에는 한계가 있다. 당뇨병의 지역별 변이에 관한 기존 연구들은 단편적인 분석기법에 의존하여 당뇨병의 변이를 분석함에 따라 지역별 특성에 맞는 당뇨병 관리 사업방안 수립에 활용하는 데 있어서 제한점이 있다(Akerblom et al., 1985; Gupta et al., 2007; 김동현 외, 2010; Barker et al., 2011; Schipf et al., 2012). 이러한 한계점을 극복하고 지역 간 상이한 요인을 도출하는데 의사결정나무모형과 같은 데이터마이닝 기법을 활용할 수 있다.

본 연구에서는 당뇨병 유병률의 지역별 변이 요인을 규명하기 위해 당뇨병 유병률의 지역별 변이 정도를 파악하고, 다양한 방법론을 활용한 당뇨병 유병률의 지역 간 변이 요인 분석을 수행하였다. 이를 기반으로 지역사회 중심의 맞춤형 당뇨병 관리 사업이 효과적으로 이루어지도록 하기 위한 방안과 기초자료를 제시하고자 한다.

II. 연구방법

1. 자료 수집

본 연구는 질병관리본부의 2011년 지역사회건강조사와 통계청 인구주택총조사, 국토해양부 자료를 이용하였다. 지역사회건강조사 자료는 2008년부터 매년 전국 시·군·구 단위의 지역주민 건강행태, 만성질환 이환 및 의료이용 등을 표본가구에 대해 지역 보건소와 대학에서 조사하는 체계적이고 신뢰성있는 지역기반의 대표적인 보건통계이다. 특히 지역사회건강조사 결과자료는 지역 간 비교가 가능하도록 조사방법, 조사내용을 표준화하였으며, 주요 보건지표에 대해서는 성, 연령을 직접표준화 방법에 따라 보정한 표준화 지표를 제공하고 있음에 따라 성, 연령의 영향을 제외한 다른 요인에 의한 당뇨병 유병률의 변이를 분석하기에 적합한 자료이다(김영택 외, 2012).

2011년 지역사회건강조사 자료를 230개 시·군·구를 기준으로 하여 당뇨병 평생 의사진단 경험률과 이에 관련된 위험요인이 될 수 있는 고혈압 평생 의사진단 경험률, 흡연율, 중등도 신체활동률, 비만율 등의 자료를 추출하였다. 추출한 자료에 사회경제학적인 지표인 유배우자 비율, 경제활동비율, 인구밀도 자료를 타 자료원에서 수집하여 추가 하였다. 유배우자 비율과 경제활동비율은 통계청 인구주택총조사 자료에서 수집하였고, 인구밀도 자료는 국토해양부 통계 자료에서 수집하였다.

2. 변수 정의

본 연구에서 사용된 변수의 정의는 이환 및 의료이용, 건강행태, 정신건강, 사회인구학적 특성 등으로 구성되며 <표 1>과 같다. 주요 변수는 인지된 고혈압 유병률, 인지된 당뇨병 유병률, 흡연율, 고위험음주율, 중등도 신체활동률, 비만율, 스트레스 인지율, 우울증 경험률, 주관적 건강상태인지율, 유배우자비율, 경제활동비율, 인구밀도이다.

변수 중 고혈압 유병률과 당뇨병 유병률은 지역사회건강조사의 고혈압·당뇨병 평생 의사진단 경험률로서 의사에게 고혈압·당뇨병을 진단받은 30세 이상 성인의 분율(%)로 인지된 고혈압·당뇨병을 나타내는 지표이다. 우리나라의 당뇨병 유병률은 10.1%,

당뇨병 인지율은 73.4%로서 10명 중 3명이 본인이 당뇨병이 있다는 사실을 모르고 있다(대한당뇨병학회, 2013). 본 조사자료인 2011년도 지역사회건강조사 당뇨병 평생 의사진단 경험률은 7.0%로서 대한당뇨병학회의 통계와 일치하는 결과를 보여준다. 본 연구에서는 고혈압당뇨병 평생 의사진단 경험률을 인지된 고혈압당뇨병 유병률로 정의하였다.

본 연구에서 성, 연령에 관한 변수는 사용하지 않았는데 지역사회건강조사의 지표인 고혈압 유병률, 당뇨병 유병률 및 각종 건강행태 지표는 성, 연령을 보정한 지표이기 때문이다. 분석 목적에 따라 구간척도인 각 지표를 그룹화 하였다. 그룹화의 기준은 사분위수에서 상위 25% 이상을 ‘높음’, 하위 25% 미만을 ‘낮음’, 나머지는 ‘보통’으로 그룹화 하였다. 당뇨병 유병률과 고혈압 유병률의 경우 상위 25%는 ‘높음’, 그 나머지는 ‘높지 않음’으로 그룹화 하였다.

표 1. 변수정의

변수명	정의	그룹화	자료원
인지된 당뇨병 유병률	의사에게 당뇨병을 진단받은 30세 이상 성인의 분율(%)로 인지된 당뇨병을 나타내는 지표	사분위수 기준 상위 25%는 높음, 하위 75%는 높지 않음	질병관리 본부 지역사회 건강조사 (2011)
인지된 고혈압 유병률	의사에게 고혈압을 진단받은 30세 이상 성인의 분율(%)로 인지된 고혈압을 나타내는 지표	사분위수 기준 낮음은 하위 25% 미만, 높음은 상위 25% 이상, 나머지는 보통	
흡연을	평생 5갑 이상 흡연한 사람으로서 현재 흡연하는 사람(“매일피움” 또는 “가끔피움”)의 분율(%)		
고위험 음주율	최근 1년 동안 음주한 사람 중에서 남자는 한 번의 술자리에서 7잔 이상, 여자는 5잔 이상을 주 2회 이상 마신다고 응답한 사람의 분율(%)		
중등도 신체활동률	최근 1주일 동안 격렬한 신체활동을 1회 20분 이상, 주 3일 이상 실천한 사람 또는 최근 1주일 동안 중등도 신체활동을 1회 30분 이상, 주 5일 이상 실천한 사람의 분율(%)		
비만율	체질량 지수가 25 이상인 사람의 분율(%)		

