

Canonical correlation between body information and lipid-profile: A study on the National Health Insurance Big Data in Korea

Han-Gue Jo*, Young-Heung Kang*

*Assistant Professor, School of Computer Information and Communication Engineering, Kunsan National University, Gunsan, Korea

*Professor, School of Computer Information and Communication Engineering, Kunsan National University, Gunsan, Korea

[Abstract]

This study aims to provide the relevant basis upon which prediction of dyslipidemia should be made based on body information. Using the National Health Insurance big data (3,312,971 people) canonical correlation analysis was performed between body information and lipid-profile. Body information included age, height, weight and waist circumference, while the lipid-profile included total cholesterol, triglycerides, HDL cholesterol and LDL cholesterol. As a result, when the waist circumference and the weight are large, triglycerides increase and HDL cholesterol level decreases. In terms of age, weight, waist circumference, and HDL cholesterol, the canonical variates (the degree of influence) were significantly different according to sex. In particular, the canonical variate was dramatically changed around the forties and fifties in women in terms of weight, waist circumference, and HDL cholesterol. The canonical correlation results of the health care big data presented in this study will help construct a predictive model that can evaluate an individual's health status based on body information that can be easily measured in a non-invasive manner.

▶ **Key words:** Healthcare big-data, Canonical correlation, lipid-profile, Dyslipidemia, Sex difference

[요 약]

본 연구에서는 2009-2016 보건의료 빅데이터를 활용하여 신체 정보와 혈액 내 지질 농도의 연관성을 다변량 분석하여 주요 요소들 사이의 영향력 정도를 비교 분석함으로써, 한국인의 이상지질혈증 예측 모델 개발을 위한 근거를 제시하고자 한다. 3,312,971 명의 건강검진정보 자료를 통해서 신체 정보(나이, 신장, 체중, 허리둘레) 항목과 혈액검사(총콜레스테롤, 중성지방, HDL 콜레스테롤, LDL 콜레스테롤) 항목 간의 다차원적 선형상관관계를 도출하고, 항목 간의 영향력을 정준변량(canonical variate)으로 분석하였다. 그 결과 허리둘레가 크고 체중이 많이 나가면 중성지방이 높아지고 HDL 콜레스테롤 수치가 낮아지는 다차원적 상관관계를 도출하였다. 또한, 나이, 체중, 허리둘레, HDL 콜레스테롤 항목은 그 영향력 정도가 성별에 따라 유의미한 차이를 보였다. 특히, 나이에 따른 영향력의 정도는 체중, 허리둘레, HDL 콜레스테롤 항목이 40-50 전후 여성에게서 뚜렷한 변화를 보였다. 보건의료 빅데이터를 정준상관 분석 기법을 적용하여 분석한 다차원적 상관관계는 비침습적인 방법으로 간편하게 측정 가능한 신체 정보를 바탕으로 건강 상태를 평가할 수 있는 예측모델을 개발하는 데 활용될 수 있을 것이다.

▶ **주제어:** 보건의료 빅데이터, 정준상관, 신체 정보, 이상지질혈증, 성별 차이

• First Author: Han-Gue Jo, Corresponding Author: Han-Gue Jo

*Han-Gue Jo (hgjo@kunsan.ac.kr), School of Computer Information and Communication Engineering, Kunsan National University

*Young-Heung Kang (yhkang@kunsan.ac.kr), School of Computer Information and Communication Engineering, Kunsan National University

• Received: 2020. 11. 12, Revised: 2020. 12. 23, Accepted: 2020. 12. 26.

I. Introduction

이상지질혈증은 지방 대사의 조절 이상으로, 혈관 내 지질 또는 지방 성분의 양이 비정상적으로 변환된 상태로, 협심증, 심근경색, 뇌졸중 등 합병증을 초래할 수 있는 혈관질환이다. 우리나라 30세 이상에서 이상지질혈증 유병률은 2005년에 8%에서 2013년에는 15.9%로 증가하였고, 2016년에는 22.1%로 보고되어 지속해서 증가하는 추세이다[1, 2]. 한편, 세계보건기구(WHO)에 따르면 세계 주요 사망원인 1, 2위는 심혈관질환, 뇌졸중 및 기타 뇌혈관질환으로 혈관 계통 질환 때문인 것으로 알려져 있다[3]. 이에 따라 심혈관질환 및 뇌혈관질환의 원인으로 알려진 이상지질혈증의 조기 발견 및 예방, 위험성 관리에 대한 중요성이 높아지고 있다.

이상지질혈증에 대한 진단은 보편적으로 받아들여지고 있는 미국의 NCEP (National Cholesterol Education Program)에 따라 총콜레스테롤(total cholesterol), 저밀도지단백(LDL, low density lipoprotein) 콜레스테롤, 고밀도지단백(HDL, high density lipoprotein) 콜레스테롤, 중성지방(triglyceride) 등의 지질 지표를 통해 진단하고 있다[4-6]. 이 지표들을 정상 수치와 비교함으로써 간단하게 진단할 수 있으나 국가와 인종, 성별, 나이 등의 차이를 고려하지 않고 정상 범위를 일괄적으로 적용하기에는 어려움이 있다[7-11]. 또한, 혈액검사를 동반해야 하고 채혈 전 가벼운 음식물 섭취도 검사 결과에 영향을 줄 수 있으며, 진단받기까지 시간과 비용이 뒤따른다. 따라서, 비침습적인 방법으로 일상생활에서 간편하게 측정 가능한 정보로 혈액 내 지질 지표를 판단할 수 있는 한국인의 고유한 신체적 특성이 반영되고 성별과 나이를 고려한 이상지질혈증 예측모델 개발의 중요성이 강조되고 있다.

