creative commons
COMMONS DEED

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

dCollection

의학박사 학위논문

교육 수준이 주요 심혈관계질환 위험요인과 심혈관계질환에 미치는 인과적 영향: 멘델 무작위 분석

Causal Effect of Education on Major Cardiovascular Risk Factors and Cardiovascular Diseases: a Mendelian Randomization Analysis

2023 년 8 월

서울대학교 대학원

의학과 의료관리학 전공

장 영 수

교육 수준이 주요 심혈관계질환 위험요인과 심혈관계질환에 미치는 인과적 영향: 멘델 무작위 분석

지도교수 강 영 호

이 논문을 의학박사 학위논문으로 제출함
2023 년 4 월

서울대학교 대학원
의학과 의료관리학 전공
장 영 수

장영수의 박사 학위논문을 인준함
2023 년 7 월

위 원 장 김 윤 (인)

부위원장 강 영 호 (인)

위 원 박 수 경 (인)

위 원 도 영 경 (인)

위 원 박 보 영 (인)

국문초록

건강불평등 연구에서 사회경제적 요인이 건강에 미치는 인과적 영향은 중요한 주제다. 사회경제적 위치가 높을수록 건강 상태가 더 좋게 나타나는 결과는 기존의 많은 관찰연구를 통해 알려졌다. 지금까지 반복된 관찰연구 결과에서 사회경제적 위치와 건강 상태 사이의 강한 연관성이 알려졌지만, 사회경제적 위치가 건강에 미치는 인과성을 탐구하는 연구가 필요하다. 연관성은 인과성을 의미하는 것이 아니기 때문이다.

개인의 건강에 영향을 주는 원인은 다양하다. 연령, 성별, 거주지역, 유전적 요인, 건강 행태 등의 수많은 요인들이 서로 복잡하게 작용한다. 우리가 알고자 하는 특정 요인만 골라내어 건강에 미치는 인과적 영향을 파악하는 것은 어려운 일이다. 특정 원인과 결과 사이의 인과성을 살피는 데 가장 큰 걸림돌은 역인과성과 교란작용이다.

역인과성과 교란작용을 통제하기 위한 연구 방법은 무작위 배정 연구가 있다. 하지만 무작위 배정 연구는 시간과 비용이 많이 들고 윤리적 문제가 존재하기 때문에 모든 연구를 무작위 배정 연구로 진행할 수 없다. 인과성을 탐구하는 또 다른 연구 방법에는 도구변수를 이용하는 방법이 있다. 도구변수는 3가지 조건을 만족시키는 변수를 말한다. 첫째, 도구 변수는 독립변수와 연관되어 있다. 둘째, 도구 변수는 오직 독립변수를 통해서만 결과변수에 영향을 준다. 셋째, 도구

변수는 독립변수와 결과변수 사이의 어떠한 교란변수와도 연관되지 않아야 한다.

멘델 무작위 분석은 독립변수와 관련된 단일염기서열 다형성(single nucleotide polymorphisms, 이하 SNP)를 도구 변수로 놓고, 독립변수와 결과변수 사이의 인과성을 분석하는 방법이다. 유전자는 부모로부터 50% 확률로 무작위하게 전달되기 때문에, 독립변수와 연관이 있는 것으로 알려진 유전자 변이를 도구변수로 설정할 수 있다. 유전자가 무작위하게 전해지는 것을 무작위 배정과 유사하다고 가정할 수 있다. 멘델 무작위 분석에서 도구변수로 사용하는 유전자 변이는 도구변수의 조건을 만족시켜야만 한다.

이 연구는 질병관리청 국립보건연구원 산하 국립중앙인체자원은행에서 제공하는 한국인유전체역학조사사업(Korean Genome and Epidemiology Study, 이하 KoGES) 코호트 자료를 활용하였다. KoGES 코호트는 지역사회 기반 코호트, 농촌 코호트, 도시 코호트, 쌍둥이-가족 코호트 등으로 구성되어 있다. 각각의 코호트는 역학 정보와 유전 정보를 함께 제공하고 있으며, 본 연구는 지역사회 기반 코호트, 농촌 코호트, 도시 코호트 자료를 사용하였다.

본 연구에서 다루게 될 주요 분석 방법은 크게 2가지로 구분할 수 있다. 첫째, 교육 수준을 예측하는 유전 변이를 GWAS 분석을 통해 찾는다. 둘째, 교육 수준을 예측하는 유전 변이를 도구변수로 놓고 결과변수인 주요 심혈관계질환 위험요인과 심혈관계질환에 대한 멘델

무작위 분석을 실시한다. 멘델 무작위 분석은 KoGES 코호트 통합 자료를 이용한 1-표본 멘델 무작위 분석(one sample Mendelian randomization), 도시 코호트와 지역사회/농촌 코호트를 나누어 2-표본 멘델 무작위 분석(two sample Mendelian randomization)을 실시하였다.

교육 수준은 교육 연한으로 변환하여 분석했다. 무학은 0년, 초등학교 중퇴는 3년, 초등학교 졸업 및 중학교 중퇴는 6년, 중학교 졸업 및 고등학교 중퇴는 9년, 고등학교 졸업은 12년, 전문대 졸업은 14년, 대학교 졸업은 16년, 대학원 이상은 18년의 연한으로 변환하였다. 전체 연구집단의 평균 교육 연한은 11.1년이었으며, 1 표준편차에 해당하는 교육 연한은 4.0년이었다.

GWAS 분석에 앞서 자료의 질 관리를 실시하였다. 전체 유전 정보는 국립자원인체자원은행에서 대체(imputation)을 실시하여 제공되어 유전 정보의 결측치는 없었다. 전체 SNP은 8,056,211개에서 마이너 대립유전자 빈도 1% 이하인 9,770개를 제외하였다. 하디-와인버그 검정을 실시해 멘델 유전법칙을 따르는 SNP를 남겼으며, 5e-6을 기준으로 72,098개 SNP을 제외하였다. 마지막으로 LD 전정(pruning)을 실시해 7,135,046개의 SNP을 제외해 총 839,297개의 SNP를 남겼다.

위 과정을 통해 추출한 839,297개 SNP 중에서 교육 연한을 예측하는 SNP을 찾기 위해 도시 코호트와 지역사회/농촌 코호트 두 경우에

서 추출하였다. 이는 교육 수준을 예측하는 SNP의 재현가능성(replication)을 위해 실시하였다. 도시 코호트에서는 126개의 SNP가 산출되었고, 지역사회/농촌 코호트에서는 163개 SNP가 산출되었다. 두 경우에서 중복되는 SNP 55개를 본 연구의 멘델 무작위 분석에 이용하였다. P-value는 5e-5를 기준으로 했으며, p-value가 5e-6일 때는 33개의 SNP를 추출하였고, 5e-7일 때는 13개의 SNP를 추출하였다.

멘델 무작위 분석은 p-value에 따라 계산하였다. 심혈관계질환에서 1-표본 멘델 무작위 분석을 통해 역분산가중치(inverse variance weighted, 이하 IVW)로 계산된 오즈비(Odds ratio, 이하 OR)는 p-value 5e-5일 때 0.76, 5e-6일 때 0.74, 5e-7일 때 0.73이었다. 이는 1 표준편차의 교육 연한인 4.0년이 증가하면 심혈관계질환 위험이 24% 낮음을 뜻한다. 2-표본 멘델 무작위 분석에서 역분산가중치로 계산된 오즈비는 p-value 5e-5일 때 0.72, 5e-6일 때 0.68, 5e-7일 때 0.69이었다. 이는 1 표준편차의 교육 연한인 4.0년이 증가하면 심혈관계질환 위험이 28% 낮음을 의미한다.

심혈관계질환 위험요인에 대한 멘델 무작위 분석 결과, 1-표본 멘델 무작위 분석에서 고혈압의 오즈비는 0.72, 당뇨병의 오즈비는 0.69, 고지혈증의 오즈비는 0.89, 25 kg/m^2 이상 비만의 오즈비는 0.90, 30 kg/m^2 이상 비만의 오즈비는 0.85, 흡연은 0.69로 나타났다. 2-표본 멘델 무작위 분석에서 고혈압의 오즈비는 0.76, 당뇨병의

오즈비는 0.76, 고지혈증의 오즈비는 0.88, 25 kg/m^2 이상 비만의 오즈비는 0.90, 30 kg/m^2 이상 비만의 오즈비는 0.85, 흡연은 0.66으로 나타났다.

해외에서 알려진 교육 수준에 대한 SNP와 KoGES 자료 내에서 비교하였다. 해외에서 알려진 SNP은 162개, 74개 SNP이 있었으며, KoGES 자료를 검토했을 때 162개 SNP 중 4개 SNP이 KoGES 자료에 존재했으며, 74개 SNP 중 2개의 SNP이 KoGES 자료에 존재하였다. 본 연구에서 추출한 교육 수준에 대한 55개의 SNP가 각 염색체 내에 어떤 유전자에 위치하는지 살펴보았다. 55개 SNP들은 다양한 유전자에 위치했으며, 1번 염색체에 위치한 MCOLN2 유전자는 음이온 채널 활성에 관여하는 유전자이며, 2번 염색체의 NCMAP 유전자는 마이엘린 관련 단백질과 연관되어 있다. 55개 염색체 중에서 8번 염색체와 12번 염색체에서 SNP들이 많이 분포했는데, 해당 위치에 있는 유전자는 DLGAP2, GIT2, CUX2, ATXN2, ACAD10, NAA25, HECTD4, PTPN11, RPH3A 등이다. 이 유전자들은 시냅스 구성, 신경세포 발달, 대사 관련 효소 등에 대해 기능을 갖고 있다. 특히 8번 염색체의 DLGAP2 유전자는 신경세포 전달과 시냅스 구성에서 중요한 역할을 하는 단백질과 관련이 있으며, CUX2 유전자는 신경세포 증식과 분화에 관련된 요인이다. 비록 이들 변이가 인트론에 위치하였지만, 인트론 변이가 있을 경우 해당 유전자의 영향을 받을 수 있다.

따라서 KoGES에서 추출한 SNP들이 위치한 유전자들의 기능을 고려했을 때, 이 SNP들이 신경전달물질, 신경 발달 등의 유전자 기능에 영향을 받는다고 볼 수 있다. 이 기능들은 교육에 대한 직접적인 경로에 해당하며, KoGES에서 산출한 SNP들이 교육을 예측하는 기전과 연관성이 있다고 해석할 수 있다.

이 연구는 우리나라에서 교육 수준이 심혈관계질환 위험요인과 심혈관계질환에 미치는 인과적 영향을 처음으로 탐구하였다는 점에서 의의가 있다. 심혈관계질환은 중요한 사망원인과 질병부담 원인 중 하나인데, 이번 연구를 통해 교육 수준의 증가가 심혈관계질환 위험도를 감소시키는 크기가 인과적으로 측정되었다. 이 결과를 바탕으로 우리나라에서 사회경제적 불평등을 완화하기 위한 정책의 근거를 세우는 데 도움을 줄 수 있다.

이 연구의 제한점이 있는데, 표본 수가 해외 연구 사례에 비해 부족했다는 점이다. 이러한 제한점을 극복하기 위한 후속 연구와 새로운 자료 구축이 필요하다. 사회경제적 위치 지표를 교육 외에 다른 지표들을 사용할 수 있는지도 후속 연구가 진행해야 할 과제다. 자료 구축이 지속적으로 이루어지면서 표본 수도 더 늘어날 것으로 생각되며 추후에 진행하는 멘델 무작위 분석 연구에도 도움이 될 것이다.

사회경제적 위치가 건강에 미치는 인과적 영향을 밝히는 것은 학문적으로도 중요한 의의가 있으며, 나아가 사회정책적으로 건강불평등 개선을 위한 근거를 제시하는 토대가 된다.

