

국민건강영양조사 자료를 활용한 라이프스타일 위험요인과 다중이환간의 연관관계분석

이현주*, 명성민**†

*증원대학교 보건행정학과, 부교수, **증원대학교 보건행정학과, 교수

Association Rule Analysis between lifestyle risk behaviors and multimorbidity: Findings from KHANES

Hyun-Ju Lee*, Sungmin Myoung**†

*Dept. of Health Administration, Jungwon University, Associate Professor

**Dept. of Health Administration, Jungwon University, Professor

ABSTRACT

Objectives: This study used an efficient data mining algorithm to explore association rules between the lifestyle risk behaviors and multimorbidity (having more than one chronic disease) in Korean adults.

Methods: We used data from the 8th Korean National Health and Nutrition Examination Survey(2019-2020) for 7,609 adults aged ≥ 19 years. This study was undertaken where 6 lifestyle risk behaviors and 11 morbidities were analyzed using R and Rstudio for the ARM.

Results: Among 117 association rules, combinations of hypertension, dyslipidemia and diabetes, hypertension were important role in inadequate sleep, physical inactivity and inadequate weight.

Conclusion: The findings of this study are significant because they demonstrate the importance of lifestyle risk factors and the role of multiple chronic diseases using big data analytics such as association rule mining. We recommend developing selective and focused health education programs, such as exercise programs to address physical inactivity, dietary interventions to address inadequate weight, and mental health education programs to address inadequate sleep.

Key words: Association Rule Mining, Chronic Disease, Lifestyle Risk Factor, KHANES, Health Education

* 본 과제(결과물)는 2023년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다(2021RIS-001(1345370811)).

접수일 : 2023년 12월 15일, 수정일 : 2023년 12월 18일, 채택일 : 2023년 12월 19일

교신저자 : 명성민(28024, 충북 괴산군 괴산읍 문무로 85)

Tel: 043-830-8604, Fax: 043-830-8675, E-mail: smmyoung@jwu.ac.kr

I. 서론

인간의 건강 수준에 영향을 미치는 것은 개인적 특성과 상황적 맥락에서 실천하는 일상적인 행동 유형이라 알려져 있다(Saint Onge 외, 2017). 이러한 일상적인 행동은 흡연, 음주, 식습관 등과 같은 라이프스타일로 설명될 수 있다. 이는 건강증진 및 개선을 위한 결정요인으로 구분되며, 수정가능하다고 제시되었다(WHO, 2021). 또한 Li 등(2020)에 의하면, 라이프스타일이 일상에서 실천되는 행동이 건강한 방식으로 나타날수록 기대수명에 긍정적인 영향을 미치며 만성질환의 발병률을 낮춘다고 주장하였다.

라이프 스타일과 관련된 선행연구로, 건강한 라이프스타일 패턴을 더 많이 준수할수록 인지 기능저하가 느려지고 건강에 이점이 나타난다고 Mamalaki 등(2022)이 노인코호트 연구를 통하여 제안하였다. 또한 Hekatpou 등(2013)의 연구에서 라이프스타일 교육프로그램이 노인의 독립적인 활동 수준과 유의미한 영향이 존재하는 것으로 보고되었다.

우리나라의 경우 이러한 라이프스타일 건강위험 행위에 대하여 정영호 등(2006)이 우리 국민의 질병 부담의 31.3%를 차지하고, 박정근(2010)에 의하면 이러한 질병 부담 중 비만과 음주에 대한 부분이 큰 폭으로 상승했다는 것을 연구한 사례 등이 있다. 또한 Park 등(2014)은 제4기 국민건강영양조사 자료를 이용하여 건강위험행위에 대한 군집분석을 통하여 이를 기반으로 유형분석 연구를 통해 건강행위방법을 제시하였다.

이와 같은 연구사례들을 볼 때 개인의 특성과 상황에 따라 영향을 미치는 라이프스타일 위험요인의 특성을 고려하여 만성질환에 영향을 미치는 라이프스타일 위험요인을 좀더 구체적인 제시하는 분석방법이 필요하다. 또한 WHO의 2021년 보고에 의하면 라이프스타일은 수정 가능한 영역이므로 만성질환에 영향을 미치는 라이프스타일 위험요인을 구체적으로 도출한다면 보건교육적 접근을 통한 수정 즉, 라이프

스타일 개선이 가능할 것이다.

따라서 본 연구에서는 이러한 라이프스타일 위험요인과 만성질환에 대한 다중이환(multimorbidity)을 빅데이터분석기법 중 연관성 분석(association rule mining: 이하 ARM)을 이용하여 제시하고자 한다.

동반질환(comorbidity)은 개인이 한 지표질환(index disease)에 동반되어 가지는 다른 질환들을 의미하며, 이에 따라 지표질환의 임상경과가 바뀔 수 있으며, 동반질환을 미인지 시 역학연구에서 비뚤림(bias)이 발생할 수 있어 예후의 원인을 잘못 해석할 수 있다(정희원, 2014). 다중이환(multimorbidity)은 한 개인이 여러 질환에 동시에 이환되어 있는 현상을 말하며, 동반질환과는 달리 여러 질환이 전반적으로 환자에게 포괄적인 영향, 예를 들면 환자의 건강수준, 삶의 질에 대해 보다 관심을 갖는다(Van den Akker M 등, 1996). 일반적으로 다중이환의 작업적 정의방법은 다음과 같은 세 가지의 유형으로 분류한다고 알려져 있다(정희원, 2014). 첫 번째 유형은 개인에서 발견되는 질병의 개수(통상 2-3개이상)로 나타내는 방법으로서 일반 역학연구에서 손쉽게 정의할 수 있으나, 개인에 미치는 각각의 질병의 기능적 영향을 반영하지 못하는 단점이 있다. 두 번째 유형은 Charson Comorbidity Index(Charlson ME 등, 1987)와 Cumulative Illness Rating Scale(Linn BS 등 1968)와 같은 질병의 개수와 중증도를 함께 반영한 지표로서 중증도를 고려하여 다중이환의 예후와 연관도를 관리 할 수 있다. 세 번째는 질병유무, 증상유무, 인지/기능 상태 등을 동시에 평가하는 방법이다.