변수명	정의	그룹화	자료원
스트레스 인지율	평소 일상생활 중 스트레스를 “대단히 많이” 또는 “많이” 느끼는 사람의 비율(%)	사분위수 기준 낮음은 하위 25%미만, 높음은 상위 25%이상. 나머지는 보통	질병관리 본부 지역사회 건강조사 (2011)
우울증 경험률	최근 1년 동안 연속적으로 2주 이상 일상생활에 지장이 있을 정도의 우울감을 경험한 사람의 비율(%)		
주관적 건강상태 인지율	주관적 건강수준이 “매우 좋음” 또는 “좋음”이라고 응답한 사람의 비율(%)		
유배우자 비율	$(\text{기혼자} \div \text{15세 이상 내국인}) \times 100$		인구주택 총조사
경제활동 비율	$(\text{경제활동인구수} \div \text{15세 이상 인구}) \times 100$		
인구밀도	지역별 인구/면적(명/km ²)	국도해양 부통계	

3. 분석 방법

분석대상 지역의 당뇨병 유병률 변이 정도를 알 수 있는 기술통계를 살펴본 후, 지역별 당뇨병 유병률의 변이 요인을 파악하기 위해 상관관계분석, 다중회귀분석, 의사결정나무(decision tree) 분석을 실시하였다. 다중회귀분석은 당뇨병 유병률의 전국적인 요인(즉, 다른 요인을 통제한 후 각 요인이 당뇨병 유병률에 미치는 평균 수준)은 알려주지만, 지역별로 어떠한 변수가 두드러진 영향을 주는 요인인지는 확인하기 어렵다. 이를 보완할 수 있는 방법으로 데이터마이닝의 의사결정나무모형을 이용하면 하위 계층으로 내려가며 각 계층 내에서 주요 요인을 도출해낼 수 있기 때문에 각 지역별 당뇨병 유병률의 변이 요인을 확인할 수 있다. 의사결정나무모형은 특정한 분류 기준에 따라 목표변수와 가장 관련성이 높은 독립변수를 선정한 후 의사결정규칙(decision rule)을 몇 개의 소집단으로 분류하여 나무구조로 표현하는 것으로 요인의 규명, 분류, 예측에 유용하다. 즉 독립변수간의 관계를 도식화하여 보여주기 때문에 연구자가 분석과정을 쉽게 이해하고 설명할 수 있는 장점을 가지고 있다(Liu et al., 2007).

분석 시 다중회귀분석은 구간척도를 그대로 사용하였고, 데이터마이닝에 이용한 분석은 변수를 그룹화한 자료를 사용하였다. 데이터마이닝을 이용한 모형개발의 평가지표는 Lift값을 이용하였다. 통계분석 및 데이터마이닝은 SAS 9.2(SAS Institute, North Carolina, USA)를 이용하였으며 통계적 유의성의 판단기준은 $\alpha=0.05$ 로 하였다.

III. 연구 결과

1. 분석대상 지역의 일반적 특성

전국 230개 시군구의 고혈압당뇨병 유병률과 건강행태 특성 등은 지역적 차이가 큰 것으로 나타났다. 인구밀도 외에 특히 우울증경험률(43.8%), 중등도 신체활동률(41.3%), 고위험음주율(20.8%)의 변동계수는 다른 변수들의 변동계수보다 높은 것으로 나타나 지역 간 건강행태에 상당한 차이가 남을 알 수 있었다. 당뇨병 유병률은 평균 7.2%이었으며 경남 함안군이 최소 4.2%, 경기 양주시가 최대 10.2%로 6.0%p까지 격차를 보이는 것으로 나타났다.

표 2. 분석대상 지역의 일반적 현황

변수	평균	표준 편차	변동 계수	최소값	최대값
인지된 당뇨병 유병률(%)	7.2	1.1	15.3	4.2	경남 함안군 10.2 경기 양주시
인지된 고혈압 유병률(%)	18.7	2.3	12.3	12.8	경북 군위군 25.6 강원 영월군
흡연율(%)	24.9	2.6	10.4	17.5	경기 과천시 32.4 전북 진안군 강원 태백시
고위험 음주율(%)	18.3	3.8	20.8	5.5	전남 장흥군 29.0 인천 강화군
중등도 신체활동률(%)	24.0	9.9	41.3	11.6	전북 익산시 67.7 전남 영암군
비만율(%)	23.4	3.0	12.8	14.6	전남 장흥군 32.7 강원 정선군
스트레스 인지율(%)	27.3	4.4	16.1	10.9	전남 장흥군 36.6 경남 산청군
우울증 경험률(%)	4.8	2.1	43.8	0.4	전남 진도군 11.4 서울 구로구

변수	평균	표준 편차	변동 계수	최소값		최대값	
				지역명	지역명	지역명	지역명
주관적 건강상태 인지율(%)	48.1	6.9	14.3	34.9	경북 칠곡군	72.5	전남 고흥군
유배우자 비율(%)	60.5	4.4	7.3	48.1	서울 관악구	70.1	충남 계룡시
경제활동 비율(%)	57.9	5.1	8.8	45.6	부산 영도구	74.0	전남 신안군
인구밀도(명/km ²)	4037.8	6440.9	159.5	19.4	강원 인제군	28735.7	서울 양천구

2. 지역별 당뇨병 유병률의 차이

지역별 당뇨병 유병률 차이를 보다 자세히 알아보기 위해 시군구 지역을 당뇨병 유병률의 사분위 계수를 기준으로 높은 지역(상위 25%), 낮은 지역(하위 25%), 보통 지역(상위 25~75%)으로 분류하였을 때, 경기 14개(45.2%) 시군, 인천 4개(40.0%) 구, 대구 3개(37.5%) 구, 충북 4개(33.3%) 시군 순으로 당뇨병 유병률이 높은 지역으로 분류되었다. 반대로 당뇨병 유병률이 낮은 지역(하위 25%)의 비율이 높은 곳은 경남 9개(50.0%) 시군, 광주 2개(40.0%) 구, 경북 8개(34.8%) 시군 순이었다 (제주 제외).

