



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

보건학 박사학위논문

연관성 분석을 이용한

한국인의 건강위험행위 군집현상 연구

— 제4기 국민건강영양조사 자료를 이용하여 —

An Association Rule Mining-Based Framework
for Understanding Lifestyle Risk Behaviors

2014년 8월

서울대학교 대학원

보건학과 보건학 전공

박 소 현


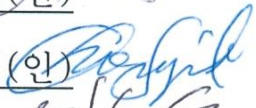

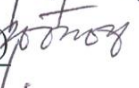

연관성 분석을 이용한
한국인의 건강위험행위 군집현상 연구
-제4기 국민건강영양조사 자료를 이용하여-
An Association Rule Mining-Based Framework
for Understanding Lifestyle Risk Behaviors

지도교수 김 호

이 논문을 보건학 박사학위논문으로 제출함
2014년 4월

서울대학교 대학원
보건학과 보건학 전공
박 소 현

박소현의 박사학위논문을 인준함
2014년 6월

위	원	장	<u>조 영 태(인)</u> 
부	위	원	<u>장 조 성 일(인)</u> 
위		원	<u>이 승 욱(인)</u> 
위		원	<u>김 혜 영(인)</u> 
위		원	<u>김 호(인)</u> 

국 문 초 록

배경 및 목적: 건강한 생활습관은 질병 예방과 조기사망의 예방뿐만 아니라 정상시의 신체기능을 적절히 유지하기 위해 중요하다. 그러나 지금까지의 연구에 의하면 이에 반하는 건강위험행위는 줄어들지 않고, 그로 인한 질병 부담은 점점 증가하고 있는 실정이다.

건강행위 및 건강위험행위는 개별적, 독립적이지 않고 상호 연관되어 행해진다. 이를 군집현상이라고 하는데, 지금까지의 건강행위 군집 연구는 흡연, 음주, 신체활동 부족 등 일부 행위만을 분석하거나, 건강행위 개수의 단순한 합을 이용한 분석에 그쳐 상호연관성이 있는 구체적인 위험행위를 알 수 없었다. 이에 본 연구에서는 연관성분석(Association rule mining: ARM)을 이용하여 건강위험행위 군집의 유형을 분석하는 연구를 하고자 하였다.

건강위험행위 군집 유형 분석 연구를 통해 여러 가지 위험행위를 함께 하는 대상자들을 위한 건강행위방법을 제시할 수 있으며, 특정 위험행위를 하는 대상자의 또 다른 건강위험행위(co-occurring behavior)의 예측에 유용하게 활용할 수 있다.

연구의 대상 및 방법: 본 연구는 분석을 위하여 제4기 국민건강영양조사 자료(2007~2008) 중 건강행위 설문에 응답한 20세 이상의 성인 14,833명을 대상으로 하였다.

본 연구에서는 건강행위의 지침을 지키지 않는 경우를 건강위험행위로 정의하였으며, 여러 문헌과 연구에서 중요하게 제시하고 있는 흡연, 과도한 음주, 신체적 비활동, 부적절한 체중, 부적절한 수면시간, 아침 결식, 잦은 간식 섭취 등 7가지 위험행위를 분석하였다.

첫째, 전체 대상자들의 일반적 특징과 건강위험행위에 대한 분포를 남녀로 나누어 살펴보았다. 대상자들이 행하고 있는 건강위험행위 개수의 합을 이용하여, 그리고 전체 대상자들의 건강위험행위의 확률을 근거로 한 관찰치 수와 기대치 수의 비(O/E ratio)를 이용하여 군집현상을 확인하였다.

둘째, ARM을 이용하여 7가지 건강위험행위가 함께 일어나는 유형을 분석하였다. ARM 연관규칙의 통계적 유의성 평가는 chi-square test를 이용하였다.

연관규칙에 포함되는 건강위험행위들을 건강위험행위 군집이라고 정의하고 건강위험행위 군집의 여러 가지 형태를 유형화하였다. 대표적인 군집을 찾아 건강위험행위 군집의 인구사회학적 특성, 건강위험행위 특성, 건강 수준 특성 등을 살펴보았다. 다중 로지스틱 회귀모형을 이용하여 건강위험행위 군집을 예측하는 모델링을 하였다.

셋째, 본 연구의 ARM 결과를 다른 건강조사 자료(지역사회 건강조사)에 적용하여 결과값이 유의하게 나타나는지 확인하였다.

결과: 한국인에게 가장 많은 건강위험행위는 신체활동 부족 및 부적절한 수면시간이었다. 또한 전체 대상자의 69.8%는 2가지 이상의 건강위험행위를 하고 있었다. 그러나 7가지를 모두 하는 대상자는 없었으며, 건강위험행위를 하나도 하지 않는 가장 이상적인 대상자는 5.0%였다. 남자 대상자들의 건강위험행위 비율이 더 높았다.

건강위험행위 2개, 3개, 4개, 5개, 6개 등으로 이루어진 군집현상이 관찰되었으며, 위와 마찬가지로 남자 대상자에서 건강위험행위 군집현상이 더 빈번하게 나타났다.

건강위험행위 군집에서 가장 많이 나타나는 행위는 부적절한 수면, 신체적 비활동, 현재 흡연 등이었다. 군집마다 특정 건강위험행위들이 있었으며, 그것들은 2~3개의 다른 위험행위와 함께 군집을 이루고 있었다.

건강위험행위 연관규칙으로 살펴본 대표적인 군집은 신체적 비활동, 현재 흡연, 부적절한 수면시간의 군집으로 이는 남녀에서 공통적으로 나타났다(여: Lift 1.06, OR 1.29, 95% CI 1.05-1.58; 남: Lift 1.01, OR 1.26, 95% CI 1.09-1.45). 남자 대상자들에서는 신체적 비활동, 부적절한 체중, 아침 결식, 흡연의 군집(Lift 1.52, OR 2.87, 95% CI 2.23-3.70)이 나타났다.

건강위험행위에 대한 연관규칙은 지역사회 건강조사 자료에서도 유의하게 나타났으며, 건강위험행위 군집의 분포도 비슷하게 나타났다.

결론: 건강위험행위는 독립적이지 않고 상호 연관성을 갖고 함께 나타난다. ARM을 이용하여 건강위험행위 군집의 유형을 파악할 수 있고, ARM을 다른 건강조사 자료 분석에도 활용할 수 있다. 건강위험행위 군집에 대한 정보를 이용하여 건강행위 대상자들에 대한 이해를 높이고, 그들에게 필요한 위험행위 예방에 대한 중재를 계획할 수 있다.

주요어: 건강위험행위, 군집, 연관성 분석, 건강위험행위 예방

학번: 2007-30617

목 차

I. 서론.....	1
1. 연구의 배경 및 필요성.....	1
2. 연구의 목적.....	4
II. 문헌고찰.....	6
1. 건강행위 군집현상.....	6
2. 건강행위 군집현상의 연구방법.....	7
2.1 Correlation(상관관계).....	7
2.2 Accumulation(행위의 개수 합을 이용하는 방법).....	7
2.3 O/E ratio(관찰치와 기대빈도의 비).....	8
2.4 K-means cluster analysis.....	9
2.5 외국의 건강행위 군집 연구.....	10
2.6 우리나라에서 건강행위 군집을 이용한 연구.....	13
3. 연관성 분석(ARM: Association Rule Mining).....	15
3.1 ARM(Apriori algorithm & minimum support).....	16
3.2 Support, Confidence, and Lift.....	17
3.3 Lift와 OR, RR의 비교.....	18
3.4 ARM의 활용.....	20
3.5 ARM의 장점 및 단점.....	20
4. 건강결정 요인.....	21
5. 건강 수준.....	22
5.1 주관적 건강.....	23
5.2 만성질병.....	24
6. 스트레스.....	26
7. 건강행위와 건강.....	27
7.1 흡연.....	29

7.2 과도한 음주.....	3 0
7.3 신체적 비활동.....	3 1
7.4 부적절한 체중(저체중 또는 비만).....	3 2
7.5 부적절한 수면 시간.....	3 3
7.6 아침 결식.....	3 3
7.7 잦은 간식 섭취.....	3 4
8. 사회인구학적 특성에 따른 건강행위.....	3 5
8.1 성별.....	3 5
8.2 연령.....	3 5
8.3 결혼.....	3 6
9. 사회경제적 요인과 건강행위.....	3 7
9.1 교육.....	3 9
9.2 직업.....	3 9
9.3 소득.....	4 0
III. 연구방법.....	4 1
1. 자료 및 연구의 대상.....	4 1
2. 변수 정의.....	4 2
2.1 건강위험행위 변수.....	4 2
2.2 건강 수준 변수.....	4 3
2.3 개인의 특성 변수.....	4 5
3. 연구의 모형.....	4 6
4. 분석방법.....	4 7
4.1 대상자들의 특징.....	4 8
4.2 건강위험행위 군집현상.....	4 8
4.3 ARM 결과의 평가 및 다른 데이터에 적용.....	5 3
4.4 건강위험행위 군집 예측 요인.....	5 3
4.5 분석에 사용한 통계 패키지.....	5 3

IV. 결과.....	5 4
1. 대상자들의 특징.....	5 4
1.1 대상자들의 일반적 특징.....	5 4
1.2 대상자들의 건강위험행위 특징.....	5 6
1.3 대상자들의 건강 수준 특징.....	5 8
2. O/E ratio로 본 건강위험행위 군집현상.....	6 0
3. 건강위험행위 ARM 결과.....	6 3
4. ARM 결과의 평가.....	6 7
4.1 여자 대상자에서 건강위험행위.....	6 7
4.2 남자 대상자에서 건강위험행위.....	6 8
5. 건강위험행위 군집이 나타나는 대상자들의 특징.....	7 0
5.1 여자 대상자에서 세 가지 행위 군집의 특성.....	7 0
5.2 남자 대상자에서 세 가지 행위 군집의 특성.....	7 4
5.3 남자 대상자에서 네 가지 행위 군집의 특성.....	7 8
6. 건강위험행위 군집을 예측하는 요인.....	8 4
6.1 여자 대상자에서 PICSIS, PIIWIS 군집의 예측 요인.....	8 4
6.2 남자 대상자에서 PICSIS, PICSBS 군집의 예측 요인.....	8 6
6.3 남자 대상자에서 PIIWBSCS, PIISBSCS 군집의 예측 요인.....	8 8
7. 단일 건강위험행위와 건강위험행위의 군집.....	9 0
7.1 여자 대상자.....	9 0
7.2 남자 대상자.....	9 1
8. 다른 데이터에 적용.....	9 4
8.1 분석데이터.....	9 4
8.2 변수.....	9 4
8.3 분석방법.....	9 7
8.4 대상자들의 특징.....	9 7
8.5 KNHANES IV의 연관규칙을 CHS 2010에 적용.....	1 0 1

V. 논의.....	1 0 4
1. 가장 빈도가 높은 건강위험행위.....	1 0 4
2. 건강위험행위 군집.....	1 0 5
3. 흡연.....	1 0 6
4. 신체적 비활동.....	1 0 6
5. 부적절한 수면.....	1 0 7
6. 아침 결식.....	1 0 8
7. 건강위험행위 군집의 특성 및 군집을 보이는 예측 요인.....	1 0 8
8. 단일 건강위험행위와 건강위험행위 군집에서의 예측 요인.....	1 1 0
9. 본 연구의 강점.....	1 1 1
10. 건강위험행위 군집 연구의 보건학적 의의.....	1 1 3
VI. 연구의 제한점.....	1 1 6
VII. 결론.....	1 1 7
참고문헌.....	1 1 9
부록.....	1 3 5
1. O/E ratio로 본 건강위험행위 군집현상.....	1 3 5
1.1 여자 대상자에서 건강위험행위.....	1 3 5
1.2 남자 대상자에서 건강위험행위.....	1 3 9
2. ARM(연관성 분석)의 과정.....	1 4 3
2.1 데이터 세트부터 Link analysis.....	1 4 3
3. 단일 건강위험행위 예측 요인.....	1 4 6
3. 1 여자 대상자에서 단일 건강위험행위를 예측하는 요인.....	1 4 6
3. 2 남자 대상자에서 단일 건강위험행위를 예측하는 요인.....	1 4 7
Abstract.....	1 4 8

표 목 차

Table 1. Clustering of cigarette smoking, excessive alcohol consumption, and lack of physical exercise	4
Table 2. Co-occurrence of health related behavior and measurement of association in previous studies	1 2
Table 3. Co-occurrence of health related behavior and measurement of association in previous studies, Korea	1 4
Table 4. 2×2 contingency table for evaluation of association rule A⇒B	1 6
Table 5. 2×2 Contingency table of current smoking and heavy drinking	1 8
Table 6. Lift vs OR vs RR.....	1 9
Table 7. Lifestyle risk behavior and health status in previous studies	2 5
Table 8. Study populations.....	4 1
Table 9. Diseases included in the Chronic Disease Variable.....	4 5
Table 10. Characteristics of study population.....	5 5
Table 11. Characteristics of lifestyle risk behaviors.....	5 7
Table 12. Characteristics of health status.....	5 9
Table 13. Lifestyle risk behavior group according to the included behaviors	6 2
Table 14. Results of association rule mining of lifestyle risk behaviors	6 4
Table 15. Results of association rule mining* and chi-square test of lifestyle risk behaviors among women.....	6 7
Table 16. Results of association rule mining* and chi-square test of lifestyle risk behaviors among men.....	6 9
Table 17. Lifestyle risk group among women.....	7 2
Table 18. Lifestyle risk group among women.....	7 3
Table 19. Lifestyle risk group among men	7 6

Table 20. Lifestyle risk group among men	7 7
Table 21. Lifestyle risk group among men	7 9
Table 22. Lifestyle risk group among men	8 0
Table 23. Multiple logistic regression analysis for predicting patterns of three lifestyle risk behaviors among women	8 5
Table 24. Multiple logistic regression analysis for predicting patterns of three lifestyle risk behaviors among men	8 7
Table 25. Multiple logistic regression analysis for predicting patterns of four lifestyle risk behaviors among men	8 9
Table 26. Variables from the KNHANES IV and CHS 2010.....	9 6
Table 27. Characteristics of study population	9 8
Table 28. Characteristics of lifestyle risk behaviors.....	9 9
Table 29. Characteristics of health status.....	1 0 0
Table 30. Rules from the KNHANES IV and chi-square test with the CHS 2010 data among women.....	1 0 1
Table 31. Rules from the KNHANES IV and chi-square test with the CHS 2010 data among men.....	1 0 2
Table 32. Distribution of lifestyle risk behaviors among women ..	1 3 5
Table 33. Distribution of lifestyle risk behaviors among women ..	1 3 6
Table 34. Distribution of lifestyle risk behaviors among women ..	1 3 7
Table 35. Distribution of lifestyle risk behaviors among women ..	1 3 8
Table 36. Distribution of lifestyle risk behaviors among men.....	1 3 9
Table 37. Distribution of lifestyle risk behaviors among men.....	1 4 0
Table 38. Distribution of lifestyle risk behaviors among men.....	1 4 1
Table 39. Distribution of lifestyle risk behaviors among men.....	1 4 2
Table 40. Risk factors for predicting lifestyle risk behaviors among women.....	1 4 6
Table 41. Risk factors for predicting lifestyle risk behaviors among men.....	1 4 7

그림 목 차

Figure 1. Apriori algorithm.....	1 6
Figure 2. Trends in economic burden of disease in Korea(10 billion won).....	2 9
Figure 3. The study model.....	4 6
Figure 4. Data analysis framework.....	4 7
Figure 5. Minimum support, confidence, and number of rules among women under confidence 50%.....	5 1
Figure 6. Minimum support, confidence, and number of rules among men under confidence 60%.....	5 1
Figure 7. A link diagram derived from association rule mining(ARM).....	6 6
Figure 8. Lifestyle risk group among women.....	7 1
Figure 9. Lifestyle risk group among men.....	7 5
Figure 10. Lifestyle risk group among men.....	7 8
Figure 11. Number of chronic disease of three behavior group among women.....	8 1
Figure 12. Number of chronic disease of three behavior group among men.....	8 2
Figure 13. Number of chronic disease of four behavior group among men.....	8 3
Figure 14. Data preparation(sas data step).....	1 4 3
Figure 15. Health behavior data.....	1 4 3
Figure 16. Transpose.....	1 4 4
Figure 17. Data set after transpose.....	1 4 4
Figure 18. SAS Enterprise Miner 4.3.....	1 4 5
Figure 19. Association rule.....	1 4 5
Figure 20. Link analysis.....	1 4 5

I. 서론

1. 연구의 배경 및 필요성

흡연, 과도한 음주, 비만, 그리고 신체적 비활동 등의 생활습관과 관련된 건강위험행위들은 만성질환과 사망에 영향을 준다고 알려져 있다(WHO, 2002). 이러한 건강위험행위들이 최근 증가하고 있는 우리 국민의 질병 부담의 31.3%를 차지하고 있으며(정영호 et al., 2006), 특히 비만과 음주에 의한 질병 부담이 큰 폭으로 상승하였다(박정근, 2010).

최근 고령화 사회에 접어들면서, 질병 예방과 조기 사망의 예방뿐 아니라 생애주기에 필요한 신체기능을 유지하면서 건강한 생활을 유지하는 것이 점차 중요해지고 있다. 이를 위해서는 올바른 건강행위를 실천하는 생활습관이 중요하다.

건강행위는 독립적으로 일어나기도 하지만 여러 가지 행위들이 복합적으로 나타나며, 서로 양의 상관관계 혹은 음의 상관관계를 보이며 함께 나타난다(J.K. Langlie, 1979). 마찬가지로 건강위험행위에 있어서도 한 가지 행위보다는 대부분 두 가지 이상의 건강위험행위를 하고 있다. 이처럼 각각의 건강행위가 함께 나타나는 수준이 각 행위가 독립적이라고 가정했을 때 나타나는 것보다 더 빈번하게 나타나는 것을 군집(clustering)이라고 한다(Schuit et al., 2002).

건강위험행위에 대한 군집연구는 다음과 같은 의의를 가진다.

첫째, 건강위험행위는 여러 가지를 함께 하는 경우가 한 가지 행위를 하는 것보다 위험하다. 두 가지 이상의 위험행위가 함께 존재하는 경우 심혈관질환의 위험을 높이는 것과 관련이 있고(Yusuf et al., 2004), 암 발생(Marchand et al., 1997), 그리고 전반적인 사망률(Johansson and Sundquist, 1999)을 높인다고 알려져 있다. 위험행위들은 건강 손상에 부가적이고 잠재적으로 상승 작용을 한다(Schlecht et al., 1999).

둘째, 군집된 건강위험행위의 교정은 많은 이점들을 가져온다. 건강위험행위를 교정하지 않은 대상자는 교정한 대상자보다 건강 수준이 악화될 가능성이 높다(Johansson and Sundquist, 1999). 여러 가지의 건강위험행위를 함께 교정하면 잠재적으로 급성심근경색의 위험을 75% 이상 줄여줄 수 있고(Yusuf et al., 2004), 금연과 체중조절에 대한 프로그램에서 두 가지를 함께 중재한 프로그램을 받은 경우 흡연이 재발할 위험이 줄어들었다(Hall et al., 1992)는 보고가 있었다. 여러 가지 행동에 대한 복합적인 중재를 하게 되면 한 가지 행동에 대한 단순 중재보다 국민 보건에 더 큰 영향을 가져다 줄 것이다(Goldstein et al., 2004; Nigg et al., 2002).

셋째, 건강증진 프로그램 계획에 보다 구체적인 정보를 제공할 수 있다. 보건전문가가 대상자들의 위험행위 군집에 대한 속성을 이해하게 되면 중재전략을 더욱 효과적으로 세울 수 있을 것이다. 위험한 행위 군집을 가지고 있는 대상자를 구별해내어 그들을 위한 종합적 예방 전략을 세울 수 있기 때문에 대상자에게 더욱 적합하고 계획적인 예방프로그램을 수행할 수 있다(Schuit et al., 2002).

그래서 건강위험행위를 복합적으로 가지고 있는 대상자들을 기준으로 적절한 건강증진을 추진할 수 있는 방안을 연구하는 것이 중요하다.

건강위험행위 군집현상에 대한 이해의 필요성이 중요한 사업과제로 인식됨에 따라, 관련 연구자들이 함께 모여 연구하고, 여러 가지 건강위험행위를 바르게 수정하려고 하는 변화를 추구하고 있다(Prochaska et al., 2008).

건강위험행위 종류와 개수는 각각 다르지만 많은 연구자들이 건강행위 군집현상에 대해 연구해왔다(Fine et al., 2004; Laaksonen et al., 2001; Poortinga, 2007; Takakura et al., 2001; 강기원 et al., 2010; 강은정, 2007).

그러나 이 연구들은 주로 흡연, 음주, 신체활동, 비만이나 식이 등 한정적인 건강위험행위의 연관성을 분석한 것으로서, 포괄적인 건강위험행위의 연관성에 관심을 둔 연구는 찾아보기 어렵다. 우리의 일상생활에는 흡연, 음주와 같은 기호식품 사용이나 여가 시간을 충분한 운동 없이 보내는 것 이외에도 건강에 영향을 주는 위험행위들이 있다.

Alameda County Study(Housman and Dorman, 2005)에서는 건강 수준에 영향을 주는 요인들을 분석하면서 7가지의 건강행위를 연구하였는데 각 행위를 지키지 않는 경우를 위험행위로 규정할 수 있다. 7가지 건강위험행위는 흡연, 과도한 음주, 신체적 비활동, 체중 조절을 안 하는 것, 부적절한 수면 시간, 아침 결식, 잦은 간식 섭취 등의 행위이다. 이는 모든 사람들이 일상생활에서 매 순간 선택 또는 회피 등의 결정을 하는 건강행위들로서 좀 더 포괄성을 갖는다.

다음의 표는 건강위험행위 보유현황을 보여주고 있다(Table 1). 예를 들어 흡연, 음주, 신체활동 등의 세 가지 위험행위만 살펴보았을 때, 남자 대상자 중 단순히 흡연만 하는 대상자는 10.27%에 지나지 않으나 흡연과 함께 다른 복합적인 건강위험행위를 함께 하는 대상자는 42.50%나 된다.

Table 1. Clustering of cigarette smoking, excessive alcohol consumption, and lack of physical exercise

Health behavior	No PI	PI	HD	CS	PI, HD	PI, CS	HD, CS	PI, HD, CS
	No HD							
	No CS							
Women	26.96	61.57	2.24	1.06	3.44	3.01	0.61	1.08
Men	11.86	20.66	6.91	10.27	7.75	16.90	10.53	15.07

Source: 강기원 et al., 2010

CS: current smoking, HD: heavy drinking, PI: physical inactivity.

이처럼 건강위험행위는 군집을 이루고 있으며, 나아가 건강위험행위 간의 다양한 조합에 따라 군집의 특성이 달라지는 것을 알 수 있다.

2. 연구의 목적

개인의 건강 수준은 여러 가지 요인들의 상호작용으로 나타난 결과이며, 개인의 건강 수준에 영향을 미치는 다른 요인에 비해 건강위험행위는 여러 노력에 의해 수정이 용이하다. 건강위험행위를 하는 대상자에 대한 건강 증진 정책을 마련하기 위해서는 관련 위험요인의 분포를 구조적으로 분석할 필요가 있다.

건강위험행위 군집에 대한 연구는 많이 진행되었지만, 구체적 행위의 조합에 대한 연구는 많지 않았다.

본 연구에서는 건강위험행위의 분포와 군집을 확인하는 한편, 군집의 특성을 살펴보고자 한다. 이전 연구에서 건강 수준을 높이기 위해 하지 않도록 권고하고 있는 흡연, 과도한 음주, 신체적 비활동, 부적절한 체중, 부적절한 수면시간, 아침 결식, 잦은 간식 섭취 등 7가지 건강위험행위를 이용하여 한국 성인의 건강위험행위 군집의 유형을 분석하는 연구를 하고자 하였다.

그 구체적인 목적은 다음과 같다.

1. 건강위험행위의 분포를 확인하고, 건강위험행위 군집현상이 어떻게 일어나고 있는지 알아본다.
2. 건강위험행위들 간의 연관성 분석(ARM: Association Rule Mining)을 통하여 건강위험행위 군집의 유형을 알아본다.
3. 건강조사에서 ARM의 활용 가능성을 확인함으로써, 이러한 조사의 결과를 보건 중재 프로그램 개발을 위한 근거자료로 제공하고자 한다.

II. 문헌고찰

1. 건강행위 군집현상

한 가지 건강위험행위를 하는 경우도 있지만, 대부분의 개인은 흡연과 음주, 흡연과 신체적 비활동, 신체적 비활동과 잦은 간식 섭취 등 여러 가지 행위를 함께 한다. 이와 같이 함께 행해지는 여러 가지 위험행위들은 서로 관련 없는 단순한 조합이라기보다는 건강위험행위들 간의 상호 작용 및 연관성을 보여주는 것이다.

네덜란드인의 20%(Schuit et al., 2002), 미국인의 17%(Fine et al., 2004), 홍콩 노인의 5%(Chou, 2008), 한국 남성의 15.2%(강기원 et al., 2010)는 세 개 이상의 건강위험행위를 하고 있다.

이처럼 건강행위는 각각의 행위 분포가 다양하고 건강행위에 영향을 주는 요인들이 서로 상호작용을 일으켜 함께 일어난다. 즉 한 가지 행위와 함께 다른 행위들도 같이 나타난다(Steele and McBroom, 1972).

2. 건강행위 군집현상의 연구방법

2.1 Correlation(상관관계)

Steele(1972)은 질병 예방을 위한 건강행위를 연구하면서 각각의 행위 간에 상관관계를 분석하였고, 분석 결과 한 가지 행위를 보이는 대상자는 다른 행위에도 참여하는 경향이 있다고 하였다(Steele and McBroom, 1972). 위 연구에서는 각각의 건강행위에 점수를 주고, 그 점수를 기준으로 변수들간의 상관관계를 분석하였다.

건강행위에 대한 이전의 연구에서는 흡연과 음주, 흡연과 신체적 비활동 등의 변수를 1:1 상관관계로 분석하고 있다(Jean K. Langlie, 1979). 본 연구에서는 좀 더 많은 행위변수를 이용하여 연관성을 동시에 분석하고자 하였으므로 1:1 상관관계로 분석하기에는 부족한 점이 있다.

2.2 Accumulation(행위의 개수 합을 이용하는 방법)

많은 연구자들이 건강행위 개수의 합을 이용하여 건강행위의 분석을 시도해 왔다(Chiolero et al., 2006; Fine et al., 2004; Keller et al., 2008; 김지미, 2003; 김혜련, 2003; 박경용, 2010; 윤태호, 2000; 이정민, 2011). 행위 개수의 합에 따라 건강 수준의 차이를 살펴보는 이들 연구의 대부분은 건강행위를 많이 실천하면 건강 수준이 좋고 건강위험행위를 많이 행하면 건강 수준이 나쁘다는 가설을 검정하고 있다.

건강위험행위를 포괄적으로 분석하여 예방전략을 세우고자 할 때에는 여러 가지 건강행위를 동시에 고려하여야 할 필요가 있다. 건강행위 중에는 장기적인 영향을 미치는 변수, 즉각적인 영향을 미치는 변수 등 다양한 변수가 존재하므로 특정 행위에만 관심을 두어서는 안 된다.

박경용(2010)은 그의 연구에서 건강 수준에 따라 나타나는 건강행위도

차이가 있음을 밝혔다. 금연, 절주, 식습관, 건강검진은 만성질환 이환군에서 수행률이 높고, 체중조절, 운동, 적정수면 등은 만성질환 비이환군에서 높았다고 하였다. 박경용(2010)의 연구를 통해서 건강행위를 개수의 합으로 구분하는 경우, 대상자의 특성별로 포함된 건강행위의 종류가 달라질 수 있다는 것을 알 수 있다.

건강위험행위 개수의 합으로만 군집을 분석하면 2개 군집, 3개 군집, 5개 군집 등 단순히 숫자로만 표현되어, 막상 군집에 포함되는 건강위험행위가 어떠한 행위인지 알 수 없어서 군집의 특성이 모호하게 드러난다.

2.3 O/E ratio(관찰치와 기대빈도의 비)

특정 건강위험행위가 어떤 행위와 함께 일어나는지 분석한 연구들이 있다. 건강위험행위 군집현상은 관찰빈도를 기대빈도로 나누어(O/E ratio) 그 값이 1보다 큰 경우 군집이라고 할 수 있다(Chou, 2008; Laaksonen et al., 2001; Schuit et al., 2002; Takakura et al., 2001; 강기원 et al., 2010).

분석에 사용할 건강위험행위를 표로 제시하면서 가능한 조합을 만들고, 건강위험행위가 일어나는 빈도에 대해 관찰치와 기대빈도의 비를 분석한다. 연관구조를 연구하기 위해, 전체 2^n 개의 가능한 레벨의 분할표를 분석하게 된다. 각 건강위험행위가 독립적이라면 그 행위들이 일어날 확률은 각각의 확률을 곱해준 것과 같으며, 건강행위 조합의 기대빈도는 건강위험행위의 확률에 전체 대상자 수를 곱하여 구한다(이승욱, 1998).

예를 들어, 관찰 대상자 수가 N일 때, A, B, C라는 건강위험행위가 일어날 확률을 각각 a, b, c 라고 한다면, A, B, C 모든 위험행위를 하는 대상자의 기대 빈도는 $(a*b*c)*N$ 이 되고, A와 B 행위만 하는 경우 $a*b*(1-c)*N$ 의 기대빈도를 갖게 된다(Schuit et al., 2002; 강기원 et al., 2010).

이 방법을 사용할 경우 분석하고자 하는 변수의 개수가 늘어나면 관찰치가 매우 작은 셀(cells)이 나타나게 된다. 본 연구에서도 O/E 값을 구하기 위해 분할표를 작성하였다. 확고한 결론을 얻기 위해 건강위험행위의 가능

한 조합을 만들어서 2^7 개인 128개의 표를 분석하였다. 분명 많은 빈 셀(empty cell)들이 나타났다. 결과적으로 표에 대한 전체적인 분석은 매우 어려웠고 통계값을 계산할 수 없거나 통계적으로 유의한 값을 가지고 있는 셀(cell)은 매우 희박하였다. 따라서, 변수의 개수가 늘어나면 O/E 값을 이용하여 발생의 개연성이 높은 군집을 발견하기는 쉽지 않다.

2.4 K-means cluster analysis

강은정(2007)의 연구에서는 K평균 군집분석을 이용하여 건강행위 전반에 대한 군집분류를 하였다. 흡연량, 음주량, 운동량의 세 가지 변수를 사용하여 군집 분석을 하였고, 분석 결과 흡연군, 음주군, 운동군, 건강 증진군, 수동적 태도군 등 다섯 가지의 매우 흥미로운 군집들을 발견하였다(강은정, 2007).

이 분석은 대상자들의 다양한 모습을 건강행위의 측정값으로 분류하는 방법이다. 사전에 결정된 군집 수 k에 기초하여 각 관측값을 군집의 중심들 중에서 가장 가까운 군집에 할당하는 방법이다.

이 분석을 실시하기 위해서는 모든 변수에 대해서 관측값들이 연속변수로 측정이 되어있어야 한다. 분석 결과 건강행위의 양(예를 들어 음주량, 흡연량 등)에 따라 군집이 분류되었고, 건강한 행위와 건강하지 않은 행위가 혼합되어 있는 군집이 나타났다. 예를 들어 수동적 태도군은 흡연량과 음주량은 낮지만 운동량도 낮은 군이었으며, 흡연군은 운동을 많이 하지만 흡연이라는 위험행위도 많이 하는 군이었다.

본 연구에서는 건강위험행위를 하는 것 자체에서 건강위험행위 군집을 찾고자 하였고, 포괄적인 이해를 위해 흡연, 과도한 음주, 신체적 비활동을 포함한 7가지 행위를 분석하고자 하였다. 또한 행위에 대한 측정을 이분변수로 하였고, 위험한 행위의 집합만을 찾아 군집으로 정의하고자 하였으므로 이 분석방법은 본 연구에 적합하지 않다고 판단하였다.

2.5 외국의 건강행위 군집 연구

Table 2에는 건강행위 군집현상에 대한 외국의 연구를 정리하였다.

Steele(1972)의 연구는 건강검진, 구강검진, 안과검진, 건강보험 가입 등의 변수들을 이용하여 각 건강행위들간의 상관관계를 분석하고, 건강행위들의 합을 건강행위지수로 사용하여 사회경제적 수준에 따라 건강행위지수의 분포에 차이가 있는지를 분석하였다.

Burke 등(1997)은 남녀 및 흡연 여부에 따라 그룹을 나누어 흡연하는 대상자들에서 과도한 음주, 신체적 비활동,식이섬유 섭취, 지방 섭취 등 4가지 건강위험행위의 관련성을 행위별 chi-square test로 제시하였다.

Johansson 등(1999)은 스웨덴 국민의 10년 간의 건강자료를 조사하여 흡연량에 따라 카테고리를 만들고 신체적 비활동과 체질량지수 등을 함께 매치(matching)하여 건강행위변화에 따른 사망률을 비교하였다.

Laaksonen 등(2001)은 흡연, 음주, 신체적 비활동, 건강하지 않은 식이 등의 변수를 이용하였고, 이들의 가능한 조합인 16가지 조합 중, 13가지 조합에 대해 각 조합이 나타날 가능성을 예측하는 모델링을 하였다.

Schuit 등(2002)은 흡연, 신체적 비활동, 부족한 과일 섭취, 과도한 음주 등의 변수를 이용하여 16가지 가능한 조합의 O/E를 구한 후, 흡연과 과도한 음주가 나타날 가능성, 흡연과 신체적 비활동이 나타날 가능성 등 두가지 행위가 포함된 군집이 나타날 가능성을 예측하는 모델링을 하였다.

Fine 등(2004)은 대상자들이 행하고 있는 흡연, 과체중, 신체적 비활동, 과도한 음주 등의 건강위험행위의 평균개수를 인구학적 변수, 사회경제적

변수, 만성질병의 보유 등을 고려하여 추정하는 모델링을 하였다.

Pronk 등(2004)은 금연, 건강한 식이, 정상 체질량지수, 신체활동 등의 변수를 이용하여 가능한 16가지의 조합을 구하여 분포를 살펴보고, 건강행위의 개수가 늘어날 가능성에 대한 모델링을 하였다.

Chiolero 등(2006)은 흡연량을 보정하면서 신체적 비활동, 부족한 채소/과일 섭취, 과도한 음주 등의 행위 개수가 늘어날 가능성에 대한 모델링을 하였다.

Poortinga W.(2007), Chou K.-L.(2008)은 흡연, 음주, 채소/과일 섭취 부족, 신체적 비활동의 변수를 이용하여 16가지의 조합의 분포를 O/E로 확인 후 흡연과 음주, 흡연과 신체적 비활동 등 두 가지의 건강위험행위들이 함께 일어날 가능성을 예측하는 모델링을 하였다. 또한 건강위험행위 개수를 한 가지 할 가능성, 두 가지 할 가능성, 세 가지 할 가능성, 네 가지 할 가능성 등의 multinomial multilevel regression modeling을 하였다.

위 연구들은 흡연, 과도한 음주, 운동, 비만, 바람직하지 않은 식습관, 채소나 과일 섭취 부족, 과도한 지방섭취 등의 건강위험행위 변수, 건강검진, 치과나 안과 검진 등의 행위변수를 이용하였다.

각 연구에서는 3개에서 5개까지의 건강행위 변수를 이용하여 분석하였으며, 변수들간의 상관관계에서부터 다중 로지스틱 회귀분석, O/E값을 계산 후 이에 대한 오즈비를 구하는 방법 등을 이용하였다.

Table 2. Co-occurrence of health related behavior and measurement of association in previous studies

Study	CS	HD	PI	OB	UD	LFV	EF	PC	DV	EV	Measure of association
Steele et al, 1972								O	O	O	correlation accumulation
Burke et al, 1997		O	O				O				logistic regression
Johansson et al, 1999	O		O	O							Regression (interaction term)
Laaksonen et al, 2001	O	O	O		O						accumulation, log-linear model
Schuit et al, 2002	O	O	O			O					O/E(POR)
Fine et al, 2004	O	O	O	O							accumulation, multinomial logit analysis
Pronk et al, 2004	O	O	O	O				O			accumulation, proportional odds model
Chiolero et al, 2006	O	O	O						O		logistic regression
Poortinga W., 2007	O	O	O						O		O/E(POR), multinomial multilevel logistic regression
Chou KL, 2008	O	O	O						O		O/E(POR), multinomial logistic regression

CS: current smoking, HD: heavy drinking, PI: physical inactivity, OB: obesity, UD: undesirable diet, LFV: lack of fruit and vegetables, EF: excessive fat intake, PC: physical check up, DV: dental visit, EV: eye doctor visit, O/E: observed/expected frequency, POR: prevalence odds ratio

또한 각 연구에서는 건강행위들 간의 연관성이 있음을 살펴보기 위해 건강행위 개수의 합을 이용하여 건강행위 개수를 예측하는 모델 (regression)(Chiolero et al., 2006; Chou, 2008; Emmons et al., 1994; Fine et al., 2004; Poortinga, 2007), 건강행위 개수의 합을 서열화하여 건강행위 개수를 예측하는 모델 (Pronk et al., 2004), 한 가지 행위를 기준으로 다른 행위들이 함께 일어나는지를 예측하는 모델(Burke et al., 1997; Laaksonen et al., 2001), 건강행위 간에 독립성을 가정하고 기대한 빈도보

다 더 많은 빈도가 함께 일어나는지에 근거한 방법을 이용하는 방법(O/E ratio) 등을 이용(Laaksonen et al., 2001; Poortinga, 2007)하였다.