한편 연관성분석(association rule mining: ARM)은 '장바구니 분석'이라고도 하며 데이터에 빈발하는 속성을 찾아내고 이 중에서 서로 연관이 있는 규칙을 발견하는 기법을 의미한다(이정진, 2011). 이 방법은 패턴 마이닝 분야에서 많이 활용되며, 특히 보건학 및 역학분야에서도 많은 선행연구들이 제시되었다. 박소현 등(2014)은 국민건강영양조사 자료를 이용하

여 건강위험행위에 대한 패턴 분석을 통하여 군집을 제시하였으며, Burton SH 등(2014)은 공공보건설문조사에 대한 ARM 기반 알고리즘을 제시하여 분석결과를 제시하였으며, Kim 등(2018)은 ADHD 동반질환에 대하여 보험공단자료를 이용하여 분석결과를 제시하였다. 하지만, 건강증진 및 개선을 위한 결정요인인 라이프스타일 위험요인과 만성 동반 질환의 패턴을 살펴본 연구는 거의 전무한 상황이다.

이에 착안하여, 본 연구는 한국의 질병관리청의 국민건강영양조사자료를 이용하여 19세 이상 성인의 라이프스타일 위험요인에 따른 만성 동반 질환 패턴을 ARM 및 시각적분석을 통하여 제시하고자 한다. 이를 통해 국내 성인 라이프스타일 위험요인에 따른 영향력을 근거자료로 제공하고 건강증진과 보건교육의 필요성을 제시하고자 한다.

II. 연구방법

1. 연구 설계

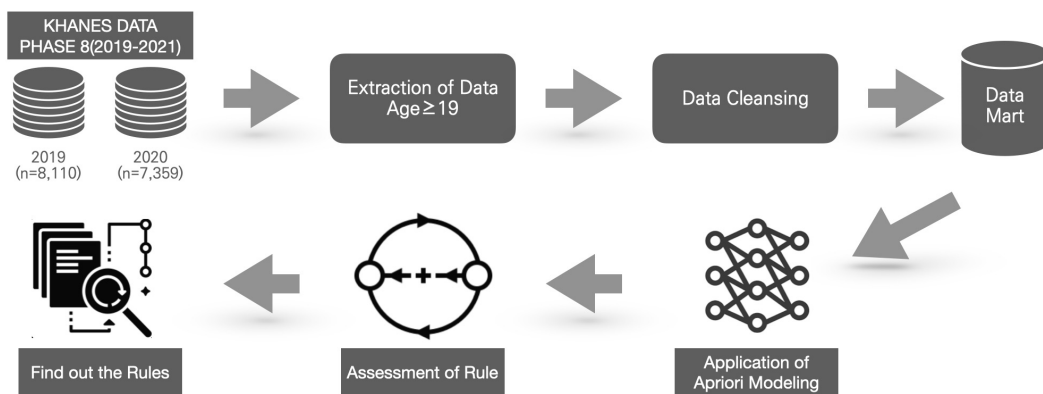
본 연구는 19세 이상 성인의 라이프스타일 위험요인과 만성질환에 대한 다중이환을 연관성 분석을 수

행하기 위해 질병관리청에서 제공하는 2019년-2020년 국민건강영양조사자료를 이용한 이차자료 분석연구로서 단면조사연구(cross-sectional study)이다.

2. 자료수집

본 연구는 질병관리청 국민건강영양조사 자료로서 제8기 1차-2차년도(2019-2020년)기간 동안의 이차 자료이다. 국민건강영양조사는 지정통계로서 통계법 제 17조에 근거하며, 국민의 건강행태, 만성질환 유병현황, 식품 및 영양섭취실태에 관하여 국민건강증진법 제16조에 근거하는 법정조사이다. 국민건강영양조사의 모든 결과는 가중치를 적용한 가중표본평균 또는 가중 표본분율로 제시되었으며, 가중치는 표본추출률, 응답률, 해당연도 모집단 인구구성비를 반영하여 산출되었다. 본 연구에서 수집된 제8기 조사에서 1차년도(2019년) 참가가구는 3,670가구, 2차년도 3,314가구로 전년대비 약 360가수가 감소하였으며, 참여자는 1차년도 8,110명, 2차년도 7,359명으로 전년대비 약 750명 감소하였다.

본 연구에서는 2년동안 조사된 15,469명에서 19세 이상 7,609명을 선정하고 이를 최종적인 분석대상으로 하였다.



<Figure 1> Analytical Framework for Association Rule Mining

3. 측정도구 및 분석방법

1) 라이프스타일 건강위험행위

라이프스타일 위험요인은 앨러미다 카운티 연구(Alameda county study)에서 제시된 7가지 건강실천행위변수를 정의하였으며, 이는 건강행태의 종합적인 수준을 파악하기 위한 지표로 알려져 있다(Schoenborn, 1986). 7가지 라이프스타일 위험요인은 이분형 자료로 구성되어 있으며, 세부적으로 현재 흡연(이하 CS: current smoking), 과도한 음주(이하 HD: heavy drinking), 신체적 비활동(이하 PI: physical inactivity), 부적절한 체중(이하 IW: inadequate weight), 부적절한 수면(이하 IS: inadequate sleep), 빈번한 아침 결식(이하 BS: breakfast skipping), 그리고 간식 섭취(이하 FS: frequent snacking)이다. 본 연구에서는 7개의 건강위험행위에 대하여 국민건강영양조사에서의 자료를 아래와 같은 기준으로 정상과 질환의 2가지로 변환하였다(Park 등, 2014).