표 3. 지역별 당뇨병 유병률

시도 (시군구수)	높은 지역(상위25%)		낮은 지역(하위25%)		보통 지역(상위25~75%)	
	N(%)	시군구	N(%)	시군구	N(%)	시군구
서울(25)	7 (28.0)	종로구, 성동구, 동대문구, 중랑구, 은평구, 서대문구, 강서구	6 (24.0)	강북구, 노원구, 구로구, 영등포구, 관악구, 송파구	12 (48.0)	중구, 용산구, 광진구, 성북구, 도봉구, 마포구, 양천구, 금천구, 동작구, 서초구, 강남구, 강동구
부산(16)	5 (31.3)	영도구, 남구, 해운대구, 사하구, 사상구	5 (31.3)	중구, 서구, 동래구, 연제구, 수영구	6 (37.5)	동구, 부산진구, 북구, 금정구, 강서구, 기장군
대구(8)	3 (37.5)	중구, 서구, 남구	1 (12.5)	북구	4 (50.0)	동구, 수성구, 달서구, 달성군
인천(10)	4 (40.0)	동구, 남구, 부평구, 서구	2(20.0)	강화군, 옹진군	4 (40.0)	중구, 연수구, 남동구, 계양구

시도 (시군구수)	높은 지역(상위25%)		낮은 지역(하위25%)		보통 지역(상위25~75%)	
	N(%)	시군구	N(%)	시군구	N(%)	시군구
광주(5)	0 (0.0)	-	2(40.0)	남구, 광산구	3 (60.0)	동구, 서구, 북구
대전(5)	1 (20.0)	동구	0(0.0)	-	4 (80.0)	중구, 서구, 유성구, 대덕구
울산(5)	0 (0.0)	-	1 (20.0)	동구	4 (80.0)	중구, 남구, 북구, 울주군
경기(31)	14 (45.2)	수원시, 의정부시, 광명시, 동두천시, 구리시, 남양주시, 파주시, 이천시, 화성시, 양주시, 포천시, 여주군, 연천군, 양평군	4 (12.9)	안양시, 과천시, 시흥시, 용인시	13 (41.9)	성남시, 부천시, 평택시, 안산시, 고양시, 오산시, 군포시, 의왕시, 하남시, 안성시, 김포시, 광주시, 가평군
강원(18)	5 (27.8)	속초시, 평창군, 정선군, 철원군, 고성군	3 (16.7)	강릉시, 양구군, 양양군	10 (55.6)	춘천시, 원주시, 동해시, 태백시, 삼척시, 홍천군, 횡성군, 영월군, 화천군, 인제군
충북(12)	4 (33.3)	청주시, 보은군, 음성군, 증평군	1 (8.3)	제천시	7 (58.3)	충주시, 청원군, 옥천군, 영동군, 진천군, 괴산군, 단양군
충남(16)	5 (31.3)	천안시, 논산시, 연기군, 서천군, 청양군	4 (25.0)	아산시, 계룡시, 금산군, 태안군	7 (43.8)	공주시, 보령시, 서산시, 부여군, 홍성군, 예산군, 당진군
전북(14)	3 (21.4)	정읍시, 김제시, 부안군	4 (28.6)	익산시, 남원시, 장수군, 순창군	7 (50.0)	전주시, 군산시, 완주군, 진안군, 무주군, 임실군, 고창군
전남(22)	4 (18.2)	목포시, 담양군, 곡성군, 함평군	6 (27.3)	순천시, 고흥군, 강진군, 해남군, 완도군, 신안군	12 (54.5)	여수시, 나주시, 광양시, 구례군, 보성군, 화순군, 장흥군, 영암군, 무안군, 영광군, 장성군, 진도군

시도 (시군구수)	높은 지역(상위25%)		낮은 지역(하위25%)		보통 지역(상위25~75%)	
	N(%)	시군구	N(%)	시군구	N(%)	시군구
경북(23)	4 (17.4)	안동시, 영주시, 청도군, 울진군	8 (34.8)	포항시, 경주시, 군위군, 의성군, 청송군, 고령군, 성주군, 울릉군	11 (47.8)	김천시, 구미시, 영천시, 상주시, 문경시, 경산시, 영양군, 영덕군, 칠곡군, 예천군, 봉화군
경남(18)	1 (5.6)	양산시	9 (50.0)	통영시, 밀양시, 거제시, 함안군, 고성군, 산청군, 함양군, 거창군, 합천군	8 (44.4)	진주시, 사천시, 김해시, 창원시, 의령군, 창녕군, 남해군, 하동군
제주(2)	0 (0.0)		1 (50.0)	제주시	1 (50.0)	서귀포시
전체 (230)	60 (26.1)		57 (24.8)		113 (49.1)	

3. 다중회귀분석을 이용한 당뇨병 유병률의 변이요인

1) 상관분석

당뇨병 유병률과 관련 변수들 간의 상관분석을 한 결과, 고혈압 유병률(0.45), 비만율(0.29), 스트레스 인지율(0.21), 흡연율(0.18), 우울증 경험률(0.15), 고위험 음주율(0.14), 유배우자 비율(-0.15), 주관적 건강상태 인지율(-0.16), 경제활동 비율(-0.19) 순으로 유의한 상관관계가 있는 것으로 나타났다.

표 4. 당뇨병 유병률과 관련 변수들간의 상관분석

	상관계수	p
인지된 고혈압 유병률	0.45	<0.0001
흡연율	0.18	0.0078
고위험 음주율	0.14	0.0342
중등도 신체활동률	-0.07	0.2791

	상관계수	p
비만율	0.29	<0.0001
스트레스 인지율	0.21	0.0011
우울증 경험률	0.15	0.02
주관적 건강상태 인지율	-0.16	0.0131
유배우자 비율	-0.15	0.0242
경제활동 비율	-0.19	0.0034
인구밀도	0.05	0.4084

2) 단계적 다중회귀분석

단계적 다중회귀분석을 실시한 결과 고혈압 유병률(0.18)과 경제활동 비율(0.06)이 높고, 비만율(-0.033)이 낮을수록 당뇨병 유병률이 높아지는 것으로 나타났다($R^2=0.22$). 일반적으로 비만과 당뇨병 유병률은 양의 상관관계를 가지는데 비해 본 연구는 대조적인 결과를 나타냈었다.

