

# 보건의료 빅데이터를 활용한 생활습관 군집현상과 만성질환 유형의 연관성 분석: 지역사회건강조사 원시자료를 활용하여

노을희<sup>1</sup>, 박상찬<sup>2</sup>

<sup>1</sup>경희대학교 일반대학원 의료경영학과 연구원, <sup>2</sup>경희대학교 경영대학 교수

## Association between Clustering of Lifestyle and Chronic Disease Using Healthcare Big Data

Eul Hee Roh<sup>1</sup>, Sang Chan Park<sup>2</sup>

<sup>1</sup>Researcher, Medical Management of Graduate School, Kyung Hee University, Seoul; <sup>2</sup>Professor, Department of Business Administration, Kyung Hee University, Seoul, Korea

**Objectives:** In this study, the correlation between the clustering of lifestyle and the prevalence of chronic diseases was identified with the association rule analysis utilizing raw data of the Korean Community Health Survey by Korea Centers for Disease Control and Prevention (KCDC). **Methods:** It identified the characteristics of the subject and selected the representative type of cluster through the association rule analysis. In addition, through the chi-square test, the sensitivity and specificity value, it tested the significance and validity of the association rule. Through the logistic regression analysis, identified the individual characteristic factors having influence on the cluster of lifestyle. **Results:** The statistically significant association rule was mainly composed of two or three lifestyle clusters and that composed of five lifestyle clusters the most was shown. In addition, there was lifestyle cluster shown commonly in the their age groups in all the chronic diseases and on the contrary, there was the lifestyle cluster shown in specific age group only. Also the specificity of all lifestyle clusters were represented greater than the sensitivity. In the results of identifying the association of individual factors by selecting the lifestyle cluster shown commonly in all chronic diseases in all age groups, the gender, residential area, and occupation showed the significant association in all clusters. **Conclusions:** It presented that drawing the meaningful results by applying the association rule analysis on the health care Big Data is possible. Particularly, the fact that the association rule showing specific pattern between the lifestyle and the chronic disease was discovered provides the important implication in correcting the lifestyle for chronic disease management and prevention.

**Key words:** Healthcare big data, Chronic disease, Clustering of lifestyle, Association rule analysis, Korean Community Health Survey

## 서론

고령화, 의료기술 발전, 기대여명 증가, 생활습관 등의 변화로 인해 만성질환 유병률이 급증하고 있으며 이에 따른 사회와 개인의 부담도 증가하고 있다[1]. 질병관리본부의 보고서에 따르면 국내의 경우 만성

질환으로 인한 사망은 전체 사망의 80.8%를 차지하며, 사망원인 상위 10위 중 7개가 만성질환이다[2]. 특히 고혈압과 당뇨병 진료비가 지속적으로 증가하고 있는데, 고혈압은 2012년 7,693억 원에서 2016년 9,010억 원으로 연평균 4% 증가했고, 당뇨병은 2012년 5,382억 원에서 2016년 7,258억 원으로 연평균 7.8% 증가하였다[3].

**Corresponding author:** Sang Chan Park

26 Kyunghedae-ro, Dongdaemun-gu, Seoul 02447, Korea  
E-mail: sangchan@khu.ac.kr

Received: January 22, 2020 Revised: February 23, 2020 Accepted: March 13, 2020

No potential conflict of interest relevant to this article was reported.

**How to cite this article:**

Roh EH, Park SC. Association between clustering of lifestyle and chronic disease using healthcare big data. J Health Info Stat 2020;45(2):113-123. Doi: <https://doi.org/10.21032/jhis.2020.45.2.113>

© It is identical to the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permit unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2020 Journal of Health Informatics and Statistics

만성질환은 흡연, 과도한 음주, 비만, 그리고 신체적 비활동 등의 생활습관과 관련된 건강위험행위들과 밀접한 관계가 있다[4]. 1974년에 캐나다에서 발간한 라론드 보고서(Lalonde's report)에서는 장기간의 추적조사결과를 통해 개인의 건강과 만성질환을 결정하는 요인이 크게 유전적 요인(20%), 환경적 요인(20%), 개인의 생활습관(52%), 의료서비스(8%)로 구분된다는 것을 밝혀냈다[5]. 또한 한 가지 생활습관보다는 여러 가지 요인이 복합적으로 작용하는 것으로 알려져 있다[6]. 개별적인 위험요인들의 기대빈도보다 실제 발생빈도가 높은 경우를 건강위험행위 군집(cluster)현상이라고 하며[7], 생활습관 위험요인들은 서로 군집하는 경향이 있다고 알려져 있다[1]. 기존 연구에 따르면 네덜란드인의 20%[7], 미국인의 17%[8], 홍콩 노인의 5%[9], 한국 남성의 15.2%[10]는 세 개 이상의 건강위험행위를 하고 있다. 생활습관 요인에 대한 군집 경향을 파악한 기존 연구에서는 주로 'Alameda 7'에 포함된 7가지 생활습관을 대상으로 분석하였다. 'Alameda 7'은 1960년대 중반 이래 미국의 캘리포니아 주 Alameda County에서 실시한 대규모 역학 조사를 통해 파악된 7가지 생활습관을 의미한다. 해당 연구에서는 건강수준과 연관성이 있는 건강행위 7가지 건강행위(Alameda 7), 즉 흡연, 음주, 비만(체중조절), 운동, 아침식사 및 간식섭취 여부, 수면을 제시하였고, 추적조사를 통하여 이들 개인의 건강행위가 질병 및 사망과 높은 연관성이 있음을 밝혔다[11].

최근에는 인구 고령화에 따른 만성 질환 및 퇴행성 질환의 증가 등의 사회적 문제를 해결하기 위해 보건의로 분야에서 빅데이터를 활용하고자 하는 다양한 연구들이 시도되고 있다. 보건의로 빅데이터의 활용은 개인이 건강을 관리하고 의료를 선택하는 방식을 변화시켜 보건 의료시스템 전반에서 건강 결과와 지출의 효율을 높이는 혁신의 동력으로 기대되고 있다[12]. 미국의 한 연구에서는 의료 분야에서 빅데이터를 활용할 시 연 3,000억 달러의 경제 비용 절감 효과가 있을 것이라고 예측하였다[13]. 이에 많은 국가에서 보건의로 빅데이터의 수집, 연계, 통합, 분석을 통한 새로운 지식 발견을 위한 정부 차원의 의료 빅데이터 사업을 진행하고 있다. 국내에서는 2019년 9월에 보건의로 분야 공공기관의 의료데이터를 정책연구 등 공공 목적으로 활용할 수 있도록 '보건의로 빅데이터 플랫폼'을 개통하였다. 이러한 추세에 맞춰서 학계에서도 보건의로 빅데이터를 활용한 연구가 증가하고 있으며, 이를 통해 보건의로분야 정책결정 및 근거생성에 기여하고 있다. 최근에는 국가 차원의 만성질환 관리 필요성이 대두됨에 따라 보건복지부에서 '일차의료 만성질환관리 시범사업'을 시행하고 있는데, 해당 사업을 효과적으로 진행하기 위해서는 환자별 특이성을 고려한 맞춤형 서비스를 제공할 필요가 있다. 이를 위해서는 보건의로 빅데이터를 활용하여 기존 만성질환자의 특성을 파악하는 것이 필수적이다.