- CS는 현재 흡연하는 경우로 정의함
- HD는 건강설문에서 음주량을 계산하여 음주량 25g/day 이상으로 정의함
- PI는 중등도 신체활동을 최소 30분 이상 일주일에 5일 이상 실천하거나 격렬한 신체활동을 최소 20분 이상 일주일에 3일 이상 실천하는 지침을 지키지 않는 경우로 정의함
- IW는 체중조절을 안 해서 발생하는 부적절한 체중에 대한 항목으로, 저체중(BMI(18.5 kg/m²)) 또는 과체중, 비만(BMI ≥ 25 kg/m²)인 경우로 정의함
- IS는 수면시간이 하루 7시간 미만 또는 8시간 이상인 경우로 정의함
- BS는 건강설문에서 하루라도 아침식사를 안 하였다고 응답한 경우로 정의함
- FS는 빈번한 간식 섭취로서 하루 평균 3번 이상 섭취하는 경우로 정의함

2) 만성질환요인

만성질환은 고혈압, 이상지질혈증, 뇌졸중, 울혈성 심부전, 심근경색, 협심증, 관절염, 폐결핵, 천식, 당뇨, 우울증의 11개 질환으로 설정하였으며, 국민건강영양조사 건강설문에서 아래와 같은 경우로 정의하였다(Goodman 등, 2016).

- 건강설문에서 각 질환에 대하여 현재 앓고 있거나 치료받고 있는 경우
- 건강설문에서 최근 1년간 3개월 이상 앓았던 경우

3) 연관성 분석

연관성 분석(이하 ARM: Association Rule Mining)은 '장바구니 분석'이라고도 불리는데, 데이터마이닝에서 자주 활용되는 분석방법 중 하나이다. ARM은 이분형자료의 형태로 구성된 변수들이 데이터셋 내에서 동시에 나타나는 빈도를 확인하기 위해 이용된다. 이 분석의 목적은 변수들 간에 상관성이 높은 그룹을 찾아내거나 특정한 타겟과 관련해 상관성이 높은 변수들을 찾아내는 것이다(Tan 등, 2004).

ARM은 대량 데이터에 숨어 있는 연관규칙(association rule)을 찾는다. 예를 들면 '흡연⇒음주'와 같은 규칙은 흡연을 하는 대상자들이 음주도 할 가능성이 많은 것을 의미한다. 연관 규칙 마이닝은 '장바구니 분석'이라는 용어에서와 같이 유통업에서의 판매시점의 거래내역을 분석하는 것에서 시작되었지만 유통업 이외에 환자의 의무기록에서 여러 치료가 같이 이루어진 경우 합병증 발생의 징후를 알아보는 데 유용하다(이인희 등, 2010). 연관 규칙 마이닝을 의학 분야에 최근 접목을 시도한 Tai 등(2009) 및 Kim 등(2018)는 연관 규칙 마이닝을 셋 또는 그 이상의 항목 사이의 연관성을 확인하는 데 사용하였고, 연관 규칙에서 나타난 신뢰도(confidence)는 역학에서 동반질환(comorbidity)과 수학적으로 같은 의미라고 하였으며, 연관 규칙 마이닝 중 선행적 알

고리즘 모델링(Apriori Algorithm Modeling)을 이용하여 주의력 결핍 과다활동 장애(attention deficit hyperactivity disorder, ADHD)와 동반되는 정신과 적 질환 사이의 연관성을 제시하였다.

연관규칙을 판단하는 기준은 지지도(support), 신뢰도(confidence), 향상도(lift) 3가지이다 (임제순 등, 2010). 지지도는 규칙 'A⇒B'의 경우 전체에 대한 A항목과 B항목이 동시에 일어나는 확률을 의미하며 다음과 같다.

$$\text{Support}(A \Rightarrow B) = \Pr(A \cap B)$$

신뢰도는 A항목이 발생한 경우 중 B가 발생하는 경우의 조건부 확률을 의미하고 다음과 같다. 일반적으로 신뢰도는 지지도와는 달리 대칭적이지 않다고 알려져 있다.

$$\text{Confidence}(A \Rightarrow B) = \Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

향상도는 지지도, 신뢰도가 높은 연관성 규칙을 나타내기 위한 지표로 쓰이며 1보다 크다면 A항목과 B항목의 연관도가 높음을 의미하며 1보다 작다면 A항목과 B항목의 연관도가 작다는 것을 의미한다.

$$\text{Lift}(A \Rightarrow B) = \frac{\Pr(B|A)}{\Pr(B)} = \frac{\Pr(A \cap B)}{\Pr(A)\Pr(B)}$$

일반적으로 연관분석을 수행하는 모형으로 선택적 규칙(Apriori algorithm)을 적용하는데, 그 특성은 다음과 같다(이정진, 2010).

- 최소지지도 이상을 갖는 항목집합인 빈발항목집합(frequent item set)을 생성하기 위하여 먼저 1-항목으로 이루어진 집합에 대하여 지지도를 계산한다.
- 최소지지도를 넘는 1-항목 빈발항목집합을 이용하여 불필요한 2-항목집합에 대한 가지치기를 하고, 2-항목 빈발항목집합 후보를 생성한다.
- 위의 빈발항목집합 후보들 중 최소지지도 기준

을 넘는 2-항목 빈발항목집합을 확정하고, 이를 이용하여 3-항목집합에 대한 가지치기를 하고 3-항목 빈발항목집합 후보를 생성한다.

- 위의 방법을 k-항목 빈발항목집합을 확정할때 까지 반복한다.

4. 자료 분석 방법

자료 분석은 R 4.3.2 및 RStudio 2023.09.1.+494를 이용하였다. 대상자의 일반적 특성은 빈도 또는 평균, 표준편차로 구성된 기술통계(descriptive statistics) 분석을 수행하였다. 대상자의 성별, 라이프스타일, 만성질환 등의 분포는 빈도와 백분율을 구하였고, 성별에 따른 분포에 통계적 차이여부를 확인하기 위해 카이제곱 검정을 수행하였다. 통계적 검정은 유의수준 5%로 설정하였다.