그러나 지금까지 6가지 이상의 건강행위의 연관성을 포괄적으로 동시에 분석한 연구는 없었으며, 또한 건강행위의 조합 또한 두 가지, 세 가지 행위의 조합으로 연구자가 사전에 설정한 가설에 근거하여 분석함으로써 건강행위간의 다른 연관성의 가능성을 열어두지 못하였다.

2.6 우리나라에서 건강행위 군집을 이용한 연구

우리나라의 연구는 Table 3에 정리하였다.

윤태호(2000)의 연구에서는 비흡연, 적절한 음주, 운동, 비만 조절, 적절한 수면, 규칙적인 아침 식사, 간식 섭취 등의 7가지 건강행위 각각을 실천하는 것을 1로 하여 그 합을 건강행위 실천지수로 나타내 사회계층 간 건강행위의 차이를 분석하였다.

김혜련(2003)의 연구에서는 비흡연, 적절한 음주, 체중 조절, 운동, 적절한 수면, 아침 식사 등의 6가지 행위 중 실천하는 개수의 합을 건강행위 실천지수로 하여 사회계층별 분석 및 만성질환 유병과의 연관성 분석을 하였다.

김지미(2003)의 연구에서는 비흡연, 적절한 음주, 운동, 체중 조절, 식습관, 건강검진, 적절한 수면 등 건강행위들 간의 1:1 상관관계를 spearman's rho값으로 분석하고, 건강행위의 합을 이용하여 관련 요인에 대한 분석을 하였다.

강은정(2007)의 연구에서는 흡연량, 음주량, 신체활동량을 이용하여 군집분석을 하였고, 그 결과 5가지의 특징적인 군집을 발견하였다.

박경용(2010)의 연구에서는 금연, 체중조절, 절주, 운동, 식습관, 걱정 수면, 건강검진 등의 변수를 이용하여 건강행위의 합이 1개, 2개, 3개, 4개 등으로 나타날 가능성에 대한 로지스틱 회귀분석을 하였다.

강기원 등(2010)의 연구에서는 흡연, 음주, 신체적 비활동 등의 건강위험행위의 가능한 8가지 조합에 대해 O/E 를 통해 분포를 살펴보고 각각의 조합이 나타날 가능성에 대한 모델링을 하였다.

이정민(2011)의 연구에서는 금연, 절주, 체중 조절, 운동, 아침식사, 간식 절제, 걱정 수면 등의 7가지 건강행위를 실천하는 것을 건강행위 실천지수로 하여 건강행위 실천지수와 주관적 건강 수준을 분석하였다.

Table 3. Co-occurrence of health related behavior and measurement of association in previous studies, Korea

Study	CS	HD	PI	OB	IS	BS	FS	WC	SD	RD	PE	Measure of association
Yoon, 2000	O	O	O	O	O	O	O					accumulation
Kim, 2003	O	O	O	O	O	O						accumulation
Kim, 2003	O	O	O		O					O	O	correlation, accumulation
Kang, 2007	O	O	O									k-means cluster analysis
Park, 2010	O	O	O		O			O	O		O	accumulation
Kang et al, 2010	O	O	O									POR, ordinal logistic regression
Lee, 2011	O	O	O	O	O	O	O					accumulation

CS: current smoking, HD: heavy drinking, PI: physical inactivity, OB: obesity, IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking, WC: weight control, SD: salty dietary habit, RD: regular dietary habit, PE: physical exam. POR: prevalence odds ratio

우리나라의 연구에서는 건강 수준을 분석하기 위해서 건강행위 개수의 합(김지미, 2003; 김혜련, 2003; 박경용, 2010; 윤태호, 2000; 이정민, 2011), 건강행위들 간의 상관관계(김지미, 2003)를 분석하였다. 또한 군집 분석(강은정, 2007), O/E 비교(장기원 et al., 2010) 등의 방법으로 군집현상을 확인하였다.

외국의 연구에서나 우리나라 연구에서 주로 3~4가지 건강행위는 O/E 값을 사용하여 군집이 형성된 건강행위 종류를 확인할 수 있었다. 그러나, Alameda County Study에서 분석한 건강행위처럼 7가지의 건강행위를 이용한 경우에는 군집현상을 건강행위 개수의 합으로만 확인하였을 뿐, 군집현상에 포함되는 건강행위 종류나 빈번하게 나타나는 군집의 형태는 확인할 수 없었다.

이처럼 지금까지의 건강행위 군집에 대한 연구에서는 건강행위나 건강위험행위를 몇 가지 하는지에 대한 내용만을 알 수 있었다.

3. 연관성 분석(ARM: Association Rule Mining)

ARM은 ‘장바구니 분석’이라고도 불리는데, 데이터 마이닝에서 자주 활용되는 분석 방법 중 하나이다. ARM은 아이템이라고 불리는 이분형 변수들이 데이터베이스 내에서 함께 나타나는 빈도가 높은지를 발견하기 위해 이용된다. 이 분석의 목적은 변수들간에 상관성이 높은 그룹을 찾아내거나 특정한 타겟과 관련해 상관성이 높은 변수들을 찾아내는 것이다(Tan et al., 2004).

ARM은 대량 데이터에 숨어 있는 연관 규칙(association rule)을 찾는다. 예를 들면 ‘흡연⇒음주’와 같은 규칙이다. 이 규칙은 흡연을 하는 대상자들이 음주도 할 가능성이 많은 것을 나타낸다.

3.1 ARM(Apriori algorithm & minimum support)

ARM의 알고리즘 중 가장 잘 알려진 Apriori algorithm은 가장 빈번한 아이템 세트를 찾아낸다.

이 알고리즘은 두 가지 단계를 거친다. 우선, 데이터베이스 내에서 몇 번의 반복을 통해서 사전에 주어진 minimum support에 의해 frequent behavior set가 다음의 그림과 같이 추출된다.

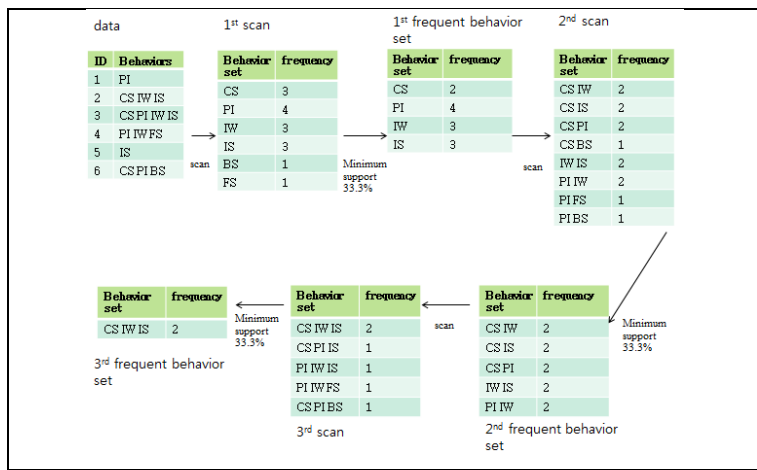


Figure 1. Apriori algorithm

다음은 이들 behavior set로부터 minimum confidence를 만족하는 규칙 (rule)들을 최종 규칙으로 찾아낸다.

연관규칙 예시를 위해 건강행위의 빈도를 a, b, c, d로 표시하였다(Table 4).

Table 4. 2x2 contingency table for evaluation of association rule $A \Rightarrow B$

		Behavior B		Row total
		B=1	B=0	
Behavior A	A=1	a	b	a+b
	A=0	c	d	c+d
Colum total		a+c	b+d	a+b+c+d

3.2 Support, Confidence, and Lift

Support($A \Rightarrow B$)는 A와 B의 두 가지 행위가 함께 나타날 확률을 뜻한다. 즉, support는 전체 transaction에서 연관규칙에 해당하는 데이터의 비율을 의미한다.

$$\begin{aligned}\text{Support}(A \Rightarrow B) &= P(A \cap B) \\ &= \frac{\text{number of persons doing A and B behaviors simultaneously}}{\text{number of all persons included}} \\ &= \frac{a}{a + b + c + d}\end{aligned}$$

Confidence($A \Rightarrow B$)는 A라는 행위를 하고 있는 대상자가 B행위를 추가하는 경우를 뜻하는 조건부확률에 해당된다. 즉, confidence는 항목 A를 포함하는 transaction 중에서 항목 B도 포함하는 transaction의 비율을 의미한다.

$$\begin{aligned}\text{Confidence}(A \Rightarrow B) &= P(B|A) = \frac{P(A \cap B)}{P(A)} \\ &= \frac{\text{number of persons doing A and B behaviors simultaneously}}{\text{number of persons doing A behavior}} = \frac{a}{a + b}\end{aligned}$$

Lift는 연관규칙($A \Rightarrow B$)의 support와 confidence를 동시에 고려하는 측도로 사용된다. ($A \Rightarrow B$)라는 규칙(rule)의 lift는 두 가지 행위가 일어날 확률이 서로 독립이라는 가정하에 나타나는 expected confidence로 confidence를 나눈 것이다. 어떠한 연관 규칙에 대하여 조건이 없을 때 기대되는 결과에 비해, 조건이 추가될 때의 결과가 얼마나 더 향상되는지를 알려주는 측도이다.

$$\begin{aligned} \text{Lift}(A \Rightarrow B) &= \frac{\text{confidence}(A \Rightarrow B)}{P(B)} = \frac{P(A \cap B)}{P(A) \cdot P(B)} \\ &= \frac{\text{Confidence}(A \Rightarrow B)}{\text{Expected Confidence}(A \Rightarrow B)} = \frac{(a + b + c + d) \times a}{(a + b) \times (a + c)} \end{aligned}$$

Lift는 행위들 간의 연관성을 설명할 수 있다. 그 값이 1보다 크면 양의 상관관계, 1보다 작으면 음의 상관관계, 1과 같으면 상관관계가 없음을 의미한다(SAS Education, 2012).

3.3 Lift와 OR, RR의 비교

다음은 흡연과 음주의 예를 들어 표시하였다(Table 5).

각 셀의 숫자는 각 행위에 해당되는 대상자의 숫자로 가정한다. 예를 들어 (a)셀은 흡연을 하는 대상자가 과도한 음주를 하는 경우에 해당되며, 대상자의 수는 65명이라고 가정한다. 이 경우 Lift 값과 OR, 코호트 연구를 가정할 경우의 RR을 비교하였다(Table 6).

Table 5. 2×2 Contingency table of current smoking and heavy drinking

	Heavy drinking yes	Heavy drinking no	Row total
Smoking yes	65 (a)	450 (b)	515 (a+b)
Smoking no	86 (c)	8,324 (d)	8,410 (c+d)
Colum total	151 (a+c)	8,774 (b+d)	8,925 (a+b+c+d)

$$\text{Lift} = \frac{a}{(a+b)} \div \frac{(a+c)}{(a+b+c+d)} = \frac{(a+b+c+d) \times a}{(a+b) \times (a+c)}$$

$$\text{OR} = \frac{a}{b} \div \frac{c}{d} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

$$\text{RR} = \frac{a}{(a+b)} \div \frac{c}{(c+d)} = \frac{a/(a+b)}{c/(c+d)} = \frac{(c+d) \times a}{(a+b) \times c}$$

Table 6. Lift vs OR vs RR

	Lift	OR	RR
Values	7.46	13.98	12.34

Lift값은 기준이 되는 건강위험행위가 있는 대상자들에서 또 다른 건강위험행위가 있는 비율을, 전체 대상자에서 해당 건강위험행위가 있는 행위의 비율로 나누어 준 값이다.

OR값은 기준이 되는 행위가 있는 경우 다른 행위를 하고 있는지 여부의 비를, 기준이 되는 행위가 없을 때와 비교하는 값이다.

RR값은 시간적 선후관계를 알 수 있을 때, 기준이 되는 건강위험행위가 있는 경우, 건강위험행위가 있는 대상자들에서 또 다른 건강위험행위가 있는 비율을, 기준이 되는 건강위험행위가 없는 대상자들에서 또 다른 건강위험행위가 있는 비율로 나누어 준 값이다.

위의 예시로 계산된 값에서는 Lift값이 가장 낮게 나타나고 있다. 기준이 되는 건강위험행위 여부와 상관없이 전체 대상자를 기준으로 하기 때문에 값이 더 낮게 나타나고 있다. 낮은 값으로도 건강위험행위의 연관성을 분석하는데 이용될 수 있음을 알 수 있다.

3.4 ARM의 활용

ARM은 경영 분야뿐 아니라 의학이나 역학 연구에서도 많이 활용되고 있다. 예를 들어 병원감염(Brossette et al., 1998), 심장질환의 예측(Lee et al., 2013), 암 사망률에 대한 사회경제적 패턴 연구(Vinnakota and Lam, 2006), 약물 부작용(Jin et al., 2008), 그리고 ADHD comorbidity(Tai and Chiu, 2009) 등의 연구에 활용되었다.

회귀분석 모델은 독립변수들 간의 상호작용을 통계적으로 테스트하여 하나 이상의 독립변수들이 다른 변수들의 수준에 따른 차이를 분석할 수 있다(Lemon et al., 2003). 그런데 변수들을 추출하는 것이 목적이라면 분석을 통해서 변수들의 조합을 찾아낼 수 있다. 하지만 그 변수들의 조합의 의미를 해석하기 쉽지 않다(Imamura et al., 2007).

Decision tree 구조도 좋은 시스템이긴 하나 tree의 위계가 매우 중요하여 변수들간의 단순한 조합을 찾아내기에는 적절하지 않다(Imamura et al., 2007).

데이터 내에서 건강위험행위의 패턴, 즉 다중건강위험행위를 각 행위마다 동등하게 취급하여 찾아내기 위해서는 ARM이 좀더 적합한 방법이라고 판단된다.

3.5 ARM의 장점 및 단점

ARM은 여러 변수들이 있지만 찾고자 하는 특정한 연관성을 모를 때 분석의 초기 작업으로 종종 이용된다.

장점으로는 첫째, 결과가 분명하여 이해하기 쉽다. 'A \Rightarrow B'라는 연관규칙으로 표현된다. 둘째, 대규모 자료에 접근하려 할 때 초기 탐색적 단계에 적용하여 다음 분석 절차에 방향을 제시할 수 있다. 셋째, 계산이 용이하다.

변수들의 빈도를 나타내는 분할표를 만들어 support, confidence, lift를 구할 수 있다.

단점으로는 첫째, 변수가 늘어남에 따라 규칙을 생성 하는데 필요한 계산량이 많아진다. 둘째, minimum support로 인해 중요한 규칙들을 고려대상에서 제외시킬 가능성이 있다(배화수 et al., 2008). 이에 연구자들은 희소하게 나타나는 변수이지만 특정 변수와는 높은 비율로 동시에 나타나는 변수를 누락시키지 않기 위해 minimum support를 조절하는 알고리즘을 개발하기도 하였다(현영준, 2006). 셋째, 단순 ARM은 중요한 변수에 대한 결과를 누락시킬 수 있다. 이에 연구자들은 가중치를 부여하여 weighted ARM을 적용하기도 하였다(김원석, 2007).

4. 건강결정 요인

생물학적 요인(human biology), 환경 요인(environment), 생활습관(life style) 및 보건의료체계(health care organization)의 요인은 인간의 질병과 그로 인한 사망에 영향을 주며 개인의 건강 수준에도 영향을 미친다(Lalonde, 1974). 이 중 생활습관 요인은 체계적인 교육과 자기관리의 실천으로 비교적 개선 가능성이 큰 요인이다(김혜련, 2009).

1992년부터 4년 여간 수행된 국제 질병 부담 연구에서 전체 사망 중 영양 불량, 물이나 위생 부족, 불건강한 성생활, 흡연, 음주, 직업, 고혈압, 운동 부족, 약물 남용, 대기오염 등의 잘 알려진 10가지 위험요인이 전체 사망의 39.9%를 차지하였고, 영양 불량, 흡연, 음주, 운동 부족 등의 중요한 생활습관 요인은 23.1%를 차지하였다. 생활습관 요인이 알려진 39.9%의 위험요인 중 57.9%에 해당하였다(Murray and Lopez, 1996).

2005년도 OECD 건강데이터를 이용하여 건강 수준(잠재생명상실년수:

Potential Years of Life Loss(PYLL))에 영향을 미치는 요인을 분석한 연구에서는 1970년대에 비해 2000년대는 흡연량과 음주량이 감소하고, 칼로리 섭취와 야채 섭취가 늘어나고 있는 결과를 보여주면서도 여전히 흡연과 음주, 칼로리 섭취 등의 건강행위가 모두 건강 수준(PYLL)에 영향을 주고 있다고 하였다(최윤정 et al., 2009).

최근 우리나라의 연구에서는 국민건강영양조사 자료를 이용하여 분석한 결과 성별, 연령, 결혼 상태, 교육수준, 가구소득, 입원 이용, 외래 이용, 예방접종, 검진수검 등의 요인뿐 아니라, 체질량지수, 운동, 수면시간, 흡연, 음주, 스트레스 인지율 등의 건강행위 관련 변수들이 기대수명과 삶의 질(EQ-5D) 등에 모두 유의하게 영향을 주는 것으로 나타났다(최령 and 문현주, 2011).

암, 뇌졸중, 심장질환, 당뇨, 신장질환, 만성 폐질환, 간 질환, 고혈압성 질환 등은 우리 국민들의 사망의 많은 부분을 차지한다(통계청, 2012). 지역사회에서의 심혈관질환 중재, 조기검진, 금연, 당뇨조절 프로젝트 등은 만성질환을 줄여줄 수 있다고 알려져 있다. 암으로 인한 사망의 28%, 심혈관 질환으로 인한 사망의 45%, 당뇨병 부작용으로 인한 사망의 50% 이상은 건강행위를 변화시킴으로써 줄일 수 있다고 알려져 있다(Mason et al., 1987).

5. 건강 수준

신체의 건강 수준을 측정하는 방법으로 주관적 건강 수준이나 만성질병 보유 여부를 이용한다.

5.1 주관적 건강

주관적 건강 수준은 임상검사에 의해 진단된 건강 수준은 아니지만, 사망과의 관련성이 높아 인구집단의 건강 수준을 측정할 때에 광범위하게 활용된다(Idler and Benyamini, 1997; Kaplan and Camacho, 1983).

또한 주관적 건강 수준과 만성질환 유병상태의 상관관계가 통계적으로 유의한 정도로 나타나 건강 수준을 측정하는데 활용할 수 있다(박경용, 2010; 윤강재, 2006).

어떤 행위를 하는데 객관적인 판단보다 주관적인 인지가 더 큰 영향을 미칠 수 있고, 건강행위를 예측하고 이들의 건강행위를 변화시키기 위해서는 주관적 인식이나 태도 등이 매우 중요한 요소이다(정현주, 2005). 또한 지각된 건강상태가 건강증진행위에 영향을 주어 주관적 건강 수준에 따라 건강실천행위가 달라질 수 있다(정현주, 2005). 이처럼 주관적 건강인식은 개인의 건강행위에 미치는 영향이 크며, 주관적 건강인식이 좋다고 생각하는 집단일수록 건강행위를 할 가능성이 높다(문상식, 2001).

반대로 주관적 건강 수준은 개인이 처한 사회경제적 요인들과 더불어 건강행위에 의해서도 영향을 받는다(Subramanian et al., 2001). 사회경제적 계층이 높을수록 건강하다고 느끼는 경향이 높고, 또한 건강행위 실천율 또한 높다(김진희, 2011).

여러 연구에서 여성, 고령, 저학력의 대상자들이 주관적으로 건강하지 않다고 하였으며(김진희, 2011; 윤강재, 2006), 질병의 종류와는 상관없이 만성질환 이환 집단에서 주관적 건강 수준이 낮다(박경용, 2010).

건강행위별로 살펴보면 흡연하는 경우와 신체활동이 부족한 경우, 비만한 경우에 주관적 건강 수준이 낮고, 음주군에서는 상반된 결과를 보인다(김진희, 2011; 양현진, 2003).

5.2 만성질환

만성질환 보유 여부를 건강 수준을 측정하는 방법으로 사용하기도 한다. 만성질환은 주로 심혈관 질환, 뇌혈관 질환, 당뇨병 등의 질병을 대상으로 한다. 본 연구에서는 기존 연구를 참고하여 7가지 건강위험행위와 관련된 만성질환을 사용하였다.

흡연(Katanoda et al., 2008), 과도한 음주(Bagnardi et al., 2001; Corrao et al., 2004), 신체적 비활동(Berlin and Colditz, 1990; Haskell, 1994; Helmrich et al., 1991; Kannel and Sorlie, 1979; Kool et al., 2004; Paluska and Schwenk, 2000; Sesso et al., 2000; Wolff et al., 1999), 부적절한 체중(Flegal et al., 2005; Larsson and Wolk, 2007; Luppino et al., 2010; McGee, 2005), 부적절한 수면(Cappuccio et al., 2011; Cole and Dendukuri, 2003; Gallicchio and Kalesan, 2009), 아침 결식(Horikawa et al., 2011), 잦은 간식 섭취(Mekary et al., 2012) 등이 건강 수준에 영향을 미치는 정도가 연구되었다.

연구에 의하면, 건강위험행위에 의해서 전반적인 사망률이 높아질 뿐 아니라 암과 심혈관질환 등으로 인한 사망률도 높아진다. 또한 악성신생물, 뇌혈관 질환, 심혈관 질환, 당뇨병, 간질환, 고혈압, 만성 폐색성 폐질환, 이상지질혈증, 골다공증, 위궤양, 요통, 우울증 등의 질병의 발생 위험이 높아지거나 질병 진행이 악화될 수 있다(Table 7).

Table 7. Lifestyle risk behavior and health status in previous studies

	CS	HD	PI	IW	IS	BS	FS
Overall death	+		+	+	+		
Heart disease mortality	+		+	+	+		
Cardiovascular disease mortality	+		+	+	+		
Coronary heart disease		±	+				
Cancer related death					+		
Gastric cancer	+	+					
Liver cancer	+	+		+			
Colon cancer	+	+		+			
Lung cancer	+	+					
Breast cancer		+					
Cervix cancer	+						
Liver cirrhosis		+					
Diabetes			+			+	+
Hypertension		+					
GI ulcer	+						
Dyslipidemia	+	+	+				
COPD	+						
Osteoporosis	+		+				
Back pain			+				
Depressive disease				+	+		
Obesity						+	

A plus sign indicates a risk factor; a minus sign indicates a protective factor; blank cells indicate the factor was not studied. Each of the risk factors shown was examined in two or more studies.

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight,

IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking

COPD: chronic obstructive pulmonary disease

CS: Brischetto et al, 1983; Craig et al, 1989; Katanoda et al, 2008

HD: Bagnardi et al, 2001; Corrao et al, 2004

PI: Kannel et al, 1979; Berlin et al, 1990; Helmrigh et al, 1991; Haskell, 1994; Mayer-Davis et al, 1998; Wolff et al, 1999; Sesso et al, 2000; Paluska et al, 2000; Kool et al, 2004, Pedersen and Saltin, 2006

IW: Flegal et al, 2005; Mcgee et al, 2005; Larsson et al, 2007; Luppino et al, 2010

IS: Cole et al, 2003; Gallicchio L et al, 2009, Cappuccio et al, 2010

BS: Horikawa et al, 2011; Merkary et al, 2012

FS: Merkary et al, 2012

6. 스트레스

스트레스는 항상성이나 알로스타시스(allostasis)를 방해하는 요소들을 포함하는데, 스트레스 반응으로 HPA(hypothalamic-pituitary-adrenal) 축과 교감신경계의 작용으로 설명되고 있다. 그러나 교감신경 자극이 지속되면 HPA 축이 활성화되고 혈압이 올라가는 것은 아직 명확히 밝혀지지 않았다.

여러 연구들에서 부신 호르몬이 비만과 체지방을 복부로 집중시키는 것과 관련이 있다고 하였다. 코티졸이 올라가면, 특히 성호르몬, 성장호르몬 분비가 억제되면서 대사에 이상이 오고 지방을 내장에 축적시키게 된다. 비슷한 기전으로 교감신경의 활성화로 인해 고혈압이 발생한다.

우울이나 스트레스 노출 후에 중추와 말초에도 비슷한 이상 반응이 오게 된다. 글루코코르티코이드 노출 후 음식 섭취가 증가하고 ‘렙틴 저항성’ 비만이 생겨, 렙틴과 뉴로펩타이드 Y의 균형이 깨진다. 그 결과로 스트레스-폭식이 나타난다. 스트레스 반응을 일으키는 요인들은 심리사회적·사회경제적 문제, 우울하고 불안한 성향, 음주와 흡연 등인데 성격과 성에 따라 약간의 차이가 있다(Björntorp, 2001).

스트레스는 뇌에서 시작하여 뇌뿐만 아니라 신체 전체에 영향을 미친다. 급성 스트레스 반응은 신경계, 심혈관계, 자율신경계, 면역계, 대사계에 반응을 일으켜 적응과 생존을 돕는다. 그러나 만성 스트레스는 위와 같은 시스템이 조절되지 않게 하며 병태생리적인 상황으로 악화시킨다. 만성 스트레스와 그에 따르는 개인의 행위 변화(흡연, 과식, 음주, 수면 질 저하)로 인한 부담을 allostatic overload라고 하는데, 만성 스트레스에 적응하는 기전에 글루코코르티코이드, 아미노산, 여러 가지 내분비 물질들이 함께 관여하게 된다(McEwen, 2008).

만성적인 스트레스는 혈압을 상승시키고 동맥경화를 촉진시키는 혈중 지질 성분을 증가시키고 혈전증을 초래한다. 이러한 경향은 사회경제적 수준과도 관련성을 가져서 사회경제적 수준이 낮을수록 더 높다고 한다. 즉 스트레스를 많이 받을수록 기존의 심혈관계 질환의 위험요인들을 많이 가지고 있으면서 동시에 직업과 관련된 사

회심리적 지표들이 나뉘었다. 이러한 스트레스는 식이와는 무관하게 교감신경의 항진, 혈관내막 세포의 손상을 통해 관상동맥 질환을 진행시킨다(김대희, 2002).

성별, 결혼 상태, 직업 계층 등에 따라 스트레스 노출 정도의 분포가 다르게 나타났다. 그러나 스트레스 노출 정도가 이처럼 다르게 나타나는 것은 사회경제적 상태에 따라 스트레스에 취약한 정도가 다르기 때문이라고 하였다(Turner et al., 1995).

근로자의 직무 스트레스 분석 연구에서 남자에서는 직무자율성, 부적절한 사회적 지지, 직무 불안정, 조직불평등이 높을수록 직무 스트레스 수준이 높아지고, 문제 음주가 증가하여 심혈관계 질환에 영향을 미칠 수 있다고 하였다. 비정규직 근무, 직무 요구도가 높을수록 현재 흡연자 비율이 유의하게 증가하였으며, 여자 근로자에서는 직무 스트레스와 건강 위험행위 간에 유의한 현상이 나타나지 않았다(전수현 et al., 2010). 또 다른 연구에서는 직무 스트레스별로 건강행위의 유의한 차이는 보이지 않았으며, 남자근로자의 순환 근무, 여자 근로자의 순환 근무, 결혼 여부 등이 규칙적 식사와 관련이 있었다(박현주 and 정혜선, 2010).

우울감이 건강증진행위를 덜 하게 하고(Chang et al., 2013), 운동을 잘 하지 않는 것과 우울은 양의 상관관계가 있다(Teychenne et al., 2010).

7. 건강행위와 건강

Kasl and Cobb(1966)는 건강행위를 “질병을 예방하거나 증상이 없는 단계에서 질병을 조기 발견하기 위한 목적으로, 개인이 스스로 건강해지기 위해 행해지는 모든 활동”이라고 하였다(Kasl and Cobb,1966, in(Jean K. Langlie, 1979)). 여기에는 건강검진, 백신 접종 등 의료기관 이용과 관련된 행위들의 개념이 포함되었다. 이후 건강을 유지하거나 손상을 예방하는 활동 등을 건강행위에 포함하는 일련의 연구들이 이어졌고(Jean K. Langlie, 1979) 최근에는 건강행위를 주로 생활습관(Life style)의 개념으로 받아들이게 되었다.

건강행위에 대한 대표적 연구로 1965년부터 미국 캘리포니아주의 Alameda County에서 6,928명의 지역주민을 대상으로 건강행위와 건강 수준에 대해 추적 조사를 실시하고 이를 분석한 코호트 연구를 들 수 있다. 이 연구에서는 일반적으로 건강과 관련이 있다고 생각되는 7가지 건강행위와 삶의 질, 만성 질병, 사망률 등의 건강수준과의 연관성을 밝혔다. 이 연구에 포함된 7가지 건강행위는 첫째, 식사를 규칙적으로 하는 것, 둘째, 아침 식사를 하는 것, 셋째, 밤에 7~8시간의 수면을 취하는 것, 넷째, 자신의 키에 대한 적당한 체중을 유지하는 것, 다섯째, 흡연을 하지 않는 것, 여섯째, 소량의 적절한 음주를 하는 것, 일곱째, 규칙적인 운동을 실시하는 것 등이었다. 그 후 일련의 연구에서 이러한 건강행위들이 심혈관계 질환, 뇌졸중, 암 등의 건강 수준에 미치는 영향을 다뤄왔다(Housman and Dorman, 2005).

건강행위를 많이 할수록 건강 수준이 좋아지고, 수명이 길어진다. Belloc(1973)은 Alameda county연구의 5년 반 동안 추적결과를 바탕으로 한 연구에서, 건강행위를 6~7개 실천한 남성은 4개 미만 실천한 그룹에 비해 기대수명이 11년이 더 길고, 여성에서는 기대 수명의 차가 7년이라고 하였다. 또한 건강행위가 사망에 미치는 영향에서 연령, 소득, 질병 보유 등을 보정하고서도 사망률에 미치는 영향이 유의하다고 하였다(Belloc, 1973).

2011년 우리나라 사망원인 통계(통계청, 2012)에서 전체 사망 중 악성신생물, 뇌혈관 질환, 심장질환 등이 가장 높은 순으로 나타났다. 이 질환과 당뇨병, 간 질환, 고혈압성 질환 등이 전체 사인에서 차지하는 비율이 56.5%에 이른다. 이는 건강습관과 관련이 있는 질환들로서 건강행위 교정으로 질병을 예방하거나 조기사망을 예방할 수 있는 질환들이다.

Michael(1999)은 선진국에서 발생하는 사망의 50%는 생활습관과 관련된 것이며, 건강증진 사업을 통해 건강습관과 건강을 향상시키면, 이는 곧 의료비 지출을 줄이고 생산성을 올릴 것이라고 연구 자료를 통해 밝혔다(Michael, 1999).

정영호 등(2006)은 흡연, 음주, 운동부족, 영양결핍 등의 4가지 건강 관련 행위만으로 인한 질병 비용만 해도 우리나라 국민의 전체 질병 비용에서 18.46%를 차지한다고 하였으며, 건강행위 요인이 전체 건강위험요인 기여도에서 차지하는 부분이 31.3%에 이른다고 하였다(정영호 et al., 2006). 박정근(2010)은 건강행위로 인한

질병부담 중 비만과 음주에 의한 질병 부담이 가장 많이 차지하는데, 최근 큰 폭으로 상승하였다고 하였다(Figure 2).

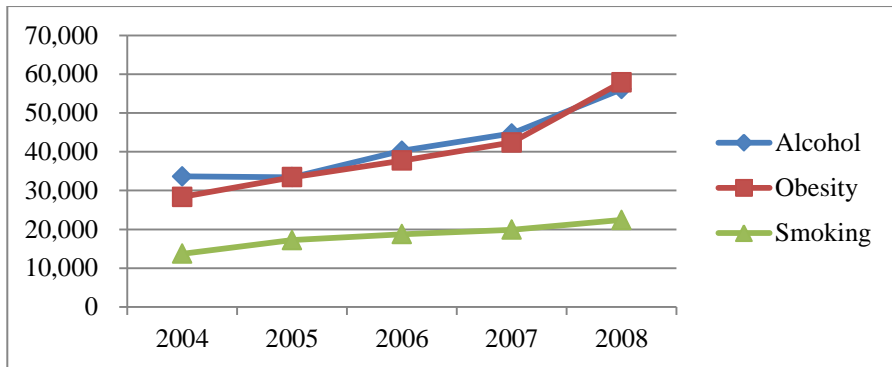


Figure 2. Trends in economic burden of disease in Korea(10 billion won)
(박정근, 2010)

7.1 흡연

담배는 주요 사망원인에 대한 위험요인으로 연구가 이루어져왔다. 전체 사망의 63%는 비감염성 질환으로 인한 것이며, 이 중 흡연은 가장 중요한 위험요인이다. 담배로 인해 매해 6백만 명이 사망하는데, 이 중 흡연하지 않는 사람이 간접 흡연으로 인한 사망은 60만 명에 이른다(WHO, 2011).

담배 연기에는 유해한 화학물질이 포함되어 있어 악성종양, 호흡기계 질환, 순환기계 질환 등을 유발하고, 대사기능에 영향을 미치며, 임신과 출산과 관련된 위험을 증가시킨다. 또한 간접흡연을 통해 비흡연자에게도 악영향을 미친다는 연구가 발표되면서 사회적 관심이 되고 있다.

흡연을 하면 비흡연자에 비해 총콜레스테롤, 중성지방, 저밀도 지단백콜레스테롤 등이 높아지고, 고밀도 지단백콜레스테롤 저하가 나타날 확률이 유의하게 높으며(Craig et al., 1989), 하루 25개피 이상의 담배를 피는 흡연자는 그 이하를 피는 흡연자나 비흡연자에 비해 고밀도 지단백콜레스테롤 수치는 낮았고, 중성지방 수치는 높게 나타났다고 하였다(Brischetto et al., 1983).

우리나라의 연구에서는 주로 남자에서 높은 흡연율을 보이고, 흡연 시작 연령이 점차 낮아지고 있다. 최근 여성 흡연이 늘면서 문제가 되고 있다. 흡연의 위해성에 대해서는 인식하고 있으나 흡연을 시작한 후 의존을 유발하여 금연하는 데에 어려움을 겪고 있다.

흡연율은 연령이 증가할수록 낮아지는 경향을 보이며 사회경제적으로 취약한 계층에서 높게 나타난다. 국민건강영양조사 분석 결과 성별, 거주지, 결혼 상태, 학력, 직업, 소득 등과 관련성을 보이며, 여자보다 남자에게서, 동 지역에 거주하는 경우보다 읍·면 지역에서, 결혼한 경우에 비해 사별한 경우, 학력이 낮은 경우, 직업이 기능직이거나 단순 노무직인 경우, 소득이 낮은 경우 흡연율이 높게 나타났다(강은정, 2007).

7.2 과도한 음주

알코올 음료를 마시는 것은 사교 모임에서 흔한 모습으로 나타나고 있지만 음주를 함으로써 알코올에 취하고, 그 독성에 영향을 받고, 의존성이 생기는 특성으로 인하여 건강과 사회적인 결과에 위험이 따르고 있다.

만성 질환은 많은 양의 알코올을 섭취하는 경우에 발생하는데, 알코올은 교통사고 등의 상해와 같이 건강상태를 급격히 악화시키기도 한다(WHO, 2011). 최근 우리나라에서도 음주로 인한 교통사고, 음주 후 폭력 등이 개인의 건강이나 지역사회의 안전 수준을 저하시키는 것을 볼 수 있다.

일부 연구에 의해 적정량의 음주를 하는 경우 인슐린 감수성을 증가시키고, 고밀도 지단백콜레스테롤 수치를 높여 심혈관질환을 예방하는 효과가 있다고 알려져 있고(Facchini et al., 1994; 김창희 et al., 2010), 관상동맥 질환으로 인한 사망, 여성의 골다공증 위험을 감소시킨다고 알려져 있다(Sacco et al., 1999).

그러나 알코올을 과량 지속적으로 마실 경우 평균 공복혈당을 상승시키고, 혈압을 높이며, 중성지방 및 혈소판 응집력을 증가시켜 허혈성 심질환의 위험을 높이게 된다(Corrao et al., 2000).

알코올로 인해 신경계 질환, 간경화와 같은 위장질환, 대장암, 직장암, 인두나 후두암, 간암, 식도암, 구강암 등의 악성 종양, 장에 손상을 줄 수 있다. 또한 의도하지 않은 상해를 입을 수 있으며, 허혈성 심질환, 뇌졸중, 고혈압과 같은 심혈관계 질환 위험을 높이며, 태아 증후군이나 조기 출산 부작용 같은 위험이 있으며 당뇨 발생에도 악영향을 미친다.