주요어 : 교육, 심혈관계질환 위험요인, 심혈관계질환, GWAS, 멘델 무작위 분석

학 번 : 2018-34360

목 차

I. 서론 ····· 1

1. 연구 배경 ····· 1

1.1. 연구의 필요성 ····· 1

1.1.1. 기존 건강불평등 연구에서 교육이 건강에 미치는 영향 ····· 1

1.1.2 건강불평등 연구에서 인과성에 대한 논의 ····· 3

1.1.3. 역인과성 ····· 4

1.1.4. 교란작용 ····· 5

1.1.5. 역인과성과 교란작용을 해결하는 연구 방법: 무작위 배정 ····· 6

1.1.6. RCT 외 다른 연구 방법: 준실험적 연구, 도구변수 분석 등 ····· 7

1.2. 멘델 무작위 분석 ····· 9

1.2.1. 멘델 무작위 분석 연구 사례 – 해외 ····· 11

1.2.2. 멘델 무작위 분석 연구 사례 – 국내 ····· 15

1.2.3. 측정 가능한 중요한 건강 지표 – 심혈관계질환 ····· 17

II. 연구 목적 ····· 19

1. 연구 목적 ····· 19

2. 연구의 세부 목표 ····· 19

III. 연구 방법 ····· 20

1. 연구 자료 ····· 20

1.1. 자료원 ····· 20

1.1.1. 지역사회 기반 코호트 ····· 20

1.1.2. 농촌 코호트 ····· 21

1.1.3. 도시 코호트 ····· 21

1.1.4. 자료의 질 관리 ····· 22

1.2. 연구 대상자 선정 기준 ····· 26

2. 분석 방법 ····· 27

2.1. 연구의 주요 지표 ····· 27

2.1.1. 독립변수 - 교육 ····· 27

2.1.2. 결과변수 - 심혈관계질환 ····· 29

2.1.3. 결과변수 - 주요 심혈관계질환 위험요인 ····· 30

2.2. 분석 ····· 36

2.2.1. Genome-wide association study ····· 37

2.2.2. 교육 수준을 예측하는 SNP 찾기 ····· 38

2.2.3. 멘델 무작위 분석 ····· 43

2.2.4 멘델 무작위 분석의 민감도 ····· 46

2.2.4. 분석 프로그램 ····· 47
2.3. 연구 윤리 ····· 48

IV. 연구 결과 ····· 49

1. 연구 대상자의 특성 ····· 49
1.1. 연구 대상자의 일반적 특성 ····· 49
1.2. KoGES 코호트에서 계산한 심혈관계질환 발생 위험도(로지스틱 회귀분석) ····· 60
2. GWAS 결과 – 교육 수준 ····· 62
3. 멘델 무작위 분석 결과 – 심혈관계질환 ····· 99
3.1. 1-표본 멘델 무작위 분석 – KoGES 통합 ····· 99
3.2. 2-표본 멘델 무작위 분석 – 도시 코호트, 지역사회/농촌 코호트 ····· 109
4. 멘델 무작위 분석 결과 –심혈관계질환 위험요인 ····· 117
4.1. 1-표본 멘델 무작위 분석 – KoGES 통합 ····· 117
4.1.1. 고혈압 ····· 117
4.1.2. 당뇨병 ····· 126
4.1.3. 고지혈증 ····· 135
4.1.4. 비만 – 체질량지수 25kg/m^2 이상 ····· 144
4.1.5. 비만 – 체질량지수 30kg/m^2 이상 ····· 153
4.1.6. 흡연 ····· 162

4.2. 2-표본 멘델 무작위 분석 - 도시 코호트, 지역사회·농촌 코호트 ····· 171
4.2.1. 고혈압 ····· 171
4.2.2. 당뇨병 ····· 178
4.2.3. 고지혈증 ····· 185
4.2.4. 비만 - 체질량지수 25kg/m^2 이상 ····· 192
4.2.5. 비만 - 체질량지수 30kg/m^2 이상 ····· 199
4.2.6. 흡연 ····· 206

V. 연구 결과의 요약 ····· 213
1.1. 연구 대상자의 일반적 특성 ····· 213
1.2. KoGES 코호트에서 계산한 심혈관계질환 발생 위험도(로지스틱 회귀분석) ····· 214
1.3. 교육 수준에 대한 GWAS 결과 ····· 215
1.4. 멘델 무작위 분석 결과 ····· 216
1.4.1. 심혈관계질환 ····· 216
1.4.2. 심혈관계질환 위험요인 ····· 219

VI. 고찰 ····· 223
1. 교육 수준에 대한 GWAS 연구 ····· 223
1.1. 외국의 교육에 대한 SNP와 비교 ····· 223

1.2. 교육을 예측하는 유전 변이에 대한 기전 ····· 225
1.2. KoGES 결과에 대한 교육을 예측하는 유전 변이에 대한 기전 ····· 227
2. 멘델 무작위 분석에 대한 해석 ····· 231
2.1. 기존 연구 결과와 비교 ····· 232
2.2. 관찰 연구 결과와 비교 ····· 237
2.3. 심혈관계질환 위험요인에 따른 OR 차이 ····· 238
2.4. 내재적, 외재적 타당성 ····· 239
3. 연구의 정책적, 학문적 의의 ····· 242
3.1. 정책적 의의 ····· 242
3.2. 학문적 의의 ····· 245
3.3. 연구의 제한점 ····· 248

VII. 결론 ····· 251

부록 ····· 254
참고문헌 ····· 277
Abstract ····· 296

표 목 차

[표 1] 멘델 무작위 분석을 통해 교육 수준이 건강에 미치는 영향을 살펴본 해외 연구 사례 ····· 13
[표 2] 멘델 무작위 분석을 이용한 국내 연구 사례 ····· 16
[표 3] KoGES 코호트에 대한 요약표 ····· 24
[표 4] 본 연구에 사용한 변수에 대한 설문 정보 ····· 31
[표 5] 본 연구의 수행 절차 ····· 36
[표 6] 지역사회 기반 코호트 대상자의 일반적 특성 ····· 51
[표 7] 농촌 코호트 대상자의 일반적 특성 ····· 54
[표 8] 도시 코호트 대상자의 일반적 특성 ····· 58
[표 9] KoGES 코호트에서 교육 수준에 따른 심혈관계질환 위험도(로지스틱 회귀분석) ····· 61
[표 10] 도시 코호트에서 GWAS 분석을 통하여 교육 수준을 예측한 목록(p-value 〈5e-5) ····· 64
[표 11] 지역사회/농촌 코호트에서 GWAS 분석을 통하여 교육 수준을 예측한 목록(p-value 〈5e-5) ····· 75
[표 12] 도시 코호트와 지역사회/농촌 코호트에서 중복되는 SNP 목록(p-value 〈5e-5) ····· 89
[표 13] 도시 코호트와 지역사회/농촌 코호트에서 중복되는 SNP 목

록(p-value 〈5e-6) ····· 94
[표 14] 도시 코호트와 지역사회/농촌 코호트에서 중복되는 SNP 목록(p-value 〈5e-7) ····· 97
[표 15] 이질성 검정 결과(교육 수준 - 심혈관계질환) ····· 99
[표 16] 다형성 검정 결과(교육 수준 - 심혈관계질환) ····· 100
[표 17] 1-표본 멘델 무작위 분석 결과(교육 수준 - 심혈관계질환) ····· 101
[표 18] 교육 수준과 심혈관계질환에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 103
[표 19] 교육 수준과 심혈관계질환에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 106
[표 20] 이질성 검정 결과(교육 수준 - 심혈관계질환) ····· 109
[표 21] 다형성 검정 결과(교육 수준 - 심혈관계질환) ····· 110
[표 22] 2-표본 멘델 무작위 분석 결과(교육 수준 - 심혈관계질환) ····· 110
[표 23] 교육 수준과 심혈관계질환에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 113
[표 24] 이질성 검정 결과(교육 수준 - 고혈압) ····· 117
[표 25] 다형성 검정 결과(교육 수준 - 고혈압) ····· 118
[표 26] 1-표본 멘델 무작위 분석 결과(교육 수준 - 고혈압) ···· 118
[표 27] 교육 수준과 고혈압에 대한 멘델 무작위 분석에서 각 SNP의 연결

결과 ····· 120
[표 28] 교육 수준과 고혈압에 대한 멘델 무작위 분석에서 confounder 검정 결과 ····· 123
[표 29] 이질성 검정 결과(교육 수준 – 당뇨병) ····· 126
[표 30] 다형성 검정 결과(교육 수준 – 당뇨병) ····· 127
[표 31] 1-표본 멘델 무작위 분석 결과(교육 수준 – 당뇨병) ··· 127
[표 32] 교육 수준과 당뇨병에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 129
[표 33] 교육 수준과 당뇨병에 대한 멘델 무작위 분석에서 confounder 검정 결과 ····· 132
[표 34] 이질성 검정 결과(교육 수준 – 고지혈증) ····· 135
[표 35] 다형성 검정 결과(교육 수준 – 고지혈증) ····· 136
[표 36] 1-표본 멘델 무작위 분석 결과(교육 수준 – 고지혈증) 136
[표 37] 교육 수준과 고지혈증에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 138
[표 38] 교육 수준과 심혈관계질환에 대한 멘델 무작위 분석에서 confounder 검정 결과 ····· 141
[표 39] 이질성 검정 결과(교육 수준 – 비만, 체질량지수 25kg/m^2 이상) ····· 144
[표 40] 다형성 검정 결과(교육 수준 – 비만, 체질량지수 25kg/m^2 이상) ····· 145

[표 41] 1-표본 멘델 무작위 분석 결과(교육 수준 - 비만, 체질량지수 25kg/m^2 이상) ····· 145
[표 42] 교육 수준과 고지혈증에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 147
[표 43] 교육 수준과 고지혈증에 대한 멘델 무작위 분석에서 confounder 검정 결과 ····· 150
[표 44] 이질성 검정 결과(교육 수준 - 비만, 체질량지수 30kg/m^2 이상) ····· 153
[표 45] 다형성 검정 결과(교육 수준 - 비만, 체질량지수 30kg/m^2 이상) ····· 154
[표 46] 1-표본 멘델 무작위 분석 결과(교육 수준 - 비만, 체질량지수 30kg/m^2 이상) ····· 154
[표 47] 교육 수준과 비만에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 156
[표 48] 교육 수준과 비만에 대한 멘델 무작위 분석에서 confounder 검정 결과 ····· 159
[표 49] 이질성 검정 결과(교육 수준 - 흡연) ····· 162
[표 50] 다형성 검정 결과(교육 수준 - 흡연) ····· 163
[표 51] 1-표본 멘델 무작위 분석 결과(교육 수준 - 흡연) ····· 163
[표 52] 교육 수준과 흡연에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 165

[표 53] 교육 수준과 흡연에 대한 멘델 무작위 분석에서 confounder 검정 결과 ····· 168
[표 54] 이질성 검정 결과(교육 수준 – 고혈압) ····· 171
[표 55] 다형성 검정 결과(교육 수준 – 고혈압) ····· 172
[표 56] 2-표본 멘델 무작위 분석 결과(교육 수준 – 고혈압) ··· 172
[표 57] 교육 수준과 고혈압에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 174
[표 58] 이질성 검정 결과(교육 수준 – 당뇨병) ····· 178
[표 59] 다형성 검정 결과(교육 수준 – 당뇨병) ····· 179
[표 60] 2-표본 멘델 무작위 분석 결과(교육 수준 – 당뇨병) ··· 179
[표 61] 교육 수준과 당뇨병에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 181
[표 62] 이질성 검정 결과(교육 수준 – 고지혈증) ····· 185
[표 63] 다형성 검정 결과(교육 수준 – 고지혈증) ····· 186
[표 64] 2-표본 멘델 무작위 분석 결과(교육 수준 – 고지혈증) 186
[표 65] 교육 수준과 고지혈증에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 188
[표 66] 이질성 검정 결과(교육 수준 – 비만, 체질량지수 25kg/m^2 이상) ····· 192
[표 67] 다형성 검정 결과(교육 수준 – 비만, 체질량지수 25kg/m^2 이상) ····· 193

[표 68] 2-표본 멘델 무작위 분석 결과(교육 수준 - 비만, 체질량지수 25kg/m^2 이상) ····· 193
[표 69] 교육 수준과 비만에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 197
[표 70] 이질성 검정 결과(교육 수준 - 비만, 체질량지수 30kg/m^2 이상) ····· 199
[표 71] 다형성 검정 결과(교육 수준 - 비만, 체질량지수 30kg/m^2 이상) ····· 200
[표 72] 2-표본 멘델 무작위 분석 결과(교육 수준 - 비만, 체질량지수 30kg/m^2 이상) ····· 200
[표 73] 교육 수준과 비만에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 202
[표 74] 이질성 검정 결과(교육 수준 - 흡연) ····· 206
[표 75] 다형성 검정 결과(교육 수준 - 흡연) ····· 207
[표 76] 2-표본 멘델 무작위 분석 결과(교육 수준 - 흡연) ····· 207
[표 77] 교육 수준과 흡연에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과 ····· 209
[표 78] 심혈관계질환에 대한 관찰 연구 결과와 멘델 무작위 분석 연구 결과 요약 ····· 217
[표 79] 심혈관계질환 위험요인에 대한 멘델 무작위 분석 연구 결과 요약 ····· 219