이에 본 연구에서는 지역사회건강조사 원시자료를 활용하여 생활

습관 군집현상과 만성질환 유병의 연관성을 연관규칙분석으로 파악함으로써 예방 측면의 맞춤형 만성질환 관리에 도움이 될 수 있는 기초자료를 제공하고자 한다.

이에 대한 구체적인 내용은 다음과 같다.

- 연령층별로 생활습관과 만성질환의 특성 및 분포를 비교하고자 한다.
- 연관규칙분석을 통해 연령층 및 만성질환별로 생활습관 군집 중 대표성이 있는 군집 유형을 선정하고자 한다.
- 카이제곱검정과 민감도, 특이도 값을 통해 본 연구의 연관규칙 유의성을 평가하고 타당성을 검증하고자 한다.
- 로지스틱 회귀분석을 통해 생활습관 군집에 영향을 미치는 개인 특성 요인을 파악하고자 한다.

## 연구 방법

### 연구대상 및 자료수집

본 연구에서는 질병관리본부에서 2016년 8월 16일부터 10월 31일까지 시행한 지역사회건강조사 원시자료를 활용하였다. 해당 조사는 국내에 거주하는 만 19세 이상의 성인을 대상으로 훈련된 조사원이 표본으로 선정된 가구에 직접 방문하여 1대1 전자설문조사로 진행되었으며, 18개 영역 161개 문항으로 구성되어있다. 세부적으로는 가구조사, 건강행태, 예방접종 및 검진, 이환, 의료이용, 사고 및 중독, 활동제한 및 삶의 질, 보건기관 이용, 사회 물리적 환경, 심폐소생술, 교육 및 경제활동 영역이 포함된다[14].

본 연구에서는 「지역사회건강조사 원시자료 공개 및 활용규정」에 따라 질병관리본부의 승인을 받은 후 홈페이지[15]에서 다운로드하여 사용하였다. 원시자료에 포함된 228,452명의 데이터 중 서울특별시와 각 지역의 대표성이 있는 6개 광역시에 거주 중인 25세 이상의 성인 60,584명의 데이터를 활용하였다. 기존 연구의 연령 분류기준에 제시된 청장년층이 25세 이상[16,17]이기 때문에 본 연구에서는 25세 미만의 성인 데이터는 제외하였다.

본 연구는 보건복지부의 2차 자료를 활용했으며, 경희대학교 생명윤리심의위원회의 심의 면제 승인을 받았다(IRB No. KHSIRB-19-323).

### 변수의 조작적 정의

#### 개인특성

본 연구에서는 개인특성 변수로 응답자의 성별, 연령, 거주 지역, 직업, 교육수준을 사용하였다.

성별은 남성과 여성으로 분류하였다. 연령은 청장년층(25-44세), 중년층(45-64세), 노년층(65세 이상)으로 분류하였다[16,17]. 거주 지역은

서울특별시, 부산광역시, 대구광역시, 인천광역시, 광주광역시, 대전광역시, 울산광역시로 나누었다.

직업은 한국표준직업분류에 따라 10개로 분류되어 있는 직업군을 3개로 재분류하고, 지역사회건강조사 응답 항목의 주부, 학생/재수생, 무직을 추가하여 총 6개 직업군으로 정의하였다. 세부적으로는 사무직(관리자, 전문가 및 관련 종사자, 사무 종사자), 서비스판매직(서비스 종사자, 판매 종사자), 생산기능직(농·림·어업 종사자, 기능원 및 관련 기능 종사자, 장치·기계조작 및 조립 종사자, 단순노무 종사자), 주부, 기타(군인, 학생/재수생), 무직으로 구성된다.

교육 수준은 지역사회건강조사 응답 항목 기준에 따라 무학, 초등학교, 중학교, 고등학교, 2년/3년제 대학, 4년제 대학, 대학원 이상으로 분류하였다.

### 부정적 생활습관

본 연구에서는 Alameda County 연구에서 사용한 7가지 건강행위 변수를 이용하여 보건 지침을 지키지 않는 경우를 만성질환에 영향을 미치는 생활습관 요인으로 정의하였다. 모든 생활습관 요인을 실천함과 실천하지 않음의 두 가지로 분류하여 건강행위를 실천하지 않는 것, 즉 부정적 생활습관을 분석하였다[18]. 본 연구에서는 지역사회건강조사 설문 항목에 제시되어 있지 않은 간식 섭취를 제외한 나머지 6가지 부정적 생활습관을 대상으로 하며 흡연, 과도한 음주, 신체적 비활동, 부적절한 체중 조절, 부적절한 수면, 아침식사 결식이 해당된다.

흡연(current smoking, CS)은 현재 흡연을 하는 경우로 정의했으며 [8], 과거에는 피웠으나 현재는 피우지 않는 경우는 건강행위를 실천하는 것으로 분류해[11] 부정적 생활습관으로 포함하지 않았다.

과도한 음주(heavy drinking, HD)는 국가통계포털 100대 지표에서 제시한 음주 고위험군 기준을 따랐으며, 남성의 경우 주 2회 이상, 1회 평균 7잔 이상, 여성의 경우 주 2회 이상, 1회 평균 5잔 이상을 마시는 경우로 정의하였다.

신체적 비활동(physical inactivity, PI)은 질병관리본부의 국민건강영양조사에서 제시한 중등도 이상 신체활동 실천 기준을 따랐다. 최근 1주일 동안 평소보다 몸이 매우 힘들거나 숨이 많이 가쁜 격렬한 신체 활동을 1일 20분 이상, 주 3일 이상 실천하거나 평소보다 몸이 조금 힘들거나 숨이 약간 가쁜 중등도 신체활동을 1일 30분 이상, 주 5일 이상 실천하는 지침을 지키지 않는 경우로 정의하였다.

부적절한 체중 조절(inappropriate weight control, IW)은 World Health Organization (WHO) 아시아-태평양 지역 체질량지수(body mass index, BMI) 기준에서 제시한 병적비만인 25 이상으로 정의하였다.

부적절한 수면(inappropriate sleep, IS)은 일반적인 성인의 적정 수면 시간인 하루 7-8시간을 실천하지 않는 하루 7시간 미만 또는 8시간 이

상 수면하는 경우로 정의하였다[19].

아침식사 결식(breakfast skipping, BS)은 국가통계포털 100대 지표에서 제시한 아침 식사 결식 기준을 따랐으며, 최근 1주일 동안 아침식사를 5일 이상 먹지 않은 경우로 정의하였다.

### 만성질환 유병

만성질환의 범위는 국가 또는 기관마다 제시하는 기준에 차이가 있으며, 우리나라에서는 만성질환의 범위를 명확하게 규정하고 있지는 않다. 본 연구에서는 만성질환의 범위를 파악하기 위해 국민건강통계, 국민건강영양조사, 의료서비스경험조사, 지역별 의료이용통계, 건강통계연보 6개 연구의 만성질환 범위를 비교하였다. 최종적으로는 6개 연구에서 모두 만성질환으로 포함하는 고혈압, 당뇨병과 3개 연구에서 공통적으로 포함하는 관절염을 본 연구의 만성질환으로 선정하였다. 또한 만성질환 유병변수로 고혈압, 당뇨병, 관절염에 대해 의사에게 진단받은 경험이 있는 경우로 정의하였다.