우리나라 연구에서 음주자의 특성을 살펴보면 여자보다 남자가 음주율이 높고, 스트레스 정도가 높을수록 음주율이 높다. 미혼보다 유배우자나 이혼 또는 별거, 사별을 한 대상에서 높고, 소득이 낮고 교육수준이 낮은 집단에서 높았다. 직업별로는 사무직에 비해 농어업이나 기능단순 노무직에서 그 비율이 높았고, 의료보험에서는 지역의료보험이 직장의료보험이나 의료급여 대상자에 비해 높았다(강은정, 2007).

7.3 신체적 비활동

신체활동이 증가하면 체중 감소와 체지방 감소가 동시에 일어나며, 혈청 중성지방(triglyceride) 농도와 고밀도 지단백(high density lipoprotein, HDL) 콜레스테롤 수치에 변화를 유발하고, 인슐린 분비가 향상된다(Mayer-Davis et al., 1998). 이와 달리 운동을 하면 당불내인성 대상자가 당뇨병으로 진행되는 것을 줄여주고, 인슐린 저항성을 감소시켜 고혈당을 개선시킨다(Pan et al., 1997).

신체활동을 하면 고혈압 환자에서 말초혈관 저항이 낮아지고, 심장지수가 좋아지며, 혈압이 낮아지고 총콜레스테롤이 낮아지는 효과가 있다(Jennings et al., 1986).

최근 연구에 의하면 신체활동으로 심혈관 질환뿐 아니라 다른 요인으로 인한 사망률도 크게 감소시킨다고 알려져 있다. Myers 등(2004)은 운동을 중등도 이상 하거나 일을 하거나 여가시간을 활동적으로 보내는 경우, 사망 위험이 50% 이상 감소한다고 한다.

신체활동에 필요한 에너지 요구량을 MET로 표시하는데¹ (버드캐런저 and 김성옥 역, 2004), 일주일에 신체활동으로 소모되는 칼로리를 평소보다 1,000칼로리 증가시키거나 운동을 1MET(metabolic equivalent) 증가시키면 사망률의 20%를 감소시키는 이득이 있다고 한다(Myers et al., 2004).

암, 인슐린 저항성, 만성피로, 이상지질혈증, 고혈압, 비만, 만성 폐색성 폐질환, 골관절염 및 골다공증 등의 만성질환 환자에 운동을 처방함으로써 질병의 진행 억제, 체력, 삶의 질 등에 좋은 결과를 나타내었다(Pedersen and Saltin, 2006).

노인들을 대상으로 한 연구에서도 권장운동을 실천하는 노인일수록 고혈압, 관절염, 신경통, 위장질환, 빈혈 등의 유병률이 낮았다(진정권, 2008).

신체활동은 남성에 비해 여성이 낮고, 청년이나 노인층에서 낮고, 배우자가 있는 대상자에 비해 미혼이나 이혼/별거/사별의 경우 낮고, 소득 수준이 낮은 대상자에서도 낮게 나타났다(강은정, 2007).

7.4 부적절한 체중(저체중 또는 비만)

정상 체질량지수에 비해 체질량지수 18.5 미만의 저체중에서는 사망 위험이 더 높아지며(Flegal et al., 2005; Visscher et al., 2000), 체질량지수 30 이상의 비만에 대해서는 사망 위험에 대해서 이견이 있었으나, 체질량지수 35 이상의 초고도 비만에서는 25세에서 69세까지 대상자의 사망 위험이 1.16~2.6배로 높았다(Flegal et al., 2005).

¹ MET(metabolic equivalent): 신체활동에 필요한 에너지 요구량을 산소소비량을 기준으로 표시. 운동 중에 사용하는 에너지가 휴식 시 에너지 소비량의 몇 배인가로 표시한다. 1MET는 휴식(편하게 앉아있기) 시 산소 소비량을 말하며, 예를 들어 평지에서 천천히 걷기는 2MET, 계단 내려가기는 3MET에 해당된다.
출처: 버드캐런저, 김성옥 역. 운동심리학. 2004. 대한미디어.

7.5 부적절한 수면 시간

7시간 미만 또는 8시간을 넘게 자는 수면 습관은 비만(Buxton and Marcelli, 2010; Gangwisch et al., 2005), 당뇨, 고혈압, 심혈관질환 위험을 높이는 위험요인이다(Buxton and Marcelli, 2010). 20년동안 추적한 코호트 연구에서 5시간 미만 또는 10시간 이상에서 사망률이 높아졌으며, 7~8시간 수면에서 사망률이 가장 낮게 나타났다(Yeo et al., 2013).

적절한 수면을 한 대상자일수록 건강한 식품을 선택할 가능성이 높아 당이 함유된 간식, 설탕이 함유된 음료수 등은 덜 섭취하고 과일이나 채소를 더 섭취하는 경향이 있다(Buxton et al., 2009).

하룻밤 동안의 수면을 방해 받으면, 아침이나 저녁 식사 전 배고픔을 느끼는 것이 더욱 심해지고, 에너지를 더욱 많이 섭취하게 되며, 그러함에도 음식 섭취 후의 즐거움은 변화가 없어 비만을 촉진할 수 있다(Brondel et al., 2010).

7.6 아침 결식

일본인들을 대상으로 한 연구에서, 젊은 대상자일수록 아침 결식을 하는 비율이 높는데, 여자보다 남자에게서 두 배 정도 높게 나타난다고 하였다. 아침 결식을 하는 사람은 에너지나 칼슘, 철분, 그리고 비타민 D 등을 덜 섭취하게 된다고 한다. 또한 혈압이 높고, 혈중 총콜레스테롤이 높다. 아침을 결식하는 사람은 흡연할 가능성이 높고, 특히 아침 결식을 하는 여성은 알코올을 더 섭취하고 운동은 덜하는 경향이 있다고 하였다(Sakata et al., 2001).

미국 NHANES 데이터를 분석한 연구에서도 아침 식사를 안 하는 경우 micronutrients 섭취가 sodium을 제외하고는 낮게 나타난다고 하였다(Kerver et al., 2006).

타이완 서베이 데이터 연구에서 아침식사를 안 하는 대상자들은 비만할

위험도가 1.34로 높았으며, 아침 식사를 하는 비율이 높을수록 비만 유병률도 낮아졌다고 하였다(Huang et al., 2010).

영국인을 대상으로 한 코호트 연구에서는 아침 식사에서 에너지의 대부분을 섭취하는 대상자는 하루 에너지 섭취량이 많은 그룹에 속한다 하더라도, 평균 체질량지수가 가장 낮았다고 하였다. 또한 관찰기간 동안 모든 대상자가 체중이 늘어났어도, 아침 식사에서 에너지의 많은 부분을 섭취하는 대상자들은 체중 증가가 비교적 적었다고 하였다(Purslow et al., 2008).

미국인을 대상으로 한 코호트 연구에서 아침 식사를 안 하는 대상자는 간식이나 다른 식사 횟수를 늘리더라도 당뇨병의 위험이 증가한다고 하였다(Mekary et al., 2012)

노년층이 아침 식사를 거르지 않는 경향이 있으며, 아침 결식은 사회경제적 요인과의 관련이 있다. 아침 결식을 자주하는 사람은 운동을 덜하는 경향이 있고, 흡연, 음주, 비활동적인 일상생활을 하는 대상자일 가능성이 높다(Anna Keski-Rahkonen et al., 2003).

7.7 잦은 간식 섭취

간식을 자주 하는 사람들은 간식을 하는 시간과 관계 없이 우유나 섬유질이 든 음식 섭취를 덜하는 경향이 강했고, 커피, 차, 콜라, 아이스크림 등을 간식으로 자주 섭취하였다(Hampl et al., 2003). 간식으로 섭취하게 되는 식품의 종류를 고려한다면 자주 간식하는 사람들은 에너지를 더욱 많이 섭취하고, 몸에 좋은 음식은 덜 섭취한다고 볼 수 있다.

또한 코호트 연구에서 저녁 식사 후 간식을 섭취하는 대상자는 그렇지 않은 대상자에 비해 제 2형 당뇨병 발생 위험이 증가하였다(Mekary et al., 2012).

8. 사회인구학적 특성에 따른 건강행위

8.1 성별

예방적 건강행위는 질병, 손상, 죽음에 노출되는 행위부터 질병을 감지하거나 예방하는 적극적인 행위까지 포함된다. 남자들은 흡연, 음주, 약물 복용, 운전 등의 위험감수 행위에 더 많이 노출되어 있고, 예방적 행위를 여자보다 덜 하고 있다.

반면 여자들은 위험을 피하려 하고, 예방적인 행동을 더 많이 취하는데, 이러한 행동들이 건강검진, 구강검진, 예방접종, 안과검진 등 의학적인 중재를 필요로 할 때 더욱 그렇다. 특히 여자들은 자신들의 문제를 의학적 용어로 이해하고 의료시스템 내에서 도움을 찾도록 권장받고 있다 (Nathanson, 1977).

8.2 연령

청년기에서 장년기, 중년기, 노년기 등으로 연령이 높아질수록 만성질환이 발생되어 유병률이 높아진다. 중년기는 건강증진행위에 대한 인식도가 높음에도 불구하고 건강증진행위보다는 건강위험행위를 보이는 대상자가 많고 실제로 긴 근로시간과 스트레스로 직장 내에서 흡연, 문제성 음주를 함께 한다(정소연 et al., 2007).

평균수명의 연장과 저출산의 영향으로 우리 사회는 빠른 속도로 고령화가 진행되고 있다. 2007년 우리나라 평균수명은 79.56세인데 비해 건강수명은 71.0세로(통계청 and e-나라지표) 노년기에는 노화로 인한 신체기능 저하뿐 아니라 질병을 겪으며 살아가게 된다.

한국 노인들을 분석한 한 연구는 노인의 삶은 건강행위 추구로 이루어졌

으며 신체적인 쇠약함을 지각하면서 건강의 중요성을 인식하게 되는 현상을 보여준다고 하였다. 그리고 흡연, 음주, 식이 등 단편적인 지표 대신, 노인들에 맞는 건강행위의 지표로 적절한 일하기, 관계망 형성하기, 체력 유지하기, 적당한 몸 움직이기 등을 제시하였다(최영희 and 김순이, 1999). 주관적 건강상태나 규칙적 운동, 연령 등이 건강행위를 잘 설명하는 요인이라고 하면서 노인들도 건강행위를 실천하도록 기회를 제공해야 한다고 하였다(김남진, 2000).

오현주(2011)의 연구에서는 사회적 지지와 지역사회환경은 직간접적으로 노인의 건강증진행위에 영향을 미치며, 특히, 사회적 지지의 경우 노인의 건강증진행위에 대한 직접적인 영향력이 매우 높았다고 하였다. 사회적 지지와 지역사회 환경은 인지된 장애성과 자기 효능감, 인지된 유익성에 영향을 미치고 이를 매개로 노인의 건강증진행위에 영향을 미친다. 그래서 잘 조성된 지역사회 환경과 높은 사회적지지 환경에 노출될수록 자기 효능감이 높아지며, 이로써 건강증진행위 실천도 강화될 수 있다(오현주, 2011).

노인의 건강증진행위 예측모형을 구축한 연구에서는 생물학적 요인, 사회문화적 요인, 기능 수준, 인지상황적 특성, 인간관계뿐 아니라 상황 관련 요인인 거주지역이나 건강증진서비스의 접근성이 모형에 포함되었다(이경자 et al., 2001).

8.3 결혼

혼인한 사람의 건강 수준이 더 좋을 수 있다(Goldman, 1993; Gove et al., 1983; 이성용, 2005). 그 이유로 혼인이 주는 선별 효과 때문에 건강한 사람과 결혼을 할 가능성이 높고, 또한 결혼해서 배우자가 있는 사람은 배우자의 돌봄으로 좋은 건강행위를 유지할 가능성이 높다는 것이다.

한편 이혼한 남성은 배우자가 없어서의 문제도 있지만 이혼으로 인한 경제적 어려움 때문에 건강 수준이 나빠진다(Lillard and Panis, 1996). 미국 은퇴자 남성 자료에서 혼인 상태의 은퇴 고령 남성이 독신 상태의 은퇴 고

령 남성과 유의미한 사망률 차이를 보이지 않지만, 이혼하거나 별거하거나 혹은 배우자의 사망이라는 삶의 사건을 겪은 은퇴 고령남성보다는 낮은 사망률을 보인다고 하였다. 그러나 이 효과는 건강상태나 소득 및 교육수준을 통제하였을 때에는 고소득층과 저소득층의 차이가 있다고 하여, 이를 혼인 상태 효과라기보다는 개인의 소득상태나 정부의 재정적 지원에 기인한 것일 수 있다고 논의하였다(이성용, 2005).

9. 사회경제적 요인과 건강행위

과거와 달리 감염병이 줄어들고 심혈관 질환 같은 만성질환이 사망에 중요한 영향을 미치는 중요 보건문제가 되었고, 발병기전이 감염병과 다름에도 여전히 사회 경제적 수준이 낮은 계층에서 건강 위험이 증가하는 것은 사회적 조건이 질병의 근본적인 원인으로서 다뤄야 하는 문제임을 의미한다(Link and Phelan, 1995).

Link 등(1995)은 주요 질환에 대한 위험요인들을 규명해 온 대부분의 연구들이 식이, 콜레스테롤 수준, 운동 등과 같은 비교적 근접한 요인들에 중점을 두고 있는 것에 문제를 제기하고, 사회적 조건과 개인적 위험요인을 함께 살펴보았다. 개인이 왜 건강 위험요인을 가지고 있을 위험이 높은지 그 이유를 간과하거나, 사회경제적 수준이나 지지 등이 질병 발생에서 근본적인 것이어서, 질병 발생 기전이 변하더라도 지속적으로 건강에 영향을 준다는 점을 간과한다면, 개인을 중심으로 한 중재만으로 국민건강을 증진하려 하는 오류를 범할 수 있다고 하였다(Link and Phelan, 1995).

Marmot 등(1978)은 영국 자료를 분석하여 관상동맥질환이 육체노동자 계층에 점점 더 많아지고 있으며, 이로 인한 사망률이 상류층보다 더 높아진다고 주장하며 그 변화를 살펴보았다. 특히 이러한 현상은 남성에서 더 심하게 나타나는데, 1930~1950년대에는 상류층에서 심장질환이 높았지만

1961년경에는 하위 계층에서 더 높아졌고, 이러한 현상에 대한 사회계층 분포의 차이는 측정(진단) 방법에 의해 나타난 것이라고 하였다.

하위 계층의 사망률을 악화시키는 것은 이 계층이 흡연을 더 많이 하고, 당을 더 많이 섭취하며, 덜 정제된 밀로 된 빵을 적게 섭취하는 것과 관련 있다고 하였다. 또한 지방 섭취와 관련해서는 사회계층과 관련이 없다고 하면서 사회계층으로 인한 사망률의 차이를 건강행위의 차이로 설명하기도 하였다(Marmot et al., 1978).

또 다른 연구에서는 하위 계층이 상류 계층보다 사망률이 더 높은 현상은 여성에게 더욱 크게 나타났는데, 하위 계층이 흡연을 더 많이 하고, 운동을 덜 하며, 더 비만하고, 내당능(glucose tolerance)이 떨어지는 것과 관련 있다고 하면서 상류계층이 이러한 위험에 덜 노출되는 기전에 대해 관심을 가져야 한다고 하였다(Rose and Marmot, 1981).

최근 질병 발생에 있어 이러한 사회경제적 요인의 중요성을 인식하고 건강행위가 이러한 기전에서 완충 효과, 매개 효과 역할을 할 수 있다는 연구가 있었다. 대사증후군에서 사회경제적 요인의 교육 수준과 흡연, 교육 수준과 음주의 유의미한 상호작용 효과가 있다는 것이다. 백경원(2004)은 교육 수준이 높은 집단에서만 흡연과 음주가 대사증후군에 영향을 미치며, 부의 상관관계를 보였고, 교육 수준이 낮은 집단에서는 이 변수에 대하여 통계적으로 유의하지 않았다고 하였다. 즉, 낮은 교육 수준에서는 건강행위의 역할보다 낮은 교육 수준 자체가 대사증후군에 영향을 미친다는 것이다. 대사증후군 유병과 중재변수(사회심리적 요인, 건강행위 요인)를 함께 분석하여 건강행위 요인과 사회경제적 요인의 상호작용이 대사증후군의 유병 여부에 통계적으로 유의하게 영향을 미쳐 완충 효과가 있음을 밝혔다(백경원, 2004).

중고령 노인에 대한 연구에서, 중년기는 노년기에 비해 사회경제적 지위에 따른 건강 수준에 운동과 음주라는 건강행위 매개 효과가 유의미하여, 운동을 안 하거나 과도한 음주를 하는 군에 비해 운동을 하거나 적정 음주를 하는 군은 주관적, 신체적, 정신적 건강 수준이 높았다고 하였다(서연숙,

2011).

건강 수준과 건강행위에 영향을 주는 변수들을 고려할 때에 사회경제적 상황을 분석하기 위해 교육, 직업, 소득 수준이 사용된다.

교육, 직업, 소득 중 한 가지 지표만 확인한다면 세 가지 다 이용할 때에 비해 정보가 부족하여 잘못된 판단을 할 우려가 있으므로 세 가지 정보를 함께 이용하는 것이 좋다(Wnkleby et al., 1992).

9.1 교육

교육 수준에 대한 자료는 초기 성인기 후부터는 고용 상태와 상관없이 얻을 수 있고 직업이나 소득에 비해서 심하게 변하지 않는 비교적 안정적인 자료이다. 연속변수로 얻을 수 있으며, 계다가 쉽게 응답을 얻을 수 있고, 역학연구에서 자주 사용되며 비교가 가능한 지표이다. 교육받은 양과 지식은 개인의 생활양식과 네트워크에 영향을 줌으로써 건강행위와 습관에 영향을 미친다(Liberatos et al., 1988).

교육 수준 변수가 안정적이기는 하지만 개인의 변화 상황을 반영하지 못할 수 있고 또한 여러 연령대의 대상자들을 왜곡하는 코호트 효과가 일어날 수 있으므로 사회계층 수준을 고려할 때 교육 수준을 단독으로 사용하는 것은 주의가 필요하다(Wnkleby et al., 1992).

9.2 직업

교육 수준에 따라 직업 선택이 영향을 받고 또한 소득으로 연결되는 구조를 이루고 있다. 여러 경로를 통해 직업이 건강 수준에 영향을 미치게 된다. 화학적, 생물학적, 물리적 유해환경과 의사결정력이 낮은 지위, 사회적 지지를 덜 받는 상황 등의 사회심리적 환경 등이 직업으로 인한 건강 수준에 영향을 준다(Lynch and Kaplan, 2000).

직업을 가지고 있다가 실업을 당한 경우에는 건강이 악화될 수 있는데,

구조적 실업으로 인해 청년층이나 노년층보다 중년층에서 건강에 미치는 영향이 크고, 또한 실업으로 인해 불규칙적인 식사, 흡연량의 증가, 음주량의 증가 등 건강위험행위를 하게 될 확률이 높아진다고 하였다(홍대균 et al., 2009).

9.3 소득

직업은 그 종류가 다양하고 교육, 소득과 관련이 있다. 소득은 단일한 기준으로 측정할 수 있어 측정이 쉬운 장점이 있다. 그러나 시간이 지나면서 개인의 경제상황이 달라질 수 있고, 직업적 능력이 향상됨으로써 소득이 증가하게 되는 점을 고려할 때 비교적 불안정한 변수일 수 있다. 또한 개인의 소득으로만 측정하지 말고 가족의 크기를 고려하여 가구 소득으로 사용할 필요가 있다(Liberatos et al., 1988).

소득에 의해 건강, 의식주, 교통, 의료서비스, 문화와 여가·운동·육아 등에 대한 기회 및 환경 오염 물질에의 노출 등이 영향을 받게 되므로 소득은 건강의 다양한 요소들과 관련이 있다. 소득은 개인의 건강에 영향을 주는 물질적 조건과 관련 있는 유용한 변수이다(Lynch and Kaplan, 2000).

III. 연구방법

1. 자료 및 연구의 대상

본 연구는 2007년에서 2009년까지 시행한 제4기 국민건강영양조사 자료(KNHANES IV)(질병관리본부)를 활용하였다. KNHANES는 읍면동, 조사구, 가구 등의 표본추출단위가 단계적으로 구성된 전국을 대표하는 확률 표본으로서 다단계 층화집락 표본추출방법으로 이루어졌다. 대한민국에 거주하는 모든 가구와 국민을 목표모집단으로 하며, KNHANES IV는 「2005년 인구주택총조사(통계청)」의 조사구를 추출틀로 활용하여 총 600개 조사구의 선정된 가구내 만 1세이상 가구원 31,705명을 조사 대상으로 계획되었다. 조사 대상자 중에서 건강설문조사, 검진조사, 영양조사 중 1개 이상 참여자는 24,871명으로 전체조사 참여율은 78.4%였다.

본 연구는 KNHANES IV 자료의 20세 이상 대상자 18,210명 중에서 건강행위 설문에 결측값이 없이 응답한 14,833명(81.46%)을 대상으로 하였으며(Table 8), 건강설문조사와 검진조사, 영양조사의 식습관 조사 등을 이용하였다.

Table 8. Study populations

Year	2007	2008	2009	Total
Women	1,052	3,586	3,837	8,925
Men	983	2,256	2,669	5,908
Weight(year)	0.2	0.4	0.4	1.0

본 연구의 자료는 「국민건강영양조사 원시자료 공개 및 활용규정」에 따라 질병관리본부의 승인을 받은 후 홈페이지(<http://knhanes.cdc.go.kr>)에서 다운로드하여 이용하였다. 다운로드 받은 자료는 「공공기관의 개인정보 보호에 관한 법률」에 근거하여 조사 대상자 개인을 추정할 수 있는 자료는 삭제된 원시자료이다.

본 연구는 서울대학교보건대학원 기관생명윤리위원회의 승인을 받았다 (IRB No: 12-2013-04-03).

2. 변수 정의

2.1 건강위험행위 변수

건강위험행위 군집은 건강위험행위를 2가지 이상 함께 하고 있는 경우로 정의하였다.

Alameda county Study에서 사용한 7가지 건강행위 변수를 이용하여 보건 지침을 지키지 않는 경우를 건강위험행위로 정의하였다(Schoenborn, 1986). 모든 건강행위 변수를 실천함과 실천 안함의 두 가지로 분류하여 건강행위를 실천하지 않는 것, 즉 위험행위를 분석하였다. 건강위험행위는 흡연(CS: current smoking), 과도한 음주(HD: heavy drinking), 신체적 비활동(PI: physical inactivity), 부적절한 체중(IW: inadequate weight), 부적절한 수면(IS: inadequate sleep), 아침 결식(BS: breakfast skipping), 그리고 간식 섭취(FS: frequent snacking) 등의 7가지이다.

- CS는 현재 흡연하는 경우로 정의하였다(Fine et al., 2004; Poortinga, 2007).
- HD는 건강설문에서 음주량을 계산하여 음주량 25g/day 이상으로 정의하였다(Corrao et al., 2004; V Bagnardi et al., 2001).
- PI는 중등도 신체활동을 최소 30분 이상 일주일에 5일 이상 실천하거나 격렬한

신체활동을 최소 20분 이상 일주일에 3일 이상 실천하는 지침을 지키지 않는 경우로 정의하였다(Haskell et al., 2007).

- IW는 체중조절을 안 해서 발생하는 부적절한 체중에 대한 항목으로, 저체중($BMI < 18.5 \text{ kg/m}^2$) 또는 비만($BMI \geq 25 \text{ kg/m}^2$)인 경우로 정의하였다(Weisell, 2002).

- IS는 수면시간이 하루 7시간 미만 또는 8시간 이상인 경우로 정의하였다(Schoenborn, 1986; Yeo et al., 2013).

- BS는 건강설문에서 어제와 오늘 아침식사를 하였는지 질문하였고, 하루라도 아침식사를 안 하였다고 응답한 경우 아침 결식으로 정의하였다.

- FS는 잦은 간식 섭취를 정의한 것으로 간식의 횟수가 증가하면 차나 당이 함유된 간식을 섭취하는 경향이 강하다는 연구에 따라(Hampl et al., 2003), 간식의 종류와 양에 상관 없이 횟수로 분류하였다. 전체 대상자에서 간식 섭취 분포를 확인하고, 분포에서 95%이상에 해당되는 하루 평균 3번 이상 섭취하는 경우로 정의하였다.

2.2 건강 수준 변수

2.2.1 주관적 건강

주관적 건강이 보통 또는 좋다고 응답한 경우 좋음으로 구분하였고, 나쁘다고 응답한 경우를 나쁨으로 구분하였다.

2.2.2 만성질병

만성질병을 진단받은 적이 있는 경우와 없는 경우로 나누어 구분하였다. 만성질병 종류는 건강행위와 관련된 연구(Table 7)에 근거하였으며 표에 정리하였다(Table 9).

포함된 질병 변수는 암(Botteri et al., 2008; Katanoda et al., 2008; V Bagnardi et al., 2001), 요통(Kool et al., 2004), 골다공증(Wolff et al., 1999), 소화기 궤양

(Katanoda et al., 2008), 간질환(Corrao et al., 2004), 당뇨병(Myers et al., 2004), 고혈압(Corrao et al., 2004), 뇌혈관질환, 만성폐색성 폐질환(Chronic obstructive pulmonary disease(Katanoda et al., 2008; Myers et al., 2004), 우울증(Strawbridge et al., 2002), 그리고 이상지질혈증(Myers et al., 2004; Taylor et al., 2004) 등의 질병이다.

또한 만성질환을 종류에 따라 암, 심혈관질환, 뇌혈관질환, 만성 호흡기질환, 내분비질환, 위장질환, 근골격계질환, 정신질환 등으로 나누어 살펴보았다.

Table 9. Diseases included in the Chronic Disease Variable

Variables	Definition	Cluster
Cancer	Diagnosed gastric cancer, liver cancer, colon cancer, lung cancer, breast cancer, cervix cancer	cancer
CVD	diagnosed cerebro-vascular disease by doctor	cerebrovascular disease
CAD	diagnosed myocardial infarction or angina pectoris	cardiovascular disease
Diabetes	FBS \geq 126 mg/dl or diagnosed diabetes by doctor or taking prescribed medicine	endocrinologic disease
Liver disease	Diagnosed liver cirrhosis by doctor	gastro-intestinal disease
Hypertension	SBP \geq 140 or DBP \geq 90mmHg or Diagnosed by doctor or taking prescribed medicine	cardiovascular disease
Dyslipidemia	Total cholesterol \geq 240mg/dl or LDL \geq 160mg/dl or HDL $<$ 40mg/dl or Triglyceride \geq 200mg/dl or diagnosed dyslipidemia by doctor or taking prescribed medicine	endocrinologic disease
COPD	Diagnosed chronic obstructive pulmonary disease by doctor	chronic respiratory disease
Gastric ulcer	Diagnosed by doctor	gastro-intestinal disease
Osteoporosis	Diagnosed by doctor	musculoskeletal conditions
Back pain	Diagnosed by doctor	musculoskeletal conditions
Depressive disorder	Diagnosed by doctor	psychiatric disorders

2.3 개인의 특성 변수

스트레스 인지 정도는 많이 느낀다와 적게 느낀다로 나누었다. 우울감은 최근 2주 이상 연속으로 우울감을 느낀적이 있다와 없다로 나누었다. 사회 인구학적 변수로 성별, 연령, 거주 지역, 결혼상태, 교육, 직업, 소득 수준 등을 사용하였다. 연령은: (1)

20~44, (2) 45~64, (3) 65세 이상으로 나누었다. 거주 지역은: (1) 대도시, (2) 중소도시, (3) 읍·면 등으로 나누었다. 결혼상태는: (1) 결혼, (2) 별거/이혼/사별, (3) 미혼 등으로 나누었다. 교육은: (1) 초등학교 졸업 이하, (2) 중학교 졸업, (3) 고등학교 졸업, (4) 대학교 졸업 등으로 나누었다. 직업은: (1) 비육체노동, (2) 육체노동, (3) 무직 등으로 나누었다. 소득은: 월평균 가계균등화 소득으로서 가구소득을 가족수의 제곱근으로 나누어 (OECD), (1) 1분위, (2) 2분위, (3) 3분위, (4) 4분위 등으로 나누었다.

3. 연구의 모형

본 연구는 개인의 특성, 개인의 건강 수준 등이 건강위험행위 군집에 영향을 미칠 것이라는 가설 하에 다음과 같은 모형으로 연구를 진행하였다 (Figure 3).

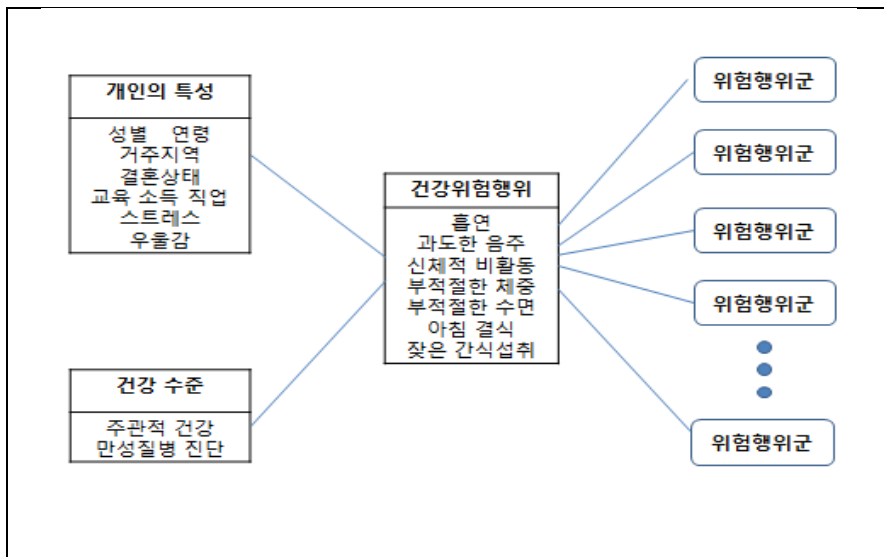


Figure 3. The study model

4. 분석방법

건강위험행위와 관련하여 성별에 따라 차이가 있다는 기존 연구(Chiolero et al., 2006; Poortinga, 2007)에 따라 여자와 남자 그룹으로 나누어서 분석하였으며, 분석 과정은 다음의 그림과 같다(Figure 4).

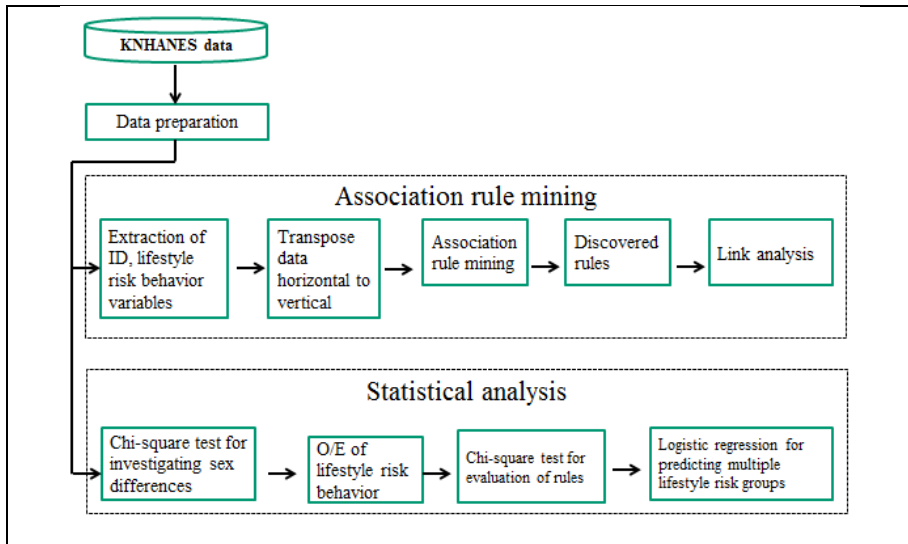


Figure 4. Data analysis framework

자료를 정리하여 ARM의 과정과 통계적 분석 과정의 두 부분으로 분석하였다.

ARM에서는 연구자가 새로 부여한 개인 아이디와 건강위험행위 변수만을 이용하여 자료의 형태를 변환하는 과정을 거친 후 연관규칙을 생성하였다. Minimum support와 confidence의 threshold를 만족하는 연관규칙을 확인한 후, 전체 건강위험행위 변수의 변수별 연관관계를 도식화하는 link diagram을 만들었다.

통계분석에서는 남, 여 대상자의 변수별 분포와 연관성 분석 결과 나타난 연관규칙에 대한 유의성을 확인하고, 건강위험행위 군집을 예측하는 로지스틱 회귀분석을 하였다.

4.1 대상자들의 특징

건강위험행위의 차이를 확인하기 위하여 대상자들을 여자와 남자 그룹으로 나누어 인구사회학적 특징, 건강위험행위 특징, 건강 수준 특징 등을 살펴보았다.

4.2 건강위험행위 군집현상

4.2.1 Accumulation(건강위험행위의 개수)

전체 대상자의 분포에서 건강위험행위의 개수를 살펴보았다. 흡연, 과도한 음주, 신체적 비활동, 부적절한 체중, 부적절한 수면, 아침 결식, 잦은 간식 섭취 등의 7가지 건강위험행위 중 0은 건강위험행위를 한 개도 하지 않는다는 의미이고, 7은 7가지 건강위험행위를 모두 한다는 것을 의미한다.

4.2.2 O/E ratio

건강위험행위의 분포를 살펴보기 위하여 7가지 위험행위의 가능한 조합을 만들어 빈도를 구하였다. 군집을 확인하기 위하여 전체 대상자에서 건강위험행위의 개수가 0인 경우를 제외하고 분석하였다. 여자 대상자는 8,925명 중 8,418명을, 남자 대상자는 5,908명 중 5,670명을 대상으로 건강위험행위의 조합을 분석하였다.

건강위험행위 군집현상은 관찰빈도를 기대빈도로 나누어(O/E ratio) 1보다 큰 경우 군집으로 판단하고(Chou, 2008; Laaksonen et al., 2001; Poortinga, 2007; Schuit et al., 2002; Takakura et al., 2001; 강기원 et al., 2010), 결과 표(Table 13)에 정리하였다.

4.2.3 ARM

건강위험행위 간 연관성을 분석하기 위해 ARM을 하였다.

본 연구에서는 ARM 결과 연관성이 높아 연관규칙에 포함되는 건강위험 행위들은 대상자들이 함께 행하고 있는 건강위험행위이므로 건강위험행위 군집으로 정의하였다.

ARM에서는 분석의 알고리즘이 건강위험행위가 없는 대상자는 분석에서 제외되도록 되어 있다. 건강위험행위가 있는 경우만 분석하기 위하여 전체 대상자에서 건강위험행위의 개수가 0인 경우를 제외하고 분석하였다. 여자 대상자는 8,925명 중 8,418명, 남자 대상자는 5,908명 중 5,670명의 대상자의 건강위험행위를 ARM 하였다.

ARM 시에 연관규칙에 포함되는 건강위험행위 개수를 최대 7개까지 분석 가능하도록 설정하여 건강위험행위의 조합이 어떻게 만들어지는지 분석하였다. Minimum support, confidence 등을 고려하면서 연관규칙을 살펴 보았다. Figure 5, 6에서는 각각의 threshold 수준에서 발견되는 연관규칙의 개수, 연관규칙에 포함된 위험행위 등을 그래프로 제시하였다.

ARM을 위한 minimum support 수준을 높일수록 발견되는 연관규칙의 개수는 점점 줄어들게 된다. 즉, 많은 대상자가 나타내고 있는 군집만 나타나게 되며, 군집에 포함되는 행위의 개수도 줄어들게 된다. 반면 minimum support 수준을 낮게 하면 발견되는 연관규칙의 수가 매우 많아지게 되며, 군집을 나타내는 대상자가 매우 적은 경우에 대한 연관규칙도 나타내게 된다. 이 때에는 비슷한 연관규칙들이 많고 이를 모두 해석해야 하는 단점이 있다.

여자 대상자의 ARM:

전체 대상자의 분포(Table 11)에서 여자 대상자에서는 건강위험행위의 빈도가 낮았고, 또한 건강위험행위 군집의 빈도도 낮았다. 전체 대상자의 분포에서 여자 대상자에서는 1개 또는 2개의 건강위험행위 군집이 가장 많은 빈도를 차지하므로 연관규칙에 최소 2개의 건강위험행위가 포함되고, support와 confidence가 너무 낮아지지 않는 수준에서 ARM을 하는 것이 적절하다고 판단하였다.

Figure 5에서 살펴보면, 연관규칙에 포함되는 건강위험행위의 개수가 minimum support 2%에서 3개 행위가 포함된 연관규칙이 생성되는 것을 알 수 있다.