한국인을 대상으로 한 기존 연구에서는 신체질량지수(body mass index), 허리둘레-신장 비율, 허리둘레 등 신체 정보가 체내 지질 수치 등 혈관질환 위험 요소와 연관성이 있다고 보고했다[7, 12-13]. 하지만, 위 특정 지표가 정상 범위라 하더라도 체내 지방량이 다를 수 있고 여러 혈관질환 위험인자가 발생할 수 있기 때문에[14], 주요 지표들을 다차원적으로 고려하여 각 지표의 영향력 정도를 상호비교할 수 있는 빅데이터를 기반으로 한 다변량 분석 연구가 필요하다.

한국 정부는 국민의 건강 향상, 의료 질 및 보건의료 정책 개선을 목표로 보건의료 빅데이터를 공개하고 있다. 특히, 매년 건강검진을 수진한 국민건강보험가입자 100만 명을 무작위로 선별하여 그들의 건강검진정보를 공개하고 있는

데(National Health Insurance Sharing Service)[15], 나이, 성별, 신장, 체중, 허리둘레 등의 신체 정보와 총콜레스테롤, LDL 콜레스테롤, HDL 콜레스테롤, 중성지방 등 이상지질혈증을 진단하는 지표들을 포함한다. 이에 본 연구는 국민건강보험공단의 건강검진정보 자료를 이용하여 빅데이터를 기반으로 신체 정보와 혈액 내 지질 지표들 사이의 다차원적 연관성을 분석하여, 한국인의 이상지질혈증 예측모델 개발을 위한 근거를 제시하고자 한다. 이를 위해 비침습적인 방법으로 일상생활에서 간편하게 측정 가능한 다변량 신체 정보와 혈액검사를 통해 측정된 체내 지질 지표들 사이의 다차원적 상호 연관성 정도를 정준상관(canonical correlation) 기법으로 분석하였다.

II. Methods

1. Subjects

‘국가중점 개방데이터’는 정부 3.0 정책 기조에 발맞추어 국민건강보험공단이 보유하고 있는 건강검진정보 데이터를 개방하고 있다(국민건강보험공단 빅데이터운영실)[15]. 본 연구에서는 2009년부터 2016년까지 매년 건강검진을 수진한 국민건강보험가입자 중 100만 명을 무작위로 선별한 건강검진정보 자료를 이용하였다. 총 34개의 항목 중에서 신체 정보와 혈액검사 항목에 해당하는 8개의 항목을 추출하였다(표 1). 각 8개 항목에 대해서 데이터를 왜곡할 수 있는 정규분포에서 벗어나는 데이터값(평균에서 표준편차의 세 배 이상)을 보이는 검진자는 제외했으며, 흡연 및 음주자도 연구에서 제외하고, 각 항목을 Z-score 정규화하였다. 결과적으로 3,312,971명(남성 1,131,442명, 여성 2,181,529명)의 건강검진정보 자료를 본 연구에 사용하였다. 여성 2,181,529명 중 20(20~29세)대, 30대, 40대, 50대, 60대, 70세 이상은 각각 194,790명, 254,960명, 565,691명, 564,260명, 363,750명, 238,078명이었으며, 남성 1,131,442명 중 각 연령대는 85,560명, 193,899명, 243,563명, 270,188명, 200,305명, 137,927명이었다.

2. Research model and methods

본 연구를 위한 연구모형은 그림 1과 같다. 신체 정보(나이, 신장, 체중, 허리둘레)와 혈액검사(총콜레스테롤, 중성지방, HDL 콜레스테롤, LDL 콜레스테롤) 항목들 사이에 유의한 다차원적 연관성이 있을 것으로 가정하였고(H1), 각 항목의 영향력이 성별에 따라 차이가 있을 것으로 예측하였다. 다변수 대 다변수, 즉 두 변수 집단의 연관

성을 분석하기 위해 정준상관분석(canonical correlation analysis) 기법을 적용하였다. 정준상관분석은 여러 변수로 구성된 두 집단의 관계를 분석할 때 사용되는 통계기법으로 1930년대에 처음으로 소개되었으나[16], 복잡한 다변량의 분석 결과를 해석하기가 쉽지 않고 높은 계산력을 요구하기 때문에 다른 통계기법에 비해 비교적 덜 사용돼온 측면이 있다. 하지만 최근 들어서 빅데이터의 관심이 높아지고 있고, 여러 변수를 다차원적으로 분석하는 다변수 대 다변수의 관계, 즉 다변량 분석의 중요성이 강조됨에 따라 여러 학문 분야에서 새롭게 관심을 받고 있는 빅데이터 분석을 위한 통계기법이다[17].