11개의 만성질환 및 라이프스타일 위험요인에 대한 연관성을 결정하기 위하여 선택적 알고리즘 모델링(Apriori Algorithm Modeling)을 이용한 연관성 분석을 수행하였다.

III. 연구결과

1. 대상자의 성별 및 연령

대상자의 성별은 남자가 3,474명(45.66%), 여자가 4,135명(54.34%)로 여자가 더 많았다. 연령은 평균 50.32±16.24세 였다.

2. 라이프스타일 위험요인 및 만성질환 발생 빈도 분석

성별로 라이프스타일 위험요인 만성질환에 대한 빈도분석 및 이에 대한 카이제곱검정결과는 <Table 1>에 제시하였다. 부적절한 수면, 아침 결식, 빈번한 간식섭취의 3개 요인을 제외한 나머지 흡연, 음주, 신

체적 비활동, 부적절한 체중 요인은 모두 통계적으로 유의하게 나타났다(p -value<0.001). 현재 흡연의 경우 남자가 32.21%, 여자가 5.73%로 남성이 높게 나타났다으며, 음주의 경우에도 남자가 여자보다 약 3배 정도 높게(남자 18.62%, 여자 5.78%) 나타났다. 또한 부적절한 체중도 남자가 42.63%, 여자가 29.14%로 나타났다으나, 신체적 비활동의 경우는 여성(59.32%) 이 남성(52.58%)보다 높게 나타났다.

만성질환의 경우 천식을 제외한 나머지 10개 질환 모두 통계적으로 유의하게 나타났다. 전반적으로 남자가 여자보다 높은 비율로 나타났으나, 이상지질혈증은 여자가 20.87%, 남자가 18.80%로서 여자가 높게 나타났으며, 관절염의 경우에는 남자가 5.99%, 여자가 16.44%로서 약 2.74배 정도 여자가 높았다. 또한 우울증의 경우에는 여자가 6.80%, 남자가 3.08%로서 약 2배 정도 높게 나타났다.

〈Table 1〉 Descriptive statistics of the measurement variables

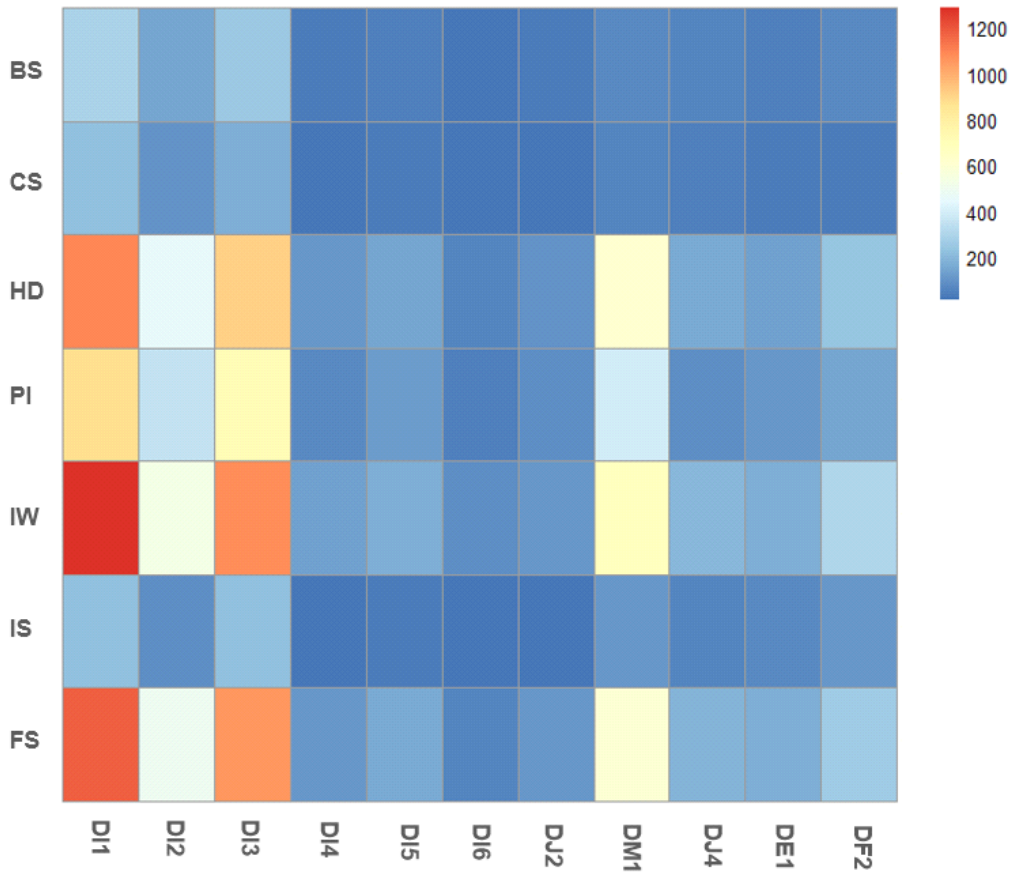
Variables	Male n(%)	Female n(%)	Chi-Sq (p -value)	
Lifestyle Risk Factor	Current Smoking(CS)	1119 (32.21%)	237 (5.73%)	903.84 (<0.001)
	Heavy Drinking(HD)	647 (18.62%)	239 (5.78%)	33.89 (<0.001)
	Physical Inactivity(PI)	1830 (52.68%)	2453 (59.32%)	33.89 (<0.001)
	Inadequate Weight(IW)	1481 (42.63%)	1205 (29.14%)	150.41 (<0.001)
	Inadequate Sleep(IS)	2421 (69.69%)	2920 (70.62%)	0.78 (0.3783)
	Breakfast Skipping(BS)	966 (27.81%)	1118 (27.04%)	0.56 (0.4537)
	Frequent Snacking(FS)	2423 (69.75%)	2810 (67.96%)	2.82 (0.0932)
Chronic Diseases	Hypertension(DI1)	907 (26.11%)	829 (20.05%)	39.37 (<0.001)
	Dyslipidemia(DI2)	653 (18.80%)	863 (20.87%)	5.09 (<0.0241)
	Stroke(DI3)	100 (2.88%)	62 (1.50%)	17.23 (<0.001)
	Congestive Heart Failure(DI4)	154 (4.43%)	75 (1.81%)	44.37 (<0.001)
	Myocardial Infarction(DI5)	67 (1.93%)	21 (0.51%)	33.34 (<0.001)
	Angina Pectoris(DI6)	99 (2.85%)	59 (1.43%)	18.80 (<0.001)
	Arthritis(DM1)	208 (5.99%)	680 (16.44%)	200.29 (<0.001)
	Pulmonary Tuberculosis(DJ2)	163 (4.69%)	98 (2.37%)	30.73 (<0.001)
	Bronchitis(DJ4)	100 (2.88%)	130 (3.14%)	0.4535 (0.5007)
	Diabetes Mellitus(DE1)	402 (11.57%)	317 (7.67%)	33.65 (<0.001)
Depression(DF2)	107 (3.08%)	281 (6.80%)	53.86 (<0.001)	