Minimum support 2%에서는 2개 행위가 포함된 연관규칙, 3개 행위가 포함된 연관규칙이 생성된다는 의미이다. minimum support가 높아짐에 따라 연관규칙이 생성되는 개수가 작아지고, 생성된 연관규칙들 중 가장 높은 confidence를 나타내는 연관규칙의 confidence도 조금씩 낮아진다.

생성되는 연관규칙들에 포함되는 건강위험행위의 개수와 confidence를 조절하고, 연관규칙들이 의미하는 대상자들의 분포를 너무 낮지 않은 정도로 조절하기 위해 support의 threshold를 2%로 정하였고, confidence의 threshold는 50%로 정하였다.

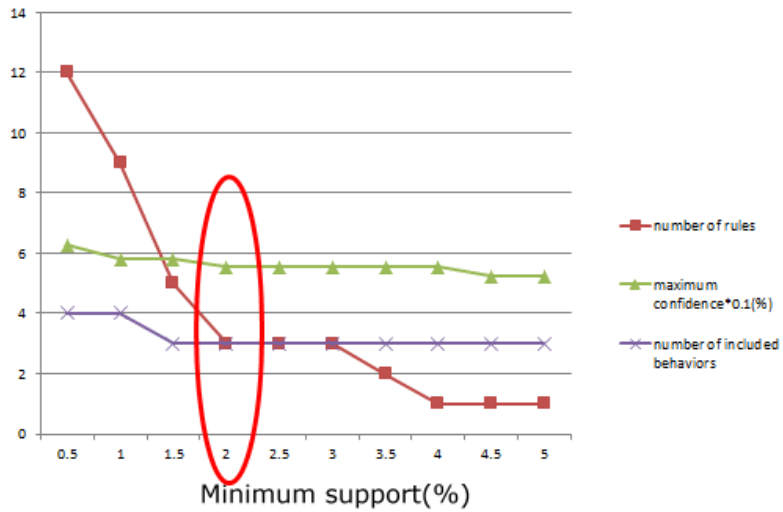


Figure 5. Minimum support, confidence, and number of rules among women under confidence 50%

남자 대상자의 ARM:

반면, 남자 대상자에서는 2개와 3개 건강위험행위 군집의 비율이 가장 높기 때문에(Table 11) 이러한 군집이 포함되는 수준을 고려하였고, 이에 support threshold를 2%, confidence의 threshold를 60%로 정하였다 (Figure 6).

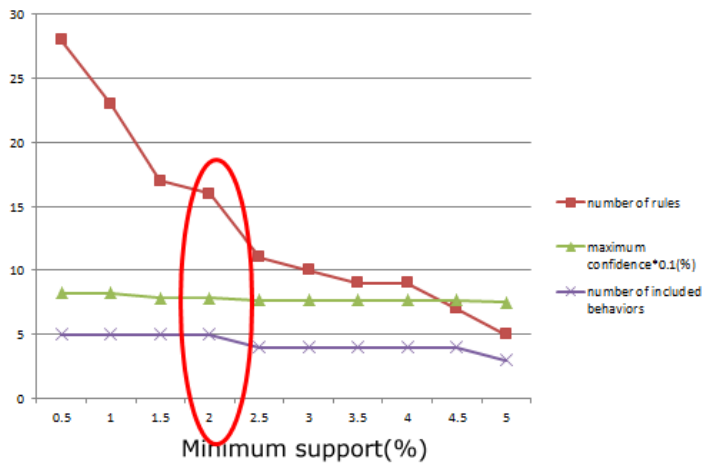


Figure 6. Minimum support, confidence, and number of rules among men under confidence 60%

Support 2%라는 것은 ARM으로 얻어진 연관규칙 중 건강위험행위 조합이 전체 빈도에서 2% 이상이 되는 경우만 허용하겠다는 것을 의미한다.

Confidence 60%는 어떠한 건강위험행위들이 있을 때 다른 건강위험행위를 하는 경우를 말하는 조건부 확률이 60% 이상 되는 연관규칙만 허용하겠다는 것을 의미한다.

Lift의 threshold를 “1보다 큰 경우”라고 정한 것은 건강위험행위 간에 양의 상관관계가 있는 규칙(rules)만 허용하겠다는 것을 의미하며, 건강위험행위 군집현상을 확인할 수 있다.

연관규칙을 통하여 건강위험행위의 연관성, 연관성의 방향, 빈번하게 나타나는 건강위험행위 군집 등을 확인하였다.

4.3 ARM 결과의 평가 및 다른 데이터에 적용

Chi-square test를 하여 ARM 결과로 나타난 연관규칙이 통계적으로 유의한지 평가하였다. ARM 결과의 lift값과 Chi-square test결과 OR값을 비교하였다.

다른 건강조사 데이터에서도 본 연구의 연관규칙이 유의하게 나타나는지 확인하기 위하여 2010년 지역사회 건강조사자료를 이용하였다. 연관규칙에 대해 Chi-square test를 하였고, 연관규칙에 의해 만들어진 건강위험행위 군집이 비슷한 분포로 나타나는지 확인하였다.

4.4 건강위험행위 군집 예측 요인

연관규칙에서 함께 나타난 건강위험행위들을 군집으로 정의하고, 발생 빈도가 높고 연관성 강도가 높은 군집들 중 대표적인 군집을 예시로 하여 로지스틱 회귀분석을 이용하여 예측 요인을 살펴보았다.

4.5 분석에 사용한 통계 패키지

KNHANES는 복합표본설계로 이루어진 서베이 조사이므로 전체 대상자의 분포를 확인하는 빈도 분석과 건강위험행위의 군집 발생에 대한 예측 모델을 위한 로지스틱 회귀분석을 시행할 때에 survey procedure를 사용하였다. 조사 자료에서 제시한 건강행위 설문과 검진, 영양 조사에 대한 가중치를 사용하여 surveyfreq, surveylogistic 등의 procedure를 사용하였다. 분석을 위하여 SAS 9.3과 SAS Enterprise Miner 4.3을 활용하였다.

IV. 결과

1. 대상자들의 특징

1.1 대상자들의 일반적 특징

대상자들은 연령 분포, 결혼 상태, 교육 수준, 직업, 가구 소득, 스트레스, 우울감 등의 변수에서 남녀 별로 차이를 보인다(Table 10).

여자 대상자에서는 남자 대상자에 비해 65세 이상 인구 비율이 더 높고 (16.0%), 혼인한 적이 있으나 별거/이혼/사별 등으로 혼자 사는 비율 (17.3%)이 높았다. 초등학교 졸업 이하의 학력(26.5%), 직업이 없는 대상자의 비율(60.8%)이 높았다. 스트레스를 인지한다고 답한 비율(31.3%)과, 우울감을 느낀다고 응답한 비율(18.9%)이 높았다.

반면, 남자에서는 20~44세 대상자(54.9%), 미혼인 대상자(23.2%)의 비율이 높았다. 고졸 이상의 교육 수준을 가진 대상자 비율(75.7%), 비육체 노동, 육체노동 등 직업을 가지고 있는 비율(71.1%)이 더 높았다. 스트레스가 심하지 않다고 응답한 비율(73.8%)과, 우울감을 안 느낀다고 응답한 비율(90.5%)이 높았다.

Table 10. Characteristics of study population

	Overall n=14,833(%)	Women %* (n=8,925)	Men %* (n=5,908)
Age			
20~44	6,076(41.0)	50.7	54.9
45~64	5,163(69.6)	33.3	33.9
65+	3,594(24.2)	16.0	11.2
Region			
Metropolitan	6,293(42.4)	48.7	47.2
Urban	4,750(64.0)	34.8	35.4
Rural	3,790(25.6)	16.6	17.4
Marital status			
Married	10,871(73.5)	68.2	72.0
Separated/Widowed	2,252(30.5)	17.3	4.8
Never married	1,660(11.2)	14.5	23.2
Education			
College, University	3,631(24.5)	25.3	33.8
High school	4,916(33.2)	37.7	41.9
Middle school	1,716(11.6)	10.5	10.7
Non or Elementary school	4,544(30.7)	26.5	13.6
Job			
Non manual	959(7.3)	7.1	12.1
Manual	5,857(44.4)	32.1	59.0
No Job	6,384(48.4)	60.8	29.0
Income level			
1st Quartile	3,863(26.7)	29.2	31.3
2nd Quartile	3,824(26.4)	28.4	29.8
3rd Quartile	3,599(24.8)	25.4	24.3
4th Quartile	3,199(22.1)	17.1	14.7
Stress			
No	10,703(72.2)	68.7	73.8
Yes	4,127(27.8)	31.3	26.2
Depressiveness			
No	12,480(84.1)	81.1	90.5
Yes	2,352(15.9)	18.9	9.5

* Proportions were calculated by surveyfreq procedure, using sample weights from the survey.

1.2 대상자들의 건강위험행위 특징

대상자들의 건강위험행위의 특징을 살펴보았다(Table 11).

남녀 모두 신체적 비활동인 대상자의 비율(74.6%)이 가장 높았으며, 부적절한 수면의 비율(49.1%)이 두 번째로 높았다. 본 연구에서 분석한 7가지 건강위험행위를 하나도 하지 않는 대상자는 5.0%로 나타났다. 건강위험행위를 7가지 모두 하는 대상자는 없었다.

여자 대상자들에서는 7가지 건강위험행위 중 1가지만 하거나 2가지 하는 비율이 가장 높았다(66.8%). 남자 대상자에 비해 흡연을 하거나 위험 음주를 하는 비율은 낮았으나, 신체적 비활동에 해당하는 비율(77.0%)이 높았다. 수면시간이 부적절(48.6%)하거나 간식을 자주 하는 비율(8.4%)도 높았다.

반면 남자 대상자들에서는 건강위험행위를 2가지, 또는 3가지 하는 비율이 가장 높았으며(58.1%), 5가지 위험행위를 하는 대상자도 5.1%로 나타났다. 흡연을 하는 대상자(46.5%)와 과도한 음주를 하는 대상자(18.4%)의 비율이 높았으며, 체중이 저체중이거나 비만인 부적절한 체중(40.2%)의 비율도 높았다.

Table 11. Characteristics of lifestyle risk behaviors

	Overall n=14,833(%)	Women %* (n=8,925)	Men %* (n=5,908)
Number of lifestyle risk behaviors			
0	745(5.0)	5.4	3.6
1	3,737(25.2)	28.6	17.0
2	5,401(36.4)	38.2	31.2
3	3,441(23.2)	21.3	26.9
4	1,182(10.3)	5.4	15.4
5	291(3.0)	0.9	5.1
6	36(0.4)	0.2	0.6
7	0(0.0)	0.0	0.0
Current Smoking			
No	11,836(79.8)	93.1	53.5
Yes	2,997(20.2)	6.9	46.5
Heavy Drinking			
No	13,663(92.1)	97.7	81.6
Yes	1,170(7.9)	2.3	18.4
Physical Inactivity			
No	3,774(25.4)	23.0	28.2
Yes	11,059(74.6)	77.0	71.8
Inadequate Weight			
No	9,398(63.4)	66.3	59.8
Yes	5,435(36.6)	33.7	40.2
Inadequate Sleep			
No	7,557(50.9)	51.4	53.1
Yes	7,276(49.1)	48.6	46.9
Breakfast Skipping			
No	12,525(84.4)	80.7	78.4
Yes	2,308(15.6)	19.3	21.6
Frequent Snacking			
No	13,817(93.2)	91.6	94.0
Yes	1,016(6.8)	8.4	6.0

* Proportions were calculated by surveyfreq procedure, using sample weights from the survey.

1.3 대상자들의 건강 수준 특징

다음은 대상자들의 건강 수준의 특징을 제시한 표이다(Table 12).

여자 대상자들에게서 남자 대상자들에 비해 주관적 건강 수준이 나쁘다고 응답(24.7%)한 비율이 높았다. 또한 대상자의 49.6%가 만성질환을 진단받은 적이 있다고 답하였다.

여자 대상자들이 진단받은 만성질환을 살펴보면, 고혈압, 협심증, 심근경색증 등을 포함하는 심혈관질환 비율이 22.7%, 당뇨병이나 고지혈증 등을 포함하는 내분비질환이 19.9%, 골다공증이나 요통 등을 포함하는 근골격계 질환이 20.2%로 나타났다. 우울증을 진단받은 경우는 4.4%로 남자 대상자들(1.3%)보다 높게 나타났다.

남자 대상자들에게서는 여자 대상자보다 주관적 건강 수준이 나쁘다고 대답한 대상자의 비율이 낮았으나, 만성질환을 진단받았다고 응답한 비율(58.2%)은 더 높았다.

남자 대상자들이 진단받은 만성질환을 살펴보면, 심혈관질환(28.7%), 내분비질환(25.7%), 위장질환(22.1%) 등으로 그 비율이 여자 대상자들보다 높게 나타났다.

Table 12. Characteristics of health status

	Overall n=14,833(%)	Women %* (n=8,925)	Men %* (n=5,908)
Self reported health			
Fair/Good	11,019(74.4)	75.3	83.7
Bad	3,794(25.6)	24.7	16.3
Chronic disease†			
No	5,996(40.4)	50.4	41.8
Yes	8,837(59.6)	49.6	58.2
Cancer			
No	14,544(98.1)	98.1	98.9
Yes	289(1.9)	1.9	1.1
Cardiovascular disease			
No	10,280(69.3)	77.3	71.3
Yes	4,553(30.7)	22.7	28.7
Cerebrovascular disease			
No	14,478(97.6)	98.6	98.3
Yes	355(2.4)	1.4	1.7
Respiratory disease			
No	14,711(99.2)	99.4	99.4
Yes	122(0.8)	0.6	0.6
Endocrinologic disease			
No	11,104(74.9)	80.1	74.3
Yes	3,729(25.1)	19.9	25.7
Gastro-intestinal disease			
No	12,625(85.1)	91.1	77.9
Yes	2208(14.9)	8.9	22.1
Musculoskeletal conditions			
No	11,825(79.7)	79.8	90.0
Yes	3008(20.3)	20.2	10.0
Psychiatric disorders			
No	14,294(96.4)	95.6	98.7
Yes	539(3.6)	4.4	1.3

* Proportions were calculated by surveyfreq procedure, using sample weights from the survey.

†chronic disease includes cancer, hypertension, coronary artery disease, cerebro vascular disease, chronic obstructive pulmonary disease, diabetes, dyslipidemia, liver disease, gastric ulcer, osteoporosis, back pain, and depressive disorder.

2. O/E ratio로 본 건강위험행위 군집현상

O/E ratio로 확인한 건강위험행위 군집현상의 표는 부록에 제시하였다 (Table 30~37).

각 건강위험행위에 해당되는 경우를 1로, 해당되는 않는 경우를 0으로 표기하였으며, 건강위험행위가 0인 경우를 제외하고 분석하였으므로 7개 행위의 조합이(0,0,0,0,0,0,0)에 해당되는 조합은 관찰 대상자수가 0이 되어 O/E값을 0으로 나타나도록 하였다.

건강위험행위 개수가 늘어날수록 확률값이 낮아져서 기대빈도가 낮아지게 되는데, 관찰된 값이 그에 비해 높아서 O/E ratio값이 커지게 된다.

비록 해당되는 대상자수는 매우 적지만 이러한 건강위험행위 유형을 보이는 경우, 건강위험행위 개수가 늘어날수록 나머지 위험행위도 하고 있을 가능성이 높아질 수 있다고 하겠다.

Table 13에 부록에 제시한 Table 32~39에서 O/E값이 1보다 크게 나타나 건강위험행위 군집에 해당되는 경우를 성별, 포함된 건강위험행위의 개수별로 정리하여 제시하였다. Size에서 1은 단일 건강위험행위이고, 2부터 6까지 표시된 경우는 건강위험행위 2개 이상인 건강위험행위 군집에 해당된다.

여자 대상자에서는 두 개 행위의 군집은 11가지, 세 개 행위의 군집은 19가지 등으로 나타났고, 남자 대상자에서는 두 개 행위의 군집은 11가지, 세 개 행위의 군집은 13가지, 네 개 행위의 군집은 18가지 등으로 나타났다.

Table 32~39에서 분석한 O/E값은 관찰빈도를 기대빈도로 나눈 것이므로 ARM의 lift값과 개념적으로 동일하다고 할 수 있다.

O/E값으로 확인한 건강위험행위의 군집은, 전체 대상자를 가능한 건강위

험행위의 조합을 이용하여 128가지의 경우에 대해 빈도분석을 이용하여 분석한 군집이다. 이 경우 각각의 조합이 상호배타적이어서 각 군집의 서로 다른 행위의 조합을 확인할 수 있는 장점이 있는 반면, 건강위험행위를 더 많이 늘려서 분석할 경우 각 셀(cell)의 관찰값이 작아져서 아주 작은 값으로도 O/E값이 크게 변동되는 것을 볼 수 있다.

이처럼 O/E값을 통해서 건강위험행위의 군집을 파악하는 데에는 관찰빈도에 크게 의존하게 되므로 유의하여야 한다.

Table 13. Lifestyle risk behavior group according to the included behaviors

Size	1	2	3	4	5	6
Women			ISBSFS			
			IWISBS			
			IWISFS	IWISBSFS		
			PIBSFS	PIISBSFS		
		BSFS	PIIWIS	HDIWISBS		
		ISBS	HDISBS	CSIWISFS	CSIWISBSFS	
		ISFS	HDIWIS	CSIWISBS	CSPIISBSFS	
	HD	IWBS	HDIWFS	CSPIBSFS	CSPIIWBSFS	
	PI	IWFS	HDIWBS	CSPIISBS	CSPIIWISBS	CSPIIWISBSFS
	IW	IWIS	HDPIBS	CSPIIWBS	CSHDIWISBS	CSHDIWISBSFS
	IS	PIBS	CSBSFS	CSHDBSFS	CSHDPIISBS	CSHDPIIWISBS
	BS	PIFS	CSISFS	CSHDISBS	CSHDPIIWBS	
	FS	HDIS	CSISBS	CSHDIWIS	CSHDPIIWIS	
		CSBS	CSIWFS	CSHDPIBS		
		CSHD	CSIWBS	CSHDPIIS		
			CSPIBS	CSHDPIIW		
			CSHDBS			
			CSHDIS			
			CSHDPI			
	Men				IWISBSFS	
				PIISBSFS		
				PIIWISFS		
			ISBSFS	HDIWISBS	PIIWISBSFS	
		IWIS	IWISFS	HDPIIWFS	HDPIIWISFS	
		PIBS	PIIWIS	CSIWISBS	CSPIISBSFS	
		PIFS	HDIWIS	CSPIBSFS	CSPIIWBSFS	
FS		PIIS	HDBSFS	CSPIISFS	CSPIIWISBS	
IS		PIIW	CSBSFS	CSPIISBS	CSHDISBSFS	CSHDPIIWBSFS
IW		HDIS	CSIWBS	CSPIIWBS	CSHDIWISFS	CSHDPIIWISBS
PI		HDIW	CSPIBS	CSIWBSFS	CSHDIWISBS	
CS		CSFS	CSHDBS	CSHDBSFS	CSHDPIBSFS	
		CSBS	CSHDIS	CSHDISBS	CSHDPIISFS	
		CSHD	CSHDIW	CSHDIWBS	CSHDPIISBS	
		CSISBS	CSHDPI	CSHDIWIS	CSHDPIIWBS	
			CSISBS	CSHDPIBS		
				CSHDPIIS		
				CSHDPIIW		

Size is the number of lifestyle risk behaviors included in the lifestyle risk behavior group

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight,

IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking

BSFS: group of doing breakfast skipping and frequent snacking.

ISBSFS: group of doing inadequate sleep, breakfast skipping, and frequent snacking simultaneously.

IWISBSFS: group of doing inadequate weight, inadequate sleep, breakfast skipping, and frequent snacking simultaneously.

CSIWISBSFS: group of doing current smoking, inadequate weight, inadequate sleep, breakfast skipping, and frequent snacking simultaneously.

CSPIIWISBSFS: group of doing current smoking, physical inactivity, inadequate weight, inadequate sleep, breakfast skipping, and frequent snacking simultaneously.

3. 건강위험행위 ARM 결과

Table 14에 ARM 결과를 제시하였다. 연관규칙(Rule)에 의하면 부적절한 수면, 신체적 비활동, 흡연 등에 관한 것인데, 남자와 여자에서 연관규칙이 다른 형태로 나타났다. 연관규칙(예를 들어 $rule(A \Rightarrow B)$)의 방향을 나타내는 화살표를 대신하여 화살표 왼쪽에 나타나는 변수를 predictor, 오른쪽에 나타나는 변수를 predicted로 정리하여 predicted, lift, confidence, support를 기준으로 하여 내림차순으로 정리하여 제시하였다.

우선 여자 대상자에서 살펴보면(Table 14), minimum support와 confidence의 threshold를 만족하여 ARM 결과로 나타난 연관규칙은 모두 부적절한 수면시간과 관련된 것이었다.

흡연을 하는 대상자는 수면시간이 부적절할 가능성이 높았다(confidence 55.34%, lift 1.06). 신체적 비활동과 흡연을 함께 하는 대상자는 수면시간이 부적절할 가능성이 높았다(confidence 55.14%, lift 1.06). 신체적 비활동과 체중 조절을 안 하는 대상자는 수면시간이 부적절할 가능성이 높았다(confidence 52.38%, lift 1.01).

예를 들어 연관규칙을 설명하면, 신체적 비활동과 흡연이라는 위험행위를 함께 갖고 있는 여자 대상자의 55.14%는 수면시간이 부적절하다. 전체 대상자에서 이 세 가지 위험행위를 다 하고 있는 경우는 2.6%에 해당된다. 임의의 어떠한 대상자가 수면시간이 부적절할 가능성에 비해, 신체적 비활동과 흡연을 모두 하는 대상자가 수면시간이 부적절할 가능성은 1.06배이다.

남자 대상자에서 살펴보면, 연관규칙이 여자 대상자에 비해 더 많이 나타났다. Minimum confidence를 60%로 높였음에도 불구하고, 연관규칙이 16개로 더 많이 나타났다. ARM 결과로 나타난 연관규칙은 모두 신체적 비활동 및 흡연과 관련된 것이었다.

체중조절을 안 하고, 간식을 자주하는 경우 신체적 비활동일 가능성이 높았다

(confidence 78.46%, lift 1.05). 과도한 음주와 아침 결식을 하는 대상자는 흡연을 할 가능성이 높았다(confidence 66.84%, lift 1.53).

예를 들어 연관규칙을 설명하면, 과도한 음주와 아침 결식을 하는 남자 대상자의 66.84%는 흡연을 하고 있다. 전체 남자 대상자의 2.24%는 과도한 음주를 하고, 아침 식사를 안 하며, 흡연을 함께 하고 있다. 임의의 어떠한 대상자가 흡연을 할 가능성에 비해, 과도한 음주와 아침 결식을 모두 하는 대상자가 흡연할 가능성은 1.53배이다.

Table 14. Results of association rule mining of lifestyle risk behaviors

Size	Predictors	Predicted	Count	Support (%)	Confidence (%)	Lift
Women	(n=8,418)					
2	CS	IS	285	3.39	55.34	1.06
3	PI & CS	IS	220	2.61	55.14	1.06
3	PI & IW	IS	1,253	14.88	52.38	1.01
Men	(n = 5,670)					
3	IW & FS	PI	102	1.80	78.46	1.05
3	FS & CS	PI	99	1.75	76.15	1.02
3	IS & CS	PI	890	15.7	75.04	1.01
3	CS & BS	PI	406	7.16	74.63	1.00
4	IS & CS & BS	PI	203	3.58	76.89	1.03
4	IW & CS & BS	PI	184	3.25	75.41	1.01
5	IW & IS & CS & BS	PI	85	1.50	74.56	1.00
3	HD & BS	CS	127	2.24	66.84	1.53
3	IW & BS	CS	244	4.30	63.54	1.45
3	IS & BS	CS	264	4.66	61.54	1.41
3	PI & BS	CS	406	7.16	60.51	1.38
4	PI & HD & BS	CS	90	1.59	66.67	1.53
4	PI & IW & BS	CS	184	3.25	66.43	1.52
4	PI & IS & BS	CS	203	3.58	64.86	1.48
4	IW & IS & BS	CS	114	2.01	62.30	1.42
5	PI & IW & IS & BS	CS	85	1.50	66.93	1.53

ARM results: minimum support 2%, minimum confidence 50% for women, 60% for men. Rules are listed by order of predicted variables, lift, confidence, and support values in descendent order. Size: the number of risk behaviors included in the rule.

Predictors are the variables to the left of the rule, and predicted variables are those to the right of the rule. CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight, IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking.

ARM 결과에서 lift값은 기대빈도에 비해 관찰빈도가 얼마나 높게 나타났느냐 하는 값을 보여준다. Table 30~37에서 분석한 O/E값과 Table 14의 lift값은 같은 방법으로 계산되었다. 다만, ARM에서의 연관규칙에 포함된 건강위험행위의 조합은 Table 30~37에서 제시한 건강위험행위 조합과 달리 각 연관규칙이 상호배타적이지 않다. 그래서 O/E값으로 lift값을 구하려면, 연관규칙에 포함된 건강위험행위가 해당되는 모든 경우에 대해 고려하여 계산할 수 있다.

즉, (CS⇒IS)의 연관규칙에서는 대상자들에서 다른 건강위험행위와 상관 없이 CS와 IS에 해당되는 경우를 찾아 분할표를 만들고 계산할 수 있다. Table 32~35에서 CS를 하는 대상자 515명, 안 하는 대상자 7,903명, IS를 하는 대상자 4,384명, IS를 안 하는 대상자 4,034명, 건강위험행위를 1개 이상 하는 전체 대상자 8,418명 등으로 하여 분할표를 만들면 lift값이 1.06으로 동일하게 나타난다.

이처럼 ARM에서는 연관규칙들이 상호 배타적이지 않아, 연관규칙에 포함된 건강위험행위 개수가 작은 경우 사이즈가 큰(포함된 건강위험행위 개수가 많은) 연관규칙에 흡수될 수 있다는 점이 O/E값을 구한 표에서 군집을 구한 방법과 다르다.

O/E값을 통해 건강위험행위 군집을 분석하였을 때에는 건강위험행위가 많은 경우의 조합에 대해서는 빈도가 낮아져서 연관성의 강도가 약한 군집들에 대해서 더 이상의 추가 정보를 얻기 위하여 다른 빈도 분석을 추가로 해야 한다. 이에 비해 ARM은 연구자가 대상자 분석에 필요한 minimum support, confidence의 threshold를 만족하는 연관규칙을 확인함으로써, 대상자의 수, 연관성의 강도, 군집에 포함된 건강위험행위의 개수 등을 고려한 분석을 할 수 있었다.

또한 연관성의 방향성을 얻을 수 있는데, ARM을 하면서 미리 가설을 가지고 접근하지 않았음에도 불구하고 각 위험행위의 단순한 조합이 아니라, 특정 건강위험행위를 하고 있을 때 함께 하고 있을 가능성이 높은 건강위험행위에 대해서도 설명할 수 있다.

분석에 사용한 7가지 건강위험행위들 간의 네트워크를 그림으로 제시하였다(Figure 7).

각 위험행위의 빈도와 행위간 연관성은 각 노드의 크기, 연결선의 굵기와 색으로 표시하였다. 노드가 클수록 개별 위험행위 빈도가 높고, 연결선이 굵고 붉은 색일수록 위험행위가 함께 일어나는 빈도가 높다.

여자 대상자에서 신체적 비활동, 부적절한 수면, 부적절한 체중의 빈도가 가장 높아 해당되는 노드를 가장 크게 표시하였다. 또한 이 세 위험행위들 간의 연관성이 높아 이 노드들을 연결하는 3개의 선을 붉고 가장 굵은 선으로 표시하였다.

남자 대상자에서 신체적 비활동, 부적절한 수면, 흡연, 그리고 부적절한 체중 빈도가 가장 높아서 그에 해당되는 노드를 가장 크게 표시하였다. 또한 이 위험행위들 간의 연관성이 가장 높아서 해당되는 노드들을 연결하는 6개의 선을 붉고 가장 굵은 선으로 표시하였다.

그림에서 알 수 있듯이 남자 대상자에서 건강위험행위 군집이 여자 대상자보다 더욱 빈번하게 나타남을 알 수 있다.

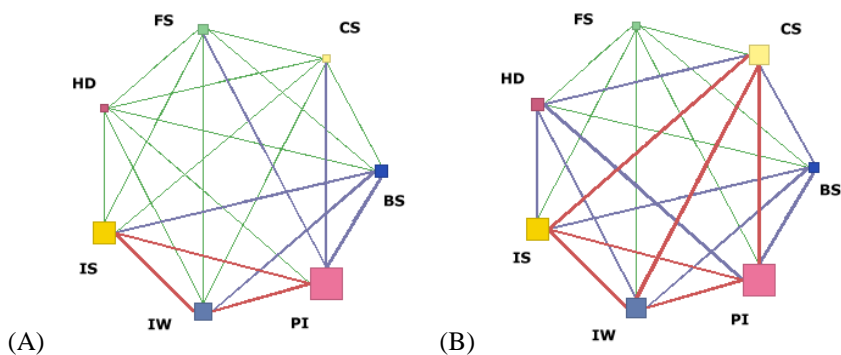


Figure 7. A link diagram derived from association rule mining(ARM)

Link Analysis of SAS Enterprise Miner. Results of Women(A), Men(B) subjects. The size of each node indicates the count number of the behavior it represents. The color of each line indicates the count of the number of the link it represents. Red links have the highest count, blue the middle, and green the lowest count.

4. ARM 결과의 평가

ARM 결과로 나타난 연관규칙이 통계적으로도 유의한지 평가하기 위해 각 연관규칙에 대해 Chi-square test 결과를 함께 제시하였다.

4.1 여자 대상자에서 건강위험행위

ARM 결과로 나온 세 개의 연관규칙에 대해 chi-square test에서도 오즈비가 1보다 크게 나타나며 5% 유의수준에서 통계적으로 유의함을 확인할 수 있다(Table 15).

즉, 흡연을 하고 있는 대상자는 수면시간이 부적절할 가능성이 높으며 이는 통계적으로 유의하다(lift 1.06, OR 1.30, 95% CI 1.09-1.56).

신체적 비활동과 흡연을 함께 하고 있는 대상자는 수면시간이 부적절할 가능성이 높으며 이는 또한 통계적으로 유의하다(lift 1.06, OR 1.29, 95% CI 1.05-1.58).

신체적 비활동과 부적절한 체중을 가지고 있는 대상자는 수면시간이 부적절할 가능성이 높으며 이도 통계적으로 유의하다(lift 1.01, OR 1.20, 95% CI 1.09-1.31).

Table 15. Results of association rule mining* and chi-square test of lifestyle risk behaviors among women

Size [†]	Predictors	Predicted	Count	Lift	<i>p</i>	OR	95% CI
Women (n=8,925)							
2	CS	IS	285	1.06	0.0037	1.30	(1.09, 1.56)
3	PI & CS	IS	220	1.06	0.0140	1.29	(1.05, 1.58)
3	PI & IW	IS	1,253	1.01	<.0002	1.20	(1.09, 1.31)

Rules are listed by order of predicted variables and lift values in descendent order.

* ARM results of Minimum support 2%; minimum confidence 50% for women.

[†] Size is the number of lifestyle risk behaviors included in the rule.

Predictors are the variables to the left of the rule, and predicted variables are those to the right of the rule.

CS: current smoking; HD: heavy drinking; PI: physical inactivity; OB: obesity;

IS: inadequate sleep; BS: breakfast skipping; FS: frequent snacking.

4.2 남자 대상자에서 건강위험행위

대부분의 연관규칙에서 chi-square test에서도 오즈비가 1보다 크게 나타나며 5% 유의수준에서 통계적으로 유의함을 확인할 수 있다(Table 16).

예를 들어 부적절한 수면시간, 흡연을 함께 하고 있는 대상자는 신체적으로 비활동일 가능성이 높으며 이는 통계적으로 유의하다(lift 1.01, OR 1.26, 95% CI 1.09-1.45).

신체적 비활동, 아침 결식을 함께 하고 있는 대상자는 흡연을 하고 있을 가능성이 높으며 이도 통계적으로 유의하다(lift 1.38, OR 2.33, 95% CI 1.98-2.75).

신체적 비활동, 부적절한 체중, 아침 결식을 하고 있는 대상자는 흡연을 하고 있을 가능성이 높으며 이 또한 통계적으로 유의하다(lift 1.52, OR 2.87, 95% CI 2.23-3.70).

신체적 비활동, 부적절한 수면시간, 아침 결식을 하고 있는 대상자는 흡연을 하고 있을 가능성이 높으며 이도 통계적으로 유의하다(lift 1.48, OR 2.69, 95% CI 2.11-3.41).

Table 16. Results of association rule mining* and chi-square test of lifestyle risk behaviors among men

Size [†]	Predictors	Predicted	Count	Lift	<i>p</i>	OR	95%CI
Men	(n=5,908)						
3	IW&FS	PI	102	1.05	0.0773	1.47	(0.96, 2.24)
3	FS&CS	PI	99	1.02	0.2404	1.28	(0.85, 1.93)
3	IS&CS	PI	890	1.01	0.0020	1.26	(1.09, 1.45)
3	CS&BS	PI	406	1.00	0.0903	1.20	(0.98, 1.46)
4	IS&CS&BS	PI	203	1.03	0.0507	1.35	(1.01, 1.80)
4	IW&CS&BS	PI	184	1.01	0.1694	1.24	(0.92, 1.66)
5	IW&IS&CS&BS	PI	85	1.00	0.5300	1.18	(0.77, 1.80)
3	HD&BS	CS	127	1.53	<.0001	2.88	(2.12, 3.91)
3	IW&BS	CS	244	1.45	<.0001	2.56	(2.06, 3.17)
3	IS&BS	CS	264	1.41	<.0001	2.35	(1.92, 2.88)
3	PI&BS	CS	406	1.38	<.0001	2.33	(1.98, 2.75)
4	PI&HD&BS	CS	90	1.53	<.0001	2.83	(1.97, 4.06)
4	PI&IW&BS	CS	184	1.52	<.0001	2.87	(2.23, 3.70)
4	PI&IS&BS	CS	203	1.48	<.0001	2.69	(2.11, 3.41)
4	IW&IS&BS	CS	114	1.42	<.0001	2.34	(1.73, 3.17)
5	PI&IW&IS&BS	CS	85	1.53	<.0001	2.86	(1.97, 4.15)

Rules are listed by order of predicted variables and lift values in descendent order.

* ARM results of Minimum support 2%; minimum confidence 60% for men.

[†] Size is the number of lifestyle risk behaviors included in the rule.

Predictors are the variables to the left of the rule, and predicted variables are those to the right of the rule.

CS: current smoking; HD: heavy drinking; PI: physical inactivity; OB: obesity;

IS: inadequate sleep; BS: breakfast skipping; FS: frequent snacking.

5. 건강위험행위 군집이 나타나는 대상자들의 특징

앞에서 건강위험행위들 간의 연관성을 설명하는 연관규칙을 찾아보았다. 부적절한 수면시간, 아침 결식, 부적절한 체중, 흡연 등과 관련된 연관규칙이었다. 부적절한 수면시간은 주로 흡연, 신체적 비활동과 관련이 있었다. 흡연은 다른 건강위험행위들과 함께 나타나면 신체적 비활동의 가능성을 높였다. 아침 결식은 신체적 비활동이나 다른 건강위험행위들과 함께 하여 흡연의 가능성을 높였다.

이처럼 연관성이 높은 건강위험행위들이 함께 나타나는 경우를 건강위험행위 군집이라고 하고, 그 중 대상자가 많거나 오즈비가 높은 대표적인 건강위험행위 군집을 예시로 하여 군집의 분포를 살펴보았다.

5.1 여자 대상자에서 세 가지 행위 군집의 특성

여자 대상자에서 신체적 비활동과 부적절한 체중을 가지고 있으면 수면시간이 부적절할 가능성이 높다는 연관규칙이 있었는데, 이 연관규칙에 해당되는 대상자가 1,253명으로 가장 많았다. 그리고 이 세 가지 행위와 비교할 수 있는 신체적 비활동과 흡연을 하면 수면시간이 부적절할 가능성이 높다는 연관규칙이 세 가지 행위가 포함된 연관규칙으로 비교할 수 있다.

신체적 비활동, 흡연, 부적절한 수면을 함께 하는 대상자를 PICSIS 군집, 신체적 비활동, 부적절한 체중, 부적절한 수면을 함께 하고 있는 대상자를 PIIWIS 군집으로 하여 두 개의 군집을 살펴보았다(Table 17~18).