Table 1. Body information and dyslipidemia factors

	Indices (unit of measure)			
Body indices	age group (1~14)			
		age group		age group
	1	20~24 years	8	55~59 years
	2	25~29 years	9	60~64 years
	3	30~34 years	10	65~69 years
	4	35~39 years	11	70~74 years
	5	40~44 years	12	75~79 years
	6	45~49 years	13	80~84 years
	7	50~54 years	14	85~ years
		height (5 cm)		
	weight (5 kg)			
	waist circumference (cm)			
Dyslipidemia factors	total cholesterol (mg/dL)			
	triglycerides (mg/dL)			
	HDL cholesterol (mg/dL)			
	LDL cholesterol (mg/dL)			

연구결과의 일관성을 확인하기 위해 위 연구모형을 2009년부터 2016년도까지 연도별, 성별, 연령별로 분석하였으며, 자료 분석을 위해 매트랩 R2019b (MathWorks, Inc) 프로그램을 사용하였다. 정준상관분석으로 도출된 정준상관계수(r)의 통계적 유의성은 순열검정법으로 검증하였다. 검진자의 신체 정보 항목을 무작위로 재정렬(shuffle)하고 항목의 정렬 순서가 혈액검사 항목과 랜덤 배치되게 하여 랜덤 정준상관계수(r_rand)를 10,000번 반복하여 계산하고 추출된 정규분포에서 정준상관계수(r)의 정규성을 검증하였다.

통계적 유의성은 $p < 0.05$ 로 정의하였으며, 다중 비교에서 생길 수 있는 오류는 Bonferroni 교정하였다. 따라서, 성별에 따른 신체 정보와 혈액검사 지표의 통계적 유의성을 $\alpha = 0.05/8$ (8개 항목)으로 교정하고, 교정된 p 값을 보고했다. 모든 지표는 평균 \pm 표준편차로 나타냈다.

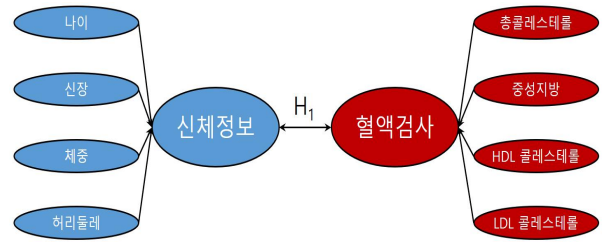


Fig. 1. Research Model

III. Results

각 항목에 대해 2009년부터 2016년까지 연도별 평균의 전체평균 나이 그룹(표1 참조)은 남성(6.92 ± 0.52)과 여성(6.43 ± 0.90 ; $p \approx 0.999$) 사이에 차이가 없었으나, 신장, 체중, 허리둘레에서 남성(신장 = 167.54 ± 3.29 ; 체중 = 66.65 ± 4.59 ; 허리둘레 = 82.83 ± 1.24)이 여성(신장 = 156.03 ± 3.24 ; 체중 = 55.70 ± 2.26 ; 허리둘레 = 75.85 ± 0.94)보다 높게 나타났다(모든 항목에 대해 $p < 0.001$). 또한, 체중과 신장의 관계로 계산된 체질량지수(BMI, body mass index)는 모든 연도에서 남성(23.7 ± 0.69)이 여성(22.9 ± 0.07)보다 높게 나타났다(uncorrected $p = 0.014$).

총콜레스테롤(남성 = 189.15 ± 2.62 ; 여성 = 193.64 ± 2.35 ; $p = 0.023$)과 HDL 콜레스테롤(남성 = 52.11 ± 1.67 ; 여성 = 59.36 ± 2.93 ; $p < 0.001$)은 남성보다 여성에게서 높게 나타났으며, 중성지방(남성 = 116.19 ± 3.27 ; 여성 = 99.13 ± 6.76 ; $p < 0.001$)은 남성이 높았다. LDL 콜레스테롤은 남성(113.75 ± 0.54)과 여성(114.20 ± 3.74 ; $p \approx 0.999$) 사이에 차이가 없었다 (표 2 참조).

Table 2. General characteristics of study population

Indices	male	female
age group (1~14)	6.92 ± 0.52	6.43 ± 0.90
height*	167.54 ± 3.29	156.03 ± 3.24
weight*	66.65 ± 4.59	55.70 ± 2.26
waist circumference*	82.83 ± 1.24	75.85 ± 0.94
total cholesterol*	189.15 ± 2.62	193.64 ± 2.35
triglycerides*	116.19 ± 3.27	99.13 ± 6.76
HDL cholesterol*	52.11 ± 1.67	59.36 ± 2.93
LDL cholesterol	113.75 ± 0.54	114.20 ± 3.74

Data are expressed as mean \pm standard deviation. *, $p < 0.001$ between male and female. Note, age group indicates the group number unit (1~14) as shown in table 1.

신체 정보와 혈액검사 항목 간의 다차원적 연관성을 정준상관으로 분석하여 가장 높은 정준상관계수(r)를 나타내는 제1 결합함수를 연도별, 성별에 따라 도출하였다. 그 결과 모든 연도 그리고 남성과 여성에서 유의한 정준상관 관계를 보였다($r = 0.401 \pm 0.065$; 범위: 0.335-0.489; 모든 r 값에 대해 $p < 0.001$). 따라서 가정 H1, 즉, 신체 정보 항목 나이, 신장, 체중, 허리둘레는 혈액검사 항목 총콜레스테롤, 중성지방, HDL 콜레스테롤, LDL 콜레스테롤과 유의한 다차원적 연관성을 나타냈다.