### 분석 방법

연령에 따라서 만성질환 유병과 건강과 관련된 생활습관의 실천 정도에 차이가 있다는 기존 연구[20-22]에 따라 청장년층, 중년층, 노년층으로 나눠서 분석하였다.

먼저 데이터를 가공하는 과정을 거친 후, R (ver. 1.0.143)과 SPSS 25.0 (IBM Co., Armonk, NY, USA) 프로그램을 사용해서 연관규칙분석과 통계적 분석 두 부분으로 분석을 진행하였다. 본 연구에서는 대상자의 생활습관 보유 유무와 질환 유병 유무를 기준으로 연관규칙분석을 진행하기 때문에 지역사회건강조사 데이터의 가중치를 고려하지 않았다. 이는 국내 보건의로 빅데이터를 활용하여 연관규칙분석을 진행한 기존연구[18,23-26]와 동일한 방법이다.

연관규칙분석에서는 연령층 기준으로 데이터를 분류해서 각 연령층별로 고혈압, 당뇨병, 관절염에 대한 연관분석을 진행하였다. 연관규칙분석에서는 두 항목 간의 향상도가 1보다 클 경우, 두 항목은 서로 양의 상관관계가 있다는 것을 의미하기 때문에[27] 최소지도도 값을 충족하며 향상도가 1보다 큰 생활습관 군집을 유의한 연관규칙으로 도출하였다. 본 연구에서 연관규칙분석을 활용한 이유는 다른 분석 방법에 비해 데이터 내에서 생활습관 요인의 군집현상을 각 행위마다 동등하게 취급해서 찾아내기 위해서는 연관규칙분석이 좀 더 적합한 방법이기 때문이다[18].

통계적 분석에서는 연령층 대상자의 변수별 분포를 파악하고 카이제곱검정을 통해 연관규칙의 통계적 유의성을 평가하였다. 그 다음으로 연관규칙 별로 민감도와 특이도 값을 구해서 연관규칙의 타당성을 검증하는 과정을 진행하였다. 마지막 분석단계에서는 연관규칙의 대

표적 생활습관 군집을 예시로 해서 로지스틱 회귀분석을 통해 생활습관 군집과 개인특성 요인의 관련성을 파악하였다.

### 연관규칙분석

본 연구에서는 생활습관 보유 유무를 조건으로, 만성질환 유병 유무를 결과로 하여 연관규칙분석의 방향을 설정해서 분석을 진행하였다. 본 연구에서는 3개의 연령층을 대상으로 3개의 단일 만성질환에 대해 총 9번의 연관규칙분석을 하였다.

본 연구에서의 지지도는 '전체 대상자 중 특정 생활습관을 가진 대상자가 특정 만성질환을 앓고 있을 확률'을 의미한다. 신뢰도는 '특정 생활습관을 가진 대상자 중 특정 만성질환을 앓고 있을 확률'을 의미한다. 또한 향상도는 '신뢰도에서 특정 만성질환을 앓고 있을 확률로 나눈 값'을 의미한다. 본 연구에서는 생활습관과 만성질환 유병이 유의한 연관성을 가지는 연관규칙을 찾는 것이 가장 중요하기 때문에 향상도가 가장 중요한 평가척도라고 할 수 있다.

### 연관규칙 유의성 평가

연관규칙분석에서 도출한 연관규칙이 통계적으로 유의한지 평가하기 위해 각 연관규칙에 대해 카이제곱검정을 진행하고, 향상도 값과 카이제곱검정의 교차비(odds ratio, OR)를 비교하였다. 이를 통해 통계적으로도 유의한 연관규칙을 최종 선정하였다. 교차비는 두 변수 사이의 연관성 척도를 의미하며, 각 대각선 셀끼리 곱하기 때문에 교차비라고 한다. 향상도 값은 기준이 되는 생활습관 여부와 상관없이 전체 대상자를 기준으로 하기 때문에 항상 교차비보다 낮은 값을 나타낸다.

### 연관규칙 타당성 검증

최종 선정된 각 연관규칙 별로 민감도와 특이도 값을 계산하여 연관규칙에 대한 타당성 검증 및 비교를 진행하였다. 민감도와 특이도는 주로 진단검사의 유용성이나 예측모델의 타당성을 평가할 때 사용되는 지표이다.

민감도는 특정 진단법을 사용했을 때 실제 양성인 사람을 양성으로 판단하는 비율[28]이며, 본 연구에서는 연관규칙이 특정 만성질환 유병 대상자 중에서 특정 생활습관 군집에 포함되는 대상으로 판단하는 비율을 의미한다.

특이도는 특정 진단법이 실제 음성인 사람을 음성으로 판단하는 비율[28]이며, 본 연구에서는 연관규칙이 특정 만성질환을 앓고 있지 않은 대상자 중에서 특정 생활습관 군집에 포함되지 않는 대상으로 판단하는 비율을 의미한다.

## 연구 결과

### 대상자의 특성

대상자의 개인 특성에서는 모든 연령층에서 여성이 남성보다 많았고, 거주 지역에서는 서울특별시 거주자가 가장 많은 것으로 나타났다. 직업에서는 연령층별로 차이를 보였는데, 청장년층에서는 사무직이 45.4%, 중년층에서는 생산기능직이 29.4%, 노년층에서는 무직이 40.2%로 가장 많았다. 교육 수준에서는 모든 연령층에서 고등학교를 졸업한 대상자가 가장 많았다(Table 1).

또한 연령대가 높아짐에 따라 만성질환을 앓고 있는 대상자의 비율이 현저하게 높아지는 현상을 보였다. 청장년층 3.9%, 중년층 23.1%, 노년층 54.2%가 고혈압을 앓고 있으며, 특히 노년층에서는 전체 노년층 대상자의 절반 이상이 고혈압을 앓고 있는 것으로 나타났다. 청장년층 1.4%, 중년층 9.3%, 노년층 23.7%는 당뇨병을 앓고 있고, 청장년층 1.1%, 중년층 8.7%, 노년층 27.4%는 관절염을 앓고 있는 것으로 나타났으며 두 질환의 연령대에 따른 유병 대상자 비율의 증가가 유사하게 나타났다(Table 1).

청장년층에서는 다른 연령대에 비해 아침식사 결식(34.9%)에 해당되는 비율이 매우 높았고, 흡연(23.9%)과 과도한 음주(30.9%)의 비율도 높은 것으로 나타났다. 반면에 부적절한 수면(48.8%)의 비율은 다른 연령대보다 낮았다. 장년층에서는 다른 연령대에 비해 신체적 비활동(77.9%)과 부적절한 체중조절(28.2%)의 비율이 약간 높은 수치를 보였다. 노년층에서는 다른 연령대에 비해 신체적 비활동(84.4%)과 부적절한 수면(58.6%)의 비율이 높은 것으로 나타났다. 반면에 과도한 음주(6.9%), 아침식사 결식(3.4%)의 비율은 다른 연령대보다 매우 낮았고, 흡연(10.0%)의 비율도 낮은 것으로 나타났다(Table 1).