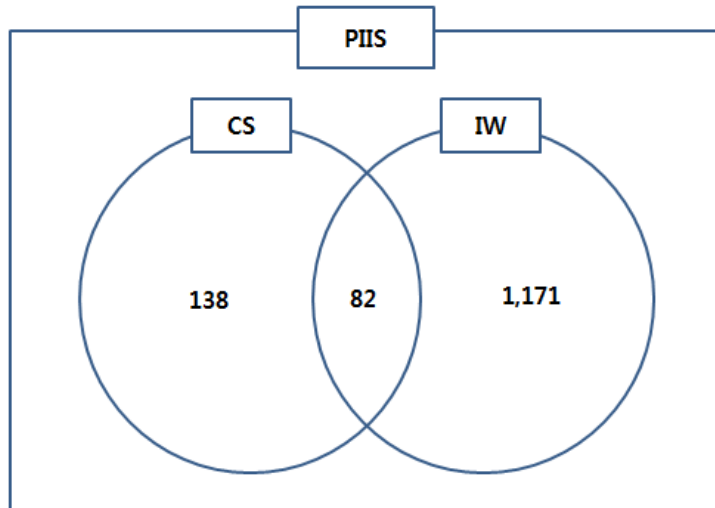


Figure 8. Lifestyle risk group among women

PICISIS: group of doing physical inactivity, current smoking, and inadequate sleep simultaneously.

PIIWIS: group of doing physical inactivity, inadequate weight, and inadequate sleep simultaneously.

연관규칙으로 만들어진 군집은 상호 배타적 그룹이 아니므로 PICISIS 군집과 PIIWIS 군집에서 PI와 IS는 공통되고, CS와 IW만 다르도록 두 군집의 공통된 부분에 대해서는 제외하고 분석하였다. 그래서 대상자 수를 CS에 138명, IW에 1,171명을 대상으로 비교하였다(Figure 8).

Table 17에서 PICISIS 군집은 PIIWIS 군집에 비해 20~44세의 젊은 연령이 많고(51.45%), 대도시 및 중소도시에 거주하는 비율(84.78%)이 높았다. 별거/이혼/사별 및 미혼의 비율이 높았다(59.56%). 고등학교 이상의 학력을 가진 대상자(50.37%)가 많았다. 비육체 노동자의 비율이 높았고(6.87%), 스트레스를 인지한다고 하는 비율(42.03%)이 높았다.

반면, PIIWIS 군집은 45세 이상의 연령(73.02%), 읍면 지역에 거주하는 비율(26.39%)이 높았다. 결혼을 한 유배우자의 비율이 높았다(65.30%). 또한 초등학교 이하의 학력을 가진 대상자(54.06%)의 비율이 높았다. 육체 노동의 비율도 30.64%로 높았다(Table 17).

여자 대상자에서 PIIWIS 군집은 PICSIS 군집에 비해 만성질환을 진단받은 비율이 높았다(70.28%)(Table 18).

Table 17. Lifestyle risk group among women

	PIIS n=3,273(%)	PICSIS % n=138	PIIWIS % n=1,171
Age^{***}			
20~44	1,134(34.65)	71(51.45)	316(26.99)
45~64	1,080(33.00)	28(20.29)	416(35.53)
65+	1,059(32.36)	39(28.26)	439(37.49)
Region[*]			
Metropolitan	1,476(45.10)	67(48.55)	510(43.55)
Urban	1,011(30.89)	50(36.23)	352(30.06)
Rural	786(24.01)	21(15.22)	309(26.39)
Marital status^{***}			
Married	2,117(64.84)	55(40.44)	762(65.30)
Separated/Widowed	875(26.80)	54(39.71)	338(28.96)
Never married	273(8.36)	27(19.85)	67(5.74)
Education^{***}			
College, University	575(17.61)	11(8.03)	140(11.98)
High school	915(28.02)	58(42.34)	281(24.04)
Middle school	324(9.92)	15(10.95)	116(9.92)
Non or Elementary school	1,452(44.46)	53(38.69)	632(54.06)
Job[*]			
Non manual	136(4.52)	9(6.87)	32(2.90)
Manual	947(31.49)	35(26.72)	338(30.64)
No Job	1,924(63.98)	87(66.41)	733(66.46)
Income level			
1st Quartile	713(22.48)	20(14.93)	209(18.46)
2nd Quartile	771(24.31)	31(23.13)	249(22.00)
3rd Quartile	788(24.84)	38(28.36)	306(27.03)
4th Quartile	900(28.37)	45(33.58)	368(32.51)
Stress[*]			
No	2,211(67.57)	80(57.97)	795(67.89)
Yes	1,061(32.43)	58(42.03)	376(32.11)
Depressiveness			
No	2,556(78.09)	103(74.64)	897(76.60)
Yes	717(21.91)	35(25.36)	274(23.40)

PICSIS: group of doing physical inactivity, current smoking, and inadequate sleep simultaneously.

PIIWIS: group of doing physical inactivity, inadequate weight, and inadequate sleep simultaneously.

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

PIIWIS 군집에서 높게 나타난 만성질병은 심혈관질환(44.41%), 내분비 질환(32.45%), 근골격계 질환(31.00%) 등이었다(Table 18).

Table 18. Lifestyle risk group among women

	PIIS n=3,273(%)	PICSIS % n=138	PIIWIS % n=1,171
Self reported health			
Fair/Good	2,189(66.96)	87(63.50)	723(61.79)
Bad	1,080(33.04)	50(36.50)	447(38.21)
Chronic disease ^{***}			
No	1,282(39.17)	65(47.10)	348(29.72)
Yes	1,991(60.83)	73(52.90)	823(70.28)
Cancer			
No	3,194(97.59)	135(97.83)	1,149(98.12)
Yes	79(2.41)	3(2.17)	22(1.88)
Cardiovascular disease ^{***}			
No	2,199(67.19)	99(71.74)	651(55.59)
Yes	1,074(32.81)	39(28.26)	520(44.41)
Cerebrovascular disease			
No	3,190(97.46)	135(97.83)	1,135(96.93)
Yes	83(2.54)	3(2.17)	36(3.07)
Respiratory disease			
No	3,248(99.24)	136(98.55)	1,162(99.23)
Yes	25(0.76)	2(1.45)	9(0.77)
Endocrinologic disease [*]			
No	2,440(74.55)	108(78.26)	791(67.55)
Yes	833(25.45)	30(21.74)	380(32.45)
Gastro-intestinal disease			
No	2,945(89.98)	126(91.30)	1,025(87.53)
Yes	328(10.02)	12(8.70)	146(12.47)
Musculoskeletal conditions ^{**}			
No	2,356(71.98)	112(81.16)	808(69.00)
Yes	917(28.02)	26(18.84)	363(31.00)
Psychiatric disorders			
No	3,078(94.04)	128(92.75)	1,096(93.60)
Yes	195(5.96)	10(7.25)	75(6.40)

PICSIS: group of doing physical inactivity, current smoking, and inadequate sleep simultaneously.

PIIWIS: group of doing physical inactivity, inadequate weight, and inadequate sleep simultaneously.

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

5.2 남자 대상자에서 세 가지 행위 군집의 특성

남자 대상자에서는 부적절한 수면시간과 흡연을 하고 있으면 신체적으로 비활동일 가능성이 높다는 연관규칙의 대상자 수가 가장 많았다. 이 연관규칙은 여자 대상자와 마찬가지로 포함하고 있는 건강위험행위의 종류는 같으나 방향성이 다른 연관규칙이었다. 또한 남자 대상자에서 신체적 비활동과 아침 결식을 하면 흡연을 하고 있을 가능성이 높다는 연관규칙의 대상자가 많았고, 앞의 부적절한 수면과 흡연, 신체적 비활동의 연관규칙과 한 가지의 행위만 다른 연관규칙으로써 비교 대상으로 정하였다.

앞의 연관규칙에서(Table 14) 흡연과 아침 결식을 하면 신체적 비활동의 가능성이 높다(confidence 74.63%, lift 1.0)는 규칙과 신체적 비활동과 아침 결식을 하면 흡연할 가능성이 높다(confidence 60.51%, lift 1.38)라는 두 가지 연관규칙이 있었다.

두 연관규칙에서 방향성과 강도는 다르지만 연관규칙에 포함된 건강위험행위의 종류가 같다는 것을 알 수 있다. 두 연관규칙 모두 행위 간에 양의 상관관계를 보여준다. 이들 행위의 조합을 건강위험행위 군집으로 하여 예시로 살펴보았다.

신체적 비활동, 흡연, 부적절한 수면을 함께 하는 군집을 PICSIS 군집, 신체적 비활동, 흡연, 아침 결식을 함께 하고 있는 군집을 PICSBS 군집으로 하여 두 개의 군집을 예시로 살펴보았다(Table 19~20).

PICSIS 군집과 PICSBS 군집에서 PI와 CS는 공통되고, IS와 BS만 다르도록 두 군집의 공통된 부분에 대해서는 제외하고 분석하였다. 그래서 대상자 수를 IS에 687명, BS에 203명의 대상자를 비교하였다(Figure 9).

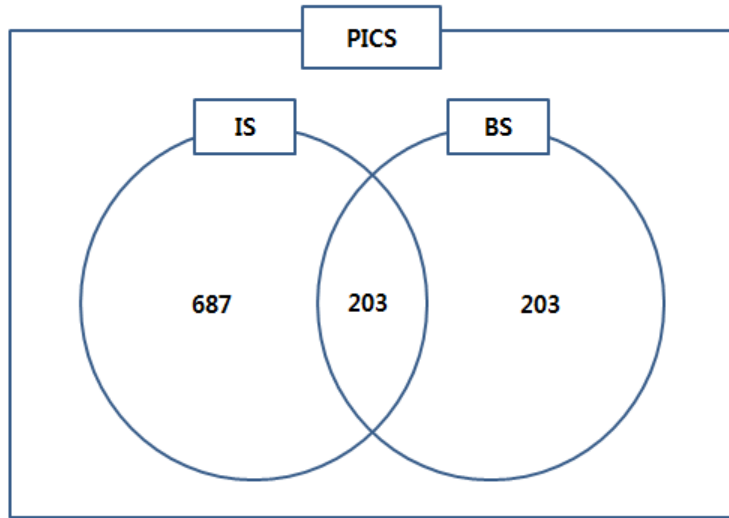


Figure 9. Lifestyle risk group among men

PICSIS: group of doing physical inactivity, current smoking, and inadequate sleep simultaneously.

PIICSBS: group of doing physical inactivity, current smoking, and breakfast skipping simultaneously.

Table 19에서 PICSIS 군집은 PICSBS 군집에 비해 중장년층, 노년층 대상자 비율이 높고(61.43%), 흡연 등의 시골에 거주하는 비율(31.59%), 결혼을 하여 배우자와 함께 사는 대상자의 비율(78.04%)이 높았다. 또한 중학교 졸업 이하의 학력을 가진 대상자(43.71%)가 많았다. 육체노동의 직업 및 직업이 없는 대상자도 많았다(91.53%).

반면, PICSBS 군집은 20~44세 젊은 연령의 대상자(81.28%), 도시에 사는 비율(87.68%)이 높았고, 미혼의 대상자 비율(34.16%)이 높았다. 고등학교 졸업 이상의 학력을 가진 대상자의 비율(91.14%), 비육체 노동의 직업(19.38%)의 비율이 높았다. 소득 4분위의 가장 낮은 소득 수준을 가진 대상자의 비율이 8.96%로 높았다. 스트레스를 인지한다고 하는 대상자(37.93%)의 비율도 높았다.

남자 대상자에서 PICSIS 군집은 PICSBS 군집에 비해 주관적 건강 수준이 나쁘다고 하는 비율(25.44%), 만성질환을 진단받은 비율(64.48%)이 높았다(Table 20).

Table 19. Lifestyle risk group among men

	PICS n=1,601(%)	PICSIS % n=687	PICSBS % n=203
Age ^{***}			
20~44	749(46.78)	265(38.57)	165(81.28)
45~64	549(34.29)	239(34.79)	32(15.76)
65+	303(18.93)	183(26.64)	6(2.96)
Region ^{***}			
Metropolitan	708(44.22)	288(41.92)	116(57.14)
Urban	471(29.42)	182(26.49)	62(30.54)
Rural	422(26.36)	217(31.59)	25(12.32)
Marital status ^{***}			
Married	1,209(75.85)	533(78.04)	121(59.90)
Separated/Widowed	125(7.84)	61(8.93)	12(5.94)
Never married	260(16.31)	89(13.03)	69(34.16)
Education ^{***}			
College, University	437(27.36)	160(23.39)	79(38.92)
High school	616(38.57)	225(32.89)	106(52.22)
Middle school	201(12.59)	99(14.47)	7(3.45)
Non or Elementary school	343(21.48)	200(29.24)	11(5.42)
Job ^{***}			
Non manual	140(10.21)	51(8.47)	31(19.38)
Manual	816(59.52)	360(59.80)	83(51.88)
No Job	415(30.27)	191(31.73)	46(28.75)
Income level ^{***}			
1st Quartile	393(25.06)	157(23.54)	56(27.86)
2nd Quartile	434(27.68)	166(24.89)	75(37.31)
3rd Quartile	386(24.62)	159(23.84)	52(25.87)
4th Quartile	355(22.64)	185(27.74)	18(8.96)
Stress ^{**}			
No	1,160(72.45)	502(73.07)	126(62.07)
Yes	441(27.55)	185(26.93)	77(37.93)
Depressiveness			
No	1,407(87.88)	587(85.44)	184(90.64)
Yes	194(12.12)	100(14.56)	19(9.36)

PICSIS: group of doing physical inactivity, current smoking, and inadequate sleep simultaneously.

PIICSBS: group of doing physical inactivity, current smoking, and breakfast skipping simultaneously.

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

PICSIS 군집에서 나타난 만성질병은 심혈관질환(32.17%), 뇌혈관질환(3.64%) 등이었다(Table 20).

Table 20. Lifestyle risk group among men

	PICS n=1,601(%)	PICSIS % n=687	PICSBS % n=203
Self reported health ^{***}			
Fair/Good	1,267(79.29)	510(74.56)	175(86.21)
Bad	331(20.71)	174(25.44)	28(13.79)
Chronic disease ^{**}			
No	603(37.66)	244(35.52)	96(47.29)
Yes	998(62.34)	443(64.48)	107(52.71)
Cancer			
No	1,589(99.25)	678(98.69)	203(100.00)
Yes	12(0.75)	9(1.31)	0(0.00)
Cardiovascular disease ^{***}			
No	1,134(70.83)	466(67.83)	165(81.28)
Yes	467(29.17)	221(32.17)	38(18.72)
Cerebrovascular disease ^{**}			
No	1,562(97.56)	662(96.36)	203(100.00)
Yes	39(2.44)	25(3.64)	0(0.00)
Respiratory disease			
No	1,586(99.06)	679(98.84)	202(99.51)
Yes	15(0.94)	8(1.16)	1(0.49)
Endocrinologic disease			
No	1,124(70.21)	488(71.03)	151(74.38)
Yes	477(29.79)	199(28.97)	52(25.62)
Gastro-intestinal disease			
No	1,208(75.45)	517(75.25)	159(78.33)
Yes	393(24.55)	170(24.75)	44(21.67)
Musculoskeletal conditions			
No	1,422(88.82)	604(87.92)	182(89.66)
Yes	179(11.18)	83(12.08)	21(10.34)
Psychiatric disorders			
No	1,584(98.94)	679(98.84)	201(99.01)
Yes	17(1.06)	8(1.16)	2(0.99)

PICSIS: group of doing physical inactivity, current smoking, and inadequate sleep simultaneously.

PIICSBS: group of doing physical inactivity, current smoking, and breakfast skipping simultaneously.

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

5.3 남자 대상자에서 네 가지 행위 군집의 특성

남자 대상자에서 네 가지 행위의 군집은 신체적 비활동, 부적절한 체중, 아침 결식, 흡연을 함께 하는 군집을 PIIWBSCS 군집, 신체적 비활동, 부적절한 수면시간, 아침 결식, 흡연을 함께 하고 있는 군집을 PIISBSCS 군집으로 하여 두 개의 군집을 예시로 살펴보았다.

PIIWBSCS 군집과 PIISBSCS 군집에서 PI와 BS, CS는 공통되고, IW와 IS만 다르도록 두 군집의 공통된 부분에 대해서는 제외하고 분석하였다. 그래서 IW에 99명, IS에 118명을 대상으로 비교하였다(Figure 10).

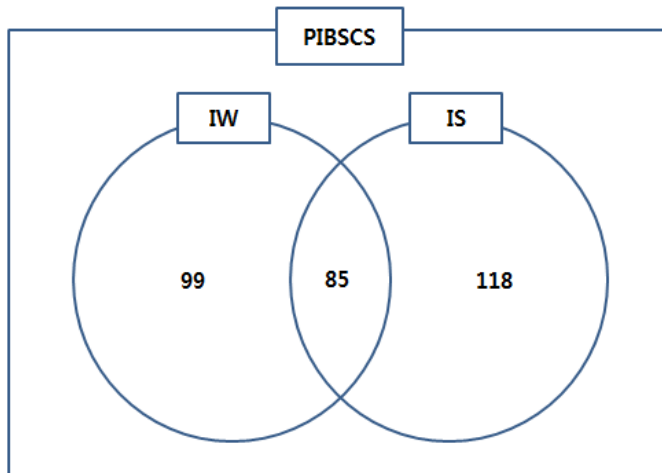


Figure 10. Lifestyle risk group among men

PIIWBSCS: group of doing physical inactivity, inadequate weight, breakfast skipping, and current smoking simultaneously.

PIISBSCS: group of doing physical inactivity, inadequate sleeping, breakfast skipping, and current smoking simultaneously.

PIIWBSCS 군집은 PIISBSCS 군집에 비해 20~44세의 젊은 대상자의 비율(89.90%)이 높았다(Table 21).

반면, PIISBSCS 군집은 PIIWBSCS 군집에 비해 45세 이상의 대상자 비율(24.57%)이 높았다.

남자 대상자에서 PIIWBSCS 군집은 PIISBSCS 군집에 비해 만성질환을 진단받은 비율(58.59%)이 높았다(Table 22).

Table 21. Lifestyle risk group among men

	PIBSCS n=321(%)	PIIWBSCS % n=99	PIISBSCS % n=118
Age*			
20~44	254(79.13)	89(89.90)	89(75.42)
45~64	58(18.07)	9(9.09)	26(22.03)
65+	9(2.80)	1(1.01)	3(2.54)
Region			
Metropolitan	164(51.09)	49(49.49)	48(40.68)
Urban	116(36.14)	38(38.38)	54(45.76)
Rural	41(12.77)	12(12.12)	16(13.56)
Marital status			
Married	187(58.62)	59(60.20)	66(56.41)
Separated/Widowed	17(5.33)	5(5.10)	5(4.27)
Never married	115(36.05)	34(34.69)	46(39.32)
Education			
College, University	125(39.06)	42(42.42)	46(39.32)
High school	164(51.25)	51(51.52)	58(49.57)
Middle school	16(5.00)	4(4.04)	9(7.69)
Non or Elementary school	15(4.69)	2(2.02)	4(3.42)
Job			
Non manual	44(17.32)	15(20.27)	13(13.83)
Manual	135(53.15)	39(52.70)	52(55.32)
No Job	75(29.53)	20(27.03)	29(30.85)
Income level			
1st Quartile	84(26.58)	24(24.49)	28(24.35)
2nd Quartile	109(34.49)	40(40.82)	34(29.57)
3rd Quartile	88(27.85)	28(28.57)	36(31.30)
4th Quartile	35(11.08)	6(6.12)	17(14.78)
Stress			
No	200(62.31)	61(61.62)	74(62.71)
Yes	121(37.69)	38(38.38)	44(37.29)
Depressiveness			
No	293(91.28)	94(94.95)	109(92.37)
Yes	28(8.72)	5(5.05)	9(7.63)

PIIWBSCS: group of doing physical inactivity, inadequate weight, breakfast skipping, and current smoking simultaneously.

PIISBSCS: group of doing physical inactivity, inadequate sleeping, breakfast skipping, and current smoking simultaneously.

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

PIIWBSCS 군집에서 높게 나타난 만성질병은 내분비질환(32.32%), 위장관계 질환(32.32%) 등이었다(Table 22).

Table 22. Lifestyle risk group among men

	PIBSCS n=321(%)	PIIWBSCS % n=99	PIISBSCS % n=118
Self reported health			
Fair/Good	278(86.60)	79(79.80)	103(87.29)
Bad	43(13.40)	20(20.20)	15(12.71)
Chronic disease**			
No	170(52.96)	41(41.41)	74(62.71)
Yes	151(47.04)	58(58.59)	44(37.29)
Cancer			
No	320(99.69)	99(100.00)	117(99.15)
Yes	1(0.31)	0(0.00)	1(0.85)
Cardiovascular disease			
No	265(82.55)	77(77.78)	100(84.75)
Yes	56(17.45)	22(22.22)	18(15.25)
Cerebrovascular disease			
No	320(99.69)	99(100.00)	117(99.15)
Yes	1(0.31)	0(0.00)	1(0.85)
Respiratory disease			
No	319(99.38)	99(100.00)	117(99.15)
Yes	2(0.62)	0(0.00)	1(0.85)
Endocrinologic disease*			
No	247(76.95)	67(67.68)	96(81.36)
Yes	74(23.05)	32(32.32)	22(18.64)
Gastro-intestinal disease***			
No	265(82.55)	67(67.68)	106(89.83)
Yes	56(17.45)	32(32.32)	12(10.17)
Musculoskeletal conditions			
No	290(90.34)	88(88.89)	108(91.53)
Yes	31(9.66)	11(11.11)	10(8.47)
Psychiatric disorders			
No	318(99.07)	99(100.00)	117(99.15)
Yes	3(0.93)	0(0.00)	1(0.85)

PIIWBSCS: group of doing physical inactivity, inadequate weight, breakfast skipping, and current smoking simultaneously.

PIISBSCS: group of doing physical inactivity, inadequate sleeping, breakfast skipping, and current smoking simultaneously.

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

다음은 여자 대상자와 남자 대상자에서 군집 간에 진단받은 만성질환의 개수가 다른지를 비교한 그림이다(Figure 11~13).

세 가지 행위 군집의 비교에서는 여자 대상자에서 PICSIS 군집과 PIIWIS 군집을 비교하였고, 남자 대상자에서는 PICSIS 군집과 PICSBS 군집을 비교하였다.

네 가지 행위 군집의 비교에서는 남자 대상자에서 PIIWBSCS 군집과 PII SBSCS 군집을 비교하였다.

여자 대상자에서 PICSIS, PIIWIS의 세 가지 행위를 하는 군집 간에 만성 질환 개수의 차가 있었고, PICSIS 군집의 진단받은 만성질환 개수가 더 많았다($p < .0001$).

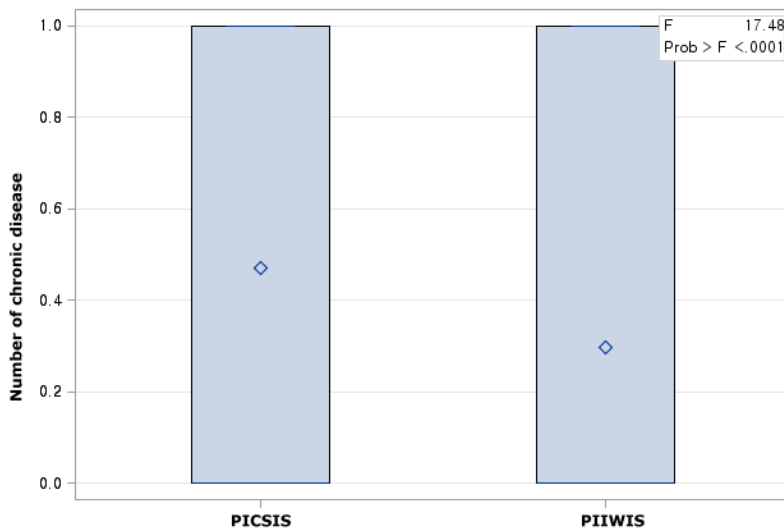


Figure 11. Number of chronic disease of three behavior group among women
Result from GLM. PICSIS: Group of doing physical inactivity, current smoking, and inadequate sleep simultaneously. PIIWIS: Group of doing physical inactivity, inadequate weight, and inadequate sleep simultaneously.

남자 대상자에서 PICSIS, PICSBS의 세 가지 행위를 하는 군집 간에 만성질환 개수의 차가 있었고, PICSIS 군집의 진단받은 만성질환 개수가 더 많았다 ($p < .0001$).

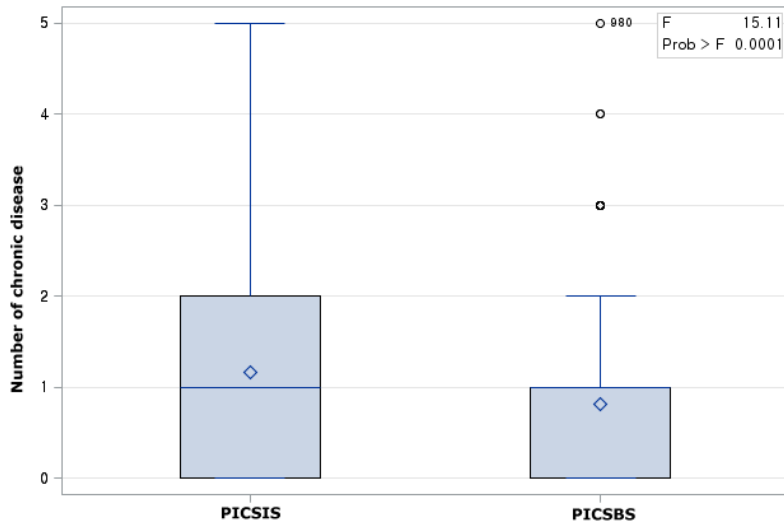


Figure 12. Number of chronic disease of three behavior group among men
 Result from GLM. PICSIS: Group of doing physical inactivity, current smoking, and inadequate sleep simultaneously. PICSBS: Group of doing physical inactivity, current smoking, and breakfast skipping simultaneously.

남자 대상자에서 PIIWBSCS 군집과 PIISBSCS 군집의 네 가지 행위를 하는 군집 간에 만성질병 개수의 차가 있었고, PIIWBSCS 군집의 진단받은 만성질병 개수가 더 많았다($p=0.0029$).

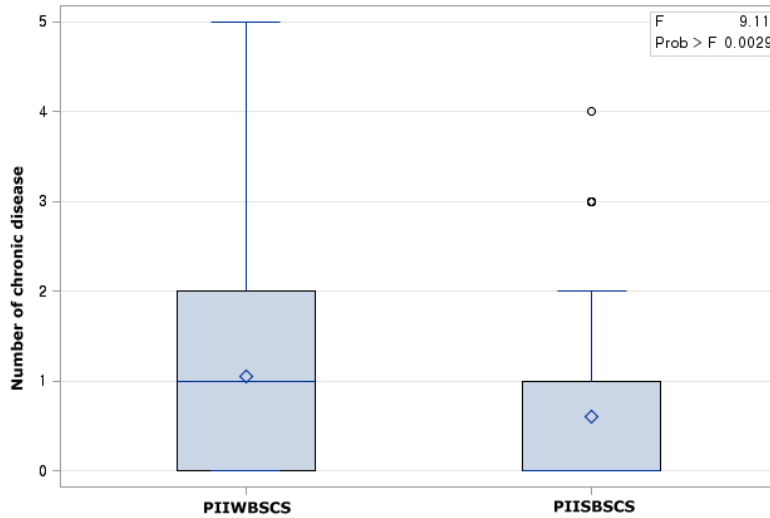


Figure 13. Number of chronic disease of four behavior group among men
 Result from GLM. PIIWBSCS: Group of doing physical inactivity, inadequate weight, breakfast skipping, and current smoking simultaneously. PIISBSCS: Group of doing physical inactivity, inadequate sleep, breakfast skipping, and current smoking simultaneously

Table 17~22와 Figure 11~13을 통해 같은 건강위험행위의 개수에 해당되는 군집이더라도 건강위험행위 군집 간에 인구학적 분포 및 건강 수준이 다를 수 있다.

6. 건강위험행위 군집을 예측하는 요인

다음은 건강위험행위 군집이 나타나는 예측 요인을 알아보기 위하여 대표적인 군집에 대하여 다중 로지스틱 회귀분석을 하였다.

반응변수는 건강위험행위의 군집이고, 설명변수에는 개인의 특성, 만성질병 진단 등의 변수들을 포함하였다. 다중 로지스틱 회귀분석으로 참조 범주에 비하여 설명변수의 영향이 얼마나 크게 미치는지 분석하였으며, 참조 범주에는 분석하고자 하는 건강위험행위 군집에 해당되지 않는 모든 경우를 적용하였다. 예를 들어 PICSIS 군집에는 대상자가 PI, CS, IS 중 하나라도 해당이 안 되면 참조 범주에 해당되도록 하였다.

6.1 여자 대상자에서 PICSIS, PIIWIS 군집의 예측 요인

Table 23에서 살펴보면, PICSIS 군집은 20~44세 대상자에 비해 연령이 높으면 위험도가 낮아지고(OR 0.27, 95% CI 0.13-0.54), 유배우 대상자에 비해 별거/이혼/사별을 하거나(OR 2.49, 95% CI 1.60-3.89), 미혼인 경우(OR 2.46, 95% CI 1.57-3.87) 위험도가 높았다. 대학교 졸업 이상의 교육 수준 대상자에 비해 교육 수준이 낮은 경우(OR 7.91, 95% CI 3.20-19.53)와 소득 수준이 1분위로 높은 경우에 비해 소득 수준이 낮은 경우(OR 1.69, 95% CI 1.07-2.66) 위험도가 높았다.

만성질병 진단의 경우 유의하지 않았으며, 스트레스를 호소하는 경우 위험도가 높았다(OR 2.03, 95% CI 1.42-2.90).

PIIWIS 군집은 연령, 혼인 여부, 직업은 유의하지 않았다. 교육 수준이 낮을수록(OR 3.11, 95% CI 2.24-4.33), 소득 수준이 낮을수록(OR 1.28, 95% CI 1.00-1.63) 위험도가 높았다.

만성질병을 진단받은 경우 위험도가 높았으며(OR 1.36, 95% CI 1.13-1.64), 스트레스 인지는 유의하지 않았다.

Table 23. Multiple logistic regression analysis for predicting patterns of three lifestyle risk behaviors among women

	Model For PICSIS (n=220) vs no PICSIS		Model for PIIWIS (n=1,253) vs no PIIWIS	
	OR	(95% CI)	OR	(95% CI)
	Age group			
20~44	1		1	
45~64	0.29	(0.16, 0.52)	0.96	(0.77, 1.20)
65+	0.27	(0.13, 0.54)	1.15	(0.85, 1.57)
Marital status				
Married	1		1	
Separated/ Widowed	2.49	(1.60, 3.89)	0.92	(0.76, 1.11)
Never married	2.46	(1.57, 3.87)	1.25	(0.91, 1.71)
Education				
College	1		1	
High school	3.51	(1.93, 6.38)	1.44	(1.13, 1.82)
Middle school	4.77	(1.89, 12.02)	1.79	(1.29, 2.47)
Elementary school	7.91	(3.20, 19.53)	3.11	(2.24, 4.33)
Job				
Non manual	1		1	
Manual	1.06	(0.58, 1.93)	0.77	(0.57, 1.04)
No job	1.18	(0.70, 2.01)	1.02	(0.79, 1.31)
Income				
1st quartile	1		1	
2nd quartile	1.03	(0.62, 1.73)	1.05	(0.83, 1.33)
3rd quartile	1.69	(1.07, 2.66)	1.29	(1.03, 1.63)
4th quartile	1.59	(0.93, 2.72)	1.28	(1.00, 1.63)
Chronic disease				
No	1		1	
Yes	1.15	(0.74, 1.78)	1.36	(1.13, 1.64)
Stress				
No	1		1	
Yes	2.03	(1.42, 2.90)	1.15	(0.97, 1.36)

Logistic regression models were constructed with survey logistic procedure using survey sample weights.

Model for PICSIS: the model predicting the probability of simultaneous PI, CS, and IS among women.

Model for PIIWIS: the model predicting the probability of simultaneous PI, IW, and IS among women.

6.2 남자 대상자에서 PICSIS, PICSBS 군집의 예측 요인

Table 24에서 살펴보면, PICSIS 군집은 20~44세 대상자에 비해 연령이 높으면 위험도가 낮아졌다(OR 0.49, 95% CI 0.35-0.68). 유배우 대상자에 비해 별거/이혼/사별을 한 경우 위험도가 높았으나(OR 1.56, 95% CI 1.12-2.16), 미혼인 경우 유의하지 않았다. 대학교 졸업 이상의 학력 대상자에 비해 중학교 졸업 이하의 교육 수준인 경우(OR 1.68, 95% CI 1.22-2.30) 위험도가 높았으며, 소득 수준이 1분위로 높은 경우에 비해 소득 수준이 낮은 경우(OR 1.48, 95% CI 1.11-1.99) 위험도가 높았다.

만성질환의 경우 유의하지 않았으며, 스트레스를 호소하는 경우 위험도가 높았다(OR 1.35, 95% CI 1.12-1.62).

PICSBS 군집은 연령이 높으면 위험도가 낮아졌다(OR 0.10, 95% CI 0.05-0.20). 별거/이혼/사별한 경우(OR 1.79, 95% CI 1.09-2.94), 미혼인 경우(OR 1.58, 95% CI 1.19-2.10) 위험도가 높았다.

교육 수준, 직업은 일관된 결과를 나타내지 않았으며, 2번째, 3번째 사분위의 소득 수준에서 위험도가 높았다(OR 1.48, 95% CI 1.07-2.04). 만성질환의 진단 여부는 유의하지 않았으며, 스트레스를 호소하는 경우 위험도가 높았다(OR 1.73, 95% CI 1.36-2.19).

Table 24. Multiple logistic regression analysis for predicting patterns of three lifestyle risk behaviors among men

	Model For PICSIS (n=890) vs no PICSIS		Model for PICSBS (n=406) vs no PICSBS	
	OR	(95% CI)	OR	(95% CI)
Age group				
20~44	1		1	
45~64	0.60	(0.47, 0.76)	0.42	(0.30, 0.58)
65+	0.49	(0.35, 0.68)	0.10	(0.05, 0.20)
Marital status				
Married	1		1	
Separated/ Widowed	1.56	(1.12, 2.16)	1.79	(1.09, 2.94)
Never married	1.09	(0.84, 1.41)	1.58	(1.19, 2.10)
Education				
College	1		1	
High school	1.10	(0.86, 1.39)	1.16	(0.88, 1.53)
Middle school	1.68	(1.22, 2.30)	0.82	(0.47, 1.43)
Elementary school	1.73	(1.26, 2.38)	0.90	(0.51, 1.58)
Job				
Non manual	1		1	
Manual	0.94	(0.74, 1.20)	0.73	(0.54, 1.00)
No job	0.75	(0.55, 1.03)	0.76	(0.52, 1.11)
Income				
1st quartile	1		1	
2nd quartile	0.99	(0.77, 1.26)	1.41	(1.03, 1.91)
3rd quartile	1.02	(0.80, 1.30)	1.48	(1.07, 2.04)
4th quartile	1.48	(1.11, 1.99)	1.35	(0.90, 2.03)
Chronic disease				
No	1		1	
Yes	0.95	(0.79, 1.13)	0.91	(0.72, 1.14)
Stress				
No	1		1	
Yes	1.35	(1.12, 1.62)	1.73	(1.36, 2.19)

Logistic regression models were constructed with survey logistic procedure using survey sample weights.

Model for PICSIS: the model predicting the probability of simultaneous PI, CS, and IS among men.

Model for PICSBS: the model predicting the probability of simultaneous PI, CS, and BS among men.

6.3 남자 대상자에서 PIWBSCS, PIIBSCS 군집의 예측 요인

Table 25에서 살펴보면, PIWBSCS 군집은 20~44세 대상자에 비해 연령이 높으면 위험도가 낮아졌다(OR 0.04, 95% CI 0.02-0.12). 유배우 대상자에 비해 별거/이혼/사별을 한 경우(OR 2.39, 95% CI 1.24-4.61) 위험도가 높았다.

교육 수준 및 직업은 유의하지 않았으며, 소득 수준이 1분위의 높은 대상자에 비해 2분위 소득의 대상자인 경우 위험도가 높았다(OR 1.68, 95% CI 1.10-2.58).

만성질환을 진단받은 적이 있는 경우(OR 1.74, 95% CI 1.26-2.41), 스트레스를 호소하는 경우(OR 1.45, 95% CI 1.05-2.00) 위험도가 높았다.

PIIBSCS 군집은 연령이 높으면 위험도가 낮아졌다(OR 0.09, 95% CI 0.04-0.22). 별거/이혼/사별한 경우는 유의하지 않았으며, 미혼인 경우(OR 1.78, 95% CI 1.18-2.70) 위험도가 높았다.

교육 수준, 직업은 일관된 결과를 나타내지 않았으며, 4번째 사분위의 낮은 소득 수준에서 위험도가 높았다(OR 1.85, 95% CI 1.06-3.20). 만성질환의 진단 여부는 유의하지 않았으며, 스트레스를 호소하는 경우 위험도가 높았다(OR 1.80, 95% CI 1.29-2.51).