그림 2는 위에서 도출된 연도별 정준상관관계에서 한 변수와 다른 집단의 변수들과의 관계를 나타내는 정준변량(canonical variate)을 나타낸다. 정준변량의 수용기준이 명확히 정해진 것은 없으나, 본 연구에서는 그 크기가 최소 |0.35| 수준 이상인 항목에 근거하여 그 영향력을 파악하였다. 그 결과 전체 검진자를 대상으로 할 때 나이, 체중, 허리둘레는 중성지방과 양의 상관 관계를, 반면 HDL 콜레스테롤과는 음의 상관 관계를 나타냈다. 특히, 신체 정보 항목에서는 허리둘레(0.495 ± 0.001), 체중(0.411 ± 0.060), 나이(0.386 ± 0.084) 순으로 그 영향력이 컸으며, 혈액검사 항목에서는 중성지방(0.592 ± 0.028), HDL 콜레스테롤(-0.533 ± 0.042) 순으로 그 영향력이 큰 것으로 나타났다. 신체 정보 항목의 신장(-0.135 ± 0.068)과 혈액검사 항목의 총콜레스테롤(0.210 ± 0.209), LDL 콜레스테롤(0.090 ± 0.126)은 그 영향력이 상대적으로 작은 것으로 나타났다.

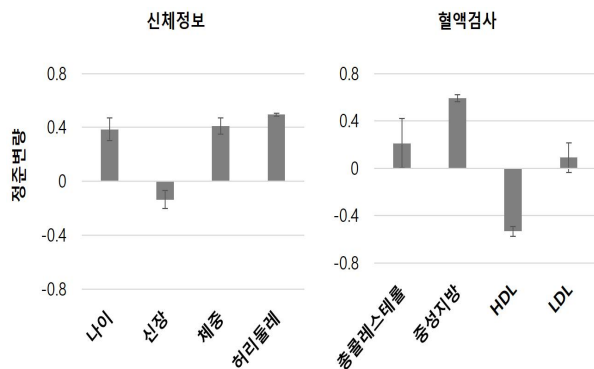


Fig. 2. Average of canonical variate

그림 3은 연도별 평균 정준변량의 성별 차이를 보여준다. 남성의 경우 허리둘레(0.626 ± 0.024), 체중(0.499 ± 0.030), 신장(-0.353 ± 0.016)순으로 신체 정보 항목이 혈액검사 항목의 HDL 콜레스테롤(-0.597 ± 0.050), 중성지방(0.576 ± 0.046)과 상호 연관성이 큰 것으로 나타났다. 반면 여성에서는 나이(0.536 ± 0.018), 허리둘레(0.360 ± 0.022)가 신체 정보 항목에서 큰 연관성을 보였으며, 이는

그 영향력을 거의 나타내지 않은 남성의 나이 항목(0.014 ± 0.034 ; $p < 0.001$)과 크게 대비된다. 허리둘레(0.360 ± 0.0217)와 체중(0.309 ± 0.0447)에서는 남성이 여성보다 더 큰 영향력을 나타냈다(두 항목에 대해 $p < 0.001$). 혈액검사 항목에서 여성은 중성지방(0.578 ± 0.048), HDL 콜레스테롤(-0.452 ± 0.081), 총콜레스테롤(0.452 ± 0.097) 순으로 그 영향력이 큰 것으로 나타났으며, HDL 콜레스테롤에서도 남성과 유의한 차이를 보였다($p=0.009$).

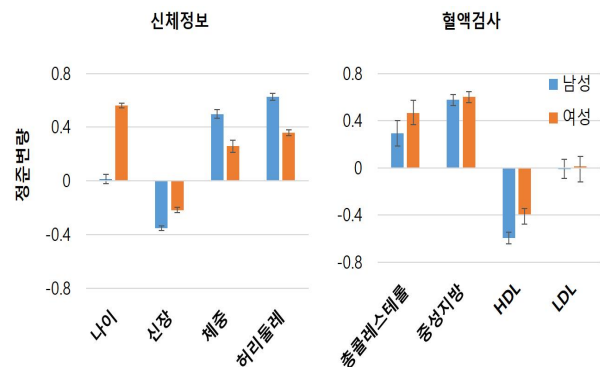


Fig. 3. Sex difference in canonical variate

그림 4는 총 3,312,971명 중 성별과 연령에 따른 신체 정보와 혈액 내 지질의 농도변화를 보여준다. 총콜레스테롤과 LDL 콜레스테롤의 경우 20~40대에서는 남성이 여성보다 높았으며, 50대 이상에서는 여성이 높게 나타났다. 중성지방의 경우 남성은 40~50대에서 가장 높았다가 60대 이상부터 서서히 감소하지만, 여성은 연령대가 증가할수록 점점 증가하는 경향을 보였다. HDL 콜레스테롤의 경우 남녀 모두에게서 연령대가 증가할수록 서서히 감소하는 경향을 보였다. 반면, 신체 정보의 경우 모든 연령대에서 남성이 여성보다 신장, 체중, 허리둘레가 높게 나타났으며, 신장, 체중 항목은 연령대가 증가할수록 서서히 줄어들고 허리둘레 항목은 증가하는 경향을 보였다.

신장, 체중, 허리둘레와 혈액검사 항목 간의 다차원적 연관성을 정준상관으로 분석하여 가장 높은 정준상관계수(r)를 나타내는 제1 결합함수를 성별, 연령에 따라 도출하였으며, 그 결과 모든 연령대 그리고 남성과 여성에서 유의한 정준상관 관계를 보였다($r = 0.312 \pm 0.069$; 범위: 0.168-0.389; 모든 r 값에 대해 $p < 0.001$). 따라서 신체 정보 항목 신장, 체중, 허리둘레는 혈액검사 항목 총콜레스테롤, 중성지방, HDL 콜레스테롤, LDL 콜레스테롤과 유의한 다차원적 연관성을 나타냈다. 하지만, 특히 50대 이상 정준상관분석에서 정준상관계수(r)가 작아지는 경향을 보였다(남성(r): 20대=0.368, 30대=0.340, 40대=0.349,

50대=0.305, 60대=0.290, 70대 이상=0.288; 여성(r): 20대=0.342, 30대=0.389, 40대=0.358, 50대=0.278, 60대=0.212, 70대 이상=0.168).