### 연관규칙분석 및 타당성 검증 결과

본 연구의 만성질환을 대상으로 한 연관규칙분석에서는 유병대상자의 수를 고려하여 최소지지도 1%, 최소신뢰도 10%를 만족하고, 향상도 값이 1보다 큰 기준을 만족하는 생활습관 군집을 유의미한 연관규칙으로 도출하였다. 또한 카이제곱검정을 통해 앞서 도출한 연관규칙의 통계적 유의성을 평가하여 최종적으로 유의미한 연관규칙을 선정하였다. 최종 선정된 연관규칙의 교차비 값은 모두 1보다 크며, 95% 유의수준에서 통계적으로 유의한 것으로 나타났다. 다음으로는 최종 선정한 만성질환 연관규칙별로 민감도와 특이도 값을 구해서 연관규칙 간의 타당성을 비교하였다.

### 고혈압

고혈압에서는 청장년층 2개, 중년층 18개, 노년층 8개의 연관규칙이

**Table 1.** Characteristics of the participants (n=60,584)

Variables	Young and middle aged (n=21,717)	Middle aged (n=25,462)	The aged (n=13,405)	p <sup>1</sup>
Gender				<0.001
Male	10,317 (47.5)	11,314 (44.4)	5,911 (44.1)	
Female	11,400 (52.5)	14,148 (55.6)	7,494 (55.9)	
Residence area				<0.001
Seoul	7,921 (36.5)	8,351 (32.8)	4,297 (32.1)	
Busan	4,251 (19.6)	5,691 (22.4)	3,325 (24.8)	
Daegu	2,167 (10.0)	2,739 (10.8)	1,565 (11.7)	
Incheon	2,723 (12.5)	3,388 (13.3)	1,960 (14.6)	
Gwangju	1,519 (7.0)	1,688 (6.6)	842 (6.3)	
Daejeon	1,523 (7.0)	1,735 (6.8)	786 (5.9)	
Ulsan	1,613 (7.4)	1,870 (7.3)	630 (4.7)	
Occupation				<0.001
Office worker	9,851 (45.4)	5,716 (22.4)	441 (3.3)	
Service worker	3,380 (15.6)	4,743 (18.6)	640 (4.8)	
Production worker	3,278 (15.1)	7,498 (29.4)	2,252 (16.8)	
Housewife	3,520 (16.2)	5,569 (21.9)	4,682 (34.9)	
Others	436 (2.0)	30 (0.1)	0 (0.0)	
Inoccupation	1,252 (5.8)	1,906 (7.5)	5,390 (40.2)	
Education level				<0.001
Ineducation	61 (0.3)	138 (0.5)	914 (6.8)	
Elementary school	881 (4.1)	2,440 (9.6)	3,605 (26.9)	
Middle school	1,454 (6.7)	4,037 (15.9)	2,426 (18.1)	
High school	7,184 (33.1)	10,895 (42.8)	3,920 (29.2)	
Junior college	3,794 (17.5)	2,032 (8.0)	552 (4.1)	
Senior college	7,101 (32.7)	4,810 (18.9)	1,650 (12.3)	
Graduate school	1,242 (5.7)	1,110 (4.4)	338 (2.5)	
Chronic disease				
Hypertension				<0.001
Yes	849 (3.9)	5,876 (23.1)	7,263 (54.2)	
Diabetes				<0.001
Yes	313 (1.4)	2,367 (9.3)	3,175 (23.7)	
Arthritis				<0.001
Yes	231 (1.1)	2,225 (8.7)	3,673 (27.4)	
Lifestyle				
Current smoking				<0.001
Yes	5,199 (23.9)	5,021 (19.7)	1,339 (10.0)	
Heavy drinking				<0.001
Yes	6,708 (30.9)	5,690 (22.3)	924 (6.9)	
Physical inactivity				<0.001
Yes	16,848 (77.6)	19,846 (77.9)	11,318 (84.4)	
Breakfast skipping				<0.001
Yes	7,584 (34.9)	3,657 (14.4)	452 (3.4)	
Inappropriate weight control				<0.05
Yes	5,854 (27.0)	7,168 (28.2)	3,749 (28.0)	
Inappropriate sleep				<0.001
Yes	10,590 (48.8)	14,169 (55.6)	7,853 (58.6)	

All values are presented as n (%).

<sup>1</sup>Chi-square test for differences among age groups.

최종 선정되었다(Table 2). 예를 들어 하나의 연관규칙을 설명하면, 중년층에서 과도한 음주, 신체적 비활동, 부적절한 체중조절, 부적절한 수면 생활습관을 가지고 있는 대상자의 36.4%는 고혈압을 앓고 있으며 이는 전체 대상자 중 1.3%에 해당된다. 임의의 대상자가 고혈압을 앓고 있을 가능성에 비해 해당 생활습관을 가지고 있는 대상자는 고혈압 유병 가능성이 1.58배이다. 본 연구에서 고혈압 유병과 생활습관의 연관성을 파악하기 위해서는 항상도 값이 가장 중요하며, 청장년층의 과도한 음주와 부적절한 체중조절 생활습관 군집의 항상도가 2.76으로 고혈압 유병과 연관성이 가장 높았다. 해당 생활습관 군집은 고혈압 연관규칙 중 교차비 값이 3.86으로 가장 크게 나타났는데, 해당 생활습관을 가진 대상자는 고혈압 유병 가능성이 높으며 이는 통계적으로 유의하다(OR 3.86, 95% confidence interval, 95% CI=3.32-4.50).

고혈압 연관규칙에서 민감도가 가장 높은 노년층의 PI, IS 군집을 예시로 해서 본 연구에서의 민감도와 특이도 값을 설명하면 다음과 같다. 연관규칙이 고혈압을 앓고 있는 노년층 중에서 신체적 비활동, 부적절한 수면 생활습관을 가지고 있다고 판단할 확률은 51.2%이며, 고혈압을 앓고 있지 않은 노년층 중에서 해당 생활습관을 가지고 있지 않다고 판단할 확률은 51.5%이다. 청장년층에서는 HD, IW 군집의 민감도가 31.1%로 가장 높았고, CS, IW 군집의 특이도가 91%로 가장 높았다. 중년층에서는 PI, IS 군집의 민감도가 44.6%로 가장 높았고, CS, HD, IW 군집의 특이도가 97.4%로 가장 높았다. 노년층에서는 PI, IS 군집의 민감도가 51.2%로 가장 높았고, HD, PI, IW 군집의 특이도가 98.9%로 가장 높았다.

### 당뇨병

당뇨병에서는 중년층 8개, 노년층 4개의 연관규칙이 최종 선정되었으며, 중년층의 신체적 비활동과 부적절한 체중조절 생활습관의 항상도가 1.51로 당뇨병 유병과 연관성이 가장 높았다(Table 3). 해당 생활습관 군집은 당뇨병 연관규칙 중 교차비 값이 1.89로 가장 크게 나타났는데, 해당 생활습관을 가진 대상자는 당뇨병 유병 가능성이 높으며 이는 통계적으로 유의하다(OR 1.89, 95% CI=1.73-2.07).