[표 80] 심혈관계질환 위험요인에 대한 멘델 무작위 분석 연구 결과 요약 ····· 221
[표 81] 교육에 대한 해외의 162개 SNP 목록 ····· 254
[표 82] 교육에 대한 해외의 72개 SNP 목록 ····· 261
[표 83] 교육에 대한 해외의 SNP와 KoGES SNP 사이의 비교 ····· 265
[표 84] UK biobank SNP와 KoGES SNP 사이의 비교 ····· 266
[표 85] Priotized genes in neural communication ····· 266
[표 86] KoGES에서 최종적으로 사용한 55개 SNP의 정보 · 267
[표 87] 교육에 대한 SNP들이 위치한 Gene과 그 기능들 ···· 272
[표 88] 교육 수준과 심혈관계질환에 대한 멘델 무작위 분석 연구 결과(해외) ····· 274
[표 89] 교육 수준과 심혈관계질환 위험요인에 대한 멘델 무작위 분석 연구 결과(해외) ····· 275

그 림 목 차

[그림 1] 사회경제적 위치가 건강에 미치는 영향을 나타낸 인과적 화살표(Causal arrow) ····· 4

[그림 2] 교란변수를 설명하는 그림 ····· 6

[그림 3] 그림 3. 도구변수를 설명하는 그림 ····· 9

[그림 4] RCT와 멘델 무작위 분석의 비교 ····· 11

[그림 5] GWAS 질 관리(quality control) 순서도 ····· 25

[그림 6] GWAS를 이용해 교육에 대한 SNP 추출하는 과정 38

[그림 7] 도구변수 조건을 만족하는 SNP 찾기 위한 인과적 경로 ····· 40

[그림 8] 멘델 무작위 분석의 과정을 설명한 그림 ····· 44

[그림 9] 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 심혈관계질환) ····· 102

[그림 10] 2-표본 멘델 무작위 분석의 산점도 (교육 수준 - 심혈관계질환) ····· 112

[그림 11] 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 고혈압) ····· 119

[그림 12] 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 당뇨병) ····· 128

[그림 13] 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 고지혈증) ····· 137
[그림 14] 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 비만, 체질량 지수 25kg/m^2 이상) ····· 146
[그림 15] 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 비만, 체질량 지수 30kg/m^2 이상) ····· 155
[그림 16] 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 흡연) ····· 164
[그림 17] 2-표본 멘델 무작위 분석의 산점도 (교육 수준 - 고혈압) ····· 173
[그림 18] 2-표본 멘델 무작위 분석의 산점도 (교육 수준 - 당뇨병) ····· 180
[그림 19] 2-표본 멘델 무작위 분석의 산점도(교육 수준 - 고지혈증) ····· 187
[그림 20] 2-표본 멘델 무작위 분석의 산점도 (교육 수준 - 비만, 체질량 지수 25kg/m^2 이상) ····· 194
[그림 21] 2-표본 멘델 무작위 분석의 산점도 (교육 수준 - 비만, 체질량 지수 30kg/m^2 이상) ····· 201
[그림 22] 2-표본 멘델 무작위 분석의 산점도 (교육 수준 - 흡연) ····· 208

약어

GWAS	Genome-wide association study
HWE	Hardy-Weinberg equilibrium
HR	Hazard ratio
IVW	Inverse-variance weighted
KoGES	Korean Genome and Epidemiology Study
LD	Linkage disequilibrium
MAF	Minor allele frequency
MR	Mendelian randomization
OR	Odds ratio
RCT	Randomized controlled trial
SNP	Single nucleotide polymorphism
2SLS	Two-stage least squares

I. 서론

1. 연구 배경

1.1. 연구의 필요성

1.1.1. 기존 건강불평등 연구에서 교육이 건강에 미치는 영향

교육과 사망률의 연관성에 대한 Kitigawa 등(Kitigawa, et al. 1973)의 연구 이후, 교육과 건강 사이의 관계를 설명하는 수많은 연구가 진행되었다. 예를 들어, 1979년부터 1985년까지 추적 관찰된 대규모 국가 종단 사망률 연구(National Longitudinal Mortality Study)에 따르면 교육 수준이 1년 증가할 때마다 사망률이 7~8% 감소하는 것으로 나타났다(Elo, et al., 1996). 22개 유럽 국가를 대상으로 한 Mackenbach 등의 연구에 따르면, 교육 수준이 낮을수록 사망률이 높았다(Mackenbach, et al., 2008; Mackenbach, et al., 2016).

국내에서 진행된 관찰 연구에서도 해외와 마찬가지로 교육 수준이 낮을수록 건강 상태가 좋지 않았다. 강영호 등에 따르면 교육 수준이 낮을수록 사망률의 위험이 높았다(Khang, et al., 2004; Khang, et al., 2005; Khang, et al., 2016). 정최경희 등에 따르면 교육 수준이 높을수록 기대수명이 높았다(Jung-Choi, et al., 2014). 박진욱 등에 따르면 40년 동안 한국

인의 교육 수준은 증가함과 동시에 사망률은 크게 감소하였지만 교육에 의한 사망률의 절대적, 상대적 불평등은 오히려 증가하였다(Bahk, et al., 2016). 임도희 등에 따르면 교육 수준의 지속적이고 급격한 증가는 저학력 집단에서 사망률의 인구집단 기여위험도(population attributable fraction, PAF)가 감소하는 데 기여했다(Lim, et al., 2015).

교육이 건강에 미치는 영향에 대한 기전에 대한 많은 관찰 연구들이 진행되었다(Winkleby, et al., 1992; Schultz, et al., 2018; Kaplan, et al., 1993; Stringhini, et al., 2017). 교육이 건강에 영향을 주는 기전은 다음과 같다. 첫째, 교육 수준이 높을수록 건강한 생활습관을 가지게 된다. 건강한 생활습관에는 건강한 음식 섭취, 운동, 비흡연이나 금주와 같은 건강 행태가 포함된다. 둘째, 높은 교육 수준을 가진 사람들은 문제 해결 능력, 감정 관리, 스트레스에 대한 저항력 등이 크다(Berkman, et al., 2014). 만성적 스트레스는 심혈관계질환 뿐만 아니라 다른 질환에도 악영향을 주는 것으로 알려져 있다. 셋째, 높은 교육 수준은 더 좋은 직업과 높은 소득을 갖게 할 연관성이 있다(Card, et al., 2001). 직업적으로 육체직보다는 사무직을 선택할 확률이 높으며, 높은 소득은 더 나은 의료서비스에 대한 접근성을 갖게 한다. 따라서 심혈관계질환 조기 발견과 치료를 가능하게 한다. 넷째, 교육이 높을수록 더 높은 교육을 받은 배우자, 동료 등을 만날 확률이 높아져 장기적인 건강 상태가 좋아질 수 있다(Berkman, et al., 2014). 주변 사람들에 의한 정보 교환이 원활하게 이루어지며 건강 정보에 대한 접근성도 높아지게 된다. 즉 교육을 더 받을수록 더 높은 “사회적 자본(social capital)”을 가지게 되므로 높은 건강 수준을 가질 확률이 증가한다.

건강불평등 연구에서 사회경제적 요인이 건강에 미치는 영향은 중요한 주제다. 지금까지 반복된 관찰 연구 결과에서 교육과 건강 상태 사이의 강한 연관성이 알려졌지만, 반드시 교육의 개선이 건강과 사망률을 개선한다는 것을 의미하지는 않는다. 이 문제를 해결하기 위해 우리는 교육이 건강에 미치는 영향을 인과적으로 탐구하는 연구가 필요하다. 연관성은 인과성을 의미하는 것이 아니기 때문이다.

1.1.2 건강불평등 연구에서 인과성에 대한 논의

Mackenbach는 건강불평등의 인과성을 논의하는 데 두 가지 질문을 제시했다(Mackenbach, 2019). 첫째, 사회경제적 위치의 차이가 건강에서의 차이를 일으키는가? 둘째, 건강에 영향을 미치는 사회경제적 위치와 관련된 특정 요인과 기전은 무엇인가? 궁극적으로 두 질문에 대해 과학적으로 타당한 정답을 찾는 것은 [그림 1]에서 보듯이 사회경제적 위치가 인과적으로 건강에 영향을 미치는지 따져보는 것이다.

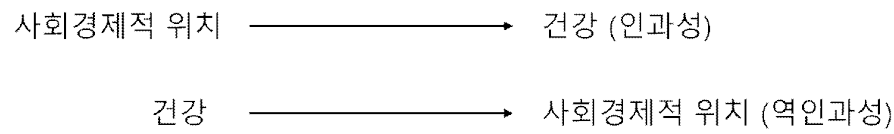

그림 1. 사회경제적 위치가 건강에 미치는 영향을 나타낸 인과적 화살표(Causal arrow)

개인의 건강에 영향을 주는 원인은 다양하다. 연령, 성별, 거주지역, 유전적 요인, 건강 행태 등의 수많은 요인들이 서로 복잡하게 작용한다. 우리가 알고자 하는 특정 요인만 골라내어 건강에 미치는 인과적 영향을 파악하는 것은 어려운 일이다. 특정 원인과 결과 사이의 인과성을 살피는 데 가장 큰 걸림돌은 역인과성과 교란작용이다 (Mackenbach, 2019; Berkman, et al., 2014).

1.1.3. 역인과성

역인과성이란 그림 1에서의 인과적 화살표가 반대인 경우를 뜻한다. 결과가 원인에 역으로 영향을 주는 것을 의미한다. 사회경제적 요인이 건강에 영향을 주기도 하지만, 반대로 건강이 나빠지면 직업적 안정성이 떨어지고 소득도 감소할 수 있다. 따라서 역인과성으로 인해 실제 효과는 관찰연구에서 보고된 효과보다 낮을 것이라고 생각할

수 있다.

교육은 소득, 직업에 비해 역인과성의 가능성이 적은 것으로 알려졌다(Kawachi, et al., 2010). 왜냐하면 학교를 졸업하고 나면 직업과 소득의 변화에 상관없이 교육 상태는 그대로 유지되기 때문이다. 하지만 교육에서도 역인과성을 완전히 배제하긴 어렵다. Case 등에 따르면 아동기에 만성적 질환을 겪게 되면 교육성취도가 내려간다고 한다(Case, et al., 2005).

1.1.4. 교란작용

교란작용은 독립변수와 결과변수 사이의 인과적 효과가 경로에 존재하지 않는 제 3의 변수에 의해 영향을 받는 것을 의미한다. 교란작용을 일으키는 변수를 교란변수라고 하는데, [그림 2]와 같다. 첫째, 독립변수와 연관성이 있으면서 결과변수의 원인이 돼야 한다. 둘째, 독립변수와 결과변수의 경로에 존재하지 않아야 한다. 두 번째 조건에서 교란변수는 매개변수와 다르다. 매개변수는 독립변수와 결과변수 경로 사이에서 매개하는 변수이기 때문이다.

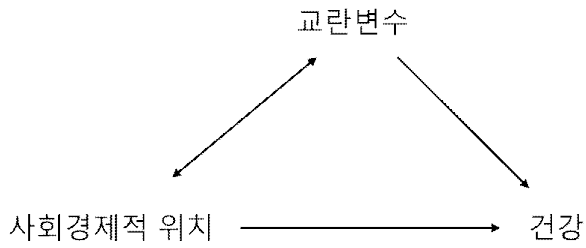

그림 2. 교란변수를 설명하는 그림

교란변수는 우리가 찾고자 하는 독립변수와 결과변수 사이의 인과적 효과를 '교란'하기 때문에 연구자는 분석을 수행할 때 적절하게 교란변수를 통제하려고 한다. 대표적인 교란변수에는 성, 연령 등이 있지만, 연구자가 모든 교란변수를 보정하는 것은 불가능하다.

1.1.5. 역인과성과 교란작용을 해결하는 연구 방법 : 무작위 배정

앞서 살펴본 것처럼 역인과성과 교란작용을 통제하는 것이 인과성 분석에 중요하다. 이론적으로 인과성을 정확하게 알 수 있는 방법은 반사실적(counterfactual) 모형 또는 잠재적 결과 접근법(potential outcome approach)을 이용하는 것이다(Rubin 2005). 동일한 대상자

가 동시에 원인변수에 노출될 경우와 노출되지 않는 경우의 결과를 비교하는 방법이지만, 현실적으로 불가능하다. 반사실적 모형의 대안으로 무작위 배정 연구(randomized controlled trial, 이하 RCT)를 수행할 수 있다(VanderWeele, 2016). RCT는 현실적으로 인과성을 밝히는 데 유용한 연구 설계 방법이다. RCT에서 역인과성과 교란작용을 통제하는 가장 중요한 수단은 무작위 배정이다.