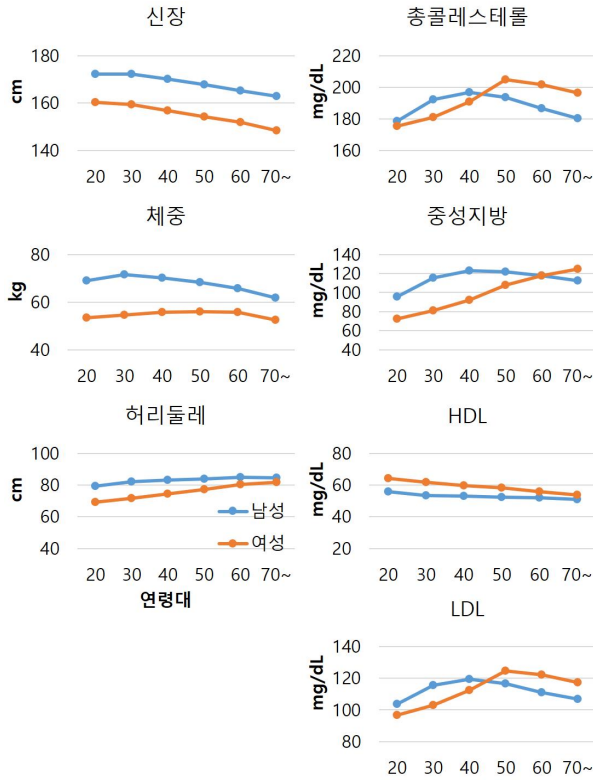


Fig. 4. Characteristics of the population for sex and age

그림 5는 성별과 연령에 따른 정준변량의 변화를 보여준다. 20-40 연령대의 경우 신체정보의 허리둘레, 체중, 신장 항목이 남성과 여성 모두에게서 높은 영향력을 보였으며 (절대값 범위: 0.316 ~ 0.796), 혈액검사 항목의 HDL 콜레스테롤(범위: -0.501 ~ -0.386), 중성지방(범위 0.453 ~ 0.700)과 상호 연관성이 큰 것으로 나타났다. 반면, 여성의 경우 40~50대 전후로 체중과 허리둘레 항목이 급격한 변화를 보여준다. 신체 정보와 혈액검사 항목 사이의 다차원적 연관성에서 체중 항목이 미치는 영향력은 낮아지지만, 허리둘레 항목의 영향력은 높아져서 고연령대일수록 허리둘레 변수의 중요도를 반영한다. 남성의 경우 체중과 허리둘레 항목이 미치는 영향력은 연령대에 따른 뚜렷한 변화를 보여주지 않았다. 남녀 모두에게서 연령대가 증가할수록 신장 항목의 영향력은 서서히 감소하는 경향을 나타냈다.

총콜레스테롤 항목은 연령대가 증가할수록 그 영향력이 점차 감소했으나, 60대 이상 연령에서는 그 영향력이 음의 상관관계를 보여준다. 중성지방의 경우 연령대가 증가할수록 그 영향력이 남녀 모두에게서 점점 증가하여, 고연령대

에서 중성지방 항목의 중요도를 나타낸다. 60대 이상 연령대에서 HDL 콜레스테롤 항목의 영향력은 여성의 경우 급격히 낮아졌으며, LDL 콜레스테롤의 경우 50대 남성에게서 그 영향력이 낮아지는 양상을 보였다.

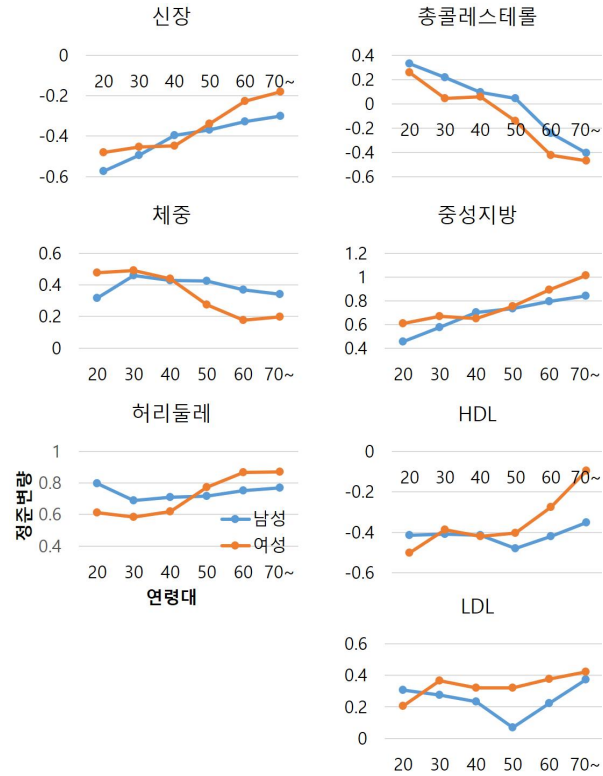


Fig. 5. Canonical variate for sex and age

IV. Discussion

본 연구는 보건의로 빅데이터 분석으로 비침습적 방법으로 일상생활에서 쉽게 측정 가능한 신체 정보와 침습적 방법으로 측정할 혈액 내 지질 수치 지표 간의 다차원적 연관성을 규명하고, 이를 통해 이상지질혈증의 예측인자로서 신체 정보 항목 간의 영향력 정도를 확인하였다.