표 5. 다중회귀분석을 이용한 당뇨병 유병률의 변이요인 분석

	회귀계수	p
상수	4.505	<0.0001
고혈압 유병률	0.180	<0.0001
경제활동 비율	0.060	0.0267
비만율	-0.033	0.0140

주: Adjusted $R^2 = 0.22$

4. 데이터마이닝을 이용한 당뇨병 유병률의 변이요인

1) 교차분석

데이터마이닝 분석을 위해 모든 변수를 그룹화하여 교차분석을 우선 실시하였다.

당뇨병 유병률을 높음(상위 25%)과 높지 않음(하위 75%)로 구분하여 제 특성별 교차분석을 실시한 결과 고혈압 유병률($p = 0.0001$), 비만율($p = 0.0027$), 유배우자 비율($p = 0.027$), 고위험 음주율($p = 0.0322$), 스트레스 인지율($p = 0.0377$)에서 유의한 차이가 나타났다.

표 6. 제 특성에 따른 당뇨병 유병률

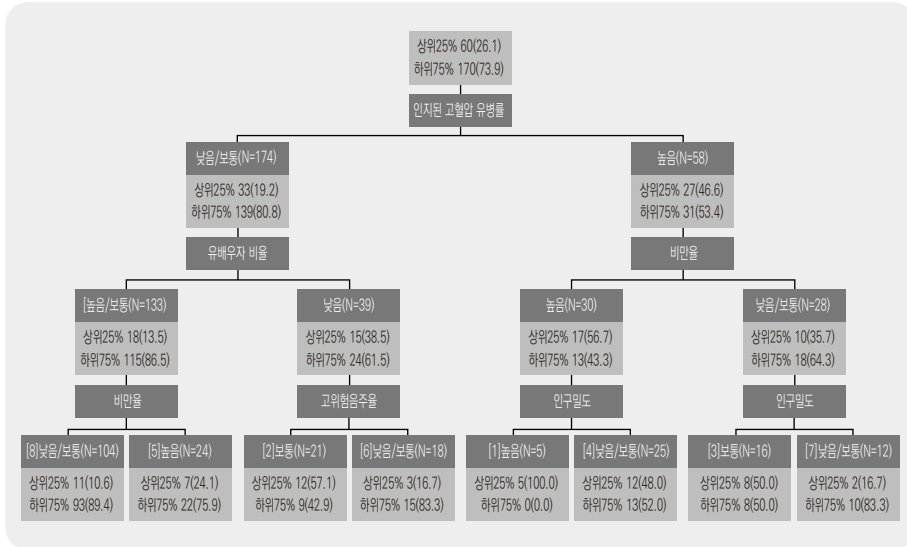
구분		높음 (상위 25%)		높지 않음 (하위 75%)		전체		p
		N	%	N	%	N	%	
인지된 고혈압 유병률	낮음	5	8.9	51	91.1	56	100.0	0.0001
	보통	28	24.1	88	75.9	116	100.0	
	높음	27	46.5	31	53.5	58	100.0	
흡연율	낮음	8	14.0	49	86.9	57	100.0	0.0518
	보통	33	28.9	81	71.1	114	100.0	
	높음	19	32.2	40	67.8	59	100.0	
고위험 음주율	낮음	7	12.7	48	82.3	55	100.0	0.0322
	보통	36	31.3	79	68.7	115	100.0	
	높음	17	28.3	43	71.7	60	100.0	
중등도 신체활동율	낮음	14	24.6	43	75.4	57	100.0	0.5900
	보통	33	28.9	81	71.1	114	100.0	
	높음	13	22.0	46	78.0	59	100.0	
비만율	낮음	6	11.0	48	89.0	54	100.0	0.0027
	보통	30	26.1	85	73.9	115	100.0	
	높음	24	39.3	37	60.7	61	100.0	
스트레스 인지율	낮음	10	17.6	47	82.5	57	100.0	0.0377
	보통	28	24.3	87	75.7	115	100.0	
	높음	22	37.9	36	62.1	58	100.0	
우울증 경험률	낮음	10	17.9	46	82.1	56	100.0	0.2534
	보통	32	27.8	83	72.2	115	100.0	
	높음	18	30.5	41	69.5	59	100.0	

구분		높음 (상위25%)		높지 않음(하위 75%)		전체		p
		N	%	N	%	N	%	
주관적 건강상태 인지율	낮음	19	34.6	36	65.5	55	100.0	0.0992
	보통	31	26.7	85	73.3	116	100.0	
	높음	10	17.0	49	83.1	59	100.0	
유배우자 비율	낮음	21	39.6	32	60.4	53	100.0	0.0270
	보통	27	24.1	85	75.9	112	100.0	
	높음	12	18.4	53	81.6	65	100.0	
경제활동 비율	낮음	19	33.3	38	66.7	57	100.0	0.0587
	보통	32	28.3	81	71.7	113	100.0	
	높음	9	15.0	51	85.0	60	100.0	
인구밀도	낮음	11	19.6	45	80.4	56	100.0	0.0546
	보통	27	23.5	88	76.5	115	100.0	
	높음	22	37.9	36	62.1	58	100.0	

2) 의사결정나무 기반의 당뇨병 유병률 변이 요인

의사결정나무 모형을 분석한 결과 당뇨병 유병률의 변이에 영향을 미치는 변수는 인지된 고혈압 유병률, 비만율, 유배우자 비율, 고위험 음주율, 인구밀도 등으로 나타났다. 인지된 고혈압 유병률, 비만율, 인구밀도가 높은 시군구 5개는 모두 당뇨병 유병률이 높은 상위 25% 지역에 속했다(노드 번호 [1]). 또 인지된 고혈압 유병률이 보통 이하이면서 유배우자 비율이 낮고 고위험 음주율은 보통인 시군구 21개 중 12개(57.1%)가 당뇨병 유병률 상위 25% 지역에 속했다. 반면 인지된 고혈압 유병률이 보통 이하, 유배우자 비율이 보통 이상, 비만율이 보통 이하인 시군구 104개 중 11개(10.6%)만이 당뇨병 유병률이 높은 상위 25% 지역에 속했다(노드 번호 [8]).