당뇨병 연관규칙에서는 중년층에서 CS, HD 군집의 민감도가 33.2%로 가장 높았고, CS, PI, IS 군집의 특이도가 97.4%로 가장 높았다. 노년층에서는 HD, PI 군집의 민감도가 27.8%, 특이도가 94.9%로 가장 높았다.

### 관절염

관절염에서는 중년층 3개, 노년층 4개의 연관규칙이 최종 선정되었으며, 노년층의 신체적 비활동, 부적절한 체중조절, 부적절한 수면 생활습관의 항상도가 1.42로 관절염 유병과 연관성이 가장 높았다(Table 4).

**Table 2.** Validation of hypertension's significant association rules

Condition	Result	Object number	Support (%)	Confidence (%)	Lift	p	OR	Sensitivity (%)	Specificity (%)
Young and middle aged									
HD, IW	Hypertension	264	1.2	10.8	2.76	0.000	3.86	31.1	90.0
CS, IW		212	1.0	10.4	2.67	0.000	3.48	25.0	91.0
Middle aged									
HD, IW	Hypertension	731	2.9	36.4	1.58	0.000	2.04	33.6	93.5
IW, IS		1,492	5.9	36.1	1.56	0.000	2.18	14.8	86.5
PI, IW		1,975	7.8	35.3	1.53	0.000	2.23	20.3	81.5
CS, IW		461	1.8	31.8	1.38	0.000	1.60	25.4	95.0
BS, IW		338	1.3	31.6	1.37	0.000	1.57	12.4	96.3
HD, IS		869	3.4	27.4	1.19	0.000	1.30	7.8	88.3
HD, PI		1,190	4.7	27.4	1.18	0.000	1.32	5.8	83.9
PI, IS		2,618	10.3	23.8	1.03	0.022	1.07	44.6	57.1
HD, PI, IW		568	2.2	36.8	1.59	0.000	2.04	4.7	95.0
HD, IW, IS		427	1.7	36.5	1.58	0.000	1.99	6.2	96.2
PI, IW, IS		1,169	4.6	36.3	1.57	0.000	2.13	7.9	89.5
CS, HD, IW		258	1.0	33.8	1.46	0.000	1.73	7.3	97.4
CS, IW, IS		275	1.1	32.1	1.39	0.000	1.60	4.4	97.0
CS, PI, IW		365	1.4	31.8	1.38	0.000	1.59	4.7	96.0
PI, BS, IW		277	1.1	31.5	1.37	0.000	1.56	9.7	96.9
HD, PI, IS		663	2.6	27.5	1.19	0.000	1.29	11.3	91.1
CS, HD, PI		467	1.8	24.9	1.08	0.048	1.12	19.9	92.8
HD, PI, IW, IS		326	1.3	36.4	1.58	0.000	1.96	5.5	97.1
The aged									
HD, IW	Hypertension	206	1.5	68.7	1.27	0.000	1.88	2.8	98.5
PI, IW		2,107	15.7	66.9	1.23	0.000	2.00	29.0	98.3
IW, IS		1,475	11.0	66.6	1.23	0.000	1.86	20.3	87.9
CS, IW		174	1.3	61.9	1.14	0.008	1.38	2.4	83.1
PI, IS		3,717	27.7	55.5	1.02	0.002	1.11	51.2	51.5
HD, PI, IW		173	1.3	71.5	1.32	0.000	2.15	2.4	98.9
PI, IW, IS		1,278	9.5	67.7	1.25	0.000	1.94	17.6	98.6
CS, PI, IW		148	1.1	63.2	1.17	0.005	1.46	2.0	90.1

OR, odds ratio; CS, current smoking; HD, heavy drinking; PI, physical inactivity; IW, inappropriate weight control; IS, inappropriate sleep; BS, breakfast skipping.

해당 생활습관 군집은 관절염 연관규칙 중 교차비 값이 1.86으로 가장 크게 나타났는데, 해당 생활습관을 가진 대상자는 관절염 유병 가능성이 높으며 이는 통계적으로 유의하다(OR 1.86, 95% CI=1.68-2.06).

관절염 연관규칙에서는 중년층에서 PI, IW 군집의 민감도가 28.9%로 가장 높았고, PI, IW, IS 군집의 특이도가 88%로 가장 높았다. 노년층에서는 PI, IS 군집의 민감도가 56.7%로 가장 높았고, PI, IW, IS 군집의 특이도가 88.2%로 가장 높았다.

### 개인 특성과 생활습관 군집의 관련성 분석 결과

각 연령층에서 모든 만성질환에 공통적으로 나타난 생활습관 군집을 대표 군집으로 선정해서 생활습관 군집과 개인특성 요인의 관련성

을 파악하기 위해 다중 로지스틱 회귀분석을 진행하였다. 청장년층에서는 CS, IW 군집, 중년층에서는 PI, IW, IS 군집, 노년층에서는 PI, IW 군집을 선정하였다. 종속변수는 생활습관 군집이며, 설명변수에는 성별, 거주 지역, 직업, 교육수준을 포함하였다.

청장년층의 CS, IW 군집에서 해당 생활습관을 동시에 가지고 있을 확률은 여성(OR 0.04, 95% CI=0.04-0.06)이 남성보다 통계적으로 유의하게 낮았고, 인천거주자(OR 1.25, 95% CI=1.07-1.47)가 서울거주자보다 높았다. 직업에서는 서비스판매직(OR 1.61, 95% CI=1.42-1.84)과 생산기능직(OR 1.43, 95% CI=1.27-1.61)은 사무직보다 해당 군집에 포함될 확률이 높게 나타난 반면에, 주부(OR 0.45, 95% CI=0.25-0.80)와 무직(OR 0.65, 95% CI=0.46-0.93)은 낮게 나타났다. 교육수준에서는 통

**Table 3.** Validation of diabetes' significant association rules

Condition	Result	Object number	Support (%)	Confidence (%)	Lift	p	OR	Sensitivity (%)	Specificity (%)
Middle aged									
PI, IW	Diabetes	787	3.1	14.1	1.51	0.000	1.89	18.9	79.2
IW, IS		531	2.1	12.8	1.38	0.000	1.56	10.6	84.4
CS, PI		492	1.9	12.4	1.33	0.000	1.47	20.8	84.9
CS, IS		339	1.3	12.1	1.30	0.000	1.40	22.4	89.3
CS, HD		251	1.0	10.5	1.13	0.037	1.16	33.2	90.7
HD, PI		447	1.8	10.3	1.10	0.015	1.14	14.3	83.1
PI, IW, IS		444	1.7	13.8	1.48	0.000	1.69	11.7	88.0
CS, PI, IS		276	1.1	12.5	1.35	0.000	1.45	18.8	91.7
The aged									
PI, IW	Diabetes	882	6.6	28.0	1.18	0.000	1.35	18.9	77.8
HD, PI		201	1.5	27.7	1.17	0.009	1.25	27.8	94.9
IW, IS		599	4.5	27.0	1.14	0.000	1.24	6.3	84.2
PI, IW, IS		523	3.9	27.7	1.17	0.000	1.28	16.5	86.7

OR, odds ratio; CS, current smoking; HD, heavy drinking; PI, physical inactivity; IW, inappropriate weight control; IS, inappropriate sleep; BS, breakfast skipping.