3. 라이프스타일 위험요인과 만성질환 간의 히트맵 분석

라이프스타일 위험요인과 만성질환간의 상관관계를 확인하기 위하여 히트맵 분석(heatmap analysis)을 수행하였다.

히트맵 분석은 수치를 색상으로 표시하여 이를 열 분포형태의 비주얼한 그래픽으로 출력하는 것(heatmap)으로 주로 웹로그분석에 많이 사용하는 분석기법이다.

히트맵 분석은 수치를 색상으로 표시하여 이를 열 분포형태의 비주얼한 그래픽으로 출력하는 것(heatmap)으로 주로 웹로그분석에 많이 사용하는 분석기법으로 파란색에서 빨간색으로 가까워질수록 다빈도로 나타나는 형태이다. 히트맵 분석결과 상대적으로 라이프스타일 위험요인과 만성질환 간의 다빈도 요인은 부적절한 체중과 고혈압이 가장 높게 나타났으며, 다음으로는 고혈압과 빈번한 간식섭취로 나타났다. 이상지질혈증의 경우 빈번한 간식섭취, 부적절한 체중, 신체적 비활동이 상대적으로 높게 나타났다.



(CS: Current Smoking, HD: Heavy Drinking, PI: Physical Inactivity, IW: Inadequate Weight, IS: Inadequate Sleep, BS: Breakfast Skipping, FS: Frequent Snacking, D11: Hypertension, D12: Dyslipidemia, D13: Stroke, D14: Congestive Heart Failure, D15: Myocardial Infarction, D16: Angina Pectoris, DM1: Arthritis, DJ2: Pulmonary Tuberculosis, DJ4: Bronchitis, DE1: Diabetes Mellitus, DF2: Depression)

<Figure 2> Heatmap Result between lifestyle risk factor and chronic diseases

뇌졸중의 경우 빈번한 간식섭취, 부적절한 체중, 음주가 상대적으로 나타났으며, 관절염의 경우 부적절한 체중, 음주, 빈번한 간식섭취가 상대적으로 높게 나타났다.

4. 연관성 분석결과

Apriori 알고리즘을 이용한 연관성 분석결과 향상도(Lift) 별로 정렬한 연관규칙은 <Table 2>와 같다.

데이터마이닝 기법에서는 최소 지지도와 신뢰도를 만족하면서 향상도가 1이상의 경우에 연관성 규칙이 있다고 판단하는데 (Lim et al, 2010), 본 연구에서는 지지도 5%이상, 신뢰도 60% 이상으로 설정하였다. 그 결과 위의 최소지지도와 신뢰도를 만족시키는 규칙은 총 117가지가 나왔으나 다중이환을 가지는 경우 즉, 만성질환이 2개 이상 나타나는 경우로 한정하면 총 13가지가 나타났다.

<Table 2> ARM results using apriori algorithm(multimorbid)

Rule	lhs	rhs	support	confidence	lift	multimorbid
1	{IW}	⇒ {DI1,DI2,IS}	0.0597	0.6475	1.6102	DI1, DI2
2	{IW}	⇒ {DI1,DI2,FS}	0.0866	0.6429	1.5965	DI1, DI2
3	{IW}	⇒ {DI1,DI2}	0.0696	0.6405	1.5896	DI1, DI2
4	{IW}	⇒ {DI1,DI2,PI}	0.0609	0.7333	1.5684	DI1, DI2
5	{PI}	⇒ {DI1,DI2,IS}	0.0582	0.6654	1.2865	DI1, DI2
6	{PI}	⇒ {DI1,DI2,IS,FS}	0.0647	0.6632	1.2822	DI1, DI2
7	{PI}	⇒ {DI1,DI2}	0.0879	0.6306	1.2193	DI1, DI2
8	{PI}	⇒ {DI1,DI2,FS}	0.0722	0.6245	1.2074	DI1, DI2
9	{PI}	⇒ {DI1,DI2,IW}	0.0609	0.6222	1.2030	DI1, DI2
10	{IS}	⇒ {DI1,DI2,PI}	0.0582	0.8333	1.1676	DI1, DI2
11	{IS}	⇒ {DI1,DI2,PI,FS}	0.0647	0.8235	1.1539	DI1, DI2
12	{IS}	⇒ {DI1,DE1}	0.0686	0.7955	1.1146	DI1, DE1
13	{IS}	⇒ {DI1,DI2,FS}	0.0824	0.7755	1.0866	DI1, DI2

IW: Inadequate Weight, PI: Physical Inactivity, IS: Inadequate Sleep, FS: Frequent Snacking DI1: Hypertension, DI2: Dyslipidemia, DE1: Diabetes Mellitus

13가지의 연관규칙에서 가장 많이 나타난 다중이환은 고혈압(DI1)과 이상지질혈증(DI2)이었으며, 고혈압(DI1)과 당뇨(DM1)은 1개로 나타났다. 고혈압과 이상지질혈증을 동시에 가지는 경우에서 관련된 라