하지만 RCT도 무결점한 연구 설계는 아니다. RCT는 시간과 비용이 많이 들기 때문에 제한된 인원으로 시행하는 경우가 많으며, RCT에 참여한 대상자의 특성이 전체 인구 집단과 다르다면 RCT의 결과를 일반화하기 어렵다. 또한, 실험군과 대조군을 설정할 때 윤리적 문제의 가능성도 존재한다.

1.1.6. RCT 외 다른 연구 방법 : 준실험적 연구, 도구변수 분석 등

RCT를 현실적으로 수행하기 어려운 이유들이 있기 때문에, 연구자들은 준실험적 연구, 도구변수 분석 등을 활용하여 인과성을 측정하고자 한다. 준실험적 연구는 RCT의 무작위 배정 없이 특정 집단에 어떤 정책이 인과적으로 효과가 있는지 알아보기 위해 수행하는 연구이다. 연구자는 '자연 실험'을 통해 실험군과 대조군을 나누게 되는데, 이때 '자연 실험'이 RCT에서 무작위 배정을 실시한 것과 유사한

효과를 가지게 된다. 대표적인 준실험적 연구는 오레곤 실험(Oregon Experiment)이 있다(Baicker, et al., 2013). 미국 오레곤(Oregon) 주에서는 메디케이드(Medicaid)의 효과를 평가하기 위해 메디케이드에 속하지 않는 저소득층을 대상으로 로또 추첨을 하여 메디케이드를 제공하였다. 메디케이드에 포함된 실험군은 대조군에 비해 자가보고건강 수준이 높았으며 경제적 안정성이 더 높아졌다.

인과성을 측정하기 위한 또 다른 연구 방법에는 도구변수를 이용하는 방법이 있다. 도구변수는 아래와 같은 3가지 조건을 만족시키는 변수를 말한다.

- 1) Relevance: 도구변수는 독립변수와 연관되어 있다.
- 2) Exclusion restriction: 도구변수는 오직 독립변수를 통해서만 결과변수에 영향을 준다.
- 3) Exchangeability or Independence: 도구변수는 독립변수와 결과변수 사이의 어떠한 교란변수와 연관되지 않아야 한다.

[그림 3]은 위의 조건을 만족하는 도구변수를 그림으로 설명한 것이다. 도구변수를 이용하게 되면 독립변수와 결과변수 사이의 인과적 효과를 보다 정확하게 알 수 있다. 하지만 사회역학과 건강불평등 연구에서 사회경제적 위치에만 영향을 주는 적절한 도구변수를 찾는 일은 쉽지 않다.

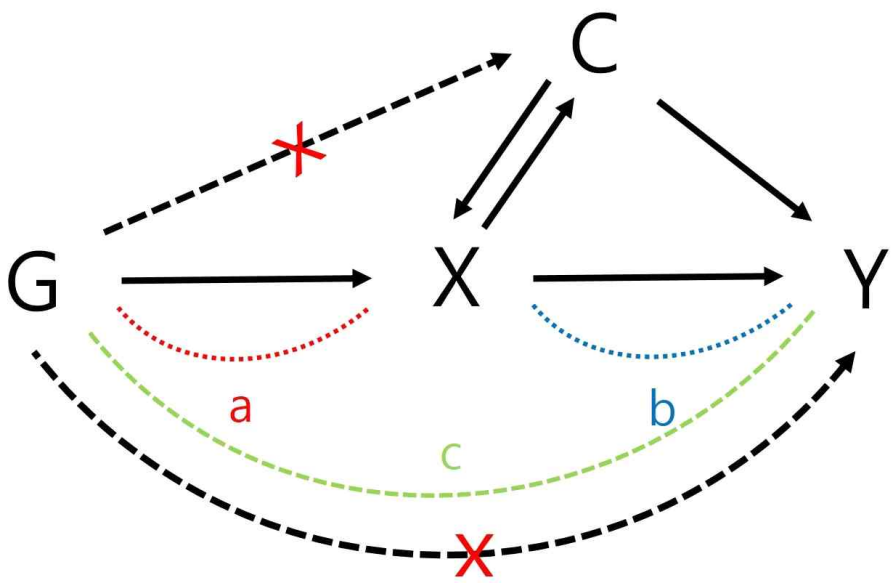

그림 3. 도구변수를 설명하는 그림

*G:유전변이(도구변수), X:독립변수, Y:결과변수, C:교란변수

1.2. 멘델 무작위 분석(Mendelian Randomization analysis, MR analysis)

멘델 무작위 분석은 독립변수와 관련된 단일염기서열 다형성(single nucleotide polymorphisms, 이하 SNP)를 도구 변수로 놓고, 독립변수와 결과변수 사이의 인과성을 분석하는 방법이다(Davey Smith, et

al., 2003; Greenland, 2000; Lawlor, et al., 2008). 유전자는 부모로부터 50% 확률로 무작위하게 내려오기 때문에, 독립변수와 연관이 있는 것으로 알려진 유전자 변이를 도구변수로 설정할 수 있다. 유전자가 무작위하게 전해지는 것을 RCT의 무작위 배정과 유사하다고 가정할 수 있다. 멘델 무작위 분석에서 도구변수로 사용하는 유전자 변이는 도구변수의 조건을 만족시켜야만 한다. RCT와 멘델 무작위 분석 방법의 비교를 [그림 4]에 도식화하였다.

멘델 무작위 분석은 도구변수 가정을 만족하는 적절한 유전자 변이를 도구변수로 이용하므로 인과성을 분석하는 좋은 연구 방법 중 하나다. 멘델 무작위 분석은 후향적 연구를 수행할 경우에 강점이 있으며, RCT에 비해 비용이 적게 든다는 장점이 있다(Burgess, et al., 2015). 후향적 연구에서는 결과가 이미 발생하고 나서 유전자 변이 여부를 측정하기 때문에, 결과가 원인에 영향을 주는 역인과성을 회피할 수 있다. 이는 환자-대조군 연구의 환경과 비슷하다. 그래서 지금까지 멘델 무작위 분석을 사용한 많은 연구들이 대부분 코호트 자료를 이용한 경우가 많았다.

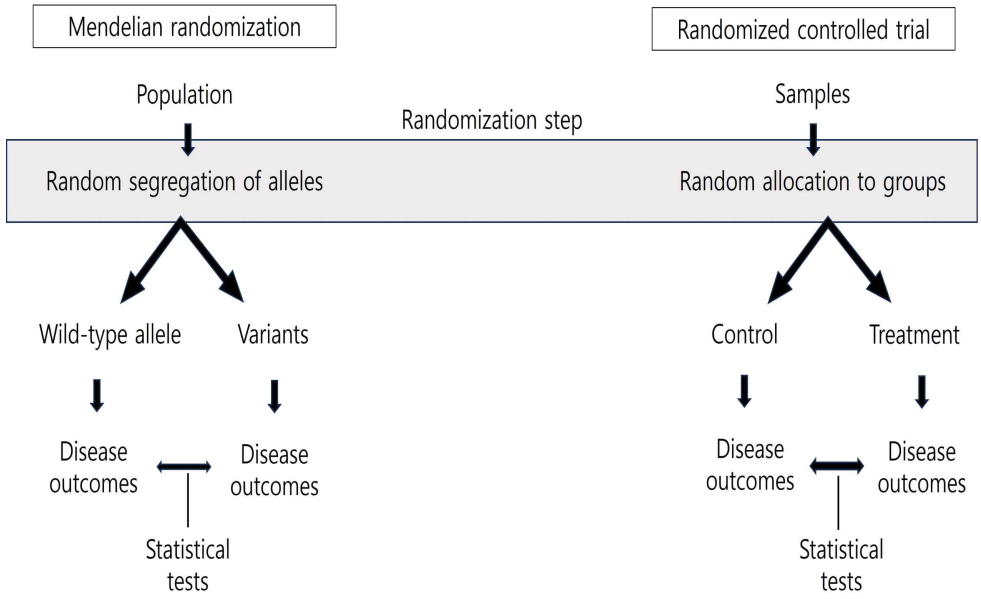

그림 4. RCT와 멘델 무작위 분석의 비교

자료원: Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. Nat Rev Methods Primers 2022.을 재구성하였음

1.2.1. 멘델 무작위 분석 연구 사례 – 해외

멘델 무작위 분석을 이용하여 교육이 건강에 미치는 인과적 영향을 분석한 연구들은 주로 해외에서 이루어졌다. 해외 연구 사례를 [표 1]에 제시하였다. Tilman 등은 교육 획득 연한이 3.6년 증가하면 관상동맥질환의 위험이 33% 감소한다고 하였다(Tilman, et al., 2017). Carter 등은 관찰연구 분석과 멘델 무작위 분석을 비교하였는

데, 관찰연구 분석에서는 3.6년의 교육 획득 연한이 증가하면 관상동맥질환 위험이 13% 감소하고, 멘델 무작위 분석에서는 37% 감소한다고 하였다(Carter, et al., 2019). Zhou 등은 3.6년의 교육 획득 연한이 증가하면 52%의 폐암 위험도가 감소한다고 밝혔다(Zhou, et al., 2019). 그 외에도 교육 연한이 증가할수록 비만(Bockerman, et al., 2017), 우울감(Viinikainen, et al., 2018), 심부전(Liao, et al., 2021), 외상후스트레스 장애(Polimanti, et al., 2019) 등의 위험이 감소하는 것으로 나타났다. Sanderson이 수행한 연구에서는 교육 연한이 1년 증가하면 현재 흡연자가 될 위험도가 4.1% 감소한다고 밝혔다(Sanderson, et al., 2019).

표 1. 멘델 무작위 분석을 통해 교육 수준이 건강에 미치는 영향을 살펴본 해외 연구 사례

연구	자료원	분석 표본 수
Lee JJ., et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature Genetics, 2018;50(8):1112-1121	1,000 Genome Project	1,131,881
Carter AR., et al. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. BMJ. 2019;365:l1855	UK Biobank	503,317
Tillmann T, et al. Education and coronary heart disease: mendelian randomisation study. BMJ. 2017;358:j3542	CARDIoGRAMplusC4D and SSGAC	543,733
Mountjoy E, et al. Education and myopia: assessing the direction of causality by mendelian randomisation. BMJ. 2018;361: k2022.	23andMe and SSGAC	67,798
Zhou, H, et al. Education and lung cancer: a Mendelian randomization study. International Journal of Epidemiology. 2019; 48(3):743-50	International Lung Cancer Consortium (ILCCO)	27,209
Sanderson E, et al. Mendelian randomisation analysis of the effect of educational attainment and cognitive ability on smoking behaviour. Nature Communications. 2019; 10(1):2949.	UK Biobank	120,050
Anderson EL, et al. Education, intelligence and Alzheimer's disease: evidence from a multivariable two-sample Mendelian randomization study. International Journal of Epidemiology. 2020;49(4):1163-1172.	International Genomics of Alzheimer's Project (IGAP)	54,132
Howe LD, et al. Effects of body mass index on relationship status, social contact and socio-economic position: Mendelian randomization and within-sibling study in UK Biobank. International Journal of	UK Biobank	378,244

Epidemiology. 2020;49(4):1173-1184.		
Viinikainen J, et al. Does education protect against depression? Evidence from the Young Finns Study using Mendelian randomization. Preventive Medicine. 2018;115:134-139.	Cardiovascular Risk in Young Finns Study	1,733
Liao L, et al. Education and heart failure: New insights from the atherosclerosis risk in communities study and mendelian randomization study. International Journal of Cardiology. 2021;324:115-121.	Atherosclerosis Risk in Communities (ARIC) study	12,315
Wen X, et al. Education and stroke: evidence from epidemiology and Mendelian randomization study. Scientific Reports. 2020;10(1):1-11.	Atherosclerosis Risk in Communities (ARIC) study	11,509
Böckerman P, et al. Does higher education protect against obesity? Evidence using Mendelian randomization. Preventive Medicine. 2017;101:195-198.	Cardiovascular Risk in Young Finns Study	2,011
Davies NM, et al. Multivariable two-sample Mendelian randomization estimates of the effects of intelligence and education on health. Elife. 2019;8:e43990.	UK Biobank	138,670
Kari JT, et al. Education leads to a more physically active lifestyle: Evidence based on Mendelian randomization. Scandinavian Journal of Medicine & Science in Sports. 2020;30(7):1194-1204.	Cardiovascular Risk in Young Finns Study	1,651

1.2.2. 멘델 무작위 분석 연구 사례 - 국내

멘델 무작위 분석을 이용한 연구는 우리나라에서도 이루어졌다. 우리나라에서 이루어진 연구 사례를 [표 2]에 제시하였다. 우리나라에서 이루어진 멘델 무작위 분석 연구들은 주로 특정 유전자와 그와 관련된 질환 사이의 연관성을 탐색하는 경우가 많았다. 혈청 감마글루타밀 전환효소 농도가 증가하면 제 2형 당뇨병의 위험이 증가하고(Lee, et al., 2016), 혈청 요산 농도가 증가하면 심방세동 위험이 증가하는(Hong, et al., 2020) 연구 결과가 있었다. 우리나라에서 수행된 기존 멘델 무작위 분석 연구들은 주로 한국인유전체역학조사사업(Korean Genome and Epidemiology Study, KoGES) 코호트 자료 중 지역사회 기반 코호트를 사용하였다. 지역사회 기반 코호트 참여자 중 유전정보를 가지고 있는 대상자 수는 최대 8,840명이며, 해외에서 수행되었던 5만 명 이상의 대상자 수에 비해 부족하다고 할 수 있다. 연구 대상자 수가 충분히 확보되지 못하면 멘델 무작위 분석 결과의 신뢰성이 감소하는 한계점을 지닌다.