그림 1. 의사결정나무 기반의 당뇨병 유병률 변이요인



주: []는 표 7의 노드 번호로 당뇨병 유병률 상위 25% 지역 비율이 높은 순임

표 7. 의사결정나무 기반의 당뇨병 유병률이 높은 지역의 특성

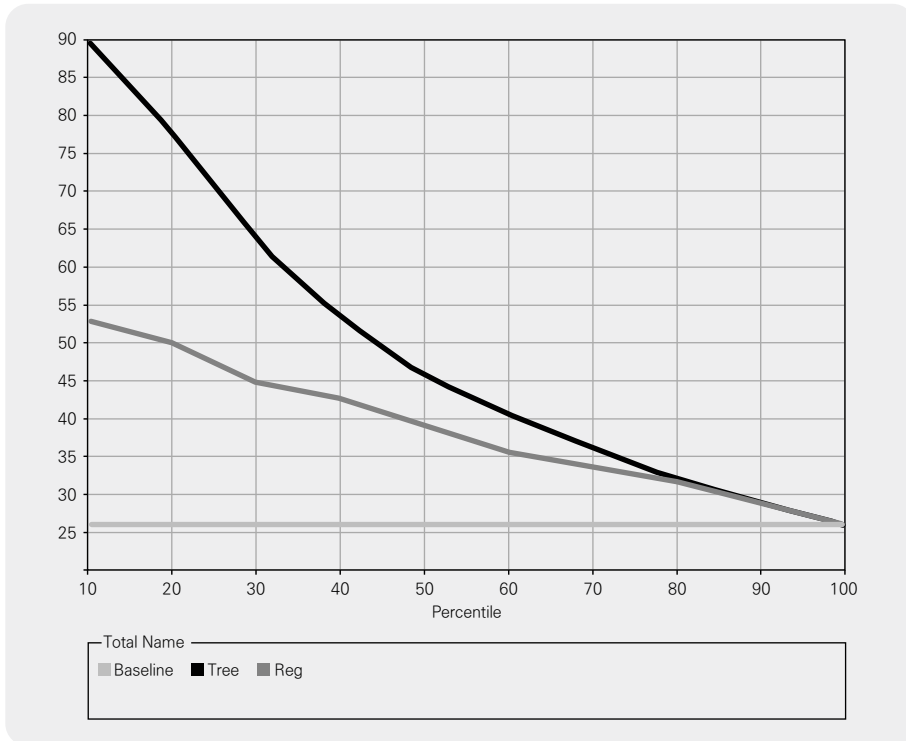
노드 번호 (지역수)	노드의 특성	당뇨병 유병률이 높은 지역(상위 25%)	
		시군구	지역수(비율)
1 (5)	인지된 고혈압 유병률(높음)-비만율(높음)- 인구밀도(높음)	서울 동대문구, 중랑구, 서대문구 인천 동구 경기 의정부시	5(100.0)
2 (21)	인지된 고혈압 유병률 (보통 이하)-유배우자비율(낮음)- 고위험음주율(보통)	서울 성동구, 은평구, 강서구 부산 남구, 사상구 대구 중구, 서구, 남구 대전 동구 경기 수원시 충북 청주시 충남 천안시	12(57.1)
3 (16)	인지된 고혈압 유병률(높음)- 비만율(보통 이하)-인구밀도(보통)	인천 서구 경기 남양주시, 파주시, 이천시, 화성시 강원 속초시 충남 서천군 전북 김제시	8(50.0)

노드 번호 (지역수)	노드의 특성	당뇨병 유병률이 높은 지역(상위 25%)	
		시군구	지역수(비율)
4 (25)	인지된 고혈압 유병률(높음)-비만율(높음)-인구밀 도(보통 이하)	경기 양주시,포천시,여주군,연천군 강원 평창군,정선군,고성군 충북 보은군,음성군 충남 논산시,연기군,청양군	12(48.0)
5 (29)	인지된 고혈압 유병률(보통 이하)-유배우자 비율(보통 이상)-비만율(높음)	경기 동두천시,양평군 전북 정읍시,부안군 전남 목포시,함평군 경북 안동시	7(24.1)
6 (18)	인지된 고혈압 유병률(보통 이하)-유배우자비율(보통 이상)-고위험음주비율(높음, 낮음) *상위 25% 지역 3개는 모두 높음	서울 종로구 부산 영도구 인천 남구	3(16.7)
7 (12)	인지된 고혈압 유병률(높음)-비만율(보통 이하)-인구밀도(높음, 낮음)	인천 부평구 강원 철원군	2(16.7)
8 (104)	인지된 고혈압 유병률(보통 이하)-유배우자 비율(보통 이상)-비만율(보통 이하)	부산 해운대구,사하구 경기 광명시,구리시 충북 증평군 전남 담양군,곡성군 경북 영주시,청도군,울진군 경남 양산시	11(10.6)

3) 모형평가

목표변수인 당뇨병 유병률을 ‘높은 지역’과 ‘높지 않은 지역’으로 이원화하고, 독립변수는 높음, 보통, 낮음 등 세 개의 그룹으로 분류하여 회귀 모형과 의사결정나무 모형을 구축한 후 모형평가를 한 결과 의사결정나무 모형의 Lift 그래프가 로지스틱 회귀 모형보다 우수한 것으로 나타났다.