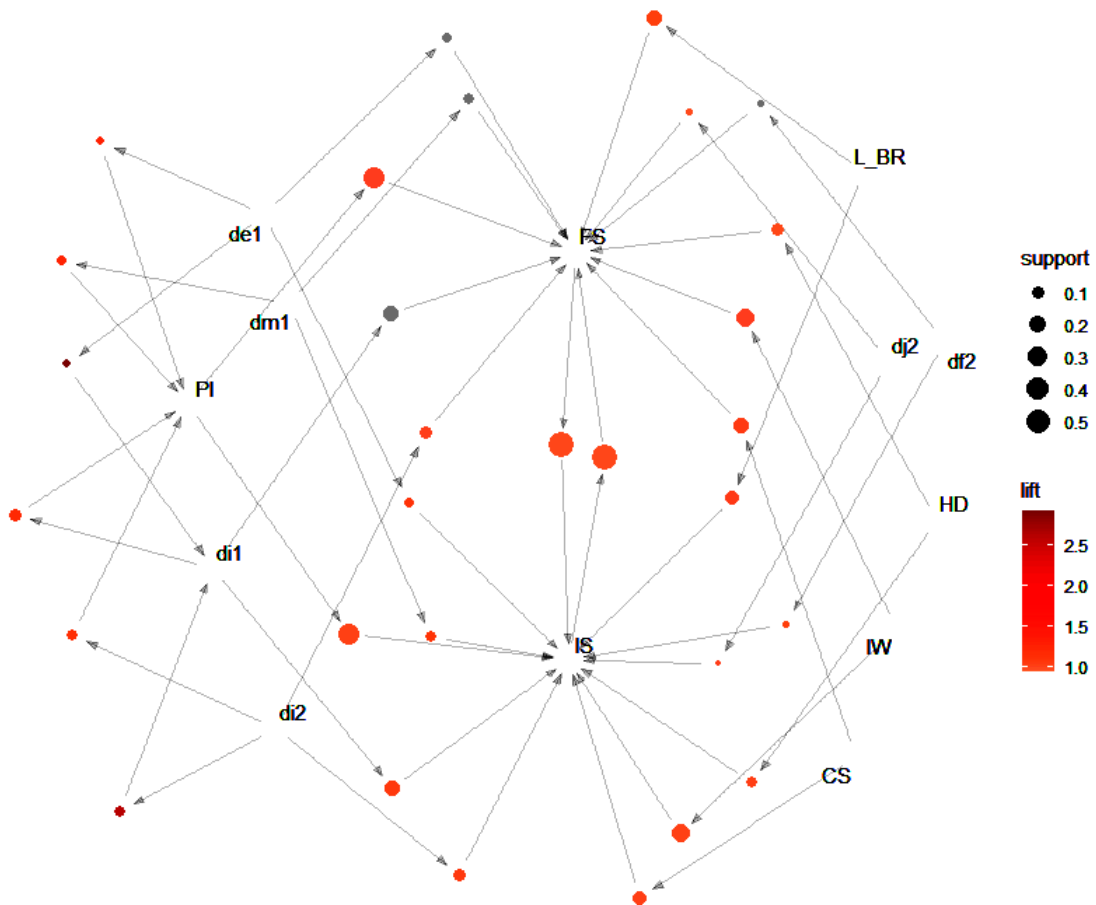
이프스타일 위험인자는 신체적 비활동, 부적절한 체중, 부적절한 수면과 관련된 것이 전반적으로 높은 향상도로 나타났다. 향상도가 가장 높은(1.6102)를 가지는 연관규칙을 살펴보면 부적절한 체중을 가지

고 있는 64.75%가 고혈압과 이상지질혈증, 부적절한 수면을 가지고 있는 사람이었다. 또한 전체집단의 5.97%가 부적절한 체중, 고혈압과 이상지질혈증, 부적절한 수면을 가지고 있다. 어떤 대상자가 부적절한 체중을 가질 가능성에 비해 고혈압과 이상지질혈증, 부적절한 수면을 가지는 대상자가 부적절한 체중을 가질 확률은 1.61배 높다는 것을 확인할 수 있다.

또한 결과에서 가장 지지도가 높은 경우는 8.79%로서 고혈압, 이상지질혈증, 신체적 비활동자인 경우로서 신체적 비활동여부가 다중질환여부에 영향이

크기 나타남을 확인할 수 있다.

라이프스타일 위험요인과 다중질환 사이의 연관성을 시각화를 위해 네트워크 그래프를 이용한 결과는 <Figure 3>과 같다. 사각형의 크기가 클수록 지지도가 높으며, 색깔이 붉어질수록 향상도가 높다는 것을 의미한다. 네트워크 그래프는 고혈압, 이상지질혈증을 중심으로 각각 신체적 비활동, 부적절한 체중, 부적절한 수면, 빈번한 간식섭취 등이 개별 네트워크들로 형성되었으며, 당뇨 또한 주요 연관규칙으로 나타남을 시각적으로 확인할 수 있다.



<Figure 3> Network graph for ARM results

IV. 결론 및 논의

본 연구에서는 국민건강영양조사 자료를 활용하여 만성질환에 대한 다빈도여부 및 이에 대한 다중질환을 파악하고 라이프스타일 위험인자를 고려한 연관성을 데이터 마이닝 기법 중 연관규칙분석(ARM)을 이용하여 분석하였다. 주요 결과는 다음과 같다.