**Table 4.** Validation of arthritis' significant association rules

Condition	Result	Object number	Support (%)	Confidence (%)	Lift	p	OR	Sensitivity (%)	Specificity (%)
Middle Aged									
IW, IS	Arthritis	481	1.9	11.6	1.33	0.000	1.48	21.6	84.3
PI, IW		644	2.5	11.5	1.32	0.000	1.50	28.9	79.0
PI, IW, IS		380	1.5	11.8	1.35	0.000	1.48	17.1	88.0
The aged									
IW, IS	Arthritis	829	6.2	37.4	1.36	0.000	1.75	22.6	52.6
PI, IW		1,139	8.5	36.2	1.32	0.000	1.73	31.0	79.4
PI, IS		2,081	15.5	31.1	1.13	0.000	1.45	56.7	85.7
PI, IW, IS		735	5.5	38.9	1.42	0.000	1.86	20.0	88.2

OR, odds ratio; CS, current smoking; HD, heavy drinking; PI, physical inactivity; IW, inappropriate weight control; IS, inappropriate sleep; BS, breakfast skipping.

계적으로 유의한 관련성을 보이는 개인특성 요인이 나타나지 않았다 (Table 5).

중년층의 PI, IW, IS 군집에서 해당 생활습관을 동시에 가지고 있을 확률은 여성(OR 0.67, 95% CI=0.61-0.74)이 남성보다 통계적으로 유의하게 낮았다. 거주 지역에서는 인천거주자(OR 1.13, 95% CI=1.00-1.27)는 서울거주자보다 해당 생활습관을 동시에 가지고 있을 확률이 높은 반면에, 대구거주자(OR 0.87, 95% CI=0.76-1.00)는 낮았다. 직업에서는 생산기능직(OR 0.80, 95% CI=0.71-0.90)이 사무직보다 해당 군집에 포함될 확률이 낮게 나타났다. 교육수준에서는 초등학교 졸업(OR 0.75, 95% CI=0.63-0.88)과 무학(OR 0.74, 95% CI=0.62-0.88)이 대학원 이상 졸업보다 포함 확률이 낮았다(Table 5).

노년층의 PI, IW 군집에서는 다른 생활습관 군집과 달리 여성(OR 1.33, 95% CI=1.18-1.51)이 남성보다 해당 생활습관을 동시에 가지고 있

을 확률이 통계적으로 유의하게 높았다. 거주 지역에서는 부산거주자(OR 0.87, 95% CI=0.79-0.97), 대구거주자(OR 0.79, 95% CI=0.69-0.91), 인천거주자(OR 0.84, 95% CI=0.73-0.95)가 서울거주자보다 해당 군집에 포함될 확률이 낮게 나타났다. 직업에서는 생산기능직(OR 0.66, 95% CI=0.51-0.84), 주부(OR 0.72, 95% CI=0.56-0.92), 무직(OR 0.73, 95% CI=0.58-0.92)이 사무직보다 해당 군집에 포함될 확률이 낮게 나타났다. 교육수준에서는 통계적으로 유의한 관련성을 보이는 개인특성 요인이 나타나지 않았다(Table 5).

## 고찰 및 결론

본 연구에서는 다양한 사회적 요인의 변화로 인해 생활습관이 큰 영향을 미치는 만성질환의 유병률 증가와 이를 해결하기 위한 보건의

**Table 5.** Logistic regression result of representative clustering

Clustering	Young and middle age's CS, IW	Middle age's PI, IW, IS	The age's PI, IW
Variables	OR	OR	OR
<b>Gender</b>			
Male	1.00	1.00	1.00
Female	0.04***	0.67***	1.33***
<b>Residence area</b>			
Seoul	1.00	1.00	1.00
Busan	1.12	0.91	0.87**
Daegu	1.17	0.87*	0.79**
Incheon	1.25**	1.13*	0.84**
Gwangju	0.97	0.89	0.85
Daejeon	1.01	0.87	0.93
Ulsan	1.12	0.95	0.98
<b>Occupation</b>			
Office worker	1.00	1.00	1.00
Service worker	1.61***	0.92	0.76
Production worker	1.43***	0.80***	0.66***
Housewife	0.45**	0.96	0.72**
Others	0.65*	1.47	
Inoccupation	0.90	0.80**	0.73**
<b>Education level</b>			
Graduate school	1.00	1.00	1.00
Senior college	0.00	0.99	1.06
Junior college	0.74	0.44	0.65
High school	1.12	1.13	0.98
Middle school	0.89	0.92	0.91
Elementary school	0.87	0.75***	0.81
Ineducation	1.01	0.84	0.64

OR, odds ratio; CS, current smoking; HD, heavy drinking; PI, physical inactivity; IW, inappropriate weight control; IS, inappropriate sleep; BS, breakfast skipping.

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

료 빅데이터 활용 연구가 증가하고 있는 추세에 주목하였다.

이에 따라 본 연구에서는 지역사회건강조사 원시자료를 활용해서 생활습관 군집현상과 만성질환 유병의 연관성을 파악함으로써 맞춤형 만성질환 관리에 도움이 될 수 있는 기초자료를 제공하고자 하였다. 이러한 연구목적 달성을 위해 연관규칙분석을 통해 연령층 및 만성질환별 생활습관 군집 유형을 파악하고, 연관규칙의 유의성 평가 및 타당성 검증 과정을 거쳐 유의한 연관규칙을 도출하였다. 그 외에도 대표적 생활습관 군집을 예시로 하여 생활습관 군집에 영향을 미치는 개인특성 요인을 파악하였다.

연령층별로 대상자의 특성을 파악한 결과, 모든 연령층에서 2가지 생활습관을 가지고 있는 대상자의 비율이 가장 높게 나타났으며, 생활습관이 군집하는 경향이 있는 것으로 주장한 기존 연구와 일치한다 [29]. 또한 생활습관별로 높은 비율을 보이는 연령층이 다르게 나타났

으며, 기존 연구에서도 연령에 따라 건강과 관련된 생활습관의 실천 정도에 차이가 있다는 것을 파악하였다[20].

또한 연관규칙분석 결과 만성질환 유병에 영향을 미치는 다양한 생활습관 군집들이 도출되었다. 본 연구의 유의한 연관규칙은 주로 두 가지 또는 세 가지의 생활습관 군집으로 구성되었고, 많게는 다섯 가지 생활습관 군집으로 구성된 연관규칙도 나타났다. 건강위험행위들의 연관성을 분석한 이전 연구에서도 대부분 1-2개[9,30] 또는 2-3개 [31]의 건강위험행위 간에 연관성을 보이는 것으로 나타났다. 이러한 연구 결과는 개별적 생활습관에 대한 관리뿐만 아니라 복합적인 생활습관에 대한 보다 적극적인 관리가 필요하다는 것을 의미한다.