그림 2. 당뇨병 유병률 변이 요인 모형의 평가



주: Lift Chart 중 %Response 도표로 당뇨병 유병률이 높은 상위 10% 지역군에서 회귀모형은 53%, 의사결정나무 모형은 90%의 당뇨병 유병률을 예측하여 의사결정나무 모형이 회귀모형에 비해 예측력이 높은 모형임

IV. 고찰 및 결론

본 연구는 지역의 특성과 요구에 맞는 당뇨병 관리 사업을 지원하기 위해 생태학적 방법론을 이용하여 당뇨병 유병률의 지역 간 변이요인을 규명하고자 하였다. 당뇨병 유병률과 상관 관계가 있는 지역 간 변이 요인은 고혈압 유병률, 비만율, 스트레스 인지율, 흡연율, 우울증 경험률, 고위험 음주율, 유배우자 비율, 주관적 건강상태 인지율, 경제활동 비율 등으로 나타났다. 이 중 고혈압 유병률, 비만율, 스트레스 인지율, 고위험 음주율, 유배우자 비율 등이 교차분석에서 유의한 변이 요인으로 나타났다. 그러나 단계

적 회귀분석에서 유의한 변이 요인으로 나타난 고혈압 유병률, 비만율, 경제활동 비율 중 비만율과 경제활동 비율은 단변량 분석과 상반된 결과를 보였다. 즉 단변량 분석에서 비만율은 당뇨병 유병률과 양의 관계, 경제활동 비율은 음의 관계가 나타났지만 다변량 분석 결과에서는 반대의 결과가 나타났다. 회귀모형에서 건강행태 변수는 상호 다중공선성을 가질 위험이 있고, 따라서 지역주민에 대한 원시 데이터가 아닌 간접 지표를 활용한 본 연구와 같은 생태학적 연구 방법은 전통적인 OLS(ordinary least squares) 방법론이 한계점을 가진다고 할 수 있다. 반면 의사결정나무 모형은 비선형 판별 모형(non-linear discrimination method)으로 생태학적 연구를 적용하기에 회귀모형보다 적합하다. 이러한 결과는 본 연구의 데이터마이닝을 이용한 모형평가 결과에서도 나타났다. 또한 의사결정나무 모형을 이용하면 지역 간 상이한 당뇨병 유병률 변이요인을 도출해낼 수 있기 때문에 본 연구목적에 부합한다.

의사결정나무 모형을 통해 살펴본 당뇨병 유병률의 지역 간 변이를 결정짓는 주요 변수는 고혈압 유병률, 비만율, 고위험 음주율, 유배우자 비율, 인구밀도 등으로 나타났다. 질병관리본부의 지역 간 건강행태의 변이요인에 대한 연구에서도 고혈압 평생 의사진단 경험률이 높을수록 당뇨병 평생 의사진단 경험률이 높은 것으로 보고되었다(김동현 외, 2010). 당뇨병 진단시점에 고혈압이 동반되어 있는 경우가 21.3%이고 당뇨병 진단 후 10년 정도 지나면 대부분의 환자에게 이상지질혈증과 고혈압 등이 동반되는 것으로 보고된 바 있기도 하다(김신곤 외, 2009). 또한 당뇨병 평생 의사진단 경험률은 건강검진 수검률과 관련이 있는데(김동현 외, 2010) 건강검진 수검률이 사회경제적 여건 및 교육수준과 관련이 있는 것으로 볼 때(여지영 외, 2012), 경기인천 지역의 당뇨병 유병률이 다른 지역에 비해 높은 것은 이를 반영한다고 추정할 수 있다. 즉 당뇨병 평생 의사진단 경험률이 당뇨병 유병률을 얼마나 반영하는 지표가 되는지 지역별 당뇨병 인지율을 확인할 필요가 있다.

고혈압 외에 비만, 고위험음주, 유배우자비율, 인구밀도가 지역 간 당뇨병 유병률의 차이를 결정짓는 요인으로 나타났는데 질병관리본부의 회귀모형에서 비만, 음주는 유의한 차이가 없는 것으로 나타나 본 연구와 대조되는 결과를 보였다(김동현 외, 2010). 그러나 최근 미국 및 유럽의 연구에서 비만은 지역 간 당뇨병 유병률의 변이의 주요 요인으로 밝혀졌다(Barker et al., 2011; Tamayo et al., 2014). 또 고혈압 유병률과 비만 수준이 높은 지역 중에서도 인구밀도가 높은 지역이 당뇨병 유병률이 높았는데 인구밀

도는 도시화(urbanization)의 정도를 반영하는 지표라 할 수 있다. 도시화는 식생활, 비만 및 신체활동 등 라이프스타일과 관련이 있으며 도시화 정도가 높은 지역이 당뇨병 유병률도 높다(Gupta et al., 2007; Barker et al., 2011). 인지된 고혈압 유병률이 보통 수준 이하인 지역 중 유배우자비율이 낮고 고위험음주율은 보통인 지역과 유배우자 비율이 보통 이상이면서 고위험음주율은 높은 지역의 당뇨병 유병률이 높은 것으로 나타났다. 본 연구는 지역 단위의 생태학적 연구로 직접적 비교는 제한점이 있지만 최근 덴마크의 한 연구에서 고위험 음주는 남자, 유배우자 비율은 여자의 경우 당뇨병 유병률과 관계가 있는 것으로 보고하고 있다(Molsted et al., 2014). 이를 통해 볼 때, 본 연구의 결과는 대체로 주요 선행연구의 결과들과 맥락을 같이한다고 여겨진다.

한편 광역자치단체 기준으로는 경기 지역이 다른 지역보다 당뇨병 유병률이 가장 높은 것으로 나타났지만, 그 안에서 시군구별로 세분화하였을 때에는 당뇨병 유병률이 낮은 지역과 보통 지역이 다양하게 분포되어 있었다. 이 변이를 바탕으로 본 연구의 의의는 제한적이지만 지역별로 상이한 당뇨병 유병률 요인을 도출해낸 데에 있다. 시군구별 지역 특성은 의사결정나무 모형 기반으로 보다 세분화하여 확인을 할 수 있었는데, 고혈압 유병률, 비만율, 인구밀도가 높은 서울 동대문구, 중랑구, 서대문구, 인천 동구, 경기 의정부시 등 5개 지역은 모두 당뇨병 유병률이 높은 지역에 속했다. 또한 인구밀도는 보통 이하이지만 고혈압 유병률과 비만율이 높은 25개 지역 중 경기 양주시, 포천시, 여주군, 연천군, 강원 평창군, 정선군, 고성군, 충북 보은군, 음성군, 충남 논산시, 연기군, 청양군 등 12개 지역 역시 당뇨병 유병률이 높은 지역에 포함되었다. 이상의 30개 지역은 고혈압과 비만 관리 사업을 집중적으로 지원해야 할 것으로 추론해 볼 수 있다.