첫째, Heatmap 분석결과 상대적으로 라이프스타일 위험요인과 만성질환 간의 다빈도 요인은 부적절한 체중과 고혈압이 가장 높게 나타났으며, 다음으로 고혈압과 빈번한 간식섭취로 나타났다. 특히, 고혈압의 경우 연구 대상자 중 1,790명이 고혈압을 앓고 있어 연구 대상자가 보유한 만성질환 총 11종 중에서 가장 유병 인원이 높았다. 따라서 중재가 우선적으로 요구되는 질환임을 확인할 수 있었다. 부적절한 체중과 빈번한 간식섭취가 고혈압 발생의 다빈도 요인임이 본 연구를 통해 확인되었을 뿐 아니라 연구 대상자들이 가장 많이 보유하고 있는 만성질환이 고혈압으로 확인되었으므로 고혈압이 있는 사람들 중 과체중, 비만, 저체중의 부적절한 체중군을 대상으로 정상체중으로 돌아갈 수 있도록 보건교육을 우선적으로 시행할 필요가 있다. 빈번한 간식섭취는 과체중, 비만과 관련이 있으므로 빈번한 간식섭취의 고혈압에 대한 위험성을 알리고 간식섭취를 조절할 수 있도록 식생활 습관 개선관련 보건교육 또한 병행하는 것이 체중조절에 보다 효과적일 것이다. 또한 고혈압 유병인구에게 보건교육을 우선하는 것안은 전 국민을 대상으로 하는 보건교육보다 한정된 자원에서 보다 비용효과를 높일 수 있는 방안으로 생각된다.

둘째, 연관규칙분석 결과 최소 지지도 및 신뢰도 및 향상도를 고려할 때, 주요 다중이환은 고혈압(DI1)과 이상지질혈증(DI2), 고혈압(DI1)과 당뇨(DE1)로 나타났다. 고혈압과 이상지질혈증을 동시에 가지는 경우에서 관련된 라이프스타일 위험인자는 신체적 비활동, 부적절한 체중, 부적절한 수면과 관련된 것이 전반적으로 높게 나타났다. 대사증후군에

서 위험요인은 고혈압, 낮은 HDL 콜레스테롤혈증, 고중성지방혈증, 복부비만, 고혈당의 총 5개 항목이 포함된다(국민건강보험공단, 2017). 이상지질혈증은 대사증후군 위험요인 중 낮은 HDL 콜레스테롤혈증, 고중성지방혈증과 관련이 있고 고혈당은 당뇨와 관련이 높다. 고혈압과 이상지질혈증, 고혈압과 당뇨가 다중이환되었다는 것은 대사증후군 주의군으로 분류되어 관리가 요구되는 집단임을 의미하는 것이다. 따라서 이들 다중이환자들에 대한 라이프스타일 위험인자를 구체적으로 도출하였다는 것은 어떤 요인을 관리하면 만성질환 관리에 효과가 있는지 명확해 졌다는 점에서 건강증진 측면, 보건교육 측면에서 매우 의미가 있는 연구이다. 고혈압과 이상지질혈증, 고혈압과 당뇨 다중이환자들에게는 흡연, 음주를 포함한 7가지 라이프스타일 전부를 개선하는 활동보다는 본 연구에서 이들 집단에 위험한 것으로 확인된 신체적 비활동, 부적절한 체중, 부적절한 수면의 개선에 집중하여 보건교육을 실시한다면 보다 효과가 클 것이다.

본 연구에서 나타난 연구결과는 라이프스타일 위험인자에 대한 중요한 역할과 다중 만성질환의 중요한 역할을 빅데이터 분석기법을 이용한 것에 의의가 있다. 즉, 동반질환 또는 다중질환연구에서 실제적인 국민건강영양조사와 같은 방대한 데이터베이스를 이용하여 연관규칙분석의 실용성을 확인한 시도는 중요한 의미가 있다고 할 수 있다.

본 연구의 제한점은 다음과 같다.

첫째, 본 연구의 경우 2년치 자료에 대한 단면 연구자료를 분석하였기 때문에 인과관계에 대한 세부 분석이 어려웠다는 점이다. 라이프스타일 위험요인과 만성질환과의 관계에 대한 연관성을 분석하면서 제시한 주요 결과에 다른 연관규칙들도 존재하였으나 다른 통제변수들에 대한 관련성을 제시하기 어려웠다는 점이다.

둘째, 라이프스타일 위험요인에 대한 종류를 좀 더 다양화 할 필요가 있다. 또한 라이프스타일 위험요인

에 대한 정의기준에 따라 결과가 상이하게 나올 가능성이 있으므로 추가적인 문헌고찰 및 지침에 따른 정교한 기준이 필요할 것이라 판단된다.

셋째, 라이프스타일 위험요인과 만성 동반질환의 패턴에 대한 선행연구가 없어 결론과의 선행연구비교가 제한된다는 점이다.

본 연구결과에서 나타난 주요 결과 중 고혈압과 이상지질혈증이 동반된 사람들에게 라이프스타일 위험인자를 줄이기 위한 보건교육프로그램 등을 개발하는데 필요한 기초자료로 활용할 수 있을 것으로 생각된다. 예를 들어 빈번한 간식 섭취를 줄이고 그 관심을 다른 건강한 쪽으로 유도하는 지역사회 보건교육프로그램을 도입할 것을 제안한다. 이 보건교육프로그램을 통해 궁극적으로는 부적절한 체중군이 정상체중군으로 전환되어 고혈압 위험을 낮추는 효과를 낼 수 있을 것이다. 또 다른 예로 다빈도 다중이환자로 확인된 고혈압과 이상지질혈증, 고혈압과 당뇨병에게는 본 연구에서 이들 집단에 위험한 것으로 확인된 신체적 비활동을 개선할 수 있는 운동프로그램, 부적절한 체중을 개선할 수 있는 식이조절프로그램, 부적절한 수면의 개선을 위한 정신보건교육프로그램 등으로 선택과 집중의 보건교육프로그램을 개발하기를 제안한다. 이들 보건교육프로그램은 다빈도 만성질환의 적절한 관리에 도움이 되며 더 나아가 개인과 국가가 만성질환에 투입하는 의료비용 감소에도 기여할 것이다.