20세 이상 한국 성인의 평균 신장, 체중, 허리둘레는 남성이 여성보다 높았다. 평균 총콜레스테롤과 HDL 콜레스테롤은 남성보다 여성에게서 높게 나타났으나 중성지방은 남성에게서 높게 나타났다. 이와 같은 성별의 차이는 선행 연구의 결과와 일치하며, 소아청소년에게서도 동일하게 관찰되었다[13, 18].

한편, 혈액 내 세부 지질의 농도변화는 성별과 연령에 따라 다른 양상을 보였다. 40~50대를 전후로 남성과 여성의 총콜레스테롤, LDL 콜레스테롤 지표가 역전되었다. 중

성지방의 경우 남성은 40~50대에서 가장 높았다가 60대 이상부터 서서히 감소하지만, 여성은 연령대가 증가할수록 점점 증가하는 경향을 보였다. 이러한 양상은 이상지질혈증 관련 위험요인이 성별과 연령에 따라 차별적으로 영향을 미치고 있음을 보여준다[8-10].

2009년부터 2016년까지 모든 연도에서 신체 정보는 혈액 내 지질 지표와 높은 상관성 나타났다. 특히, 신체 정보 항목에서 허리둘레, 체중, 나이순으로 그 영향력이 큰 것으로 나타났으며, 이는 혈액 내 지질 지표인 중성지방과 HDL 콜레스테롤 수치와 높은 상관성을 보였다. 허리둘레가 크고 체중이 많이 나가고 나이가 많을수록 중성지방 수치는 높아지고 HDL 콜레스테롤 수치가 낮아지는 상관성을 보였다. 신체 정보 항목 중에서 허리둘레 지표가 체내 지질 성분과의 상관성에서 가장 큰 영향을 나타냈는데, 이는 중성지방과 정상관계 HDL 콜레스테롤과는 음의 상관계를 보인 2015년도 국민건강영양조사 자료를 이용한 선행연구의 결과에 의해 뒷받침된다[19, 20].

위와 같은 상관관계는 남녀 모두에게서 관찰되었다. 혈액검사 항목에서 중성지방과 HDL 콜레스테롤이 남녀 모두에서 높은 영향을 보였지만, 신체 정보에서는 남녀 차이가 뚜렷이 구별됐다. 특히, 신체 정보의 나이 항목이 뚜렷한 남녀 차이를 보였다. 여성에게서는 높은 영향력 있는 것으로 관찰됐으나 남성에게는 그 영향이 미미하였다. 즉, 나이에 따른 혈액 내 지질 농도의 변화는 여성에게서 더 뚜렷했으며, 이러한 변화는 이상지질혈증 유병률에도 영향이 있을 것으로 추측된다. 실제로도 이와 같은 성별의 차이는 선행연구에서 관찰되었다. 국민건강영양조사(2010-2014년)에 참여한 건강한 성인을 대상으로 한 연구에서 여성(12,370명)은 20대 10.7%, 30대 17.4%, 40대 25.1%, 50대 42.6%, 60대 54.1%, 70대 이상 54.3%의 이상지질혈증의 유병률을 보인 반면, 남성(8,615명)은 20대 28.2%, 30대 43.1%, 40대 49.2%, 50대 56.9%, 60대 54.0%, 70대 이상 46.9%의 이상지질혈증의 유병률을 보였다[13]. 즉 남성과 비교하면 여성은 20대에는 낮았으나 40-50대 이후 급격한 증가 양상을 보였다. 이러한 나이에 따른 남녀 차이는 본 연구의 콜레스테롤과 LDL 콜레스테롤 변화와 유사한 특성을 나타낸다.

신체 정보와 혈액검사 항목 사이의 다차원적 연관성은 성별, 연령에 따라 다른 양상을 보였다. 신체 정보의 신장 항목은 연령대가 증가함에 따라 그 영향력이 작아진 반면, 혈액검사의 중성지방 항목은 그 영향력이 높아졌다. 특히 여성의 경우 연령대가 높아지면서 허리둘레 항목의 영향력이 높아졌으며, 이는 혈액검사의 중성지방과 LDL 콜레스테롤 항목

을 예측하는 데 주요 변수로 활용 가능성을 보여준다.

본 연구는 정준상관분석을 이용하여 보건 빅데이터를 분석하여 혈액 내 지방 수치를 평가할 수 있는 예측 모델을 신체 정보로 구성할 때 필요한 주요 핵심 요소들의 영향력을 도출하였다. 허리둘레는 남녀 모두에게서 중성지방과 HDL 콜레스테롤 수치를 예측할 수 있는 주요 요소임을 보여준다. 더해서, 성별에 따른 차이를 고려해야 하며 특히 여성에게는 나이 정보를 주요 예측인자로 추가할 필요가 있다. 반면, 남성에게는 체중 항목이 주요 예측인자로 활용될 필요가 있다. 다양한 신체 정보 및 병리 검사 결과 항목을 고려한 후속 연구가 지속적으로 이루어져 본 연구 결과와 종합함으로써 비침습적 방법으로 측정된 정보를 기반으로 이상지질혈증을 조기에 정확히 예측하는 모델을 구축할 필요가 있다.