또한 인지된 고혈압 유병률은 보통 이하, 고위험 음주율은 보통이면서 유배우자 비율이 낮은 21개 지역 중 서울 성동구, 은평구, 강서구, 부산 남구, 사상구, 대구 중구, 서구, 남구, 대전 동구, 경기 수원시, 충북 청주시, 충남 천안시 등 12개 지역이 당뇨병 유병률이 높은 지역에 포함되었다. 이 지역은 독신 가구 비율이 높아 지역사회 지지 필요하며 정신보건 사업과 연계하여 당뇨병 관리 사업을 시행하는 것이 필요할 것으로 파악된다. 또 인지된 고혈압 유병률이 보통 이하이고 유배우자 비율이 보통 이상이지만 고위험 음주비율이 높거나 낮은 18개 지역 중 서울 종로구, 부산 영도구, 인천 남구는 모두 고위험 음주비율이 상위 25%에 속한 지역으로 당뇨병 유병률이 높은 지역에 속했다. 이들 지역은 금주와 절주에 대한 보건교육의 실시를 통해 만성질환 발생을 예방하고

당뇨병 관리사업과 연계한다면 효과적으로 고혈압 유병율을 낮출 수 있을 것으로 추론된다. 반면 인지된 고혈압 유병률과 비만율이 보통 이하이면서 유배우자 비율이 보통 이상인 104개 지역 중 부산 해운대구, 사하구, 경기 광명시, 구리시, 충북 증평군, 전남 담양군, 곡성군, 경북 영주시, 청도군, 울진군, 경남 양산시 등 11개 지역만이 당뇨병 유병률 상위 그룹에 속해 있었는데, 이들 지역은 의사결정나무에서 밝혀진 변이 요인 외의 다른 요인의 기여 가능성을 고려하여야 한다.

성공적인 지역보건의료 정책 수행과 이를 통한 지역주민의 건강향상을 위해서는 지역의 특성과 요구에 맞는 ‘지역보건의료계획’의 수립과 시행이 필요하고 이를 가능하게 하는 국가적 지원전략이 필요하다(질병관리본부, 2013). 지역사회 차원의 당뇨병 관리 사업은 이와 같이 지역적 특성을 고려하여 선택, 집중을 통해 위험요인에 대한 예방활동을 수행할 수 있다. 당뇨병 관리 사업은 교육을 통해 이루어지는데 유배우자 비율이 낮은 지역에서는 특히 자조관리 프로그램(self-help control)과 같은 지지집단(support group) 프로그램을 통해 당뇨병 교육을 실시할 필요가 있다(김은영, 2009). 당뇨병은 식이, 운동, 음주, 흡연, 스트레스, 투약 등에 대한 지속적인 지지가 필요한 질환이므로 지역사회에서 개인별 맞춤형 평생 교육 프로그램을 개발하고 지원할 필요가 있다.

본 연구는 대표적인 만성질환 중 하나인 당뇨병의 지역 간 변이에 대한 분석을 생태학적 연구방법론과 데이터마이닝 기법을 이용하여 분석하여 지역별 맞춤형 당뇨병 관리 사업 계획을 수립할 수 있는 자료를 제공하였다. 그러나 이용 가능한 지표가 제한됨에 따라 충분한 변이 요인을 고려하지 못한 데 제한점이 있다. 따라서 본 연구 결과를 기초로 향후 다양한 후속 연구가 필요하다. 첫째, 본 연구에서는 2011년도 조사 결과만을 분석하였는데 다른 연도 조사결과를 추가하고 고려할 수 있는 변이 요인을 추가하여 본 연구 결과의 유효성을 검증하고 새로운 변이 요인을 도출하는 것이 필요하다. 둘째, 향후 개인 데이터 단위 연구를 본 연구의 지역 단위 결과와 결합하면 지역별 변이 요인 뿐만 아니라 인구통계학적 변이 요인을 규명하여 보다 효과적으로 지역 단위 당뇨병 관리 계획 수립에 활용될 수 있을 것이다. 셋째, 본 연구에서 활용한 데이터마이닝 기법의 일환인 의사결정나무 모형과 별도로 지리적 가중 회귀분석(geographically weighted regression)이나 다층모형(multi-level modeling) 등의 방법론을 활용해 지역 간 변이 요인을 체계적으로 분석할 필요가 있다.

김유미는 인제대학교에서 보건학 석·박사학위를 받았으며, 현재 상지대학교 의료경영학과의 교수로 재직 중이다. 주요 관심분야는 보건정보관리, 의료의 질관리 등이다. 현재 지역별 보건사업 수립방안 및 평가 등에 대해서 연구 중이다. (E-mail: ymkim@sangji.ac.kr)

조대곤은 미국 콜롬비아대학에서 산업공학 석사를, 카네기멜론대학에서 경영정보학으로 박사학위를 받고, 현재 포스텍 산업경영공학과 교수로 재직 중이다. 주요 관심분야는 경영정보, 의료정보 등이다. 현재 건강보험 플랫폼 설계 등에 대해서 연구 중이다. (E-mail: dgcho@postech.ac.kr)

강성홍은 서울대학교에서 보건학 석사를, 인제대학교에서 보건학 박사학위를 받고, 현재 인제대학교 보건행정학과 교수로 재직 중이다. 주요 관심분야는 의료정보, 의료의 질관리 등이다. 현재 병원간 벤치마킹 시스템, 지역보건 사업 수립방안 등에 대해서 연구 중이다. (E-mail: hcmkang@inje.ac.kr)