표 2. 멘델 무작위 분석을 이용한 국내 연구 사례

연구	자료원	분석 표본 수
Lee YS, Cho Y, Burgess S, et al. Serum gamma-glutamyl transferase and risk of type 2 diabetes in the general Korean population: a Mendelian randomization study. Human Molecular Genetics. 2016;25(17):3877-86.	KoGES Ansung Ansan Study	7,640
Yang, SJ, Kwak SY, Jo G, Song TJ, Shin MJ. Serum metabolite profile associated with incident type 2 diabetes in Koreans: findings from the Korean Genome and Epidemiology Study. Scientific Reports. 2018;8(1):1-10.	KoGES Ansung Ansan Study	7,515
Cho, Y, Shin SY, Won S, Relton CL, Davey Smith G, Shin MJ. Alcohol intake and cardiovascular risk factors: a Mendelian randomisation study. Scientific Reports. 2015; 5(1):1-13.	KoGES Ansung Ansan Study	7,152
Hong M, Park JW, Yang PS, et al. A mendelian randomization analysis: The causal association between serum uric acid and atrial fibrillation. European Journal of Clinical Investigation. 2020; 50(10):e13300.	KoGES Ansung Ansan Study	4,166
Jeon C, Lee JY, Lee SJ, Jung KJ, Kimm H, Jee SH. Bilirubin and risk of ischemic heart disease in Korea: a two-sample Mendelian randomization study. Epidemiology and Health. 2019;41.	KoGES HEXA Study	33,598
Jee YH, Lee SJ, Jung KJ, Jee SH. Alcohol Intake and Serum Glucose Levels from the Perspective of a Mendelian Randomization Design: The KCPS-II Biobank. PLoS One. 2016;11(9):e0162930.	KCPS-II Biobank	4,367
Sull JW, Lee TY, Jee SH. Effect of smoking on the association of HHEX (rs5015480) with diabetes among Korean women and heavy smoking men. BMC Medical Genetics. 2018; 19(1):68.	KoGES Ansung Ansan Study	4,240
Sull, JW, Kim G, Jee SH. Association of MC4R (rs17782313) with diabetes and cardiovascular disease in Korean men and women. BMC Medical Genetics. 2020; 21(1):1-6.	KoGES Ansung Ansan Study	4,294
Lee SJ, Jee YH, Jung KJ, Hong S, Shin ES, Jee SH. Bilirubin and Stroke Risk Using a Mendelian	KoGES Ansung Ansan Study	4,793

Randomization Design. Stroke. 2017;48(5): 1154-60.		
Choi, Y, Lee SJ, Spiller W, Jung KJ, Lee JY, Kimm H, Back JH, Lee S, Jee SH. Causal associations between serum bilirubin levels and decreased stroke risk: a two-sample Mendelian randomization study. Arteriosclerosis, Thrombosis, and Vascular Biology. 2020;40(2):437-445.	KoGES HEXA Study	25,406
Sull, JW, Kim S, Jee SH. Effects of Obesity and Family History of Diabetes on the Association of CETP rs6499861 with HDL-C Level in Korean Populations. Journal of Lipid and Atherosclerosis. 2019;8(2):252.	KoGES Ansung Ansan Study	4,294

1.2.3. 측정 가능한 중요한 건강 지표 - 심혈관계질환

교육이 건강에 미치는 영향을 분석할 때, 공중보건 분야에서 중요한 건강 위험요인을 결과변수로 선택하는 것이 중요하다. 심혈관계질환은 사망원인에 기여하는 가장 중요한 위험요인 중 하나다(Naghavi, et al., 2017). 세계질병부담연구(Global burden of disease study)에 따르면 2016년 심혈관계질환으로 인한 전체 사망자 수는 17,646,600명으로 전체 사망에 기여하는 질병부담이 가장 큰 것으로 나타났다(Naghavi, et al., 2017). 우리나라 통계청 자료에서도 2018년 심혈관계질환으로 인한 사망률은 10만 명당 122.7명으로 암에 이어 2위를 차지했다(Kim, et al., 2019).

심혈관계질환 사망은 사회경제적 요인에 따라 불평등하게 존재한다(Kivimaki, et al., 2008). 기존의 연구는 대부분 관찰연구에 해당한다는 한계가 있다. 관찰연구를 통해 심혈관계질환 양상이 어떻게 나타나는지 알 수 있으나, 불평등 결과를 일으킨 원인을 인과적으로 밝혀내는 연구가 필요하다.

II. 연구 목적

1. 연구 목적

첫째, 교육 수준이 주요 심혈관계질환 위험요인에 미치는 인과적 영향을 규명한다.

둘째, 교육 수준이 심혈관계질환에 미치는 인과적 영향을 규명한다.

2. 연구의 세부 목표

첫째, 교육 수준을 예측하는 도구 변수로 사용할 유전 변이를 전장 유전체 연관성 분석(Genome-wide association study, 이하 GWAS) 분석을 이용하여 규명한다.

둘째, 주요 심혈관계질환 위험요인과 관련된 유전 변이를 GWAS 분석을 이용하여 규명한다.

셋째, 심혈관계질환을 예측하는 유전 변이를 GWAS 분석을 이용하여 규명한다.

넷째, GWAS 분석을 통하여 얻은 유전 변이를 활용하여 멘델 무작위 분석을 실시해, 교육 수준이 주요 심혈관계질환 위험요인과 심혈관계질환에 미치는 인과적 영향을 분석한다.

III. 연구 방법

1. 연구 자료

1.1. 자료원

이 연구는 질병관리청 국립보건연구원 산하 국립중앙인체자원은행에서 제공하는 한국인유전체역학조사사업(Korean Genome and Epidemiology Study, 이하 KoGES) 코호트 자료를 활용하였다. KoGES 코호트는 지역사회 기반 코호트, 농촌 코호트, 도시 코호트, 쌍둥이-가족 코호트 등으로 구성되어 있다. 각각의 코호트는 역학 정보와 유전 정보를 함께 제공하고 있으며, 본 연구는 지역사회 기반 코호트, 농촌 코호트, 도시 코호트를 사용하였다. 3개의 코호트에 대한 요약 정보를 [표 3]에 제시하였다.

1.1.1 지역사회 기반 코호트

지역사회 기반 코호트는 안성/안산 코호트로 잘 알려져 있으며, 중소도시 안산 및 농촌지역 안성에 거주하는 40-69세 남녀가 대상이다. 생활습관, 식이, 환경요인 등이 만성질환 발병에 미치는 영향을

조사하기 위해 코호트를 구축하였고, 2001년과 2002년에 기반 조사를 실시하였다. 2년 주기로 연구 대상자들을 추적 관찰하고 있으며, 2019년-2020년까지 9차 추적 조사가 완료되었다. 1기 기반 조사에서 10,030명이 참가했으며, 9차 추적에는 5,854명이 대상으로 남았다.

1.1.2. 농촌 코호트

농촌 코호트는 농촌 지역에 거주하는 40세 이상 남녀가 대상이며, 2004년부터 2016년까지 시행되었다. 심혈관계질환 발병과 관련된 유전적, 환경적 요인의 규명을 위해 구축되었으며, 조사 지역은 경기도 양평군, 경상북도 고령군, 전라남도 남원시, 강원도 원주시, 강원도 평창군, 인천광역시 강화군이다. 2004년 기반 조사에 28,337명이 참여했으며, 2016년 1,449명이 4차 추적이 완료되었다.

1.1.3. 도시 코호트

도시 기반 코호트는 2004년부터 2016년까지 시행되었으며, 전국 대도시, 중소도시 병원 건강검진센터 내원자 40-79세 남녀가 대상이다. 도시 코호트의 목적은 대도시, 중소도시 지역의 의료기관, 보건소와 보건의료원 등을 중심으로 한국인에게 흔한 만성질환의 환경적, 유전적 위험 요인 규명을 위한 유전체 연구 자료를 구축하는 것이다.

조사 지역은 서울특별시, 부산광역시, 등을 포함한 전국 6대 광역시와 경기도 안양시, 고양시, 성남시, 강원도 춘천시, 충청남도 천안시, 전라남도 화순군, 경상남도 창원시이다. 병원 건강검진센터의 성인병 검진과 연계하여 단기간에 대규모 대상자를 모집했으며 2004년 기반 조사에서 173,195명이 참가했다. 1차 추적은 2012년부터 실시하여 65,608명이 추적 대상자였다.

1.1.4. 자료의 질 관리

GWAS 분석을 위해 먼저 질 관리를 실시하였다. 질 관리에 대한 순서는 아래 [그림 5]에 제시하였다. 본 자료는 국립중앙인체자원은행에서 한국인칩을 사용하여 자체적으로 결측치에 대해 대체를 하여 제공되었다(Moon et al., 2019). 대체 과정은 1000 Genomes Phase 3 자료를 기준으로 하여 ShapeIT v2(Delaneau et al., 2011)), IMPUTE v2(Marchini et al., 2007)를 이용해 진행되었다. 대체 후에는 대체 질 관리 점수 0.4 이하와 마이너 대립유전자 빈도 1% 이하의 SNP은 제외하였다.

전체 SNP은 8,056,211개이며, 마이너 대립 유전자 빈도, 하디-와인버그 검정, LD 전정(prune) 과정을 거쳐 질 관리를 실시하였다. 마이너 유전자 빈도는 1% 미만인 유전변이는 분석에서 제외하였다. 이는 총 9,770개였으며 제외 후 SNP은 8,046,441개였다. 하디-와인버

그 검정은 멘델 유전 법칙을 만족하는 유전변이만 남기기 위해 실시하였다. P값이 5e-6 이하인 유전변이를 제외하였으며, 총72,098개를 제거하였다.

마지막으로 LD 전정을 실시하였다. 유전자는 부모로부터 독립적으로 유전받지만, 물리적으로 가까운 거리에 있을수록 함께 유전되는 특성이 있기 때문에 가까운 거리에 있는 유전변이에 영향을 받는 SNP들을 제거해야만 한다. LD 전정으로 7,135,046개의 SNP을 제거하여 최종적으로 839,297개 SNP이 남았다.