본 연구의 결과를 해석할 때 다음과 같은 제한점을 고려해야 한다. 첫째, 건강검진정보 자료에 의존하였기 때문에 임상 결과에 대한 정보가 부족하여 건강 상태의 변화로 인해 허리둘레나 체중이 변화했을 가능성을 배제할 수 없다. 둘째, 20세 이상 건강검진을 수진한 국민건강보험가입자를 대상으로 하였기 때문에 모든 연령층에게 일반화하기에는 한계가 있다. 셋째, 단면연구의 특성상 신체 정보와 혈액 내 지질 수치의 인과관계를 단언할 수 없다. 넷째, 이상지질혈증의 위험인자인 고혈압, 당뇨, 가족력 등과의 관련성에 대한 조사가 이루어지지 않아 이에 관한 추가적인 연구가 필요하다. 마지막으로, 비록 모든 정준상관분석에서 통계적으로 유의미한 정준상관계수(r)를 도출하였지만, 50대 이상 연령대에서 계수 값이 줄어들었다. 따라서 특히 고연령대에서 다차원적 연관성을 해석하는 데 유념할 필요가 있다. 위 제한점에도 불구하고 전 국민 보건의료 빅데이터를 이용하여 신체 정보와 혈액검사 정보를 다차원적으로 분석함으로써, 한국 성인의 이상지질혈증 예측모델 개발을 위한 주요 요소들 사이의 영향력 정도를 성별과 연령에 따라 제시했다는 점에서 그 의의가 있다.

V. Conclusions

본 연구는 건강검진을 수진한 3,312,971명의 성인을 대상으로 신체 정보와 이상지질혈증 지표와의 다차원적 상관관계를 성별과 연령에 따라 분석한 결과 다음과 같은 결론을 도출하였다.

허리둘레가 크고 체중이 많이 나가면 중성지방이 높아지고 HDL 콜레스테롤 수치가 낮아지는 다차원적 상관관

계를 확인하였다. 나이에 따른 혈액 내 지질 농도 변화는 남성보다 여성에게서 뚜렷이 나타났지만, 체중에 따른 변화는 남성에게서 더 뚜렷이 나타났다. 반면, 위 다차원적 상관관계에서 신장과 LDL 콜레스테롤 항목은 남성과 여성 모두에게서 상대적으로 작은 영향력을 보였다. 특히, 여성에게서 체중, 허리둘레, HDL 콜레스테롤 항목의 영향력은 연령에 따라 급격하게 변했다. 따라서, 한국인의 이상지질혈증을 효과적으로 예방하고 관리하기 위한 예측 모델을 개발하는 데 있어서 성별과 연령에 따른 각 지표의 영향력 정도를 반영할 필요가 있겠다. 이를 통해, 이상지질혈증 조기 발견 및 예방은 물론, 위험성 관리를 통해 사회경제적 손실비용을 낮출 수 있을 것으로 기대한다.

ACKNOWLEDGEMENT

This paper was supported by research funds of Kunsan National University.