참고문헌

- 김동현 등(2010). 2008년 지역사회건강조사 자료를 이용한 지역 간 건강행태 변이요인에 대한 분석연구. 오송: 질병관리본부.
- 김신곤, 최동섭(2009). 우리나라 당뇨병의 역학과 현황. *Hanyang Medical Reviews*, 29(2), 122-129.
- 김영택, 최보율, 이계오, 김호, 전진호, 김수영, 외(2012). 지역사회건강조사의 조사 기획과 수행. *대한의사협회지*, 55(1), pp.74-83.
- 김은영, 소애영(2009). 당뇨병 환자를 위한 지역사회 중심 자조관리 프로그램의 효과. *지역사회간호학회지*, 20(3), pp.307-315.
- 대한당뇨병학회, 건강보험심사평가원(2007). 당뇨병 기초통계연구 Task Force Team 보고서 *Diabetes in Korea 2007*. 서울.
- 대한당뇨병학회화당뇨병학연구재단·질병관리본부-International Conference on Diabetes and Metabolism(2012). *Diabetes Fact Sheet in Korea*. available from http://www.diabetes.or.kr/temp/Diabetes_Fact_sheet2012.pdf (accessed 2014.05.18 인출.)
- 박경민, 김정남, 박명화, 김혜련, 신아(2010). 보건소에서 실시한 당뇨병 환자 방문간호 프로그램의 효과. *한국보건간호학회지*, 24(1), pp.71-81.
- 보건복지부·질병관리본부(2013). 2012 건강행태 및 만성질환 통계: 국민건강영양조사 제 5기 3차년도(2012) 및 제8차(2012) 청소년건강행태온라인조사. 오송: 질병관리본부.
- 신은영, 김철환, 유원섭, 김희걸, 김창엽(2003). 지역사회 중심의 당뇨 사례관리사업 효과. *지역사회간호학회지*, 14(4), pp.559-567.
- 여지영, 정형선(2012). 건강검진 수검의 결정요인 및 건강증진행위 변화 효과. *보건행정학회지*, 22(1), pp.49-64.
- 질병관리본부(2013). 2008-2012 지역건강통계 한눈에 보기. 질병관리본부.
- Akerblom, H. K., Reunanen, A. (1985). The epidemiology of insulin- dependent diabetes mellitus (IDDM) in finland and in northern europe. *Diabetes Care*, 8 Suppl 1, pp.10-16.
- Barker, L.E., Kirtland, K. A., Gregg, E.W., Geiss, L. S., Thompson, T. J. (2011).

- Geographic distribution of diagnosed diabetes in the U.S.: a diabetes belt. *Am J Prev Med*, 40, pp.434-439.
- Gupta, R., Misra, A. (2007). Review: Type 2 diabetes in india: *Regional disparities. The British Journal of Diabetes & Vascular Disease*, 7(1), pp.12-16.
- Kasl, S. (1979). Mortality and the business cycles: some questions about research strategies when utilizing macro-social and ecological data. *Am J Pub Health*, 69, pp.784-8.
- Liu, Y., Salvendy G. (2007). Visualization support to better comprehend and improve decision tree classification modelling process: *A survey and appraisal. Theoretical Issues in Ergonomics Science*, 8(1): pp.63-92.
- Macintyre, S., Ellaway, A. (2000). *Ecological approaches: Rediscovering the role of the physical and social environment*. In *Social Epidemiology*, Edited by Berkman and Kawachi: Oxford University Press.
- Molsted, S., Johnsen, N. F., Snorgaard, O. (2014). Trends in leisure time physical activity, smoking, body mass index and alcohol consumption in danish adults with and without diabetes: A repeat cross-sectional national survey covering the years 2000 to 2010. *Diabetes Research and Clinical Practice*, In Press, Corrected Proof, Available online 29 May 2014.
- Oggioni, C., Lara, J., Wells, J. C. K., Soroka, K., Siervo, M. (2014). Shifts in population dietary patterns and physical inactivity as determinants of global trends in the prevalence of diabetes: An ecological analysis. *Nutrition, Metabolism and Cardiovascular Diseases*, In Press, Corrected Proof, Available online 24 May 2014.
- Schipf, S., Werner, A., Tamayo, T., Holle, R., Schunk, M., Maier, W., et al. (2012). Regional differences in the prevalence of known type 2 diabetes mellitus in 45-74 years old individuals: Results from six population-based studies in germany (DIAB-CORE consortium). *Diabetic Medicine : A Journal of the British Diabetic Association*, 29(7), pp.e88-95.
- Schwartz, S. (1994). The fallacy of the ecological fallacy: the potential misuse of

a concept and the consequences. *Am J Pub Health*, 84(5), pp.819-824.

Tamayo, T., Rosenbauer, J., Wild, S. H., Spijkerman, A. M. W., Baan, C., Forouhi, N. G., et al. (2014). Diabetes in europe: An update. *Diabetes Research and Clinical Practice*, 103(2), pp.206-217.

An Empirical Analysis on Geographic Variations in the Prevalence of Diabetes.

Kim, Yoo Mi
(Sangji University)

Cho, Dae Gon
(POSTECH)

Kang, Sung Hong
(Inje University)

We investigate how the regional prevalence of diabetes is affected by health-related and socioeconomic factors with a special emphasis on geographic variations. We focus on the likelihood of diabetes as function of various region-specific attributes. We construct a unique set of data at the level of 230 small administrative district collected from 2011 Annual Community Health Survey by Korea Centers for Disease Control and Prevention and other government agencies. To estimate, we use several methods including correlation analysis, multiple regression and decision tree model. We find that diabetes prevalence is more likely to be associated with hypertension prevalence, obesity and economic activity rate. Further findings using decision tree model suggest that hypertension prevalence, obesity, the rate of drinking and the density of population are more likely to affect the prevalence of diabetes than other regional attributes considered. More importantly, we find significant geographic variations in factors affecting diabetes prevalence across administrative districts. This topic are not just of academic interest, but have practical implications by helping policy makers to understand and identify important regional factors relating to the prevalence of diabetes by which they can implement more effective planning for promotion of health.

Keywords: Diabetes Prevalence, Korean Community Health Survey, Geographic Variations, Ecological Study, Decision Tree Model