Table 25. Multiple logistic regression analysis for predicting patterns of four lifestyle risk behaviors among men

	Model For PIIWBSCS (n=184) vs no PIIWBSCS		Model for PIISBSCS (n=203) vs no PIISBSCS	
	OR	(95% CI)	OR	(95% CI)
	Age group			
20~44	1		1	
45~64	0.29	(0.18, 0.46)	0.55	(0.36, 0.84)
65+	0.04	(0.02, 0.12)	0.09	(0.04, 0.22)
Marital status				
Married	1		1	
Separated/ Widowed	2.39	(1.24, 4.61)	1.83	(0.97, 3.43)
Never married	1.38	(0.90, 2.11)	1.78	(1.18, 2.70)
Education				
College	1		1	
High school	1.06	(0.72, 1.55)	1.02	(0.68, 1.52)
Middle school	0.98	(0.48, 2.00)	1.12	(0.56, 2.24)
Elementary school	1.02	(0.45, 2.34)	0.91	(0.40, 2.05)
Job				
Non manual	1		1	
Manual	0.80	(0.54, 1.20)	0.75	(0.48, 1.19)
No job	0.58	(0.35, 0.97)	0.60	(0.33, 1.09)
Income				
1st quartile	1		1	
2nd quartile	1.68	(1.10, 2.58)	1.22	(0.80, 1.85)
3rd quartile	1.57	(0.96, 2.58)	1.42	(0.92, 2.19)
4th quartile	1.51	(0.83, 2.76)	1.85	(1.06, 3.20)
Chronic disease				
No	1		1	
Yes	1.74	(1.26, 2.41)	0.80	(0.58, 1.10)
Stress				
No	1		1	
Yes	1.45	(1.05, 2.00)	1.80	(1.29, 2.51)

Logistic regression models were constructed with survey logistic procedure using survey sample weights.

Model for PIIWBSCS: the model predicting the probability of simultaneous PI, IW, BS, and CS among men.

Model for PIISBSCS: the model predicting the probability of simultaneous PI, IS, BS, and CS among men.

7. 단일 건강위험행위와 건강위험행위의 군집

단일 건강위험행위와 건강위험행위 군집에서 예측 요인을 비교하여 보았다. 단일 건강위험행위에 대한 예측 요인은 부록에 제시하였다(Table 40~41).

7.1 여자 대상자

Table 40에서 여자 대상자에서 PI를 할 위험도를 살펴보면, 20~44세 연령에 비해 45~64세 연령에서 위험도가 낮고, 대학교 졸업 이상의 학력을 가진 대상자에 비해 초등학교 졸업 이하의 학력 대상자에서 위험도가 높았다. 비육체 노동의 대상자에 비해 육체 노동 대상자의 경우 위험도가 낮았다.

CS를 할 위험도를 살펴보면, 20~44세 연령에 비해 45세 이상의 연령에서 위험도가 낮았다. 결혼을 했으나 이혼/별거/사별 등으로 혼자 사는 경우와 미혼의 경우에서 위험도가 높았다. 대학교 이상의 학력에 비해 고등학교 졸업 이하의 학력 대상자 모두에서 위험도가 높았다. 소득 수준이 가장 낮은 사분위에 해당되는 경우 위험도가 높았다. 스트레스를 인지한다고 한 경우 위험도가 높았다.

IS를 할 위험도를 살펴보면, 20~44세 대상자에 비해 45세 이상의 연령대 모두에서 위험도가 높았고, 유배우 대상자에 비해 이혼/별거/사별의 대상자에서 위험도가 높았다. 대학교 졸업 이상의 학력 대상자에 비해 고등학교 졸업 이하의 모든 대상자에서 위험도가 높았다. 스트레스를 많이 느낀다고 하는 대상자가 위험도가 높았다.

IW를 할 위험도를 살펴보면, 유배우 대상자에 비해 이혼/별거/사별의 대상자에서 위험도가 오히려 낮았다. 대학교 졸업 대상자에 비해 고등학교 졸업 이하의 모든 대상자들에서 위험도가 높았다. 낮은(3분위) 소득 대상자

에서 위험도가 높았고, 만성질환을 진단받은 적이 있다고 응답한 경우 위험도가 높았다(Table 40).

단일 건강위험행위의 예측 요인을 건강위험행위 군집에 대한 로지스틱 회귀분석 결과와 비교해 볼 수 있다.

우선 PICSIS를 할 위험도를 살펴보면(Table 23), 연령이 증가할수록 위험도가 낮고, 배우자 없이 혼자 사는 대상자, 미혼의 대상자, 중학교 졸업 이하의 대상자, 소득이 낮은 경우 위험도가 높았고, 스트레스도 위험도를 높였다. 이는 PI, CS, IS 중 대상자들의 비율이 가장 낮은 CS의 예측 요인과 유사하다.

여자 대상자에서 PIIWIS를 할 위험도를 살펴보면(Table 23), 대학교 졸업에 비해 고등학교 졸업 이하의 대상자에서 위험도가 높았으며, 소득 수준이 낮은 경우 위험도가 높았다. 스트레스는 유의하지 않았으나, 만성질환을 진단받은 경우 위험도가 높았다. 이는 PI, IW, IS 중 대상자들의 비율이 가장 낮은 IW의 예측 요인과 유사하다.

단일 건강위험행위와 건강위험행위 군집에서의 예측 요인을 비교한 결과, 건강위험행위 군집의 예측 요인은 주로 군집을 구성하고 있는 위험행위 중 가장 빈도가 낮은 행위의 예측 요인과 유사하게 나타났다.

7.2 남자 대상자

Table 41에서 남자 대상자에서 PI를 할 위험도를 살펴보면, 연령이나 결혼 상태는 유의하지 않았고, 대학교 졸업 이상의 대상자에 비해 고등학교 졸업의 대상자에서 위험도가 낮았다. 비육체 노동의 직업군에 비해 육체 노동을 하는 직업군에서 위험도가 낮았다. 소득 수준이 1분위로 가장 높은 대상자에 비해 4분위의 낮은 계층에서 위험도가 높았다.

CS를 할 위험도를 살펴보면, 45세 이상의 대상자 모두에서 위험도가 낮았다. 이혼/별거/사별의 대상자에서 위험도가 높았고, 고등학교 졸업 이하

의 대상자 모두에서 위험도가 높았다. 육체노동을 하는 직업군에서 위험도가 높았고, 3, 4분위의 낮은 소득 수준의 대상자들에서 위험도가 높았다. 스트레스를 많이 느낀다고 인지한 대상자에서 위험도가 높았다.

IS를 할 위험도를 살펴보면, 65세 이상의 대상자에서 위험도가 높았고, 이혼/별거/사별의 대상자에서, 스트레스를 많이 느낀다고 인지한 대상자에서 위험도가 높았다.

BS를 할 위험도를 살펴보면, 45세 이상의 대상자들에서 위험도가 낮았다. 이혼/별거/사별의 대상자와 미혼의 대상자에서 위험도가 높았다. 육체노동을 하는 직업군에서는 위험도가 낮았다. 스트레스를 많이 느낀다고 인지한 대상자에서 위험도가 높았다.

IW를 할 위험도를 살펴보면, 65세 이상의 대상자에서 위험도가 낮았다. 중학교 졸업 이하의 모든 대상자에서 위험도가 낮았다. 2, 3분위 소득계층에서는 위험도가 높았고, 만성질환을 진단받은 경우 위험도가 높았다.

단일 건강위험행위의 예측 요인을 건강위험행위 군집에 대한 로지스틱 회귀분석 결과와 비교해 볼 수 있다.

PICISIS를 할 위험도를 살펴보면(Table 24), 연령이 높아질수록 위험도가 낮았다. 결혼을 하였으나 배우자 없이 혼자 사는 경우, 학력 수준이 중학교 졸업 이하의 대상자, 소득 수준이 낮은 경우 위험도가 높았다. 만성질환 진단 여부는 유의하지 않았으며, 스트레스를 많이 느낀다고 하는 경우 위험도가 높았다. 이는 PI, CS, IS 중 대상자들의 비율이 가장 낮은 CS의 예측 요인과 유사하다.

PICSBS를 할 위험도를 살펴보면(Table 24), 연령이 높아질수록 위험도가 낮았고, 배우자 없이 혼자 사는 경우, 미혼인 경우 위험도가 높았다. 교육 수준이나 직업의 경우 유의하지 않았고, 소득 수준이 중간에 해당되는 경우 위험도가 높았다. 만성질환 진단 여부는 유의하지 않았으며, 스트레스를 인지하는 경우 위험도가 높았다. 이는 PI, CS, BS 중 대상자들의 비율이 가장 낮은 BS의 예측 요인과 유사하다.

PIIWBSCS를 할 위험도를 살펴보면(Table 25), 연령이 높아질수록 위험도가 낮아졌고, 결혼을 하였으나 배우자 없이 혼자 사는 경우 위험도가 높았다. 학력 수준과 직업은 유의하지 않았다. 소득 수준이 2번째 사분위로 높은 경우 위험도가 높았다. 만성질환을 진단받은 경우, 스트레스를 많이 느낀다고 하는 경우 위험도가 높아졌다. 이는 PI, IW, BS, CS 중 대상자들의 비율이 가장 낮은 BS의 예측 요인과 유사하다.

PIISBSCS를 할 위험도를 살펴보면(Table 25), 연령이 높아질수록 위험도가 낮아졌고, 미혼의 경우 위험도가 높아졌다. 학력 수준과 직업은 유의하지 않았다. 소득 수준이 4분위 중 네 번째로 가장 낮은 경우 위험도가 높았다. 만성질환 진단 여부는 유의하지 않았으며, 스트레스를 많이 느낀다고 하는 경우 위험도가 높아졌다. 이는 PI, IS, BS, CS 중 대상자들의 비율이 가장 낮은 BS의 예측 요인과 유사하다.

남자 대상자에서도 단일 건강위험행위와 건강위험행위 군집에서의 예측 요인을 비교한 결과, 건강위험행위 군집의 예측 요인은 주로 군집을 구성하고 있는 위험 행위 중 가장 빈도가 낮은 행위의 예측 요인과 유사하게 나타났다.

8. 다른 데이터에 적용

분석 결과 나타난 ARM 결과를 확인하기 위하여 다른 건강조사 자료에서도 유의한지 분석하였다.

8.1 분석데이터

2010년 지역사회건강조사 자료(Community Health Survey: CHS 2010)를 이용하였으며, 건강행위 자료에 결측치가 없는 140,731명(20세 이상 남자 78,993명, 여자 61,738명)을 대상으로 하였다.

CHS는 지역보건의료계획을 수립 및 평가하고, 조사수행 체계를 표준화하여 비교 가능한 지역건강통계를 생산하기 위해 2008년부터 매년 전국 보건소에서 실시하고 있다.

자료는 질병관리본부의 이용 승인의 절차를 받은 후 홈페이지(<https://chs.cdc.go.kr>)에서 다운로드 받아 이용하였다. 본 연구는 서울대학교보건대학원 기관생명윤리위원회의 승인을 받았다(IRB No: 59-2013-10-29).

8.2 변수

건강위험행위 변수로는 흡연(CS), 과도한 음주(HD), 신체적 비활동(PI), 부적절한 체중(IW), 부적절한 수면(IS), 아침 결식(BS) 등의 6가지 변수로 하였다. CHS 2010 자료에서는 간식 섭취(FS)에 대한 설문은 없고, 한국인의 영양지침을 인지하는지에 대한 설문이 있어서 간식 섭취에 대한 변수는 제외하였다. 모든 건강행위 변수를 실천함과 실천 안함의 두 가지로 분류하여 위험행위 변수를 분석하였다.

- CS는 현재 흡연하는 경우로 정의하였다(Fine et al., 2004; Poortinga, 2007).
- HD는 건강설문에서 음주량을 계산하여 음주량 25g/day 이상으로 정의하였다(Corrao et al., 2004; V Bagnardi et al., 2001).
- PI는 중등도 신체활동을 최소 30분 이상 일주일에 5일 이상 실천하거나 격렬한 신체활동을 최소 20분 이상 일주일에 3일 이상 실천하는 지침을 지키지 않는 경우로 정의하였다(Haskell et al., 2007).
- IW는 체중조절을 안 해서 발생하는 부적절한 체중에 대한 항목으로, 저체중(BMI<18.5 kg/m²) 또는 비만(BMI≥25 kg/m²)인 경우로 정의하였다(Weisell, 2002).
- IS는 수면시간이 하루 7시간 미만 또는 8시간 이상인 경우로 정의하였다(Schoenborn, 1986; Yeo et al., 2013).
- BS는 건강설문에서 일주일간 아침식사를 하는 일수를 질문하였고, 하루라도 결식하는 경우를 아침 결식으로 정의하였다.

주관적 건강이 보통 또는 좋다고 응답한 경우 좋음으로 구분했고, 나쁘다고 응답한 경우를 나쁨으로 구분하였다.

만성질환 진단을 받은 적이 있는 경우와 없는 경우로 나누어 구분하였다. 만성질환 종류는 설문조사로 병원에서 진단받은 다음의 질병이 있는 경우로 하였다. 포함된 질병 변수는 고혈압(Corrao et al., 2004), 심근경색, 뇌졸중, 당뇨병(Myers et al., 2004), 이상지질혈증, 관절염, 골다공증(Wolff et al., 1999), 우울증(Strawbridge et al., 2002) 등의 질병이다.

일상생활 중 스트레스 인지 정도는 많이 느낀다와 적게 느낀다로 나누었다. 우울감 인지는 최근 2주 이상 일상생활에 지장이 있을 정도로 슬프거나 절망감을 느낀 적이 있다와 그렇지 않다로 나누었다. 사회 인구학적 변수로 성별, 연령, 거주 지역, 결혼 상태, 교육, 직업, 소득 수준 등을 사용하였다. 연령은: (1) 20~44, (2) 45~64, (3) 65세 이상으로 나누었다. 거주 지역은: (1) 대도시, (2) 중소도시, (3) 읍·면 등으로 나누었다. 결혼상태는: (1)

결혼, (2) 별거/이혼/사별, (3) 미혼 등으로 나누었다. 교육은: (1) 초등학교 졸업 이하, (2) 중학교 졸업, (3) 고등학교 졸업, (4) 대학교 졸업 등으로 나누었다. 직업은: (1) 비육체노동, (2) 육체노동, (3) 무직 등으로 나누었다. 소득은: 가구소득을 4분위로 나누어(OECD), (1) 1분위, (2) 2분위, (3) 3분위, (4) 4분위 등으로 나누었다.

분석에 사용된 변수를 KNHANES IV 자료와 비교하면 다음과 같다 (Table 26).

Table 26. Variables from the KNHANES IV and CHS 2010

	KNHANES IV	CHS 2010
Demographic characteristics	Age	Age
	Region	Region
	Marital status	Marital status
	Education	Education
	Job	Job
	Income level	Income level
	Stress	Stress
Lifestyle risk behaviors	Depressiveness	Depressiveness
	Current smoking	Current smoking
	Heavy drinking	Heavy drinking
	Physical inactivity	Physical inactivity
	Inadequate weight	Inadequate weight
	Inadequate sleep	Inadequate sleep
	Breakfast skipping	Breakfast skipping
Chronic disease	Frequent snacking	
	Cancer	
	CVD	CVD
	CAD	Myocardial infarction
	Diabetes	Diabetes
	Liver disease	
	Hypertension	Hypertension
	Dyslipidemia	Dyslipidemia
	COPD	
	Gastric ulcer	
	Osteoporosis	Osteoporosis
Back pain	Arthritis	
Depressive disorder	Depressive disorder	

개인적 특성에 관한 변수는 동일하였고, 건강위험행위 변수에서 간식 섭취에 대한 변수가 없었으며, 만성질병 진단 여부에 대한 항목에서 조사된

질병의 종류에 차이가 있었다.

8.3 분석방법

CHS 자료는 설문 항목, 조사 방법 등이 KNHANES와 조금 다르게 조사되었다. 그래서, KNHANES 자료에서 분석한 방법을 똑같이 적용하여 같은 결과가 나오는 것을 확인하기가 어렵다.

건강위험행위 항목에서는 간식섭취에 대한 항목이 없고, 검진조사가 없으며, 만성질환 진단 여부 등을 조사하였는데, 질병의 종류가 KNHANES 자료와 다른 부분이 있었다.

KNHANES IV 자료에서는 압, 소화기 궤양, 간질환, 만성 폐색성 폐질환 등에 대한 진단 여부가 조사되었었지만, CHS 2010 자료에서는 이 질환에 대한 질문은 없었고, 심혈관계 질환 중 심근경색에 대해 구체적 진단명으로 조사되었고 또한 관절염을 조사하였다.

만성질환 진단 여부 조사에서는 CHS 2010 자료와 KNHANES IV 자료에서 공통적인 부분을 이용하여 분석하였다.

CHS 2010 자료의 분포를 확인하고 KNHANES 자료에서 분석한 연관 규칙이 CHS 자료에서도 유의한지 chi-square test로 확인하였다.

8.4 대상자들의 특징

Table 27에서 살펴보면, 20~44세의 젊은 연령의 대상자가 많았다(49.2%). 약 63.9%의 대상자가 대도시 및 중소도시에 거주하고 있었다.

전체에서 70.4%의 대상자가 혼인한 상태였으며, 고등학교 졸업 이상의 학력을 가진 대상자가 약 72%에 해당되었다. 약 62%의 대상자가 직업을 가지고 있었다. 스트레스를 호소하는 대상자는 27.6%였으며, 우울감을 호소하는 경우는 약 5.0%에 해당되었다.

Table 27. Characteristics of study population

	Overall n=140,731 (%)	Women %* n=61,738	Men %* n=78,993
Age			
20~44	69,136(49.2)	60.7	57.2
45~64	53,234(37.8)	32.7	34.5
65+	18,361(13.0)	6.6	8.2
Region			
Metropolitan	46,445(33.0)	50.1	47.5
Urban	43,512(30.9)	35.9	35.7
Rural	50,774(36.1)	14.0	16.8
Marital status			
Married	99,064(70.4)	64.8	67.2
Separated/Widowed	15,680(11.2)	13.1	5.7
Never married	25,836(18.4)	22.1	27.1
Education			
College, University	45,864(32.7)	37.7	42.6
High school	55,129(39.2)	40.4	41.7
Middle school	15,611(11.1)	9.2	8.1
Non or Elementary school	23,833(17.0)	12.7	7.6
Job			
Non manual	32,251(30.3)	28.2	45.1
Manual	33,917(31.8)	20.3	27.2
No Job	40,403(37.9)	51.5	27.7
Income level			
1st Quartile	38,388(30.0)	36.1	35.2
2nd Quartile	35,773(28.0)	29.4	30.3
3rd Quartile	29,572(23.1)	21.4	22.0
4th Quartile	24,130(18.9)	13.1	12.4
Stress			
No	101,912(72.4)	71.4	70.2
Yes	38,762(27.6)	28.6	29.8
Depressiveness			
No	133,672(95.0)	95.0	96.6
Yes	6,986(5.0)	5.0	3.4

* Proportions were calculated by surveyfreq procedure, using sample weights from the survey.

Table 28에서 살펴보면, 약 76%의 대상자가 2가지 이상의 건강위험행위를 하고 있다. 가장 빈도가 높은 건강위험행위는 신체적 비활동으로 85.7%였다. 그 외 부적절한 수면시간이 45.6%로 많았다.

여자 대상자에서는 신체적 비활동(90.3%)이 가장 높은 빈도를 보였고, 그 다음으로 부적절한 수면시간(44.6%), 아침 결식(38.1%)으로 나타났다. 남자 대상자에서는 신체적 비활동(87.2%)이 가장 높은 빈도를 보였고, 그 다음으로 흡연(50.1%), 부적절한 수면시간(48.0%)이 그 뒤를 이었다.

Table 29에서 살펴보면 12.9%의 대상자가 주관적 건강 수준이 나쁘다고 응답하였고, 29.3%의 대상자가 만성질환 진단을 받은 적이 있다고 했다. 가장 빈도가 높은 질병은 고혈압(16.4%), 이상지질혈증(7.9%)으로 나타났다.

Table 28. Characteristics of lifestyle risk behaviors

	Overall n=140,731 (%)	Women %* n=61,738	Men %* n=78,993
Number of lifestyle risk behaviors			
0	4,010(2.8)	2.7	1.6
1	29,632(21.1)	26.8	13.2
2	48,254(34.3)	40.1	28.2
3	36,195(25.7)	23.5	29.4
4	16,697(11.9)	5.9	19.2
5	5,127(3.6)	0.8	7.2
6	816(0.6)	0.1	1.2
Current Smoking			
No	99,189(70.5)	95.9	49.9
Yes	41,542(29.5)	4.1	50.1
Heavy Drinking			
No	118,784(84.4)	95.0	75.7
Yes	21,947(15.6)	5.0	24.3
Physical Inactivity			
No	20,112(14.3)	9.7	12.8
Yes	120,619(85.7)	90.3	87.2
Inadequate Weight			
No	100,715(71.6)	76.0	68.6
Yes	40,016(28.4)	24.0	31.4
Inadequate Sleep			
No	76,591(54.4)	55.4	52.0
Yes	64,140(45.6)	44.6	48.0
Breakfast Skipping			
No	96,951(68.9)	61.9	63.1
Yes	43,780(31.1)	38.1	36.9

* Proportions were calculated by surveyfreq procedure, using sample weights from the survey.

여자 대상자의 23.3%에서 만성질환 진단을 받은 적이 있다고 응답하였고, 고혈압과 관절염을 진단받았다고 응답한 비율이 10.7%, 8.5% 등으로 높게 나타났다. 남자 대상자는 고혈압(15.0%), 이상지질혈증(8.3%), 당뇨병(5.7%) 등이 높게 나타났다.

Table 29. Characteristics of health status

	Overall n=140,731 (%)	Women %* n=61,738	Men %* n=78,993
Self reported health			
Fair/Good	122,501(87.1)	89.0	91.0
Bad	18,174(12.9)	11.0	9.0
Chronic disease†			
No	99,497(70.7)	76.7	75.3
Yes	41,234(29.3)	23.3	24.7
Hypertension			
No	117,569(83.6)	89.3	85.0
Yes	23,123(16.4)	10.7	15.0
Diabetes			
No	132,479(94.2)	96.8	94.3
Yes	8,188(5.8)	3.2	5.7
Dyslipidemia			
No	129,406(92.1)	93.4	91.7
Yes	11,155(7.9)	6.6	8.3
Myocardial Infarction			
No	139,485(99.2)	99.6	99.1
Yes	1,167(0.8)	0.4	0.9
Cerebrovascular attack			
No	139,337(99.0)	99.5	99.2
Yes	1,370(1.0)	0.5	0.8
Arthritis			
No	129,771(92.3)	91.5	97.1
Yes	10,900(7.7)	8.5	2.9
Osteoporosis			
No	135,330(96.2)	94.6	99.4
Yes	5,302(3.8)	5.4	0.6
Depressive disorder			
No	138,186(98.2)	97.2	99.2
Yes	2,515(1.8)	2.8	0.8

* Proportions were calculated by surveyfreq procedure, using sample weights from the survey.

† chronic disease includes hypertension, diabetes, dyslipidemia, myocardial infarction, cerebrovascular attack, arthritis, osteoporosis, and depressive disorder.

8.5 KNHANES IV의 연관규칙을 CHS 2010에 적용

KNHANES IV 자료에서 여자 대상자에서는 3가지 연관규칙이 나타났었다(Table 14). 모두 부적절한 수면과 관련된 연관규칙이었다.

Table 14의 연관규칙을 CHS 2010 자료에 적용하여 분석한 결과(Table 30), 흡연하는 대상자는 수면시간이 부적절할 가능성이 높았다($p<.0001$). 신체적 비활동과 흡연을 함께 하는 대상자는 수면시간이 부적절할 가능성이 높았다($p<.0001$). 신체적 비활동과 체중 조절을 안 하여 부적절한 체중을 가진 대상자는 수면시간이 부적절할 가능성이 높았다($p<.0001$).

또한 KNHANES IV 자료에서처럼 CHS 2010 자료에서도 CSIS, PICSIS, PIIWIS의 군집의 support가 건강위험행위가 한 개 이상인 여자 대상자에서 2.22%, 1.98%, 10.53%로, KNHANES IV 자료의 3.39%, 2.61%, 14.88%와 대상자의 규모와 그 순서에서 크게 다르지 않음을 알 수 있다.

Table 30. Rules from the KNHANES IV and chi-square test with the CHS 2010 data among women

Size*	Predictors	Predicted	Support(%)	<i>p</i>	OR	95% CI
Women (n=59,505)						
2	CS	IS	2.22	<.0001	1.42	(1.31, 1.54)
3	PI & CS	IS	1.98	<.0001	1.41	(1.30, 1.54)
3	PI & IW	IS	10.53	<.0001	1.08	(1.04, 1.12)

Rules are listed by order of KNHANES IV results.

*Size: the number of risk behaviors included in the rule.

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight, IS: inadequate sleep, BS: breakfast skipping.

KNHANES IV 자료에서 남자 대상자에서는 16개의 연관규칙이 나타났었다(Table 14). 신체적 비활동 및 흡연과 관련된 연관규칙이었다. 16개의 연관규칙 중 간식 섭취와 관련된 2개의 연관규칙은 변수가 없어서 chi-square값을 구할 수 없었다.

나머지 연관규칙에 대해 Table 14의 연관 규칙을 CHS 2010 자료에도

적용하여 분석한 결과(Table 31), 부적절한 수면과 흡연을 하면 신체적으로 비활동이라는 연관규칙 한 가지를 제외하고 대부분의 연관규칙이 통계적으로 유의하였다($p < .0001$).

예를 들어 부적절한 체중, 부적절한 수면시간, 흡연, 아침 결식을 함께 하고 있는 대상자는 신체적으로 비활동일 가능성이 높았다($p < .0001$).

과도한 음주, 아침 결식을 하는 대상자는 흡연을 하고 있을 가능성이 높았다($p < .0001$).

Table 31. Rules from the KNHANES IV and chi-square test with the CHS 2010 data among men

Size*	Predictors	Predicted	Support(%)	<i>p</i>	OR	95% CI
Men	(n=77,156)					
3	IW&FS	PI	NA			
3	FS&CS	PI	NA			
3	IS&CS	PI	20.14	0.1214	1.04	(0.99, 1.08)
3	CS&BS	PI	16.13	<.0001	1.37	(1.29, 1.44)
4	IS&CS&BS	PI	8.10	<.0001	1.36	(1.27, 1.47)
4	IW&CS&BS	PI	5.23	<.0001	1.41	(1.29, 1.54)
5	IW&IS&CS&BS	PI	2.76	<.0001	1.48	(1.31, 1.69)
3	HD&BS	CS	5.87	<.0001	2.93	(2.76, 3.10)
3	IW&BS	CS	5.94	<.0001	1.64	(1.56, 1.72)
3	IS&BS	CS	9.25	<.0001	1.81	(1.73, 1.88)
3	PI&BS	CS	16.13	<.0001	1.84	(1.78, 1.90)
4	PI&HD&BS	CS	5.09	<.0001	2.89	(2.72, 3.07)
4	PI&IW&BS	CS	5.23	<.0001	1.63	(1.55, 1.71)
4	PI&IS&BS	CS	8.10	<.0001	1.75	(1.68, 1.83)
4	IW&IS&BS	CS	3.11	<.0001	1.61	(1.50, 1.71)
5	PI&IW&IS&BS	CS	2.76	<.0001	1.59	(1.49, 1.71)

Rules are listed by order of KNHANES IV results.

NA: not applicable.

*Size: the number of risk behaviors included in the rule.

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight,

IS: inadequate sleep, BS: breakfast skipping.

또한 KNHANES IV 자료에서처럼 CHS 2010 자료에서도 ISCSPI, CSBSPI, ISBSCS 등의 군집의 support가 건강위험행위가 1개 이상인 남자 대상자에서 가장 높은 순서로 나타나 KNHANES IV 자료의 15.70%, 7.16%, 4.66%와 비슷하게 나타났다.

이처럼 남, 여 대상자 모두에서 KNHANES IV 자료에서의 ARM 결과는 CHS 2010 자료에서도 유의하게 나타났고 군집의 분포가 차지하는 비율도 비슷하게 나타났다.

V. 논의

이 연구의 목적은 국민건강영양조사 자료를 이용하여 한국 성인의 건강 위험행위 군집의 유형을 분석하는 것이다. 본 연구에서는 건강위험행위를 2가지 이상 행하는 것을 건강위험행위 군집이라 정의하였고, 건강위험행위의 군집현상을 확인하였다. ARM을 이용하여 부적절한 수면, 아침 결식, 부적절한 체중, 현재 흡연 등의 행위가 포함된 건강위험행위 군집의 유형을 발견하였다.

본 연구에서는 남자와 여자 대상자들의 분포가 달랐고, 건강위험행위에 있어 성별의 차이가 있었다. 건강위험행위는 여자보다 남자에게서 더욱 빈번하게 나타났다.

Pronk 등(2004)은 건강행위가 성별에 따라 유의한 차이가 없다고 하였다(Pronk et al., 2004). 그러나 건강행위와 관련한 성별 차이가 관찰되었다는 연구들도 있다(Chiolero et al., 2006; Poortinga, 2007). 연구 대상자와 연관성 측정 방법이 각각 다르기 때문에 이전 연구들과 직접적인 비교는 어렵지만(Chou, 2008; Fine et al., 2004; Laaksonen et al., 2001; Schuit et al., 2002; Takakura et al., 2001; 강기원 et al., 2010), 본 연구에서도 성별에 따라 건강위험행위에 차이가 있었다.

1. 가장 빈도가 높은 건강위험행위

한국인의 건강위험행위 중, 가장 빈도가 높은 것은 신체적 비활동과 부적절한 수면이었다.

이는 외국의 연구에서도 비슷한 결과를 보인다. Behavioral Risk Factor Surveillance System(BRFSS) data에 따르면 51%의 미국 성인은 신체활동에 대한 지침을 따르지 않았다(CDC, 2009).

부적절한 수면 역시 높게 나타났다. BRFSS data를 연구한 Strine 등(2005)에 따르면 26%의 미국 성인이 종종 수면 부족이라고 응답하였다고 했다(Strine and Chapman, 2005). National Sleep Foundation에서 실시한 "2009 Sleep in America Poll Summary Findings" data에 따르면 70% 이상의 응답자가 하루 7시간 미만 또는 8시간 이상의 수면을 취한다고 답하였고, 24%의 성인만이 적절한 시간의 수면을 취한다고 하였다(Strine and Chapman, 2005).

2. 건강위험행위 군집

본 연구에서 건강위험행위를 0개 한다고 응답한 대상자는 5.0%로 낮게 나타났고, 7개 모두 행한다고 답한 대상자는 없었다. 69.8%의 대상자가 2개 이상의 건강위험행위를 하고 있었다.

대부분의 성인은 2개에서 3개씩 건강위험행위를 한다고 한다. 이전의 연구에서도 3~5개까지의 건강위험행위들간의 연관성을 분석하였고, 대부분 1~2개의 건강위험행위(Chiolero et al., 2006; Chou, 2008), 또는 2~3개 건강위험행위(Poortinga, 2007)를 한다고 하였다.

그러나 만성질병 예방과 사망률을 낮추기 위해 여러 연구에서 교정할 것을 권고해 온 흡연, 과도한 음주, 신체적 비활동, 부적절한 체중, 부적절한 수면시간, 아침 결식, 잦은 간식 섭취 등의 7가지 건강위험행위 간의 동시적 연관성에 대한 연구는 찾아보기 어려웠다. 본 연구에서는 7가지 건강위험행위를 모두 포함하여 건강위험행위가 함께 일어날 가능성을 살펴보고, 3개에서 5개의 건강위험행위가 포함된 의미있는 연관규칙들을 발견하였다.

3. 흡연

여자 대상자에서 흡연을 하는 대상자는 부적절한 수면시간을 가질 가능성이 높았다. 이는 흡연이 수면의 질을 떨어뜨리고(Jaehne et al., 2012), 짧거나 긴 수면시간과 관련이 있다(Ryu et al., 2011)고 한 연구들에서 그 근거를 찾을 수 있다.

남자 대상자에서 과도한 음주와 아침 결식, 신체적 비활동과 과도한 음주 및 아침 결식 등이 함께 나타나는 경우 흡연할 확률이 높았다. 이전의 연구에서 흡연은 과도한 음주와 양의 상관관계를 가진다(Burke et al., 1997; Chiolero et al., 2006; Chou, 2008; Laaksonen et al., 2001; Poortinga, 2007; Schuit et al., 2002)고 하였다. 연구 결과, 흡연할 확률은 과도한 음주뿐 아니라 아침 결식, 신체적 비활동 등이 함께 나타나는 경우와도 연관이 있었다.

4. 신체적 비활동

남녀 대상자에서 신체적 비활동은 다른 건강위험행위와 함께 나타나면서, 부적절한 수면(lift>1.01), 흡연(lift>1.38) 등의 가능성을 높이는 것으로 나타났다.

연관성의 강도는 각각 다르지만 신체적 비활동이 흡연, 부적절한 체중, 아침 결식, 부적절한 수면, 과도한 음주 등의 위험행위 등과 함께 나타났다.

신체적 비활동은 흡연과 양의 상관관계(Burke et al., 1997; Chiolero et al., 2006; Laaksonen et al., 2001; Schuit et al., 2002)를 보인다고 하는 연구가 있었고, 음의 상관관계(Chou, 2008; Poortinga, 2007)를 보인다고 하는 연구도 있었다.

신체적 비활동은 과도한 음주와 양의 상관관계(Burke et al., 1997; Laaksonen et al., 2001)를 보인다고 하는 연구와 음의 상관관계

(Poortinga, 2007)를 보인다고 하는 기존 연구도 있었다.

남자 대상자들의 연관규칙에서 간식 섭취와 흡연을 함께 하는 대상자는 신체적으로 비활동일 가능성이 높았다(lift 1.02). 부적절한 체중과 간식 섭취를 하는 대상자는 신체적 비활동의 가능성이 높았다(lift 1.05).

이처럼 신체적 비활동은 흡연, 과도한 음주와의 연관성뿐 아니라 간식 섭취, 부적절한 체중과도 연관이 있다. 이러한 연관관계는 다음의 연구에서 그 근거를 찾을 수 있다. 연구자들은 현재 흡연자나 잦은 간식 섭취 대상자들의 연구에서 신체활동이 흡연 욕구를 줄여주거나(Taylor et al., 2005; Thayer et al., 1993) 또는 간식 섭취에 대한 욕구를 줄여주어(Thayer et al., 1993) 흡연이나 간식 섭취를 줄일 수 있다고 하였다.

5. 부적절한 수면

여자 대상자에서 부적절한 체중과 신체적 비활동이 함께 하는 경우 부적절한 수면의 가능성을 높였다.

이 결과는 부적절한 수면이 신체적 비활동, 비만(Strine and Chapman, 2005)과 관련이 있다는 기존 연구들과 비슷한 결과를 보인다.

이전 연구에서는 부적절한 수면이 비만이 될 수 있다(Brondel et al., 2010; Buxton et al., 2009; Nedeltcheva et al., 2009)고 밝힘으로써 본 연구에서 나타난 연관규칙을 지지하고 있다.

이 연구들은 간식 섭취의 영향을 분석하였는데, 수면을 적절히 취하지 못한 대상자들은 적절한 시간의 수면을 취한 대상자들보다 식전 배고픔이 더 심하여, 더 많은 에너지를 섭취하게 되고(Brondel et al., 2010), 당이 들어간 간식을 더 많이 섭취하게 되며(Buxton et al., 2009), 간식으로 칼로리를 더 많이 섭취하게 됨으로써(Nedeltcheva et al., 2009), 체질량지수가 높아질 수 있다(Gangwisch et al., 2005)고 하였다.

6. 아침 결식

남자 대상자에서 아침 결식을 하는 대상자가 흡연을 하고 있을 가능성이 높았다. 아침 결식과 신체적 비활동, 과도한 음주 등의 다른 행위들과 함께 나타나면서 흡연할 가능성이 높았다($\text{lift} > 1.38$).

이전 연구에서도 비슷한 결과를 찾을 수 있는데, 아침을 안 먹는 대상자들은 아침을 먹는 대상자들보다 남녀 모두에게서 흡연할 확률이 더욱 높았다(Sakata et al., 2001). 아침 결식은 흡연, 과도한 음주, 비활동적인 생활 습관과 군집현상을 보인다(A. Keski-Rahkonen et al., 2003)고 하였다.

7. 건강위험행위 군집의 특성 및 군집을 보이는 예측 요인

여자 대상자에서 신체적 비활동, 흡연, 부적절한 수면 시간의 군집과 신체적 비활동, 부적절한 체중, 부적절한 수면의 군집을 비교하였다. 세 가지 행위 중 흡연과 부적절한 체중의 위험행위만 다른 군집이다. 흡연이 포함된 군집에 비해 부적절한 체중이 있는 군집에서 만성질환을 진단받은 비율이 높았고, 진단받은 만성질환의 개수도 많았다($p < .0001$).