REFERENCES

- [1] Ministry of Health and Welfare, "Korea Health Statistics 2016: Korea National Health and Nutrition Examination Survey (KNHANES VII-1)," Cheongju: Korea Centers for Disease Control and Prevention; 2017.
- [2] H.C. Kim, "Epidemiology of dyslipidemia in Korea," *Journal of the Korean Medical Association*, Vol. 59, No. 5, pp. 352-357, May 2016. DOI: <https://doi.org/10.5124/jkma.2016.59.5.352>
- [3] World Health Organization, "Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016," Geneva, 2018.
- [4] J.P. Halcox, J.R. Banegas, C. Roy, J. Dallongeville, G. Backer, E. Guallar, J. Perk, D. Hajage, K.M. Henriksson, C. Borghi, "Prevalence and treatment of atherogenic dyslipidemia in the primary prevention of cardiovascular disease in Europe: EURIKA, a cross-sectional observational study," *BMC Cardiovascular Disorders*, Vol. 17, pp. 160, June 2017. DOI: <https://doi.org/10.1186/s12872-017-0591-5>
- [5] E. Ward, E.B. Gold, W.O. Johnson, F. Ding, P.-Y. Chang, P. Song, S.R.E. Khoudary, C. Karvonen-Gutierrez, K.R. Ylitalo, J.S. Lee, "Patterns of Cardiometabolic Health as Midlife Women Transition to Menopause: A Prospective Multiethnic Study," *The Journal of Clinical Endocrinology & Metabolism*, Vol. 104, No. 5, pp. 1404-1412, May 2019. DOI: <https://doi.org/10.1210/jc.2018-00941>
- [6] W.-H. Choi, Y.-M. Seo, M.Y. Jeon, S.Y. Choi, "Convergence Study on the Comparison of Risk Factors for Dyslipidemia by Age and Gender: Based on the Korea National Health and Nutrition Examination Survey(2013~2015year)," *Journal of the Korea Convergence Society*, Vol. 9, No. 10, pp. 571-587, October 2018. DOI: <https://doi.org/10.15207/JKCS.2018.9.10.571>
- [7] Y.G. Cho, H.J. Song, B.J. Park, "The Comparison of Guidelines for Management of Dyslipidemia and the Appropriateness of Them in Korea," *Korean Journal of Family Medicine*, Vol. 31, pp. 171-181, March 2010. DOI: <https://doi.org/10.4082/kjfm.2010.31.3.171>
- [8] B.-K. Cha, "Risk Factors affecting Dyslipidemia according to Age Group in Women: Data from the Korea National Health and Nutrition Examination Survey, 2014~2017." *Journal of Korean Academy of Fundamentals of Nursing*, Vol. 27, No. 3, pp. 310-322, August 2020. DOI: <https://doi.org/10.7739/jkafn.2020.27.3.310>
- [9] J.H. Park, M.H. Lee, J.-S. Shim, D.P. Choi, B.M. Song, S.W. Lee, H. Choi, H.C. Kim, "Effects of age, sex, and menopausal status on blood cholesterol profile in the Korean population," *Korean Circulation Journal*, Vol 45, No. 2, pp 141-148, March 2015. DOI: <https://doi.org/10.4070/kcj.2015.45.2.141>
- [10] S.W. Yi, J.J. Yi, H. Ohrr, "Total cholesterol and all-cause mortality by sex and age: a prospective cohort study among 12.8 million adults," *Scientific Reports*, Vol 9, pp. 1596, February 2019. DOI: <https://doi.org/10.1038/s41598-018-38461-y>
- [11] K.S. Kim, "Related Factors of Dyslipidemia among Pre- and Post-Menopausal Women in South Korea - Based on the data of the Sixth Korea National Health and Nutrition Examination Survey (2013~2015)," *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, Vol. 8, No. 8, pp. 139-152, August 2018. DOI: <http://dx.doi.org/10.21742/AJMAHS.2018.08.84>
- [12] J.H. Han, S.M. Kim, "Relationship between percent body fat and cardiovascular risk factors for normal weight adults," *Journal of the Korean Academy of Family Medicine*, Vol. 27, No. 5, pp. 352-357, May 2006. <https://www.kjfm.or.kr/journal/view.php?number=432>
- [13] Y.W. Yoon, H.-M. Lee, H.-W. Oh, "Correlation of Dyslipidemia with Waist Circumference and Waist-Height Ratio," *Korean Journal of Family Practice*, Vol. 6, No. 6, pp. 560-567, December 2016. DOI: <https://doi.org/10.21215/kjfp.2016.6.6.560>
- [14] H.C. Choi, H.J. Kim, S.J. Min, K. Lee, K.W. Kim, S.M. Oh, T.W. Yoo, "Changes in Blood Pressure, Blood Glucose, and Lipid Profile Caused by Changes of Weight, Percent Body Fat, and Waist Circumference in Adult Men with Normal Weight and Waist Circumference," *Korean Journal of Family Medicine*, Vol. 31, pp. 430-436, June 2010. DOI: <https://doi.org/10.4082/kjfm.2010.31.6.430>
- [15] National Health Insurance Sharing Service, <http://nhiss.nhis.or.kr>
- [16] H. Hotelling, "Relations between two sets of variates.

- Biometrika,” Vol. 28, pp. 321-337, December 1936. DOI: <https://doi.org/10.1093/biomet/28.3-4.321>
- [17] H.T. Wang, J. Smallwood, J. Mourao-Miranda, C.H. Xia, T.D. Satterhwaite, D.S. Bassett, D. Bzdok, “Finding the needle in a high-dimensional haystack: Canonical correlation analysis for neuroscientists,” *Neuroimage*, Vol. 216, pp. 116745, August 2020. DOI: <https://doi.org/10.1016/j.neuroimage.2020.116745>
- [18] S.H. Kim, B.C. Ahn, H.J. Joung, M.J. Park, “Lipid profiles and prevalence of dyslipidemia in Korean adolescents,” *Endocrinology and Metabolism*, Vol. 27, pp. 208-216, September 2012. DOI: <https://doi.org/10.3803/EnM.2012.27.3.208>
- [19] Y. Choi, C.S. Soon, S.H. Shin, W.-Y. So, H.S. Kim, “The relationship between anthropometric measurements and metabolic syndrome risk factors in Korean population,” *The Journal of Korea Society for Wellness*, Vol. 12, No. 4, pp. 647-654, December 2017. DOI: [10.21097/ksw.2017.11.12.4.647](https://doi.org/10.21097/ksw.2017.11.12.4.647)
- [20] D.H. Chun, “Association between Waist Circumference and High-Density Lipoprotein Cholesterol in Korean Adults: Based on the Data of Korea National Health and Nutrition Examination Survey 2015,” *Korean Journal of Family Practice*, Vol. 9, No. 1, pp. 3-9, February 2019. DOI: <https://doi.org/10.21215/kjfp.2019.9.1.3>

Authors



Han-Gue Jo received the B.E. in Electrical, Electronic, Information Engineering from Kongju National University, M.S. in Information System from Yonsei University, and Ph.D. in Cognitive Neuroscience from

European University Viadrina in 2007, 2009 and 2014, respectively. He is currently an Assistant Professor in the School of Computer Information and Communication Engineering, Kunsan National University. He is interested in ICT, AI, Neuroscience, and Medical Information.



Young-Heung Kang received the B.S. in Communication Engineering, M.S. in Electronics Engineering and Ph.D. in Electronics Engineering from Korean Airspace University in 1984, 1986 and 1993,

respectively. He is currently a Professor in the School of Computer Information and Communication Engineering, Kunsan National University since 1990. He is interested in ICT, Wireless Communication, Standards and Medical Information.