표 3. KoGES 코호트에 대한 요약표

	지역사회 기반 코호트	농촌 기반 코호트	도시 기반 코호트
사업명	지역사회(안산, 안성) 코호트	농촌 코호트	도시 코호트
사업명(영문)	KoGES Ansan and Ansung study	KoGES CAVAS (Cardiovascular Disease Association) study	KoGES HEXA (Health Examinees) study
조사지역	안산, 안성	양평군, 고령군, 남원시, 원주시, 평창군, 강화군	전국 광역시, 안양시, 고양시, 성남시, 춘천시, 천안시, 화순군, 창원시
조사기간	2001 – 2020년 (기반 – 9차 추적)	2005 – 2016년 (기반 – 4차 추적)	2004 – 2016년 (기반 – 1차 추적)
대상집단	중소도시 안산 및 농촌 지역 안성에 거주하는 40-69세 남녀	농촌 지역에 거주하는 40-69세 남녀	전국 대도시, 중소도시 병원 건강검진센터 내원자 40-79세 남녀
대상자 수(기반)	10,030 명	28,337 명	173,195 명
대상자 수(한국인칩)	5,493 명	8,105 명	58,696 명

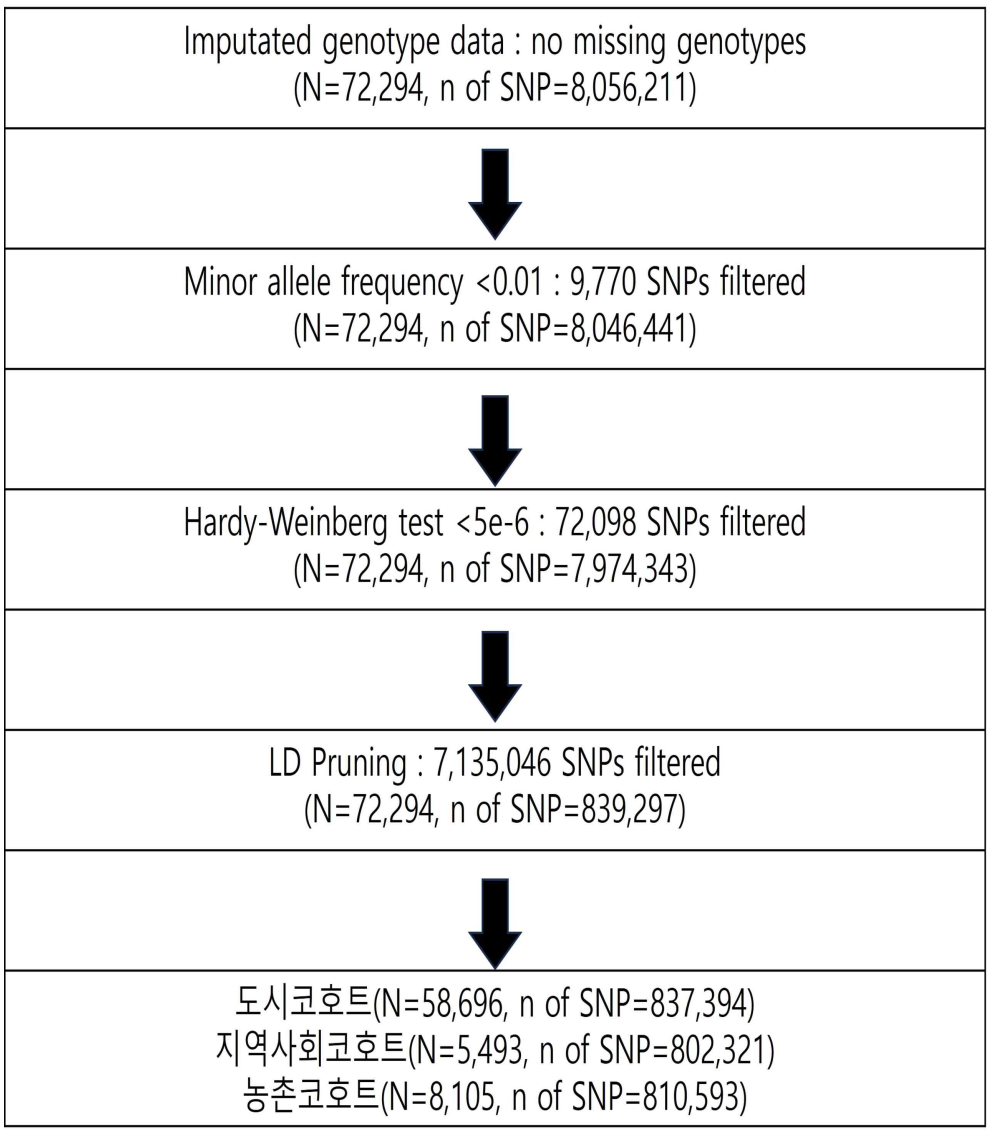

그림 5. GWAS 질 관리(quality control) 순서도

1.2 연구 대상자 선정 기준

지역사회 기반 코호트, 농촌 코호트, 도시 코호트는 각각 기반 조사 시점에서 10,030명, 28,337명, 173,195명이 참여했다. 연구 대상자는 기반 조사 시점의 참여자 중에서 유전 정보 제공에 동의한 사람으로 한정했다. 유전 정보를 얻기 위해 관련 기관에서 연구 대상자의 사전 동의를 받았으며, 연구 대상자는 의료 기관에서 채혈을 통해 유전 정보를 제공하였다. 연구 대상자의 유전 정보는 염기서열 해석을 거쳐 얻게 되며, 염기서열 분석에 반도체 칩 크기의 키트가 이용된다. 기존에 알려진 분석 키트는 Affy, Exome, 한국인칩(K-Chip) 등이 있는데 Affy, Exome 등은 서양인을 중심으로 설계되어 있어 한국인에 대한 유전체 대표성이 낮다는 단점이 있다. 국립중앙인체자원은행에서는 이러한 문제를 극복하기 위해 한국인의 유전체 대표성을 갖춘 한국인칩을 개발했다.

각 키트에 따라 연구 대상자의 수가 다르며, 본 연구는 한국인의 특성에 맞게 개발된 한국인칩을 사용한 유전 정보를 이용하였다. 기반 조사 시점을 기준으로 할 때, 한국인칩 유전 정보 제공에 동의한 인원은 지역사회 기반 코호트 5,493명, 농촌 코호트 8,105명, 도시 코호트 58,696명이었다. 연구 대상자의 유전 정보와 역학 정보를 함께 제공받았다. 역학 정보는 설문조사와 임상생물학적 검사를 통해 구성되었으며 개인의 건강 행태, 사회경제적 위치, 질병 과거력과 혈

압, 혈당, 콜레스테롤 등의 같은 임상생물학적검사 결과가 포함되었다.

2. 분석 방법

2.1 연구의 주요 지표

2.1.1. 독립변수 - 교육

독립변수는 교육 수준을 사용하였다. 건강불평등 연구에서 사용되는 사회경제적 위치 지표에는 대표적으로 교육, 소득, 직업 등이 있다. 본 연구에서 교육을 사회경제적 위치 지표로 삼은 이유는 역인과성을 최대한 배제하기 위해서다. 소득과 직업은 개인의 상황에 따라 변할 수 있지만, 교육은 개인의 사회경제적 상황과 관계없이 일정하게 유지된다는 장점이 있다.

교육 수준은 개인이 습득한 지식과 인지 기능을 측정하는 지표다. 교육 수준은 초등학교, 고등학교, 대학 등과 같은 범주형 변수로 측정할 수도 있고, 교육 연한 같은 연속형 변수로 측정할 수도 있다. 우리나라에서 교육 수준을 사회경제적 위치 지표로 사용한 관찰 연구들의 경우, 교육 연한보다는 최종 학력이라는 범주형 변수를 사용하였

다.

교육은 사회경제적 위치 지표로서 다음과 같은 특성을 가진다. 첫째, 교육은 생애주기에서 성인기 초반에 끝난다. 고등학교만 졸업한다면 10대 후반에 교육은 끝나며, 대학교 졸업의 경우에는 20대 중반에 끝나게 된다. 둘째, 성인기에 교육이 끝나게 되면 건강 상태가 교육에 영향을 주지 않는다. 이는 직업, 소득에 비해 역인과성이 낮다는 장점이 있다. 셋째, 교육 수준은 부모의 사회경제적 위치에 의해 영향을 받게 되며, 교육 수준은 해당 개인의 아동기 또는 청소년기의 사회경제적 위치 지표로도 사용될 수 있다. 또한 교육 수준은 장래 직업과 소득의 강력한 결정 요인이다. 넷째, KoGES 코호트는 지역별 특성을 고려할 필요가 있다. 지역사회 코호트와 농촌 코호트는 인구가 적은 농촌 지역과 중소 도시에서 연구 대상자를 모집했고, 도시 코호트는 인구가 많은 광역시, 도시 등에서 연구 대상자를 모집했다. 따라서 농촌 지역보다 도시 지역에서 고학력자가 더 많이 있으며, 실제 분석 결과 도시 코호트에서 대학 이상인 대상자 비율이 지역사회 코호트, 농촌 코호트에서 대학 이상의 대상자 비율보다 더 높았다.

교육 수준의 측정은 교육 연한으로 실시하였다. 연구대상자는 설문조사에서 “귀하는 어디까지 교육을 받으셨습니까”라는 질문에 “무학, 초등학교, 중학교, 고등학교, 전문대, 대학교, 대학원 이상” 중에서 선택해야 한다. 따라서 교육 수준을 교육 연한으로 변환하기 위해서 각각 보기에 대해 교육 연한을 부여했다. 무학은 교육 연한 0년, 초등

학교 중퇴는 3년, 초등학교 졸업 및 중학교 중퇴는 6년, 중학교 졸업 및 고등학교 중퇴는 9년, 고등학교 졸업은 12년, 전문대 졸업은 14년, 대학교 졸업은 16년, 대학원 이상은 18년을 부여했다.

2.1.2. 결과변수(1) - 심혈관계질환

심혈관계질환은 심장과 주요 동맥에 발생하는 질환으로 관상동맥질환, 협심증, 심근경색, 뇌혈관질환 등이 포함된다. 심혈관계질환에 대한 코호트별 정의는 [표 4]에 제시하였다. 3개 코호트별 심혈관계질환에 포함되는 질환이 차이가 있다. 지역사회 코호트의 심혈관계질환은 심근경색증, 관상동맥질환, 뇌혈관질환(뇌졸중)을 포함하며 해당 질환을 진단받고 치료 중인 사람을 유질환자로 하였다. 농촌 코호트의 심혈관계질환은 협심증 또는 심근경색증, 뇌혈관질환(뇌졸중)을 포함하며 해당 질환을 진단받고 치료 중인 사람을 유질환자로 하였다. 도시 코호트의 심혈관계질환은 협심증/심근경색증, 뇌졸중을 포함하며 해당 질환을 진단받고 치료 중인 사람을 유질환자로 하였다. 심혈관계질환 유질환자는 각 코호트의 심혈관계질환에 포함되는 개별 질환들 중 하나 이상에 해당되면 유질환자로 정의하였다. 심혈관계질환 대상자는 기반 조사에서 이미 진단 받은 사람들만 포함하였으며, 새로 발생한 인원은 제외하였다.

2.1.3. 결과변수(2) – 주요 심혈관계질환 위험요인

심혈관계질환을 일으키는 주요 위험요인은 흡연, 고혈압, 당뇨병, 고지질혈증 등이 있다. 심혈관계질환 위험요인에 대한 코호트별 정의는 [표 4]에 제시하였다. 현재 흡연자는 “지금까지 살아오면서 담배를 20갑(400개비) 이상 피웠습니까”라는 질문에 “예, 가끔 피운다”와 “예, 현재 피운다”로 응답한 사람으로 정의하였다. 고혈압은 해당 질환을 진단받고 현재 치료 중인 사람과 미치료 중인 상태여도 수축기 혈압이 140 mmHg 이상이거나 이완기 혈압이 90mmHg 이상인 사람을 유질환자로 정의하였다. 당뇨병은 해당 질환을 진단받고 현재 치료 중인 사람과 미치료 중인 상태여도 공복혈당이 126 mg/dL 이상인 사람을 유질환자로 정의하였다. 고지혈증은 해당 질환을 진단받고 현재 치료 중인 사람과 미치료 중인 상태여도 총콜레스테롤이 200 mg/dL 이상인 사람을 유질환자로 정의하였다. 비만은 국내 기준과 세계보건기구의 기준이 다르기 때문에, 국내 기준과 세계보건기구 기준에 맞게 각각 나누어 정의하였다. 국내 기준에 의한 비만의 정의는 체질량지수 25 이상인 경우이며 세계보건기구에 의한 비만의 정의는 체질량지수 30 이상인 경우다.