남자 대상자에서 신체적 비활동, 흡연, 부적절한 수면시간의 군집과 신체적 비활동, 흡연, 아침 결식의 군집을 비교하였다. 세 가지 행위에서 흡연과 아침 결식의 위험행위가 다른 군집이다. 아침 결식이 포함된 군집에 비해 흡연의 군집에서 만성질환을 진단받은 비율이 높았고, 진단받은 만성질환의 개수도 많았다($p < .0001$).

남자 대상자에서 신체적 비활동, 부적절한 체중, 아침 결식, 흡연 등의 군집과 신체적 비활동, 부적절한 수면, 아침 결식, 흡연 등의 군집을 비교하였다. 네 가지 행위에서 부적절한 체중과 부적절한 수면시간의 위험행위가 다른 군집이다. 부적절한 수면시간에 비해 부적절한 체중이 포함된 군집에서 만성질환을 진단받은 비율이 높았고, 진단받은 만성질환의 개수도

더 높게 나타나 유의한 차이를 보였다($p=0.0029$).

김혜련의 연구에서도 만성질환 유병에 공통된 건강위험행위로 과체중을 지적하였다(김혜련, 2003). 이는 본 연구 및 김혜련의 연구에서처럼 조사된 만성질환의 종류가 고혈압을 비롯한 심혈관질환, 뇌혈관질환, 당뇨병 및 이상지질혈증 등을 포함한 내분비질환 등을 포함하고 있고, 이들 질환의 유병률 또한 높기 때문이라고 판단된다.

건강위험행위 군집의 예측 요인을 살펴보면, 젊은 연령, 별거/이혼/사별한 대상자, 미혼인 대상자, 교육 수준이 낮은 경우, 소득 수준이 낮은 경우, 스트레스 인지 등의 요인은 예시를 든 여자와 남자의 신체적 비활동, 흡연, 부적절한 수면시간의 세 가지 행위를 함께하는 군집과 남자의 신체적 비활동, 부적절한 수면시간, 아침 결식, 흡연 등의 네 가지 행위를 함께하는 군집을 예측하는 요인이었다.

이 결과는 이전의 연구와도 같은 결과를(Berrigan et al., 2003; Chioloro et al., 2006; Fine et al., 2004; Laaksonen et al., 2001; Poortinga, 2007) 보인다. 이는 젊은 성인, 배우자가 없는 대상자, 사회경제적 수준이 낮은 대상자, 건강상태가 좋지 않은 대상자들의 건강위험행위를 개선할 필요가 있음을 여러 연구에서와 마찬가지로 지속적으로 보여주는 결과이다. 따라서 이들을 위한 보건 중재에 건강위험행위 교정 중재를 포함해야 한다는 것을 보여준다.

또한 공통적으로 나타난 예측 요인 외에 군집별로 예측 요인이 조금씩 다르게 나타났다.

예를 들어, 만성질환의 진단여부는 부적절한 체중이 함께 하는 위험행위의 군집을 예측하는 요인이었으며, 젊은 연령, 스트레스를 인지하는 경우는 흡연이 포함된 위험행위 군집을 예측하는 요인이었다.

부적절한 체중을 가지고 있는 대상자는 만성질환이 함께 할 가능성이 높기 때문에, 부적절한 체중을 예방하거나 체중 조절에 대한 중재를 한다면

질병의 경과를 악화시키지 않을 수 있는 장점이 있을 것이다.

젊은 연령에서 흡연이라는 위험행위를 많이 하고 있고, 만성질환은 없으나 스트레스를 많이 인지하는 그룹에서는 흡연 예방과 더불어 스트레스 완화에 대한 관리를 꾸준히 한다면 미래에 발생할 위험이 높은 만성질환을 예방하는 효과가 있을 것이다.

사회 인구학적인 요인은 개선하기 어려운 경우가 많으나, 건강행위는 개인의 노력으로 얼마간 개선이 가능하다. 건강위험행위를 교정하려는 개인의 노력이 무엇보다 필요하며, 건강위험행위 교정을 목표로 중재하는 보건 프로그램은 이들 대상자들의 특징과 이들이 함께 하고 있는 여러 가지 위험 행위에 대한 중재를 고려한 프로그램이어야 할 것이다.

8. 단일 건강위험행위와 건강위험행위 군집에서의 예측 요인

단일 건강위험행위와 군집에서 로지스틱 회귀분석을 통해 분석한 결과, 건강위험행위 군집을 예측하는 요인은 주로 군집을 구성하고 있는 가장 빈도가 낮은 건강위험행위의 예측 요인과 유사하게 나타났다. 즉, 여자 대상자에서 PICSIS 군집이라면 군집을 구성하고 있는 세 가지 행위 중 빈도가 가장 낮은 위험행위인 CS의 예측 요인을 따르는 것으로 나타났고, 남자 대상자에서 PICSIS 군집이라면 세 가지 행위 중 가장 빈도가 낮은 IS의 예측 요인을 따른다고 할 수 있다.

건강위험행위의 군집에 대해 중재를 한다면, 대상자에게 가장 낮은 빈도로 나타나는 위험행위의 예측 요인에 근거하여 중재하는 것이 중재 대상자의 특성을 고려하게 되어 효과적으로 작용할 것이라 판단된다.

9. 본 연구의 장점

본 연구는 건강위험행위 교정을 위한 중재 계획에 중요한 시사점을 준다. 대상자들의 건강위험행위 간에는 특정 유형을 보이는 연관규칙이 발견되었기 때문이다.

이전 연구에서도 건강행위간 연관성을 규명하였다. 건강행위 간 연관성을 확인하기 위하여 다양한 분석 방법을 사용하였다. 건강위험행위 개수를 사용한 연구(Laaksonen et al., 2001; Takakura et al., 2001), O/E값 및 prevalence of odds ratios(Berrigan et al., 2003; Poortinga, 2007; 강기원 et al., 2010), 상관관계 계수(J.K. Langlie, 1979; Takakura et al., 2001), 군집분석(Mistry et al., 2009), 그리고 다양한 회귀분석 방법들(Burke et al., 1997; Johansson and Sundquist, 1999; Laaksonen et al., 2001; Takakura et al., 2001)을 이용하였다.

위 연구들에서는 분석 대상이 건강위험행위의 몇몇 행위에 한정되어 있었고, 변수가 늘어나는 경우는 행위 개수의 합을 이용하여 각 군집에 포함된 구체적 행위를 알 수 없었다.

군집을 구체적 행위의 조합으로 분석하고자 한 연구들에서는 O/E값을 이용하였다. 그러나, O/E 방법으로는 건강위험행위 군집을 구할 때에 상호 배타적인 군집으로 만들기 때문에 행위 변수가 늘어나는 경우 특정 군집에서 관찰되는 대상자 수가 줄어들게 되므로 작은 값으로도 O/E값이 크게 변동되어 O/E값만으로 군집을 판단하기에 적절하지 않다.

본 연구에서는 ARM을 이용하여 연관성을 분석하였고, 상호배타적인 군집이 아니라 건강위험행위들 간의 가능한 행위의 조합을 연관규칙으로 설명하였다. 따라서 건강위험행위의 종류가 작은 군집에서부터 많은 행위를 포함하는 가능한 군집을 모두 고려할 수 있었다. O/E값 단독으로 군집을

구할 때 관찰빈도가 낮아 연관성의 강도를 평가하기 어려운 점을 ARM에서는 confidence, support, lift 등의 값으로 보완하면서 분석하였다. 이러한 방법으로 중재하는 대상자들의 군집을 대상자의 규모나 건강위험행위 연관성의 강도에 따라 어느 범위까지 정할 지의 수준을 조절할 수 있었다.

또한 ARM을 이용하여 건강위험행위 간 방향성을 확인할 수 있었다. 같은 건강위험행위의 군집이라도 방향성이 다르게 나타남을 알 수 있다. 여자 대상자에서 PICSIS 군집은 신체적 비활동과 흡연을 하는 것이 부적절한 수면시간의 가능성을 높인다는 연관규칙으로 나타났다. 그러나 남자 대상자에서 같은 행위의 군집인 PICSIS 군집은 부적절한 수면시간과 흡연이 함께 나타나면 신체적 비활동을 할 가능성이 75.04%로 높게 나타난다는 방향성을 확인할 수 있다.

이전 연구에서 흡연과 음주는 상관성이 높다고 밝혔다(Burke et al., 1997; Chiolero et al., 2006; Chou, 2008; Laaksonen et al., 2001; Poortinga, 2007; Schuit et al., 2002). ARM을 하면 두 행위간의 단순한 상관성 뿐 아니라 방향성을 확인할 수 있다.

Minimum support 2%, confidence 50%라는 threshold에 의해서 본 연구에서는 제시하지 않았지만, 과도한 음주를 하면 흡연의 가능성이 높다(confidence 57.7%, lift 1.31%). 또한 흡연하는 대상자들은 과도한 음주를 할 가능성이 높다(confidence 23.6%, lift 1.31).

이처럼 연관규칙을 통해서 연관규칙의 방향성을 파악하고 중재의 방법을 모색할 수 있다. 과도한 음주를 하는 대상자는 대부분 흡연을 하고 있을 것이라는 판단을 할 수 있다. 그러나 흡연하는 대상자들이 과도한 음주를 할 대상자라고 판단하기에는 확률적으로 가능성이 낮아지는 것이다. 흡연과 음주 예방에 대한 프로그램을 구성할 때 이러한 정보를 활용하면 프로그램에 도움이 되는 중재 방법을 연구할 수 있을 것이다.

기존 연구들은 흡연과 과도한 음주, 흡연과 신체적 비활동, 신체적 비활

동과 채소/과일 섭취 감소 등(Chiolero et al., 2006; Poortinga, 2007; Schuit et al., 2002)의 연관성을 분석하였다. 한 가지 건강위험행위를 기준으로 다른 건강위험행위와의 연관성을 보게 되면 기준이 되는 건강위험행위를 중요시하게 되어 처음에 기준이 되는 행위의 역할이 과장될 수 있다(Laaksonen et al., 2001).

본 연구에서는 ARM을 이용하였고, 건강위험행위 간 위계를 주지 않았기 때문에 건강위험행위 간 연관성을 편견 없이 분석할 수 있었다. ARM을 통해 개인이 가지고 있는 건강위험행위에 추가될 수 있는 또 다른 위험행위를 예측할 수 있다.

본 연구에 나타난 건강위험행위 군집의 일부 요소들은 이전 연구에서 그 근거를 찾을 수 있었다. ARM은 이전 연구에서 일상 생활에서 시행하고 있는 여러 건강위험행위들을 포괄적으로 다룰 수 있고, 이전 연구에서 밝힌 군집들을 연결하여 설명력을 높였다.

마지막으로, 지역사회 건강조사를 이용하여 앞에서 분석한 건강위험행위에 대한 연관 규칙이 통계적으로 유의하게 적용되고 ARM의 방법이 건강조사 자료의 분석에 활용이 가능한지 확인하였다. 중재를 해야 하는 행위가 있다면 문제 행위들을 포함시켜서 새로운 연관분석을 할 수 있다. 조사된 자료에 포함된 건강위험행위 변수에 따라 행위들 간의 상호관계를 살펴볼 수 있다.

10. 건강위험행위 군집 연구의 보건학적 의의

건강행위 중재 프로그램에서의 어려움은 위험 행위에 대한 전반적인 스크리닝이 부족하거나 효과적인 중재 방법을 계획하지 못해서이다(Goldstein et al., 2004). 보건 시스템에서 건강위험행위 군집에 대한 정보

가 있으면 건강증진 정책을 좀 더 효과적으로 시행할 수 있다.

건강위험행위 군집에 대한 중재는 한 가지 위험행위에 대한 중재보다 더 큰 효과를 가져다 줄 것이다(Nigg et al., 2002).

건강위험행위를 함께 교정한 경우 효과가 있다는 보고가 이어지고 있다 (Hall et al., 1992; Yusuf et al., 2004).

건강위험행위의 복합 중재를 위해 대상자의 건강위험행위 군집에 대한 정보를 알고 있어야 하는데, 기존의 건강위험행위의 개수를 이용하는 방법으로는 정보가 제한적일 수 있다.

건강 수준을 측정하기 위해 건강위험행위 개수의 합을 이용하는 방법은 매우 효율적으로 활용될 수 있는 방법이며 통계적으로 유의한 좋은 방법이다. 그러나 건강 수준을 평가하는 것이 목적이 아닌, 건강 수준을 향상시키기 위한 중재에 대한 정보가 필요하다면, 대상자들의 건강위험행위가 어떻게 연관되어 있는지 파악할 수 있는 통찰력이 필요하다.

건강위험행위가 어떻게 함께 일어나는지에 대한 기존 연구가 충분치 않은 상황에서 연관성 분석은 대상자들의 가능한 행위의 범위를 예상할 수 있도록 도와줄 수 있을 것이다. 만성질병 예방과 각 위험행위와 연관된 위험행위들을 포괄적으로 파악할 수 있는 이점을 얻을 수 있을 것이다.

본 연구에서 밝힌 ARM 결과를 이용하면 건강증진 프로그램을 계획하는데에 효과적일 것이다.

예를 들어 흡연하는 여자 대상자의 흡연과 신체적 비활동의 군집에는 수면시간을 조절하도록 중재할 수 있어야 한다.

남자 대상자의 신체적 비활동에 대한 중재 프로그램에는 흡연, 아침 결식, 간식 섭취, 체중에 대한 중재를 함께 포함시킬 필요가 있다.

흡연하는 남자 대상자가 중재의 타겟이라고 한다면 단순히 흡연자를 가려내기만 할 것이 아니라, 중재 프로그램에 흡연 예방뿐 아니라 아침식사에 대한 교정 프로그램을 포함하고, 과도한 음주나 체중 조절, 신체활동 등

에 대한 프로그램의 추가 여부도 고려해야 할 것이다.

2001년에 Multiple Health Behavior Change[MHBC] Special Interest Group이 컨퍼런스를 가졌고, 그 후 건강위험행위 군집에 대한 연구들이 발표되었다. 그러나 아직 연구 초기 단계이며, MHBC intervention에 대한 이점도 충분히 평가되지 못하고 있는 단계이다(Evers and Quintiliani, 2013).

연구가 더 누적되어야 하는 단계인데, 본 연구는 그러한 의미에서 건강위험행위의 군집에 대한 정보를 주어 중재 방향에 대한 정보를 제시하고 있다.

앞으로는 이러한 건강위험행위 군집에 대한 정보를 활용하여 건강위험행위에 대한 복합 중재의 효과를 평가하는 연구가 필요하다.

VI. 연구의 제한점

본 연구는 몇 가지 제한점을 가지고 있다.

첫째, 단면 연구 자료를 분석한 것이어서 인과관계를 알 수 없다. 건강위험행위의 연관성을 분석하면서 몇 가지의 행위가 또 다른 건강위험행위의 가능성을 높일 수 있다는 연관규칙들을 발견하였으나, 이 연관규칙들은 같은 시점에서 또 다른 행위를 함께 하고 있을 확률을 나타내며, 추후 일어날 건강위험행위라고 판단할 수는 없다.

둘째, 보건지침, 문헌 연구를 통하여 건강에 해로운 영향을 줄 수 있는 기준을 찾아 흡연, 과도한 음주, 신체적 비활동, 부적절한 체중, 부적절한 수면시간, 아침 결식, 잦은 간식 섭취 등의 건강위험행위를 정의하였다. 따라서 건강위험행위를 정의하는 기준에 따라 결과가 조금 달라질 수 있다.

셋째, 연관 규칙을 이용한 군집들이 상호 배타적인 군집이 아니어서, 포함된 위험행위 개수가 작은 군집의 경우, 다른 행위가 추가되면 위험행위 개수가 많은 다른 군집에도 포함될 수 있다. 행위의 가능한 조합을 만들면서 빈도에 근거하여 만들어지는 규칙이다. 만약, 상호배타적인 군집을 찾고자 한다면 각 군집에서 공통되는 대상자를 제외하고 순수한 군집으로 만들어야 하는 단점이 있다.

넷째, 건강위험행위의 종류가 많은 군집은 대상자가 많지 않았다. 개별 군집별로 건강 수준의 특징을 만성질환의 개수나 만성질환의 종류를 비교할 수 있어야 했는데, 대상자 수가 작아서 분석이 쉽지 않았다. 건강행위와 만성질환 유병의 자료를 포함하고 있는 대규모 데이터로 군집의 특성을 연구하는 것이 필요하다.

VII. 결론

흡연, 과도한 음주, 신체적 비활동, 부적절한 체중, 부적절한 수면시간, 아침 결식, 잦은 간식 섭취 등을 건강위험행위로 정의하고 분석한 결과, 한국인이 가장 많이 하고 있는 건강위험행위는 신체활동 부족, 부적절한 수면시간이었다. 대상자의 69.8%는 2가지 이상의 건강위험행위를 하고 있다. 남자 대상자에서 건강위험행위를 하는 비율이 더 높았다.

건강위험행위는 독립적이지 않고 상호 연관성을 갖고 함께 나타난다. 연관성분석을 이용하여 건강위험행위의 군집 현상을 관찰하였으며, 건강위험행위 2개, 3개, 4개, 5개, 6개로 이루어진 군집이 나타났다. 남자 대상자에서 건강위험행위 군집이 더 빈번하게 나타났다.

연관성 분석으로 나타난 건강위험행위 군집은 부적절한 수면, 신체적 비활동, 흡연 등과 관련된 것이었다. 군집마다 반복되어 나타나는 특정 건강위험행위들과 2~3개의 다른 위험행위가 함께 연관되어 군집을 이루고 있었다. 연관성 분석으로 나타난 건강위험행위의 군집 현상은 통계적으로도 유의하였다.

대표적인 군집으로는 여자 대상자에서 신체적 비활동, 현재 흡연, 부적절한 수면시간을 함께 하는 군집(Lift 1.06, OR 1.29, 95% CI 1.05-1.58), 남자 대상자에서 신체적 비활동, 현재 흡연, 부적절한 수면시간을 함께 하는 군집(Lift 1.01, OR 1.26, 95% CI 1.09-1.45)과 신체적 비활동, 부적절한 체중, 아침 결식, 흡연 등을 함께하는 군집(Lift 1.52, OR 2.87, 95% CI 2.23-3.70) 등이 나타났다.

연구의 결과 부적절한 수면과 부적절한 체중, 과도한 음주와 흡연, 아침 결식과 흡연 등은 강한 상관관계가 있고 이는 기존 역학 연구들의 결과를 통해서도 뒷받침된다.

본 연구의 결과는 건강위험행위의 연관성 분석을 통한 구체적 행위의 군집을 제시하였다. 각 군집은 포함된 건강위험행위의 개수가 같더라도 행위 변수의 포함에 따라 특성이 다른 집단일 수 있다.

또한 특정 건강위험행위를 하는 대상자가 어떠한 건강위험행위를 할 가능성이 높은지에 대한 방향성을 보여줌으로써 건강 증진 사업에서 어떠한 건강위험행위에 대한 중재 프로그램을 실시하는 것이 필요한 지에 대한 정보를 제시한다.

뿐만 아니라 본 연구는 건강조사자료에서 건강위험행위를 연구하는 경우에 연관성 분석의 활용 가능성을 증명하였다. 본 연구에서 사용한 연관성 분석의 방법은 비슷한 건강위험행위 문제에 확장, 적용할 수 있을 것이다.

건강위험행위 군집에 대한 정보를 이용하여 건강행위 대상자들에 대한 이해를 높이고, 그들에게 필요한 위험행위 예방에 대한 중재를 계획할 수 있다.

참고문헌

- 버드캐런저, 김성옥 역. 운동심리학. 대한미디어. 2004
- 배화수, 조대현, 석경하, 김병수, 최국렬, 이종언, 노세원, 이승철, 손용희.
SAS Enterprise Miner를 이용한 데이터마이닝. 교우사. 2008
- 이승욱. 통계학의 이해. 자유아카데미. 1998
- 이경자, 윤종률, 고일선. 노인의 질병예방 및 건강증진행위 실천을 위한
예측모형 구축. 연세대학교, 보건복지부. 2001
- 정영호, 서미경, 이종태, 정형선, 고숙자, 채수미, 김명희. 우리나라 국민의
건강결정요인 분석(Determinants of Health in Korea).
한국보건사회연구원. 2006
- 질병관리본부, 국민건강영양조사 제4기. <http://knhanes.cdc.go.kr>.
- 질병관리본부, 2010년 지역사회건강조사. <http://chs.cdc.go.kr>.
- 통계청, 2011년 사망원인통계. <http://kostat.go.kr>.
- 통계청, e-나라지표, 평균수명 및 건강수명 추이. <http://www.index.go.kr>.
- 김원석. 전략적 중요도를 고려한 연관규칙 연구. 성균관대학교 석사논문.
2007
- 김지미. 한국인의 건강행위 관련요인에 대한 다수준 분석. 한양대학교
박사논문. 2003
- 김혜련. 사회계층 및 건강행위와 만성질환 유병과의 연관성 분석.
한양대학교 박사논문. 2003
- 문상식. 주관적 건강인식과 건강검진 결과의 비교분석을 통한 건강행위
연구. 서울대학교 박사논문. 2001
- 박경용. 주관적 건강 수준 및 만성질환 이환여부와 건강관련행위와의
관련성. 충남대학교 박사논문. 2010
- 박정근. 건강관련 행태별 경제적 질병부담 변화추이: 2004-2008.

- 동국대학교 박사논문. 2010
- 백경원. 대사증후군의 주요 변인인 사회경제적 요인에 대한 사회심리적
요인과 건강행태 요인의 완충효과. 아주대학교 박사논문. 2004
- 서연숙. 중고령자의 사회경제적 지위에 따른 건강행위와 건강 수준 연구.
가톨릭대학교 박사논문. 2011
- 양현진. 건강 수준과 관련된 보건의식 행태 연구. 서울대학교 석사논문.
2003
- 오현주. 노인의 건강증진행위에 영향을 미치는 요인에 관한 연구.
인제대학교 대학원 박사논문. 2011
- 윤강재. 만성질환 유병상태 및 건강검진 결과와 주관적 건강인식의 관련성
연구. 서울대학교 석사논문. 2006
- 윤태호. 우리나라의 사회계층간 건강행태 차이. 서울대학교 석사논문.
2000
- 이정민. 한국인의 연령층별 사회경제적 계층에 따른 Alameda 7 건강행위
실천 수준과 이들의 주관적 건강 수준에 미치는 영향. 인제대학교
대학원 박사논문. 2011
- 정현주. 건강 수준에 따른 행위 분석. 서울대학교 석사논문. 2005
- 진정권. 고령자의 운동과 식생활습관이 활동체력, 만성질환 유병률,
건강관련 삶의 질에 미치는 영향. 국민대학교 박사논문. 2008
- 현영준. 이행적 연관규칙 탐사기법 연구. 성균관대학교 석사논문. 2006
- 강기원, 성주현, 김창엽. 흡연, 음주와 운동습관의 군집현상을 통한
건강행태의 고위험군: 국민건강영양조사. 예방의학회지.
2010;43(1):73-83.
- 강은정. 흡연, 음주, 신체활동을 사용한 한국 성인의 건강행태 군집의 분류.
보건사회연구. 2007;27(2):44-66.
- 김남진. 노인들의 건강행위 실천정도와 관련요인 분석 한국노년학.
2000;20(2):169-182.
- 김대희. 심혈관질환과 스트레스의 연관성 연구 : 건강행태 환경의 변화와

- 보건교육을 위하여. 보건과 사회과학. 2002;12:157-172.
- 김진희. 근로자의 사회경제적 지위와 건강행태 주관적 건강수준의 관련성 :
남녀차이를 중심으로. 보건교육·건강증진학회지. 2011;28(1):57-67.
- 김창희, 임수민, 박혜순. 한국 남성에서 낮은 HDL-콜레스테롤혈증 유병률
및 그 관련요인 추세: 1998년에서 2005년까지 3개의
국민건강영양조사를 자료를 이용하여. 가정의학회지.
2010;31(10):755-764.
- 김혜련. 건강형평성에서 건강행태의 함의와 정책과제. 보건복지포럼.
2009;3:36-47.
- 박현주, 정혜선. 대규모 사업장 남녀 근로자의 직무스트레스 수준에 따른
건강행태. 대한간호학회지. 2010;40(6):852-862.
- 이성용. 혼인과 은퇴 고령 남성의 사망 간의 관계: 미국 은퇴자 종단
자료의 분석 결과. 한국인구학. 2005;28(1):23-45.
- 전수현, 김지용, 조정진, 류정금. 직장인에서 작업특성 및 직무스트레스가
건강행태에 미치는 영향. 가정의학회지. 2010;31(6):444-452.
- 정소연, 김경우, 장윤녕, et al. 한국 중년에서 건강 행동의 인식도와
실천율의 관련성. 가정의학회지. 2007;28(11):845-852.
- 최령, 문현주. Lalonde Health Model을 이용한 성인의 건강결정요인.
보건의료산업학회지. 2011;5(6):77-89.
- 최영희, 김순이. 노인의 건강행위 평가 연구. 한국노년학. 1999;19(1):119-
130.
- 최윤정, 배성일, 이영호, 강민선. 건강결과와 건강결정요인간의 횡단면
시계열 연구: 주요 OECD 국가를 대상으로. 보건행정학회지.
2009;19(4):33-52.
- 홍대균, 김정원, 강동목. 구조적 실업이 건강관련행태와 건강에 미치는
영향. 대한산업의학회지. 2009;21(4):346-353.
- CDC, 2009. Prevalence and Trends Data : Nationwide (States and DC)
-2009 Physical Activity,

<http://apps.nccd.cdc.gov/BRFSS/display.asp?cat=PA&yr=2009&qkey=4419&state=UB>.

Accessed 7 Jan 2013.

OECD. Organization for Economic Cooperation and Development(OECD): What are equivalence scales? <http://www.oecd.org/dataoecd/61/52/35411111.pdf>.

Accessed 23 Jan 2013.

WHO. The world health report 2002, Reducing risks, Promoting healthy life. 2002; <http://www.who.int/whr/2002/en/>.

Accessed 7 Jan 2013.

WHO. Global status report on alcohol and health. 2011; http://www.who.int/topics/alcohol_drinking/en/.

Accessed 15 Mar 2013.

WHO. Global status report on noncommunicable diseases 2010. 2011; <http://www.who.int/tobacco/en/>.

Accessed 15 Mar 2013.

Lalonde M. New perspective on the health of Canadians a working document. Canada: Minister of National Health and Welfare Canada. 1974.

Lynch J, Kaplan G. Socioeconomic position. In Berkman LF and Kawachi I.(eds.) Social Epidemiology. pp 13-35. New York Oxford University Press. 2000.

SAS Education, ed Applied Analytics Using SAS Enterprise Miner. SAS Education(<http://www.sas.com/korea/training>). 2012.

Bagnardi V, Blangiardo M, Vecchia CL, Corrao G. A meta-analysis of alcohol drinking and cancer risk. Br J Cancer. 2001;85(11):1700-1705.

Belloc NB. Relationship of health practices and mortality. Prev Med.

1973;2(1):67-81.

Berlin JA, Colditz GA. A meta-analysis of physical activity in the prevention of coronary heart disease. *Am J Epidemiol.* 1990;132(4):612-628.

Berrigan D, Dodd K, Troiano RP, Krebs-Smith SM, Barbash RB. Patterns of health behavior in U.S. adults. *Prev Med.* 2003;36(5):615-623.

Björntorp P. Do stress reactions cause abdominal obesity and comorbidities? *Obes Rev.* 2001;2(2):73-86.

Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: A Meta-analysis. *JAMA.* 2008;300:2765-2778.

Brischetto CS, Connor WE, Connor SL, Matarazzo JD. Plasma lipid and lipoprotein profiles of cigarette smokers from randomly selected families: enhancement of hyperlipidemia and depression of high-density lipoprotein. *Am J Cardiol.* 1983;52(7):675-680.

Brondel L, Romer MA, Nougues PM, Touyarou P, Davenne D. Acute partial sleep deprivation increases food intake in healthy men. *Am J Clin Nutr.* 2010;91:1550-1559.

Brossette SE, Sprague AP, Hardin JM, Waites KB, Jones WT, Moser SA. Association Rules and Data Mining in Hospital Infection Control and Public Health Surveillance. *JAMIA.* 1998;5:373-381.

Burke V, Milligan RAK, Beilin LJ, et al. Clustering of Health-Related Behaviors among 18-Year-Old Australians. *Prev Med.* 1997;26:724-733.

Buxton OM, Marcelli E. Short and long sleep are positively associated with obesity, diabetes, hypertension, and cardiovascular disease among adults in the United States. *Soc Sci Med.*

2010;71(5):1027-1036.

Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. *Eur Heart J*. 2011;32(12):1484-1492.

Chang C-F, Lin M-H, Wang J, Fan J-Y, Chou L-N, Chen M-Y. The relationship between geriatric depression and health-promoting behaviors among community-dwelling seniors. *J Nurs Res*. 2013;21(2):75-82.

Chiolero 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.

Chou K-L. The Prevalence and Clustering of Four Major Lifestyle Risk Factors in Hong Kong Chinese Older Adults. *J Aging Health*. 2008;20(7):788-803.

Cole MG, Dendukuri N. Risk factors for depression among elderly community subjects: a systematic review and meta-analysis. *Am J Psychiatry*. 2003;160(6):1147-1156.

Corrao G, Bagnardi V, Zambon A, La Vecchia C. A meta-analysis of alcohol consumption and the risk of 15 diseases. *Prev Med*. 2004;38(5):613-619.

Corrao G, Rubbiati L, Bagnardi V, Zambon A, Poikolainen K. Alcohol and coronary heart disease: a meta-analysis. *Addiction*. 2000;95(10):1505-1523.

Craig WY, Palomaki GE, Haddow JE. Cigarette smoking and serum lipid and lipoprotein concentrations: an analysis of published data. *Br Med J*. 1989;298(6676):784-788.

Emmons KM, Marcus BH, Linnan L, Rossi JS, B.Abrams D.

- Mechanisms in multiple risk factor intervention_Smoking, physical activity, and dietary fat intake among manufacturing workers. *Prev Med.* 1994;23:481-489.
- Facchini F, Ida Chen Y, Reaven GM. Light-to-moderate alcohol intake is associated with enhanced insulin sensitivity. *Diabetes care.* 1994;17(2):115-119.
- 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.
- Flegal KM, Graubard BI, Williamson DF, Gail MH. Excess deaths associated with underweight, overweight, and obesity. *JAMA.* 2005;293(15):1861-1867.
- Gallicchio L, Kalesan B. Sleep duration and mortality: a systematic review and meta-analysis. *J Sleep Res.* 2009;18(2):148-158.
- Gangwisch JE, Malaspina D, Boden-Albala B, Heymsfield SB. Inadequate sleep as a risk factor for obesity: analyses of the NHANES I. *Sleep.* 2005;28(10):1289-1296.
- Goldman N. Marriage Selection and Mortality Patterns: Inferences and Fallacies. *Demography.* 1993;30(2):189-208.
- Goldstein MG, Whitlock EP, DePue J, Planning Committee of the Addressing Multiple Behavioral Risk Factors in Primary Care Project. Multiple behavioral risk factor interventions in primary care. Summary of research evidence. *Am J Prev Med.* 2004;27(2 Suppl):61-79.
- Gove WR, Hughes M, Style CB. Does Marriage have Positive Effects on the Psychological Well-Being of the Individual? *J Health Soc Behav.* 1983;24(2):122-131.
- Hall SM, Tunstall CD, Vila KL, Duffy J. Weight Gain Prevention and

- Smoking Cessation: Cautionary Findings. *Am J Public Health*. 1992;82:799–803.
- HAMPL J, HEATON C, TAYLOR C. Snacking patterns influence energy and nutrient intakes but not body mass index. *Hum Nutr Appl Nutr*. 2003;16(1):3–11.
- HASKELL WL, LEE I-M, PATE RR, et al. Physical Activity and Public Health: Updated Recommendation for Adults from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc*. 2007;39(8):1423–1434.
- HASKELL WL. Health consequences of physical activity : understanding and challenges regarding dose–response. *Med Sci Sports Exerc*. 1994;26:649–660.
- HELMRICH SP, RAGLAND DR, LEUNG RW, Paffenbarger RS. Physical activity and reduced occurrence of non–insulin–dependent diabetes mellitus. *N Engl J Med*. 1991;325(3):147–152.
- HORIKAWA C, KODAMA S, YACHI Y, et al. Skipping breakfast and prevalence of overweight and obesity in Asian and Pacific regions: a meta–analysis. *Prev Med*. 2011;53(4):260–267.
- HOUSMAN J, DORMAN S. The Alameda County Study: A Systematic, Chronological Review. *Am J Health Educ*. 2005;36(5):302–308.
- HUANG C, HU H, FAN Y, LIAO Y, TSAI P. Associations of breakfast skipping with obesity and health–related quality of life: evidence from a national survey in Taiwan. *Int J Obes*. 2010;34(4):720–725.
- IDLER EL, BENYAMINI Y. Self–rated health and mortality: A review of twenty–seven community studies. *J Health Soc Behav*. 1997;38(1):21–37.
- IMAMURA T, MATSUMOTO S, KANAGAWA Y, et al. A technique for

- identifying three diagnostic findings using association analysis. *Med Biol Eng Comput.* 2007;45(1):51-59.
- Jaehne A, Unbehau T, Feige B, Lutz UC, Batra A, Riemann D. How smoking affects sleep: a polysomnographical analysis. *Sleep.* 2012;13(10):1286-1292.
- Jennings G, Nelson L, Nestel P, et al. The effects of changes in physical activity on major cardiovascular risk factors, hemodynamics, sympathetic function, and glucose utilization in man: a controlled study of four levels of activity. *Circulation.* 1986;73(1):30-40.
- Jin H, Chen J, He H, Williams GJ, Kelman C, O'Keefe CM. Mining Unexpected Temporal Associations: Applications in Detecting Adverse Drug Reactions. *IEEE Trans Inf Technol Biomed.* 2008;12(4):488-500.
- Johansson S-E, Sundquist J. Changes in lifestyle factors and their influence on health status and all-cause mortality. *Int J Epidemiol.* 1999;28:1073-1080.
- Kannel WB, Sorlie P. Some Health Benefits of Physical Activity - The Framingham Study. *Arch Intern Med.* 1979;139:857-861.
- Kaplan GA, Camacho T. Perceived health and mortality: A nine-year follow-up of the human population laboratory cohort. *Am J Epidemiol.* 1983;117(3):292-304.
- Katanoda K, Marugame T, Saika K, et al. Population Attributable Fraction of Mortality Associated with Tobacco Smoking in Japan: A Pooled Analysis of Three Large-scale Cohort Studies. *J Epidemiol.* 2008;18(6):251-264.
- Keller S, Maddock JE, Hannover W, Thyrian JR, Basler HD. Multiple health risk behaviors in German first year university students.