참고문헌

- 이인희, 신아미, 손창식, 박희준, 김중휘, 박상영, 최진호, 김윤년. (2010). 데이터 마이닝을 활용한 뇌경색증과 동반되는 질환의 연관성 분석. *대한물리치료학회지*, 22(1), 75-81.
- 이윤미, 신지은. (2020). 건강행위 실천에 따른 운동실천율 분석: 국민건강영양조사 2016~2018년 자료를 이용하여. *한국체육과학회*, 29(5), 1281-1291.
- 이정진. (2011). R, SAS, MS-SQL을 활용한 데이터마이닝. 서울: 자유아카데미.
- 임영명, 박지혜. (2023). 지역사회 노인의 라이프스타일 위험요인이 일상생활 활동과 인지기능에 미치는 영향. *재활치료과학*, 12(4), 111-122.
- 임재순, 이경준, 조영석. (2010). 발생빈도를 고려한 연관성 분석 연구. *한국데이터정보과학회지*, 21(6), 1061-1069.
- Burton SH, Morris RG, Giraud-Carrier CG. (2014). Mining useful association rules from questionnaire data. *Intelligent Data Analysis*, 18(3), 479-494.
- Charlson ME, Pompei P, Ales KL, MacKenzie CR. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*, 40, 373-83.
- Feinstein AR. (1970). The pre-therapeutic classification of co-morbidity in chronic disease. *Journal of Chronic Diseases*, 23, 455-68.
- Hekmatpou D, Shamsi M, Zamani M. (2013). The effect of a healthy lifestyle program on the elderly's health in Arak. *Indian Journal of Medical Sciences*, 67(3-4), 70-77.
- Leejin K, Sungmin M. (2018). Comorbidity Study of Attention-deficit Hyperactivity Disorder (ADHD) in Children: Applying Association Rule Mining (ARM) to Korean National Health Insurance Data. *Iranian Journal of Public Health*, 47(4), 481-488.
- Linn BS, Linn MW, Gurel L. (1968). Cumulative illness rating scale. *J Am Geriatr Soc*, 16, 622-6.
- Li Y, Schoufour J, Wang D, Dhana K, Pan A,

- Liu X, Song M, Liu G, Shin HJ, Sun, Q, Al-Shaar L, Wang M, Rimm EB, Hertzmark E, Stampfer MJ, Willett WC, Franco OH, Hu FB. (2020). Healthy lifestyle and life expectancy free of cancer, cardiovascular disease, and type 2 diabetes: Prospective cohort study. *British Medical Journal*, *368*, l6669.
13. Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A. (2011). Aging with multimorbidity: a systematic review of the literature. *Ageing Res Rev*, *10*, 430-9.
 14. Mamalaki E, Charisis S, Anastasiou CA, Ntanasi E, Georgiadi K, Balomenos V, Kosmidis MH, Dardiotis E, Hadjigeorgiou G, Sakka P, Scarmeas N, Yannakouli M, M. (2022). The longitudinal association of lifestyle with cognitive health and dementia risk: Findings from the HELIAD study. *Nutrients*, *14*, 2818.
 15. Park SH, Jang SY, Kim H. (2014). An association rule mining-based framework for understand lifestyle risk behaviors. *PloS One*, *9*(2), e88859
 16. Richard AG, Shari ML, Peter AB, Gibson P, Marcel ES, Bruce SF. (2016). Multimorbidity Patterns in the United States: Implications for Research and Clinical Practice. *The Journals of Gerontology: Series A*, *71*(2), 215-220.
 17. Saint JM, Krueger PM. (2017). Health lifestyle behaviors among U.S. adults. *Social Science and Medicine - Population Health*, *3*, 89-98.
 18. Schoenborn CA. (1986). Health Habits of U.S. Adults, 1985: the "Alameda 7" Revisited. *Public Health Rep*, *101*(6), 571-580.
 19. Tai YM, Chiu HW. (2009). Comorbidity study of ADHD: Applying association rule mining (ARM) to National Health Insurance Database of Taiwan. *Int J M Inform*, *78*(12), e75-83.
 20. Tornero-Quiñones I, Sáez-Padilla J, Espina Diaz A, Abad Robles MT, Sierra Robles Á. (2020). Functional ability, frailty and risk of falls in the elderly: Relations with autonomy in daily living. *International Journal of Environmental Research and Public Health*, *17*(3), 1006.
 21. Van den Akker M, Buntinx F, Knottnerus JA. (1996). Comorbidity or multimorbidity: what's in a name? A review of literature. *European Journal of General Practice*, *2*, 65-70.

■ 이현주	https://orcid.org/0000-0003-0185-6380
■ 명성민	https://orcid.org/0000-0002-0131-7961

국문초록

국민건강영양조사 자료를 활용한 라이프스타일 위험요인과 다중이환간의 연관관계분석

이현주*, 명성민**

*중원대학교 보건행정학과, 부교수, **중원대학교 보건행정학과, 교수

목적: 본 논문에서는 대한민국 성인의 라이프스타일 위험요인과 복합만성질환간의 연관성 규칙을 탐색하여 보건교육프로그램에 필요한 방향성과 기초정보를 제공하는데 목적을 둔다.

방법: 제8기 국민건강영양조사 중 2019년부터 2020년까지 만 19세 이상 성인 7,609명을 대상으로 하였으며, 6개의 라이프스타일 위험요인과 11가지 이환질환에 대하여 R과 R 스튜디오를 이용하여 연관규칙마이닝을 수행하였다.

결과: 본 연구 결과를 통하여 연관규칙마이닝과 같은 데이터마이닝 기법을 통해 생활 습관 위험 요인의 중요성과 여러 만성 질환의 역할을 보여줬다는 점에서 의미가 있다.

결론: 상기 결과를 통하여 신체 활동 부족을 해결하기 위한 운동 프로그램, 부적절한 체중을 해결하기 위한 식이 중재, 부적절한 수면을 해결하기 위한 정신건강 교육프로그램과 같은 선택적이고 집중적인 건강교육 프로그램에 대한 개발의 필요성이 요구된다.

주제어: 연관규칙마이닝, 만성질환, 라이프스타일 위험요인, 국민건강영양조사, 보건교육