모든 만성질환과 연령대에서 신체적 비활동과 부적절한 체중조절로 구성된 생활습관 군집이 가장 많이 나타났는데, 이는 만성질환 예방 및 관리를 위해서는 운동부족과 비만을 우선적으로 개선해야 함을 의미한다. 또한 모든 만성질환에서 세 연령층에 공통적으로 나타난 생활습관 군집도 있었고, 반대로 특정 연령층에서만 나타난 생활습관 군집도 존재하였다. 고혈압의 경우에는 과도한 음주와 부적절한 체중 조절 군집과 흡연과 부적절한 체중조절 군집이 모든 연령층에서 나타난 반면에, 다른 연령대보다 유의한 연관규칙이 많이 도출된 중년층에서는 해당 연령대에서만 나타난 생활습관 군집들이 있었다. 당뇨병의 경우에는 노년층에서 나타난 생활습관 군집들이 중년층에서도 동일하게 나타났으며, 중년층에서는 이외에 몇 개의 생활습관 군집들이 더 도출되었다. 관절염의 경우에는 중년층에서 나타난 생활습관 군집들이 노년층에서도 동일하게 나타났으며, 노년층에서는 신체적 비활동과 부적절한 수면 군집이 추가적으로 나타났다. 공통적으로 나타난 생활습관 군집은 연령과 상관없이 특정 만성질환 관리를 위해 해당 생활습관을 가장 우선적으로 개선해야 함을 의미하며, 특정 연령층에서만 나타난 생활습관 군집은 연령층의 특성을 고려하여 해당 생활습관을 개선해야 함을 의미한다.

도출된 연관규칙을 대상으로 민감도와 특이도 값을 계산해서 타당성을 검증하는 과정도 진행했으며, 모든 생활습관 군집의 특이도가 민감도보다 크게 나타났다. 민감도가 낮게 나타난 이유는 전반적으로 전체 대상자 중 만성질환을 앓고 있는 대상자의 수가 적었기 때문이다. 본 연구에서는 생활습관을 동시에 가지고 있는 대상자를 생활습관 군집에 포함했는데, 해당 생활습관 중 하나라도 가지고 있는 대상자를 군집에 포함하게 되면 민감도 값을 높일 수 있다.

이외에도 로지스틱 회귀분석을 통해 개인 특성과 대표 생활습관 군집으로 선정한 세 개 군집의 관련성을 파악한 결과 개인특성 요인 별로 군집에 포함될 확률에 통계적으로 유의한 차이가 있는 것으로 나타났다. 성별, 거주 지역, 직업은 세 개 군집에서 모두 유의한 관련성을 나타낸 반면에, 교육수준은 중년층의 생활습관 군집에서만 유의한 관



련성을 보였다. 특히 직업 요인이 세 개의 생활습관 군집과 유의한 관련성을 보인 것은 사회경제적 계층요인이 만성질환 유병에 영향을 미치는 것으로 나타난 기존 연구의 결과와 유사하다[32,33]. 많은 연구들이 건강행위를 질병의 주요 결정요인으로 강조해 왔지만, 건강행위와 건강과의 관련성이 실제적으로는 건강행위와 사회경제적 지위와의 관련성 때문이라고 주장하는 연구들이 늘어나고 있다[34].

본 연구에서 생활습관 군집을 파악하는 방법으로 연관규칙분석을 활용하였다는 점은 기존 연구에서 주로 관찰치 수와 기대치 수의 비(observed to expected ratio, O/E ratio) 또는 생활습관의 개수의 합을 이용하는 방법과 차별성을 가진다. 연관규칙분석은 O/E값과 달리 상호 배타적인 군집이 아니라 건강위험행위의 종류가 적은 군집부터 많은 행위를 포함하는 군집까지 모두 고려할 수 있다는 장점이 있다. 또한 건강 수준을 향상시키기 위한 생활습관 개선이 목적일 때에는 생활습관이 어떻게 연관되어 있는지 파악하는 것이 중요하기 때문에 생활습관 개수의 합을 이용하는 방법보다 연관규칙분석이 적합한 방법이다.

또한 본 연구에서 도출된 연관규칙은 생활습관과 만성질환 간에 특정한 관계를 보인다는 것을 의미하는데, 이는 만성질환 관리 및 예방을 위한 생활습관 교정에 중요한 시사점을 준다. 기존의 생활습관 개선 프로그램은 생활습관에 대한 전반적인 스크리닝 부족으로 인해 효과적인 개선 방법을 계획하지 못하는 경우가 많았다[35]. 보건의로 시스템에서 생활습관 군집에 대한 정보가 충분하면 건강증진 정책을 좀 더 효과적으로 시행할 수 있다[18]. 한 가지 생활습관보다 생활습관 군집에 대한 교정이 더 큰 효과가 있다는 것은 여러 연구에서도 입증되었다[36,37]. 특히 2018년부터 정부에서 진행하고 있는 ‘일차의료 만성질환관리 시범사업’의 질병관리계획 및 생활개선교육 프로그램에서는 본 연구에서 도출한 만성질환별 생활습관 군집이 중요한 근거자료로 활용될 수 있다.

하지만 본 연구는 몇 가지 한계점을 가진다. 첫째, 본 연구에서는 대상자의 한 시점에서의 특성을 조사한 단면연구 자료를 활용했기 때문에 시간적 요인을 반영하지 못하였다. 일반적으로 만성질환은 다수의 생활습관 위험요인에 노출된 이후 발병시점까지 오랜 시간이 걸리는 특성이 있다. 대상자가 생활습관을 보유한 기간에 따라 만성질환에 미치는 영향에 차이가 있을 수 있다는 가능성을 고려하지 못하였다. 둘째, 본 연구에서 활용한 변수는 모두 자기기입식 설문을 통해 수집되었기 때문에 응답자의 주관성에 의존해야 한다는 단점이 있다. 응답자가 사회적으로 바람직한 생활습관을 고려해서 실제 자신의 흡연, 음주, 신체적 활동 여부 등과 다르게 응답했을 가능성이 있다. 셋째, 본 연구에서 정의한 생활습관 기준은 연구자에 따라 달라질 수 있으며, 그에 따라 연구 결과도 다르게 나타날 수 있다. 본 연구에서는 기존 보건의료 문헌 등을 통해 특정 횟수 또는 시간의 충족 여부로 부정적

생활습관을 정의했는데, 이 기준을 다르게 설정하면 도출되는 생활습관 군집의 수에 차이가 있을 수 있다.

이와 같은 한계점에도 불구하고 본 연구는 보건의로 빅데이터에 연관규칙분석을 적용하여 의미 있는 결과를 도출하는 것이 가능함을 제시하였다는 점에서 학문적 의의가 있다. 본 연구에서 적용한 분석 방안은 다른 연도, 지역, 만성질환 등에 확장하여 적용할 수 있다는 장점이 있다. 향후 다양한 방향으로 분석을 확대하여 자료를 축적한다면 더 정확하고 신뢰성 높은 생활습관 개선 프로그램의 근거자료를 제시할 수 있을 것으로 기대된다.