표 4. 본 연구에 사용한 변수에 대한 설문 정보

코호트	변수		설문 내용	설문 답가지
지역사회 코호트	교육		귀하는 어디까지 교육을 받으셨습니까?(최종학력)	1. 초등학교 이하 2. 중학교 3. 고등학교 4. 전문대 5. 대학교 6. 대학원 이상
	심혈관계질환	심근경색증	현재 심근경색증을 치료 받고 있습니까?	1. 아니오 2. 예
		관상동맥질환	현재 관상동맥질환을 치료 받고 있습니까?	1. 아니오 2. 예
		뇌혈관질환(뇌졸중)	현재 뇌혈관질환(뇌졸중)을 치료 받고 있습니까?	1. 아니오 2. 예
	흡연		현재 담배를 피우고 계십니까?	1. 전혀 피운 적 없다. 2. 흡연경력은 있으나 현재 안 피운다. 3. 가끔씩 피운다. 4. 습관적으로 계속 피운다
	고혈압		현재 고혈압을 치료 받고 있습니까? 수축기 혈압 140 mmHg 이상 또는 이완기 혈압 90mmHg 이상	1. 아니오 2. 예

농촌 코호트	당뇨병		현재 당뇨병을 치료 받고 있습니까?	1. 아니오 2. 예
			공복혈당 126mg/dL 이상	
	고지혈증		현재 고지혈증을 치료 받고 있습니까?	1. 아니오 2. 예
			총콜레스테롤 200mg/dL 이상	
	비만		BMI 25 이상(국내)	
			BMI 30 이상(세계보건기구)	
	교육		학교는 어디까지 다니셨습니까?	0. 학교에 다니지 않았다 1. 초등학교 중퇴 2. 초등학교 졸업, 중학교 중퇴 3. 중학교 졸업, 고등학교 중퇴 4. 고등학교 졸업 5. 전문대 졸업 6. 대학교 중퇴 7. 대학교 졸업 8. 대학원 이상
	심혈관계질환	협심증 또는 심근경색증	협심증 또는 심근경색증을 진단 후 치료 받고 있습니까?	1. 아니오 2. 예
		뇌혈관질환(뇌졸중)	뇌혈관질환(뇌졸중)을 진단 후 치료 받고 있습니까?	1. 아니오

			2. 예
	흡연	지금까지 담배를 통틀어 20갑(400개비) 이상 피웠습니까?	1. 아니오 2. 예(과거 흡연) 3. 예(현재 흡연)
	고혈압	고혈압을 진단 후 치료 받고 있습니까?	1. 아니오 2. 예
		수축기 혈압 140 mmHg 이상 또는 이완기 혈압 90mmHg 이상	
	당뇨병	당뇨병을 진단 후 치료 받고 있습니까?	1. 아니오 2. 예
		공복혈당 126mg/dL 이상	
	고지혈증	고지혈증을 진단 후 치료 받고 있습니까?	1. 아니오 2. 예
		총콜레스테롤 200mg/dL 이상	
	비만	BMI 25 이상(국내)	
		BMI 30 이상(세계보건기구)	
도시 코호트	교육	학교는 어디까지 다니셨습니까?	0. 학교에 다니지 않았다 1. 초등학교 중퇴 2. 초등학교 졸업, 중학교 중퇴

			3. 중학교 졸업, 고등학교 중퇴 4. 고등학교 졸업 5. 전문대 졸업 6. 대학교 중퇴 7. 대학교 졸업 8. 대학원 이상
심혈관계질환	협심증/심근경색증	협심증/심근경색증을 진단 후 치료 받고 있습니까?	1. 아니오 2. 예
	뇌졸중	뇌졸중을 진단 후 치료 받고 있습니까?	1. 아니오 2. 예
흡연		지금까지 담배를 통틀어 20갑(400개비) 이상 피웠습니까?	1. 아니오 2. 예(과거 흡연) 3. 예(현재 흡연)
고혈압		고혈압을 진단 후 치료 받고 있습니까? 수축기 혈압 140 mmHg 이상 또는 이완기 혈압 90mmHg 이상	1. 아니오 2. 예
당뇨병		당뇨병을 진단 후 치료 받고 있습니까? 공복혈당 126mg/dL 이상	1. 아니오 2. 예
고지혈증		고지혈증을 진단 후 치료 받고 있습니까? 총콜레스테롤 200mg/dL 이상	1. 아니오 2. 예

비만	BMI 25 이상(국내) BMI 30 이상(세계보건기구)

2.2. 분석

본 연구에서 다루게 될 주요 분석 방법은 크게 2가지로 구분할 수 있다. 첫째, 교육 수준을 예측하는 유전 변이를 GWAS 분석을 통해 찾고, 둘째, 교육 수준을 예측하는 유전 변이를 도구변수로 놓고 멘델 무작위 분석을 실시한다. [표 5]를 통해 이 연구의 설계와 분석 방법을 나타냈다.

표 5. 본 연구의 수행 절차

순서	수행내용
1단계	KoGES 코호트 각각에서 로지스틱 회귀분석을 통해 심혈관계질환 오즈비 산출
2단계	교육 수준, 심혈관계질환, 심혈관계질환 위험요인에 대한 GWAS 분석 실시
2-1단계	도시 코호트, 지역사회,농촌 코호트에서 각각 교육에 대한 GWAS 분석 실시하여 replication 확인
3단계	심혈관계질환, 심혈관계질환 위험요인에 대한 멘델 무작위 분석
3-1단계	1-표본 멘델 무작위 분석 - KoGES 코호트 통합(72,294명)
3-2단계	2-표본 멘델 무작위 분석 - KoGES 코호트(도시 코호트 58,696명, 지역사회·농촌 코호트 13,598명)
4단계	관찰 연구 결과와 멘델 무작위 분석 결과를 비교

2.2.1. Genome-wide association study (GWAS)

첫 번째 연구는 GWAS 분석을 통해 교육 수준을 예측하는 유전 변이를 찾는 것이다. GWAS 분석은 특정 형질이 어느 유전 변이와 연관되어 있는지 분석하는 연구 방법으로, 일반적으로 SNP와 주요 질환이나 위험요인 사이의 연관성에 초점을 둔다.

GWAS 분석의 일반적인 접근 방식은 환자-대조군 연구 설정이다. 연구 대상 집단을 대조군과 실험군으로 나누게 되는데, 우리가 찾고자 하는 표현형을 가진 집단을 실험군으로 설정한다. 실험군과 대조군 집단에 속한 모든 대상자는 일반적으로 알려진 공통 SNP를 갖고 있다. 정확한 SNP의 수는 염기서열 분석 방법에 따라 다르지만, 보통 100만 개 혹은 그 이상이다. 각 SNP에 대해 대립 유전자 빈도가 실험군과 대조군 간에 유의하게 차이를 보이는지 조사한다.

GWAS 분석을 수행하면 개별 SNP마다 실험군으로 설정한 표현형에 대한 효과 크기를 산출할 수 있다. 환자-대조군 설정이기 때문에, 특정 대립 유전자를 가진 개인이 발병할 확률과 동일한 대립 유전자를 갖지 않은 개인이 발병할 확률의 오즈비를 계산할 수 있다.

2.2.2. 교육 수준을 예측하는 SNP 찾기

위의 질 관리를 거친 839,297개 SNP을 이용해 GWAS 분석을 실시하였다. GWAS 분석은 SNP의 재현가능성을 위해 도시 코호트에서 한번 실시하고, 농촌, 지역사회 코호트를 합친 자료에서 한번 실시하여 두 결과를 비교하여 중복되는 SNP만 멘델 무작위 분석에 이용하였다. 이 때 p값을 5e-5부터 5e-7까지 다양하게 한계선을 설정하여 어떠한 SNP들이 남는지 확인하였다. GWAS를 이용해 교육에 대한 SNP를 추출하는 과정을 [그림 6]으로 나타내었다.

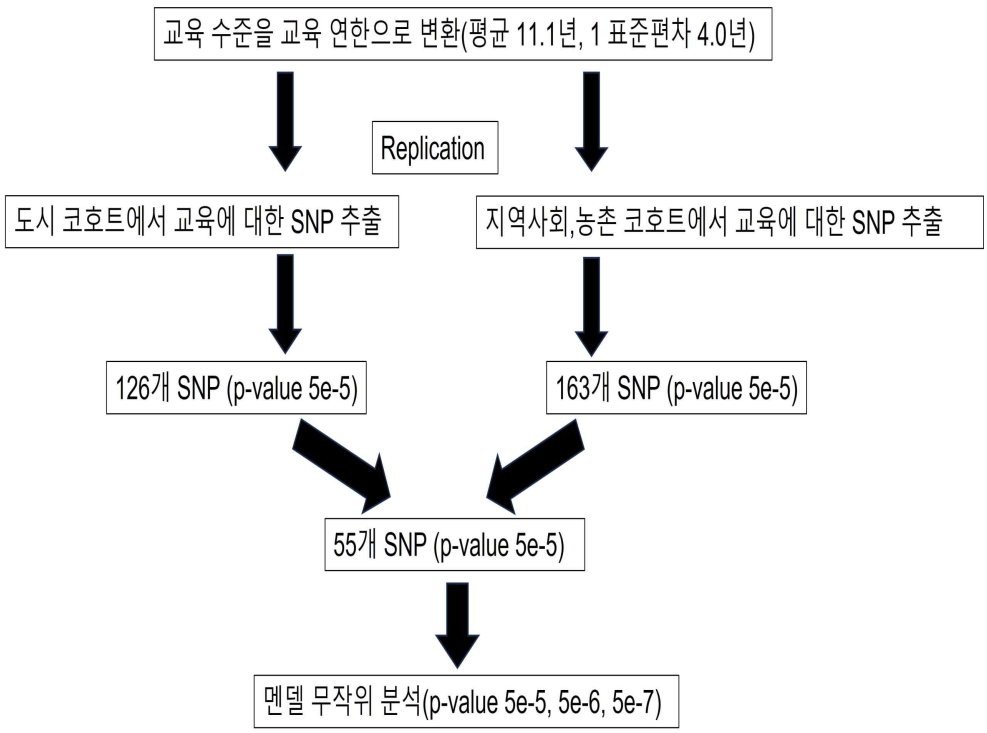

그림 6. GWAS를 이용해 교육에 대한 SNP 추출하는 과정

SNP가 도구변수 기준을 만족하는지 알아보기 위해 이질성(heterogeneity)와 다형성(pleiotropy)의 검정이 필요하다(Tam, et al., 2019; Uffelmann, et al., 2021). 특정 대립 유전자가 표현형을 나타내지 않고, 같은 표현형을 여러 다른 대립 유전자가 나타낼 경우에 이질성(heterogeneity)이 있다고 한다. 예를 들면, 특정 염색체 위치에서 G/A 변이가 있을 때만 표현형이 나타나는게 아니라 G/C, G/T 변이가 있어도 표현형이 나타나는 경우를 말한다. 유전 변이가 이질성을 가지고 있으면 도구변수로서 가지는 설명력이 감소하게 된다.

다형성(pleiotropy)은 하나의 SNP가 하나의 표현형과 연관되어 있지 않고, 결과변수에 영향을 주는 다른 표현형과도 연관이 있는 경우를 말한다. [그림 7]에 나와 있듯이 유전변이 G는 오직 독립변수인 교육(그림에서 X)과 연관성이 있어야 하며, 그 외 교란변수 C나 결과변수 Y와도 연관성이 있으면 다형성(pleiotropy)이 존재하기 때문에 유전변이 G의 설명력이 감소하게 된다. 다형성 검정을 통과하지 못하면 유전변이는 도구변수의 조건을 만족시키지 못하게 된다.

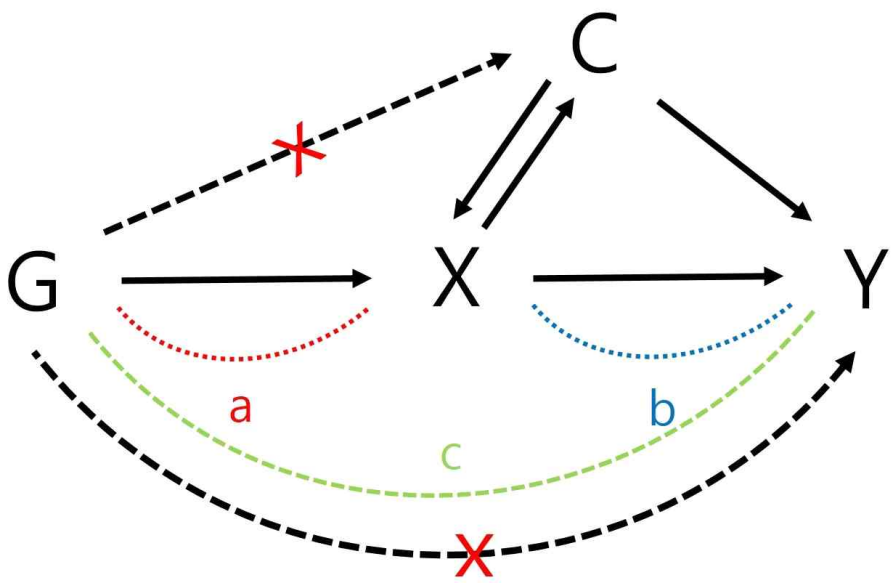

그림 7. 도구변수 조건을 만족하는 SNP 찾기 위한 인과적 경로

유전 변이를 담고 있는 연구 자료에는 결측된 유전 변이도 포함하고 있다. 결측된 부분을 그대로 제외하고 분석할 수도 있지만, 기존의 GWAS 연구 결과, 인종, 성별, 연령 등을 고려하여 결측된 유전 변이를 대체(imputation)하는 것이 중요하다. 이 연구에 사용된 KoGES 코호트의 한국인칩 유전 정보는 사전에 국립중앙인체자원은행에서 결측치에 대해 대체를 완료한 상태로 제공하였기 때문에 따로 대체 과정을 수행하지 않았다.