- Prev Med. 2008;46(3):189-195.
- Kerver JM, Yang EJ, Obayashi S, Bianchi L, Song WO. Meal and snack patterns are associated with dietary intake of energy and nutrients in US adults. *J Am Diet Assoc.* 2006;106(1):46-53.
- Keski-Rahkonen A, Kaprio J, Rissanen A, Virkkunen M, Rose RJ. Breakfast skipping and health-compromising behaviors in adolescents and adults. *Eur J Clin Nutr.* 2003;57(7):842-853.
- Kool J, Bie Rd, Oesch P, Knu̇sel O, Brandt Pvd, Bachmann S. Exercise reduces sick leave in patients with non-acute non-specific low back pain: a meta-analysis. *J Rehabil Med.* 2004;36(2):49-62.
- Laaksonen M, Prattala R, Karisto A. Patterns of unhealthy behaviour in Finland. *Eur J Public Health.* 2001;11(3):294-300.
- Langlie JK. Interrelationships Among Preventive Health Behaviors: A Test of Competing Hypotheses. *Public Health Rep.* 1979;94(3):216-225.
- Larsson SC, Wolk A. Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. *Am J Clin Nutr.* 2007;86(3):556-565.
- Lee DG, Ryu KS, Bashir M, Bae J-W, Ryu KH. Discovering Medical Knowledge using Association Rule Mining in Young Adults with Acute Myocardial Infarction. *J Med Syst.* 2014;37:1-10.
- Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W. Classification and Regression Tree Analysis in Public Health: Methodological Review and Comparison With Logistic Regression. *Ann Behav Med.* 2003;26(3):172-181.
- Liberatos P, Link BG, Kelsey JL. The measurement of social class in epidemiology. *Epidemiol Rev.* 1988;10(1):87-121.
- Lillard LA, Panis CWA. Marital Status and Mortality: The Role of

- Health. *Demography*. 1996;33(3):313–327.
- Link BG, Phelan J. Social Conditions as Fundamental Causes of Disease. *J Health Soc Behav*. 1995;35(Extra Issue):80–94.
- Luppino FS, de Wit LM, Bouvy PF, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. *Arch Gen Psychiatry*. 2010;67(3):220–229.
- Marchand LL, Wilkens LR, Kolonel LN. Associations of Sedentary Lifestyle, Obesity, Smoking, Alcohol Use, and Diabetes with the Risk of Colorectal Cancer. *Cancer Res*. 1997;57:4787–4794.
- Marmot MG, Adelstein AM, Robinson N, Rose GA. Changing social-class distribution of heart disease. *Br Med J*. 1978;2:1109–1112.
- Mason JO, Koplan JP, Layde PM. The prevention and control of chronic diseases: reducing unnecessary deaths and disability—a conference report. *Public health rep*. 1987;102(1):17–20.
- Mayer-Davis EJ, D'Agostino Jr R, Karter AJ, et al. Intensity and amount of physical activity in relation to insulin sensitivity. *JAMA*. 1998;279(9):669–674.
- McEwen BS. Central effects of stress hormones in health and disease: Understanding the protective and damaging effects of stress and stress mediators. *Eur J of Pharm*. 2008;583(2):174–185.
- McGee DL. Body mass index and mortality: a meta-analysis based on person-level data from twenty-six observational studies. *Ann Epidemiol*. 2005;15(2):87–97.
- Mekary RA, Giovannucci E, Willett WC, van Dam RM, Hu FB. Eating patterns and type 2 diabetes risk in men: breakfast omission, eating frequency, and snacking. *Am J Clin Nutr*. 2012;95(5):1182–1189.
- Michael P. *Health Promotion: An Emerging Strategy for Health*

- Enhancement and Business Cost Savings in Korea. Paper presented at: 3rd International Symposium-Development of Health Education Program for Community Health Promotion, pp. 107~ 116. 1999.
- Mistry R, McCarthy WJ, Yancey AK, Lu Y, Patel M. Resilience and patterns of health risk behaviors in California adolescents. *Prev Med.* 2009;48(3):291-297.
- Murray CJ, Lopez AD. Evidence-based health policy-lessons from the Global Burden of Disease Study. *Science* (New York, NY). 1996;274(5288):740-743.
- Myers J, Kaykha A, George S, et al. Fitness versus physical activity patterns in predicting mortality in men. *Am J Med.* 2004;117(12):912-918.
- Nathanson CA. Sex roles as variables in preventive health behavior. *J Community Health.* 1977;3(2):142-155.
- Nedeltcheva AV, Kilkus JM, Imperial J, Kasza K, Schoeller DA, Penev PD. Sleep curtailment is accompanied by increased intake of calories from snacks. *Am J Clin Nutr.* 2009;89(1):126-133.
- Nigg CR, Allegrante JP, Ory M. Theory-comparison and multiple-behavior research: common themes advancing health behavior research. *Health Educ Res.* 2002;17(5):670-679.
- Paluska SA, Schwenk TL. Physical Activity and Mental Health. *Sports Med.* 2000;29(3):167-180.
- Pan X-R, Li G-w, Hu Y-H, et al. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance: the Da Qing IGT and Diabetes Study. *Diabetes care.* 1997;20(4):537-544.
- Pedersen BK, Saltin B. Evidence for prescribing exercise as therapy in

- chronic disease. *Scand J Med Sci Sports*. 2006;16(S1):3-63.
- Poortinga W. The prevalence and clustering of four major lifestyle risk factors in an English adult population. *Prev Med*. 2007;44(2):124-128.
- Prochaska JJ, Spring B, Nigg CR. Multiple health behavior change research: an introduction and overview. *Prev Med*. 2008;46(3):181-188.
- Pronk NP, Anderson LH, Crain AL, et al. Meeting recommendations for multiple healthy lifestyle factors. Prevalence, clustering, and predictors among adolescent, adult, and senior health plan members. *Am J Prev Med*. 2004;27(2 Suppl):25-33.
- Purslow LR, Sandhu MS, Forouhi N, et al. Energy intake at breakfast and weight change: Prospective study of 6,764 middle-aged men and women. *Am J Epidemiol*. 2008;167(2):188-192.
- Rose G, Marmot MG. Social class and coronary heart disease. *Br Heart J*. 1981;45:13-19.
- Ryu SY, Kim KS, Han MA. Factors associated with sleep duration in Korean adults: results of a 2008 community health survey in Gwangju metropolitan city, Korea. *J Korean Med Sci*. 2011;26(9):1124-1131.
- Sacco RL, Elkind M, Boden-Albala B, et al. The protective effect of moderate alcohol consumption on ischemic stroke. *JAMA*. 1999;281(1):53-60.
- Sakata K, Matumura Y, Yoshimura N, et al. Relationship between skipping breakfast and cardiovascular disease risk factors in the national nutrition survey data. *Jpn J Public Health*. 2001;48:837-841.
- Schlecht NF, Franco EL, Pintos J, et al. Interaction between Tobacco

- and Alcohol Consumption and the Risk of Cancers of the Upper Aero-Digestive Tract in Brazil. *Am J Epidemiol.* 1999;150(11):1129-1137.
- Schoenborn CA. Health Habits of U.S. Adults, 1985: the "Alameda 7" Revisited. *Public Health Rep.* 1986;101(6):571-580.
- Schuit AJ, van Loon AJ, Tijhuis M, Ocke M. Clustering of lifestyle risk factors in a general adult population. *Prev Med.* 2002;35(3):219-224.
- Sesso HD, Paffenbarger RS, Lee I-M. Physical Activity and Coronary Heart Disease in Men : The Harvard Alumni Health Study. *Circulation.* 2000;102:975-980.
- Steele JL, McBroom WH. Conceptual and Empirical Dimensions of Health Behavior. *J Health Soc Behav.* 1972;13:382-390.
- Strawbridge WJ, Deleger S, Roberts RE, Kaplan GA. Physical Activity Reduces the Risk of Subsequent Depression for Older Adults. *Am J Epidemiol.* 2002;156(4):328-334.
- Strine TW, Chapman DP. Associations of frequent sleep insufficiency with health-related quality of life and health behaviors. *Sleep Med.* 2005;6(1):23-27.
- Subramanian SV, Kawachi I, Kennedy BP. Does the state you live in make a difference? Multilevel analysis of self-rated health in the US. *Soc Sci Med.* 2001;53:9-19.
- Tai YM, Chiu HW. Comorbidity study of ADHD: applying association rule mining (ARM) to National Health Insurance Database of Taiwan. *Int J Med Inform.* 2009;78(12):e75-83.
- Takakura M, Nagayama T, Sakihara S, Willcox C. Patterns of health risk behavior among Japanese high school students. *J Sch Health.* 2001;71(1):23-29.

- Tan P-N, Kumar V, Srivastava J. Selecting the right objective measure for association analysis. *Inf Syst.* 2004;29(4):293-313.
- Taylor AH, Katomeri M, Ussher M. Acute effects of self-paced walking on urges to smoke during temporary smoking abstinence. *Psychopharmacology.* 2005;181(1):1-7.
- Taylor RS, Brown A, Ebrahim S, et al. Exercise-based rehabilitation for patients with coronary heart disease: systematic review and meta-analysis of randomized controlled trials. *Am J Med.* 2004;116(10):682-692.
- Teychenne M, Ball K, Salmon J. Sedentary behavior and depression among adults: a review. *Int J Behav Med.* 2010;17(4):246-254.
- Thayer RE, Peters III DP, Takahashi PJ, Birkhead-Flight AM. Mood and Behavior(smoking and sugar snacking) following moderate Exercise : A partial test of self-regulation theory. *Pers Individ Dif.* 1993;14(1):97-104.
- Turner RJ, Wheaton B, Lloyd DA. The epidemiology of social stress. *Am Sociol Rev.* 1995:104-125.
- V Bagnardi, Blangiardo M, Vecchia CL, Corrao G. A meta-analysis of alcohol drinking and cancer risk. *Br J Cancer.* 2001;85(11):1700-1705.
- Vinnakota S, Lam NS. Socioeconomic inequality of cancer mortality in the United States: a spatial data mining approach. *Int J Health Geogr.* 2006;5(9):1-12.
- Visscher T, Seidell J, Menotti A, et al. Underweight and Overweight in Relation to Mortality Among Men Aged 40-59 and 50-69 Years The Seven Countries Study. *Am J Epidemiol.* 2000;151(7):660-666.
- Weisell RC. Body mass index as an indicator of obesity. *Asia Pacific J*

Clin Nutr. 2002;11(Suppl):S681–S684.

Winkleby MA, Jatulis DE, Frank E, Fortmann SP. Socioeconomic status and health: how education income and occupation contribute to risk factors for cardiovascular disease. *Am J Public Health.* 1992;82:816–820.

Wolff I, van Croonenborg JJ, Kemper HCG, Kostense PJ, Twisk JWR. The Effect of Exercise Training Programs on Bone Mass: A Metaanalysis of Published Controlled Trials in Pre- and Postmenopausal Women. *Osteoporos Int.* 1999;9:1–12.

Yeo Y, Ma SH, Park SK, 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. *J Prev Med Public Health.* 2013;46(5):271–281.

Yusuf S, Hawken S, Ounpuu S, 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.

부록

1. O/E ratio로 본 건강위험행위 군집현상

ARM 결과와 비교할 수 있도록 남, 여 대상자에서 건강위험행위를 하지 않는 대상자는 제외한 후 분석하였다.

1.1 여자 대상자에서 건강위험행위

Table 32. Distribution of lifestyle risk behaviors among women

CS	HD	PI	IW	IS	BS	FS	Observed	Expected	O/E
0	0	0	0	0	0	0	0	334.87	0.00
0	0	0	0	0	0	1	45	29.52	1.52
0	0	0	0	0	1	0	82	66.80	1.23
0	0	0	0	0	1	1	10	5.89	1.70
0	0	0	0	1	0	0	459	363.92	1.26
0	0	0	0	1	0	1	42	32.08	1.31
0	0	0	0	1	1	0	85	72.60	1.17
0	0	0	0	1	1	1	9	6.40	1.41
0	0	0	1	0	0	0	285	199.63	1.43
0	0	0	1	0	0	1	22	17.60	1.25
0	0	0	1	0	1	0	40	39.82	1.00
0	0	0	1	0	1	1	2	3.51	0.57
0	0	0	1	1	0	0	276	216.95	1.27
0	0	0	1	1	0	1	22	19.13	1.15
0	0	0	1	1	1	0	46	43.28	1.06
0	0	0	1	1	1	1	7	3.82	1.83
0	0	1	0	0	0	0	1,731	1,450.39	1.19
0	0	1	0	0	0	1	162	127.87	1.27
0	0	1	0	0	1	0	290	289.33	1.00
0	0	1	0	0	1	1	26	25.51	1.02
0	0	1	0	1	0	0	1,532	1,576.23	0.97
0	0	1	0	1	0	1	119	138.96	0.86
0	0	1	0	1	1	0	270	314.44	0.86
0	0	1	0	1	1	1	31	27.72	1.12
0	0	1	1	0	0	0	849	864.62	0.98
0	0	1	1	0	0	1	65	76.22	0.85
0	0	1	1	0	1	0	146	172.48	0.85
0	0	1	1	0	1	1	9	15.21	0.59
0	0	1	1	1	0	0	943	939.64	1.00
0	0	1	1	1	0	1	62	82.84	0.75
0	0	1	1	1	1	0	139	187.45	0.74
0	0	1	1	1	1	1	11	16.53	0.67

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight, IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking

Table 33. Distribution of lifestyle risk behaviors among women

CS	HD	PI	IW	IS	BS	FS	Observed	Expected	O/E
0	1	0	0	0	0	0	7	6.12	1.14
0	1	0	0	0	0	1	0	0.54	0.00
0	1	0	0	0	1	0	1	1.22	0.82
0	1	0	0	0	1	1	0	0.11	0.00
0	1	0	0	1	0	0	10	6.65	1.50
0	1	0	0	1	0	1	0	0.59	0.00
0	1	0	0	1	1	0	3	1.33	2.26
0	1	0	0	1	1	1	0	0.12	0.00
0	1	0	1	0	0	0	2	3.65	0.55
0	1	0	1	0	0	1	1	0.32	3.11
0	1	0	1	0	1	0	2	0.73	2.75
0	1	0	1	0	1	1	0	0.06	0.00
0	1	0	1	1	0	0	4	3.96	1.01
0	1	0	1	1	0	1	0	0.35	0.00
0	1	0	1	1	1	0	1	0.79	1.27
0	1	0	1	1	1	1	0	0.07	0.00
0	1	1	0	0	0	0	10	26.49	0.38
0	1	1	0	0	0	1	0	2.34	0.00
0	1	1	0	0	1	0	9	5.28	1.70
0	1	1	0	0	1	1	0	0.47	0.00
0	1	1	0	1	0	0	10	28.79	0.35
0	1	1	0	1	0	1	0	2.54	0.00
0	1	1	0	1	1	0	2	5.74	0.35
0	1	1	0	1	1	1	0	0.51	0.00
0	1	1	1	0	0	0	7	15.79	0.44
0	1	1	1	0	0	1	0	1.39	0.00
0	1	1	1	0	1	0	1	3.15	0.32
0	1	1	1	0	1	1	0	0.28	0.00
0	1	1	1	1	0	0	12	17.16	0.70
0	1	1	1	1	0	1	1	1.51	0.66
0	1	1	1	1	1	0	3	3.42	0.88
0	1	1	1	1	1	1	0	0.30	0.00

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight, IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking

Table 34. Distribution of lifestyle risk behaviors among women

CS	HD	PI	IW	IS	BS	FS	Observed	Expected	O/E
1	0	0	0	0	0	0	17	21.82	0.78
1	0	0	0	0	0	1	0	1.92	0.00
1	0	0	0	0	1	0	7	4.35	1.61
1	0	0	0	0	1	1	2	0.38	5.21
1	0	0	0	1	0	0	19	23.72	0.80
1	0	0	0	1	0	1	3	2.09	1.43
1	0	0	0	1	1	0	10	4.73	2.11
1	0	0	0	1	1	1	0	0.42	0.00
1	0	0	1	0	0	0	8	13.01	0.61
1	0	0	1	0	0	1	2	1.15	1.74
1	0	0	1	0	1	0	6	2.60	2.31
1	0	0	1	0	1	1	0	0.23	0.00
1	0	0	1	1	0	0	13	14.14	0.92
1	0	0	1	1	0	1	2	1.25	1.60
1	0	0	1	1	1	0	5	2.82	1.77
1	0	0	1	1	1	1	1	0.25	4.02
1	0	1	0	0	0	0	71	94.51	0.75
1	0	1	0	0	0	1	2	8.33	0.24
1	0	1	0	0	1	0	27	18.85	1.43
1	0	1	0	0	1	1	2	1.66	1.20
1	0	1	0	1	0	0	82	102.72	0.80
1	0	1	0	1	0	1	6	9.06	0.66
1	0	1	0	1	1	0	33	20.49	1.61
1	0	1	0	1	1	1	2	1.81	1.11
1	0	1	1	0	0	0	32	56.34	0.57
1	0	1	1	0	0	1	2	4.97	0.40
1	0	1	1	0	1	0	19	11.24	1.69
1	0	1	1	0	1	1	1	0.99	1.01
1	0	1	1	1	0	0	51	61.23	0.83
1	0	1	1	1	0	1	2	5.40	0.37
1	0	1	1	1	1	0	17	12.21	1.39
1	0	1	1	1	1	1	6	1.08	5.57

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight, IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking

Table 35. Distribution of lifestyle risk behaviors among women

CS	HD	PI	IW	IS	BS	FS	Observed	Expected	O/E
1	1	0	0	0	0	0	4	0.40	10.04
1	1	0	0	0	0	1	0	0.04	0.00
1	1	0	0	0	1	0	3	0.08	37.73
1	1	0	0	0	1	1	2	0.01	285.32
1	1	0	0	1	0	0	2	0.43	4.62
1	1	0	0	1	0	1	0	0.04	0.00
1	1	0	0	1	1	0	5	0.09	57.86
1	1	0	0	1	1	1	0	0.01	0.00
1	1	0	1	0	0	0	0	0.24	0.00
1	1	0	1	0	0	1	0	0.02	0.00
1	1	0	1	0	1	0	0	0.05	0.00
1	1	0	1	0	1	1	0	0.00	0.00
1	1	0	1	1	0	0	2	0.26	7.75
1	1	0	1	1	0	1	0	0.02	0.00
1	1	0	1	1	1	0	2	0.05	38.83
1	1	0	1	1	1	1	1	0.00	220.20
1	1	1	0	0	0	0	7	1.73	4.05
1	1	1	0	0	0	1	0	0.15	0.00
1	1	1	0	0	1	0	8	0.34	23.23
1	1	1	0	0	1	1	0	0.03	0.00
1	1	1	0	1	0	0	7	1.88	3.73
1	1	1	0	1	0	1	0	0.17	0.00
1	1	1	0	1	1	0	8	0.37	21.38
1	1	1	0	1	1	1	0	0.03	0.00
1	1	1	1	0	0	0	2	1.03	1.94
1	1	1	1	0	0	1	0	0.09	0.00
1	1	1	1	0	1	0	6	0.21	29.23
1	1	1	1	0	1	1	0	0.02	0.00
1	1	1	1	1	0	0	4	1.12	3.58
1	1	1	1	1	0	1	0	0.10	0.00
1	1	1	1	1	1	0	2	0.22	8.96
1	1	1	1	1	1	1	0	0.02	0.00

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight, IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking

1.2 남자 대상자에서 건강위험행위

ARM 결과와 비교할 수 있도록 건강위험행위를 하지 않는 대상자는 제외한 후 분석하였다.

Table 36. Distribution of lifestyle risk behaviors among men

CS	HD	PI	IW	IS	BS	FS	Observed	Expected	O/E
0	0	0	0	0	0	0	0	154.33	0.00
0	0	0	0	0	0	1	18	9.66	1.86
0	0	0	0	0	1	0	20	29.43	0.68
0	0	0	0	0	1	1	0	1.84	0.00
0	0	0	0	1	0	0	217	160.67	1.35
0	0	0	0	1	0	1	19	10.06	1.89
0	0	0	0	1	1	0	21	30.64	0.69
0	0	0	0	1	1	1	2	1.92	1.04
0	0	0	1	0	0	0	131	104.64	1.25
0	0	0	1	0	0	1	6	6.55	0.92
0	0	0	1	0	1	0	16	19.95	0.80
0	0	0	1	0	1	1	0	1.25	0.00
0	0	0	1	1	0	0	143	108.93	1.31
0	0	0	1	1	0	1	9	6.82	1.32
0	0	0	1	1	1	0	20	20.77	0.96
0	0	0	1	1	1	1	2	1.30	1.54
0	0	1	0	0	0	0	558	449.16	1.24
0	0	1	0	0	0	1	34	28.11	1.21
0	0	1	0	0	1	0	88	85.64	1.03
0	0	1	0	0	1	1	4	5.36	0.75
0	0	1	0	1	0	0	564	467.6	1.21
0	0	1	0	1	0	1	28	29.27	0.96
0	0	1	0	1	1	0	45	89.16	0.50
0	0	1	0	1	1	1	8	5.58	1.43
0	0	1	1	0	0	0	330	304.54	1.08
0	0	1	1	0	0	1	18	19.06	0.94
0	0	1	1	0	1	0	41	58.07	0.71
0	0	1	1	0	1	1	3	3.63	0.83
0	0	1	1	1	0	0	350	317.04	1.10
0	0	1	1	1	0	1	29	19.84	1.46
0	0	1	1	1	1	0	26	60.45	0.43
0	0	1	1	1	1	1	5	3.78	1.32

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight, IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking

Table 37. Distribution of lifestyle risk behaviors among men

CS	HD	PI	IW	IS	BS	FS	Observed	Expected	O/E
0	1	0	0	0	0	0	27	33.81	0.80
0	1	0	0	0	0	1	0	2.12	0.00
0	1	0	0	0	1	0	3	6.45	0.47
0	1	0	0	0	1	1	1	0.40	2.48
0	1	0	0	1	0	0	39	35.20	1.11
0	1	0	0	1	0	1	2	2.20	0.91
0	1	0	0	1	1	0	5	6.71	0.74
0	1	0	0	1	1	1	0	0.42	0.00
0	1	0	1	0	0	0	38	22.93	1.66
0	1	0	1	0	0	1	0	1.44	0.00
0	1	0	1	0	1	0	4	4.37	0.92
0	1	0	1	0	1	1	0	0.27	0.00
0	1	0	1	1	0	0	24	23.87	1.01
0	1	0	1	1	0	1	0	1.49	0.00
0	1	0	1	1	1	0	5	4.55	1.10
0	1	0	1	1	1	1	0	0.28	0.00
0	1	1	0	0	0	0	66	98.41	0.67
0	1	1	0	0	0	1	3	6.16	0.49
0	1	1	0	0	1	0	12	18.76	0.64
0	1	1	0	0	1	1	0	1.17	0.00
0	1	1	0	1	0	0	67	102.45	0.65
0	1	1	0	1	0	1	2	6.41	0.31
0	1	1	0	1	1	0	14	19.53	0.72
0	1	1	0	1	1	1	1	1.22	0.82
0	1	1	1	0	0	0	49	66.72	0.73
0	1	1	1	0	0	1	5	4.18	1.20
0	1	1	1	0	1	0	7	12.72	0.55
0	1	1	1	0	1	1	0	0.80	0.00
0	1	1	1	1	0	0	43	69.46	0.62
0	1	1	1	1	0	1	5	4.35	1.15
0	1	1	1	1	1	0	11	13.24	0.83
0	1	1	1	1	1	1	0	0.83	0.00

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight, IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking

Table 38. Distribution of lifestyle risk behaviors among men

CS	HD	PI	IW	IS	BS	FS	Observed	Expected	O/E
1	0	0	0	0	0	0	140	120.16	1.17
1	0	0	0	0	0	1	9	7.52	1.20
1	0	0	0	0	1	0	31	22.91	1.35
1	0	0	0	0	1	1	3	1.43	2.09
1	0	0	0	1	0	0	92	125.09	0.74
1	0	0	0	1	0	1	3	7.83	0.38
1	0	0	0	1	1	0	24	23.85	1.01
1	0	0	0	1	1	1	1	1.49	0.67
1	0	0	1	0	0	0	75	81.47	0.92
1	0	0	1	0	0	1	3	5.10	0.59
1	0	0	1	0	1	0	24	15.53	1.55
1	0	0	1	0	1	1	1	0.97	1.03
1	0	0	1	1	0	0	61	84.81	0.72
1	0	0	1	1	0	1	4	5.31	0.75
1	0	0	1	1	1	0	17	16.17	1.05
1	0	0	1	1	1	1	0	1.01	0.00
1	0	1	0	0	0	0	341	349.69	0.98
1	0	1	0	0	0	1	13	21.89	0.59
1	0	1	0	0	1	0	75	66.68	1.12
1	0	1	0	0	1	1	7	4.17	1.68
1	0	1	0	1	0	0	321	364.04	0.88
1	0	1	0	1	0	1	24	22.79	1.05
1	0	1	0	1	1	0	85	69.41	1.22
1	0	1	0	1	1	1	9	4.34	2.07
1	0	1	1	0	0	0	193	237.1	0.81
1	0	1	1	0	0	1	10	14.84	0.67
1	0	1	1	0	1	0	69	45.21	1.53
1	0	1	1	0	1	1	6	2.83	2.12
1	0	1	1	1	0	0	179	246.83	0.73
1	0	1	1	1	0	1	11	15.45	0.71
1	0	1	1	1	1	0	62	47.06	1.32
1	0	1	1	1	1	1	3	2.95	1.02

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight, IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking

Table 39. Distribution of lifestyle risk behaviors among men

CS	HD	PI	IW	IS	BS	FS	Observed	Expected	O/E
1	1	0	0	0	0	0	46	26.33	1.75
1	1	0	0	0	0	1	0	1.65	0.00
1	1	0	0	0	1	0	10	5.02	1.99
1	1	0	0	0	1	1	2	0.31	6.37
1	1	0	0	1	0	0	41	27.41	1.50
1	1	0	0	1	0	1	1	1.72	0.58
1	1	0	0	1	1	0	6	5.23	1.15
1	1	0	0	1	1	1	1	0.33	3.06
1	1	0	1	0	0	0	31	17.85	1.74
1	1	0	1	0	0	1	1	1.12	0.90
1	1	0	1	0	1	0	6	3.40	1.76
1	1	0	1	0	1	1	0	0.21	0.00
1	1	0	1	1	0	0	31	18.58	1.67
1	1	0	1	1	0	1	2	1.16	1.72
1	1	0	1	1	1	0	12	3.54	3.39
1	1	0	1	1	1	1	0	0.22	0.00
1	1	1	0	0	0	0	80	76.62	1.04
1	1	1	0	0	0	1	3	4.80	0.63
1	1	1	0	0	1	0	21	14.61	1.44
1	1	1	0	0	1	1	1	0.91	1.09
1	1	1	0	1	0	0	92	79.76	1.15
1	1	1	0	1	0	1	5	4.99	1.00
1	1	1	0	1	1	0	24	15.21	1.58
1	1	1	0	1	1	1	0	0.95	0.00
1	1	1	1	0	0	0	68	51.95	1.31
1	1	1	1	0	0	1	3	3.25	0.92
1	1	1	1	0	1	0	23	9.90	2.32
1	1	1	1	0	1	1	1	0.62	1.61
1	1	1	1	1	0	0	52	54.08	0.96
1	1	1	1	1	0	1	3	3.38	0.89
1	1	1	1	1	1	0	20	10.31	1.94
1	1	1	1	1	1	1	0	0.65	0.00

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight, IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking

2. ARM(연관성 분석)의 과정

2.1 데이터 세트부터 Link analysis

데이터는 두 개의 변수로 구성된다.

트랜잭션(아이디)과 아이템(건강위험행위)으로 나타낼 수 있다.

ARM을 하기 위해 우리가 가지고 있는 데이터를 아이디와 건강위험행위 변수들로만 구성되도록 데이터를 구성한다.

```

/*전체*/
/*association total*/
data nhn.aso0789_all(keep=id sex CS HD PI IW IS BS FS);
set nhn.aso0789run;

/*엑셀로 보낸 후 새로운 아이디 부여*/
PROC EXPORT DATA= NHN.ASO0789_all
  OUTFILE= "H:\now on\Wsas programming\국민건강영양조사\W0111nhn
  aso0789toexcel.xls"
  DBMS=EXCEL REPLACE;
  SHEET="Sheet";
RUN;

/*엑셀에서 newid생성 sex 컬럼 삭제*/
PROC IMPORT OUT= NHN.ASO0789t
  DATAFILE= "H:\now on\Wsas programming\국민건강영양조사\W0111nhn
  naso0789toexcel.xls"
  DBMS=EXCEL REPLACE;
  RANGE="Sheet";
  GETNAMES=YES;
  MIXED=NO;
  SCANTEXT=YES;
  USEDATE=YES;
  SCANTIME=YES;
RUN;
  
```

Figure 14. Data preparation(sas data step)

엑셀파일에서 건강위험행위를 한 경우 1, 안 한 경우를 0으로 코딩된 것을 볼 수 있다.

	A	B	C	D	E	F	G	H
1	newid	CS	HD	PI	IS	IW	BS	FS
2	1	0	0	1	1	0	1	0
3	2	0	0	1	1	0	1	0
4	3	0	0	1	0	0	0	0
5	4	1	0	1	1	0	0	0
6	5	1	0	1	0	0	1	0
7	6	0	0	0	0	1	0	0
8	7	0	0	1	1	1	0	0
9	8	0	0	1	0	0	0	0
10	9	0	0	1	1	1	0	0
11	10	0	0	1	1	1	0	0
12	11	0	0	1	1	0	0	0
13	12	0	0	0	0	1	0	0
14	13	0	0	1	0	1	0	1
15	14	0	0	1	1	0	0	0
16	15	0	0	0	1	0	0	0

Figure 15. Health behavior data

SAS data step으로 돌아와서 자료를 아이디와 건강위험행위 두 개 칼럼만 남도록 가로로 된 모양을 세로로 변환(transpose)시킨다. 그리고 나서 건강위험행위를 안 한 경우 col1에 0이라고 남아 있는 것을 data set에서 삭제시킨다. 그러면 아이디와 위험행위(문자형태) 변수들만 남게 된다.

```

proc transpose data=nhn.aso0789t out=nhn.long0789t
  by newid;
run;
data nhn.long0789a;
set nhn.long0789t(rename=(col1=yesno));
behavior=_name_;
run;

+association 사용할 데이터셋;
data nhn.long0789b;
set nhn.long0789a;
if yesno=0 then delete;
run;

```

Figure 16. Transpose

	newid	NAME OF FORMER VARIABLE	LABEL OF FORMER VARIABLE	COL1
1	1	CS	CS	0
2	1	HD	HD	0
3	1	PI	PI	1
4	1	IS	IS	1
5	1	OB	OB	0
6	1	BS	BS	1
7	1	FS	FS	0
8	2	CS	CS	0
9	2	HD	HD	0
10	2	PI	PI	1
11	2	IS	IS	1
12	2	OB	OB	0
13	2	BS	BS	1
14	2	FS	FS	0
15	3	CS	CS	0

Figure 17. Data set after transpose

SAS EM화면에서 두 개 칼럼만 남은 데이터를 불러와서 association analysis를 하여 결과 창에서 association rules을 확인하고, link analysis 실행 후 이미지로 association을 확인한다.

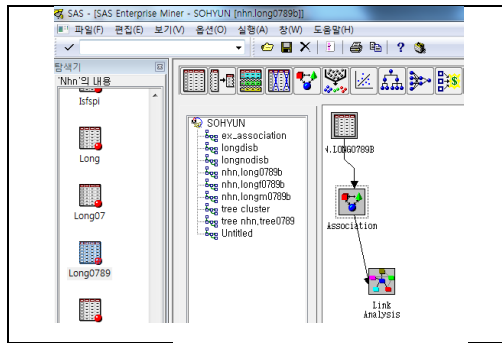


Figure 18. SAS Enterprise Miner 4.3

The screenshot shows the 'Results - Association' window with a table of association rules. The table has columns for 'Relations', 'Lift', 'Support(%)', 'Confidence(%)', and 'Transaction Count'. The rules are numbered 1 through 19.

Relations	Lift	Support(%)	Confidence(%)	Transaction Count	
1	2	2.54	6.07	54.19	853.00 HD => CS
2	3	1.00	12.50	51.83	1755.00 PI & OB => IS
3	3	2.62	4.21	55.95	592.00 PI & HD => CS
4	3	2.46	2.88	52.40	404.00 IS & HD => CS
5	3	2.40	2.25	51.05	316.00 OB & HD => CS
6	3	3.02	1.56	64.41	219.00 HD & BS => CS
7	3	1.07	1.10	55.36	155.00 OB & FS => IS
8	3	1.06	0.71	54.64	100.00 FS & BS => IS
9	3	1.04	0.63	53.94	89.00 FS & CS => IS
10	4	2.53	2.00	53.83	281.00 PI & IS & HD => CS
11	4	2.52	1.59	53.73	223.00 PI & OB & HD => CS
12	4	3.06	1.11	65.27	156.00 PI & HD & BS => CS
13	4	1.07	0.63	55.24	116.00 PI & OB & FS => IS
14	4	2.90	0.73	61.82	102.00 IS & HD & BS => CS
15	4	3.03	0.61	64.66	86.00 OB & HD & BS => CS
16	4	1.00	0.54	55.80	76.00 PI & FS & BS => IS
17	4	1.12	0.50	57.85	70.00 PI & FS & CS => IS
18	4	1.00	0.50	78.65	70.00 IS & FS & CS => PI
19	5	2.91	0.50	61.95	70.00 PI & IS & HD & BS => CS

Figure 19. Association rule

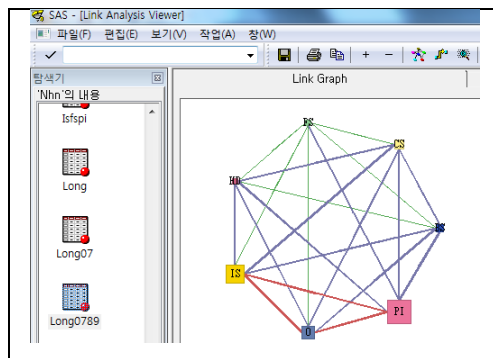


Figure 20. Link analysis

3. 단일 건강위험행위 예측요인

3. 1 여자 대상자에서 단일 건강위험행위를 예측하는 요인

Table 40. Risk factors for predicting lifestyle risk behaviors among women

	CS	HD	PI	IW	IS	BS	FS
	n=515	n=151	n=6,839	n=3,144	n=4,384	n=1,400	n=682
Age group							
20~44							
45~64	-	-	-		+	-	
65+	-	-			+	-	
Marital status							
Married							
Separated/ Widowed	+	+		-	+	+	
Never married	+	+				+	
Education							
College							
High school	+	+		+	+		-
Middle school	+	+		+	+		-
Elementary school	+	+	+	+	+	-	-
Job							
Non manual							
Manual			-				
No job							+
Income							
1st quartile							
2nd quartile							
3rd quartile				+			
4th quartile	+						
Chronic disease							
No							
Yes				+			
Stress							
No							
Yes	+	+			+	+	

+: positive association, -: negative association, blank cells: inconsistent association

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight,

IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking

3. 2 남자 대상자에서 단일 건강위험행위를 예측하는 요인

Table 41. Risk factors for predicting lifestyle risk behaviors among men

	CS	HD	PI	IW	IS	BS	FS
	n=2,482	n=1,019	n=4,220	n=2,291	n=2,892	n=908	n=334
Age group							
20~44							
45~64	-	-				-	
65+	-	-		-	+	-	+
Marital status							
Married							
Separated/ Widowed	+				+	+	
Never married		-				+	
Education							
College							
High school	+	+	-				
Middle school	+	+		-			
Elementary school	+	+		-			
Job							
Non manual							
Manual	+		-			-	
No job							
Income							
1st quartile							
2nd quartile				+			
3rd quartile	+			+			
4th quartile	+		+				
Chronic disease							
No							
Yes		+		+			-
Stress							
No							
Yes	+				+	+	

+: positive association, -: negative association, blank cells: inconsistent association

CS: current smoking, HD: heavy drinking, PI: physical inactivity, IW: inadequate weight,

IS: inadequate sleep, BS: breakfast skipping, FS: frequent snacking

Abstract

An Association Rule Mining-Based Framework for Understanding Lifestyle Risk Behaviors

So Hyun Park

Department of Public Health, Division of Biostatistics

Graduate School of Public Health

Seoul National University

Objectives: Two or more lifestyle risk behaviors are associated with an increased risk of chronic disease, more so than would result from the sum of the separate effects. The prevalence of multiple lifestyle risk behaviors has been studied and published extensively; however, few studies have examined the associations of three or more behaviors and patterns of lifestyle risk behaviors. The aim of this study was to investigate the prevalence and patterns of multiple lifestyle risk behaviors using the Fourth Korea National Health and Nutrition Examination Survey (KNHANES IV, 2007 to 2009) dataset by association rule mining (ARM).

Methods: Using data from KNHANES IV, this study analyzed 14,833 adults aged 20 years and over. This study examined lifestyle risk behaviors by characterizing non-adherence to public health recommendations according to the Alameda 7 health behaviors. The variables were current smoking, heavy drinking, physical inactivity, inadequate weight, inadequate sleep, breakfast skipping, and frequent snacking. Multiple lifestyle risk behaviors were investigated based on the ratios of observed and expected frequencies of behaviors. ARM was conducted to identify multiple lifestyle risk behavior patterns.

Results: The majority of subjects were physically inactive (77% of men and 71.8% of women). Approximately 49.1% of subjects reported inadequate sleep duration. Men were more likely to smoke (46.5%) and drink heavily (18.4%) than women. Women were more likely to be physically inactive (77.0%) than men.

This study revealed that 5.0% of subjects had no lifestyle risk behaviors. Few had 7 risk behaviors. About 69.8% of Korean adults had two or more lifestyle risk behaviors. Multiple lifestyle risk behaviors were more prevalent among men (79.4%) than among women (66.0%). Multiple lifestyle risk behavior patterns included 2, 3, 4, 5, or 6 behaviors.

This study found a number of patterns related to inadequate sleep, physical inactivity, and current smoking using ARM with a support threshold 2%, confidence threshold of 50% for women and 60% for men. Inadequate sleep was associated with physical inactivity and current smoking among women (Lift, 1.06; odds ratio [OR], 1.29; 95% confidence interval [CI], 1.05-1.58). Physical inactivity was associated with inadequate sleep and current smoking among men (Lift, 1.01; OR, 1.26; 95% CI, 1.09-1.45). Current smoking was associated with physical inactivity, inadequate weight, and breakfast skipping among men (Lift, 1.52; OR, 2.87; 95% CI, 2.23-3.70).

Conclusion: Lifestyle risk behaviors were found to be correlated with one another. Inadequate sleep, breakfast skipping, inadequate weight, and current smoking were associated with two or more additional risk behaviors. Identifying patterns of multiple lifestyle risk behaviors will help in planning future prevention strategies.

Key Words: Lifestyle Risk Behaviors, Patterns, Association Rule Mining, Prevention of Lifestyle Risk Behaviors

Student Number: 2007-30617