## ORCID

Eul Hee Roh <https://orcid.org/0000-0002-9459-9995>

Sang Chan Park <https://orcid.org/0000-0003-4488-3258>

## REFERENCES

1. World Health Organization. Global status report on noncommunicable diseases 2010. Available at [https://www.who.int/nmh/publications/ncd\\_report2010/en/](https://www.who.int/nmh/publications/ncd_report2010/en/) [accessed on November 4, 2019].
2. Korea Centers for Disease Control and Prevention. 2018 Current state and issue of chronic disease. Cheongju: Korea Centers for Disease Control and Prevention; 2018 (Korean).
3. Health Insurance Review and Assessment Service. 100 Disease statistics in living. Wonju: Health Insurance Review and Assessment Service; 2018 (Korean).
4. World Health Organization. Innovative care for chronic conditions. Available at <https://www.who.int/chp/knowledge/publications/iccreport/en/> [accessed on October 31, 2019].
5. Lalonde M. A new perspective on the health of Canadians. Ottawa: Government of Canada; 1974.
6. Lee JJ, Yang JH, Hwang TY. Clustering of lifestyle risk factors in urban poor and rural adults. Korean J Health Educ Promot 2005;22(4):167-177 (Korean).
7. Schuit AJ, Loon AJ, Tijhuis M, Ocke M. Clustering of lifestyle risk factors in a general adult population. Prev Med 2002;35(3):219-224. DOI: 10.1006/pmed.2002.1064
8. Fine LJ, Philogene GS, Gramling R, Coups EJ, Sinha S. Prevalence of multiple chronic disease risk factors. 2001 National Health Interview Survey. Am J Prev Med 2004;27(2 Suppl):18-24. DOI: 10.1016/j.

- amepre.2004.04.017
9. Chou KL. The prevalence and clustering of four major lifestyle risk factors in Hong Kong Chinese older adults. *J Aging Health* 2008;20(7):788-803. DOI: 10.1177/0898264308321082
  10. Kang KW, Sung JH, Kim CY. High risk groups in health behavior defined by clustering of smoking, alcohol, and exercise habits: National Health and Nutrition Examination Survey. *J Prev Med Public Health* 2010;43(1):73-83 (Korean). DOI: 10.3961/jpmph.2010.43.1.73
  11. Lee JM. Level of alameda 7 health behaviors execution by the age bracket and socioeconomic class and their influences on self rated health in Korean adults [dissertation]. Inje University; Korea, 2011.
  12. Kang HJ. National-level use of health care big data and its policy implications. *Health Welf Policy Forum* 2016;238:55-71 (Korean).
  13. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, et al. *Big data: the next frontier for innovation, competition, and productivity*. New York, USA: McKinsey Global Institute; 2011.
  14. Korea Centers for Disease Control and Prevention. *Community Health Survey's guidebook for 2016 raw data*. Cheongju: Korea Centers for Disease Control and Prevention; 2018 (Korean).
  15. Korea Centers for Disease Control and Prevention. *Community Health Survey*. Available at <https://chs.cdc.go.kr/chs/index.do> [accessed on February 22, 2020].
  16. Kim TI, Choi YY, Lee KH. Analysis on the differences in medical service usage in terms of income levels. *Korean Soc Secur Stud* 2008;24(3):53-75 (Korean).
  17. Jung Y, Cho YT, Oh JH. Differential effects of family income on self-rated health by age: analysis of Seoul Citizens Health Indicators Survey 2001, 2005. *J Prev Med Public Health* 2007;40(5):381-387 (Korean). DOI: 10.3961/jpmph.2007.40.5.381
  18. Park SH. An association rule mining-based framework for understanding lifestyle risk behaviors [dissertation]. Seoul University; Korea, 2014.
  19. Yeo Y, Ma SH, Park SK, Chang SH, Shin HR, Kang D, et al. A prospective cohort study on the relationship of sleep duration with all-cause and disease specific mortality in the Korean Multi-center Cancer Cohort Study. *Korean J Prev Med Public Health* 2013;46(5):271-281 (Korean). DOI: 10.3961/jpmph.2013.46.5.271
  20. Koo JO, Park S. Analysis of body composition, dietary behaviors and life style of 30-50 year old women living in Seoul. *Korean J Community Nutr* 2012;17(4):440-449 (Korean). DOI: 10.5720/kjcn.2012.17.4.440
  21. Park YR, Wee H, Kim SJ. Lifestyle, depression, and health related quality of life among women in Jeju province. *J Korean Acad Community Health Nurs* 2010;21(2):148-155 (Korean).
  22. Yoo JS, Jeong JI, Park CG, Kang SW, Ahn JA. Impact of life style characteristics on prevalence risk of metabolic syndrome. *J Korean Acad Nurs* 2009;39(4):594-601 (Korean). DOI: 10.4040/jkan.2009.39.4.594
  23. Lee JH, Kim SL, Cho WS. Analysis of associations between diseases using healthcare public data focusing on dementia. *Database Res* 2017;33(2):137-148 (Korean).
  24. Jeon SA. Analysis the association and effect variables of hypertension-related diseases based on the Korea National Health and Nutrition Examination Survey [dissertation]. Korea University; Korea, 2017.
  25. Kim SH. Analysis of associated disease in patients with chronic obstructive pulmonary disease by applying parallax association rules. *HIRA Policy Issues* 2018;12(1):53-59 (Korean).
  26. Kim SL, Lee KH, Cho WS. Using association analysis to find diseases related to childhood obesity. *Database Res* 2017;33(3):52-60 (Korean).
  27. Yoo KS, Jin SH. A study on improvement of standardized lift in association rules analysis. *J Korean Data Anal Soc* 2016;18(1):189-197 (Korean).
  28. Song SW. Using the receiver operating characteristic (ROC) curve to measure sensitivity and specificity. *Korean J Fam Med* 2009;30(11):841-842 (Korean). DOI: 10.4082/kjfm.2009.30.11.841
  29. World Health Organization. *Global status report on noncommunicable disease 2010*. Available at [https://www.who.int/nmh/publications/ncd\\_report2010/en/](https://www.who.int/nmh/publications/ncd_report2010/en/) [accessed on February 22, 2020].
  30. Chioloro A, Wietlisbach V, Ruffieux C, Paccaud F, Cornuz J. Clustering of risk behaviors with cigarette consumption: A population based survey. *Prev Med* 2006;42(5):348-353. DOI: 10.1016/j.ypmed.2006.01.011
  31. Poortinga W. The prevalence and clustering of four major lifestyle risk factors in an English adult population. *Prev Med* 2007;44(2):124-128. DOI: 10.1016/j.ypmed.2006.10.006
  32. Kim HR. The relationship of socioeconomic position and health behaviors with morbidity in Seoul, Korea. *Health Soc Welf Rev* 2005;25(2):3-35 (Korean).
  33. Son M. The relationship of social class and health behaviors with morbidity in Korea. *Korean J Prev Med Public Health* 2002;35(1):57-64 (Korean).
  34. Smith GD, Blane D, Bartly M. Explanations for socioeconomic differentials in mortality: evidence from Britain and elsewhere. *Eur J Public*

- Health 1994;4(2):131-144.
35. Goldstein MG, Whitlock EP, DePue J. Multiple behavioral risk factor interventions in primary care. Summary of research evidence. *Am J Prev Med* 2004;27(2 Suppl):61-79. DOI: 10.1016/j.amepre.2004.04.023
36. Nigg CR, Allegrante JB, Ory M. Theory-comparison and multiple behavior research: common themes advancing health behavior research. *Health Educ Res* 2002;17(5):670-679. DOI: 10.1093/her/17.5.670
37. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;364(9438):937-952. DOI: 10.1016/S0140-6736(04)17018-9