대립 유전자 빈도(allele frequency)는 모집단 내 특정한 염색체 위치에서 해당 대립 유전자의 상대적인 빈도수를 의미한다(Tam, et al., 2019; Uffelmann, et al., 2021). 분모에는 대립 유전자의 전체 수를 두고, 분자에는 해당 대립유전자의 수를 넣으면 된다. 즉, 전체 대립 유전자 수에서 특정한 대립 유전자를 가지는 염색체의 비율을 의미한다. 여기서 주의해야 할 점은 대립 유전자 빈도를 구할 때 표현형을 기준으로 하는 것이 아니다. 대립 유전자는 minor와 major 대립 유전자로 나눌 수 있는데 일반적으로 계산하는 빈도는 minor 대립 유전자 빈도(minor allele frequency, MAF)이다. GWAS 분석 수행 시 MAF가 0.01 미만인 SNP는 분석에서 제외하는 것이 일반적이다.

GWAS 분석에서 데이터 질 관리 방법 중 하나는 연관불균형(linkage disequilibrium, 이하 LD)을 이용하는 것이다(Tam, et al., 2019; Uffelmann, et al., 2021). SNP는 염색체의 여러 부위 중 하나에 속해있기 때문에 주위의 SNP나 염색체 덩어리(locus)의 영향을 받게 된다. 유전자는 부모로부터 독립적으로 전해지나, 예측빈도 이상의 고빈도로 염색체 내에서 가까이 뭉쳐있는 하플로타입(haployrpe, haploid와 genotype의 합성어)으로 유전되는 성질이 있다. 하나의 SNP가 표현형을 설명하더라도 주위의 SNP나 좌위(locus)와 더 강하게 연관되어 있다면 SNP가 표현형을 설명하는 힘이 떨어질 수밖에

없다. 따라서 근처의 SNP나 좌위에 영향을 받지 않는 SNP를 LD 상태에 있다고 말한다. SNP가 주위의 영향을 받는 정도를 R제곱으로 나타낼 수 있으며 이 값이 0.2를 넘어가면 해당 SNP는 배제한다.

데이터 정리 기준을 통과한 SNP만 남겨두고 이 SNP들이 얼마나 표현형을 잘 설명하는지 GWAS 분석을 수행하게 된다. SNP와 표현형 사이에 공변량을 넣어 회귀분석을 실시해 설명력을 구하게 되는데 아래 수식과 같다.

$$Y_{phenotype} = \beta_1 X_{SNP} + \beta_2 X_{SEX} + \beta_3 X_{AGE} + \epsilon$$

Y 값은 우리가 원하는 표현형을 넣어주면 된다. 이 연구에서는 교육 수준, 주요 심혈관계질환 위험요인, 심혈관계질환이 표현형으로 활용한다. 성과 연령은 보정을 위해 사용하였고, e는 오차항이다. 모든 유전 변이에 대해 회귀분석을 실시하며, Beta1 계수가 개별 SNP들의 표현형에 대한 효과 크기가 된다.

GWAS 분석을 통해 SNP 효과 크기와 함께 p-value를 구할 수 있다. SNP 개수는 수십만 개이므로 일반적인 p-value 0.05를 기준으로 하지 않는다. GWAS 분석에서 사용하는 p-value는 5×10^{-6} 또는 5×10^{-8}인데 이 연구에서는 p-value를 5e-5에서 5e-7까지 다양하게 설정하였다.

이렇게 추출한 SNP은 맨하탄 분포도(Manhattan plot)를 이용해 그림으로 나타낼 수 있다. 맨하탄 분포도는 가로축이 염색체 번호, 세로축은 각각 SNP의 p-value에 음의 로그를 취한 값이다. GWAS를 통해 산출된 SNP을 맨하탄 분포도로 나타내면 -log(p-value)가 8 이상인 SNP들이 한줄로 쭉 늘어서는 모양을 나타낸다. 이 모양이 맨하탄의 마천루와 비슷하다고 하여 맨하탄 분포도라고 명명한다.

2.2.3. 멘델 무작위 분석

멘델 무작위 분석은 유전 변이를 도구변수로 두고 독립변수가 결과변수에 미치는 인과적 영향을 측정하는 방법이다. [그림 8]를 보면, 유전 변이 G가 독립변수에 미치는 영향력을 a라고 할 수 있고, 독립변수가 결과변수에 미치는 영향력은 b이다. 결과변수에 직접적으로 영향을 주는 유전 변이가 있으면 그 영향력은 c이다. 영향력 c는 a와 b의 곱으로 구할 수 있다. 우리가 구하고자 하는 독립변수가 결과변수에 미치는 영향력 b는 c를 a로 나누어주면 구할 수 있다. 이 과정을 아래 수식과 그림으로 설명했다.

$$\frac{G-Y\ \ Association}{G-X\ \ Association} = \frac{c}{a} = \frac{a \times b}{a} = b$$

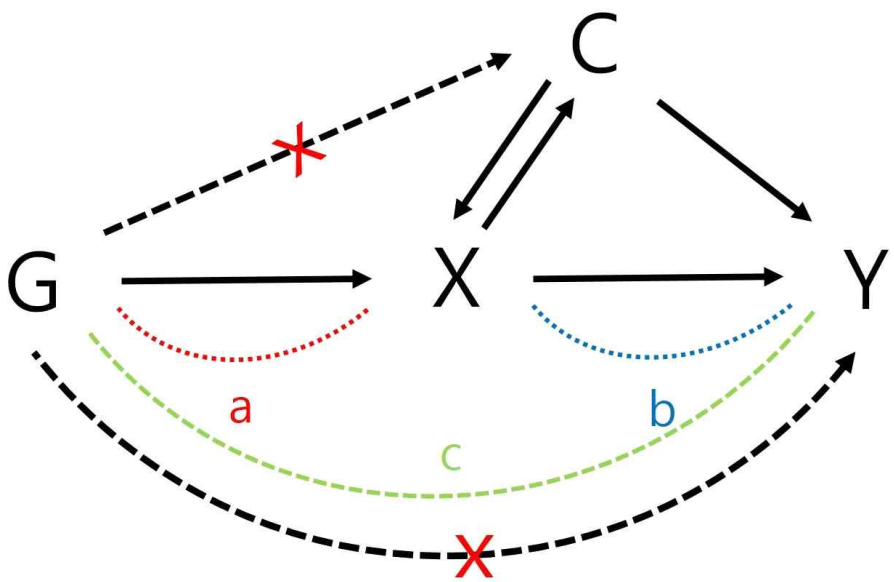

그림 8. 멘델 무작위 분석의 과정을 설명한 그림

교육 수준을 예측하는 유전 변이를 G1이라고 하면, G1은 영향력 a를 갖게 된다. 만약 G1이 교란변수 C와 결과변수 Y와 연관성이 있다면 이것은 다형성(pleiotropy)을 나타낸 것으로 도구변수 가정에 위배된다. 마찬가지로 주요 심혈관계질환 위험요인과 심혈관계질환을 예측하는 유전 변이를 G2라고 하면, G2는 영향력 c를 갖게 되는데 G1과 마찬가지로 다른 교란변수와 연관성이 없어야 한다.

위 수식과 그림을 통해, 멘델 무작위 분석에 필요한 것은 a, c라는 사실을 알 수 있다. G-X 영향력 a와 G-Y 영향력 c는 하나의 대상 집단에서 구할 수도 있고, 각각 다른 대상 집단에서 구할 수도 있다. 따라서 a, c를 동일한 대상 집단으로부터 계산하여 멘델 무작위 분석을 하게 되면 이를 1-표본 멘델 무작위 분석이라고 한다. 만약 a의 값을 A 집단에서 계산하고, c의 값은 다른 B 집단 혹은 연구에서 수행한 결과를 가져와서 멘델 무작위 분석을 수행하는 것을 2-표본 멘델 무작위 분석이라고 한다.

이 연구에서는 1-표본 멘델 무작위 분석과 2-표본 멘델 무작위 분석 모두 실시하였다. 1-표본 멘델 무작위 분석은 지역사회 기반 코호트, 농촌 코호트, 도시 코호트 자료를 모두 하나로 합쳐서 통합된 코호트 자료에서 교육 수준과 심혈관계질환에 대한 영향력을 계산하였다. 1-표본 멘델 무작위 분석을 수행한 이유는 최대한 표본 수를 확보하기 위해서다. 현재 우리나라에서 사용 가능한 유전 정보 자료 중에서는 가장 많이 표본 수를 확보할 수 있는 방법이다.

2-표본 멘델 무작위 분석은 KoGES 코호트를 두 개로 나누어 진행했다. 도시 코호트에서 교육 수준을 예측하는 SNP를 추출했으며, 지역사회 기반 코호트와 농촌 코호트를 합친 자료에서 심혈관계질환

위험요인과 심혈관계질환에 대한 SNP를 분석했다.

2.2.4 멘델 무작위 분석의 민감도

일반적으로 도구변수를 이용한 도구변수 분석 방법은 2단계 최소 제곱법(two stage least squares, 2SLS)을 사용하지만, 멘델 무작위 분석에서는 기본적으로 역분산가중치(inverse variance weighted, 이하 IVW)를 구하게 된다. 앞서 GWAS 분석을 통해 각각 SNP의 회귀계수를 구할 수 있었고, 이 회귀계수가 SNP의 효과 크기를 나타냈다. 유의미한 p-value를 가지는 개별 SNP의 효과 크기를 하나로 합쳐서 통합 효과 크기를 구할 수 있는데, 이때 IVW 방법이 사용된다. 문자 그대로 분산의 역수를 가중치로 둔다는 뜻인데 SNP의 회귀계수에서 표준오차의 제곱의 역수가 가중치가 된다. 도구변수 가정에 의해 다형성(pleiotropy)이 없다고 하면 잔차(intercept)가 0인 회귀식을 구할 수 있다.

멘델 무작위 분석에서 사용하는 민감도 분석에는 여러 종류가 있지만, 일반적으로 사용하는 방법은 MR-Egger 회귀분석과 Median 멘델 무작위 분석이다. MR-Egger 회귀분석은 도구변수로 사용된 SNP가 기본 가정을 만족하지 않고, 다형성(pleiotropy)이 존재할 때

사용한다. IVW의 회귀식에 다형성 잔차항(pleiotropic intercept)을 추가하게 되는데 IVW 직선과 MR-Egger 직선이 최대한 비슷하게 일치할수록 다형성(pleiotropy)이 적다는 뜻으로 해석할 수 있다. Median 멘델 무작위 분석은 도구변수로 사용된 SNP 중에 기본 가정을 만족하는 유효한 SNP가 적어도 50%는 된다고 놓고 수행하는 방법이다. 이 연구에서는 IVW와 함께 MR-Egger, Median 멘델 무작위 분석의 결과를 함께 제시하였다.

2.2.4. 분석 프로그램

본 연구는 교육 수준, 심혈관계질환 위험요인, 심혈관계질환에 대한 GWAS 분석을 위해 PLINK (1.9 beta, 2.0 alpha)를 이용하였다. GWAS 분석을 통해 얻은 유전 변이 정보를 이용해 멘델 무작위 분석을 시행할 때 R 프로그램을 사용했으며 버전은 4.2.0이었다. STATA 17.0을 이용하여 코호트 역학 자료에서 Cox 비례위험모형을 통해 심혈관계질환 위험도를 계산하였다. 또한 연구 결과의 해석을 돕기 위한 표와 그림은 R 프로그램과 관련 패키지들을 활용하였다.

3. 연구 윤리

이 연구에서 사용된 지역사회 기반 코호트, 도시 기반 코호트, 농촌 기반 코호트의 유전 정보와 역학 정보는 개인 정보를 식별할 수 없도록 자료 제공 전 단계에서 비공개 처리되었다. 이 자료는 국립중앙인체자원은행에서 정책 및 학술 연구용으로 제공하고 있으며, 사전 연구 계획서 신청 및 심사 승인을 거쳐야만 접근이 가능하다. 따라서, 본 연구는 절차에 따라 서울대학교의과대학/서울대학교병원 의학연구윤리심의위원회(IRB no. E-2211-012-1373)로부터 연구계획 승인을 받았으며, 국립중앙인체자원은행 분양심의위원회의 분양 심사(NBK-2021-045)를 통과하였다.

IV. 연구 결과

1. 연구 대상자의 특성

1.1. 연구 대상자의 일반적 특성

연구 대상자의 일반적 특성을 [표 6, 7, 8]에 제시하였다. 이 연구의 대상자 선정은 조사 기반 시점을 기준으로 하였으며, 각각 코호트 대상자의 특성을 살펴보았다. 지역사회 기반 코호트 참여자 중 한국인칩 유전 정보 제공자는 5,493명이었으며, 남성은 47.6%인 2,616명, 여성은 52.4%인 2,877명이었다. 연령군은 10세 단위로 산출하였으며 40-49세가 전체 집단에서 차지하는 비율이 남녀 모두에서 약 50%로 높았다.

교육 수준을 살펴보면 전체 5,493명 중 고등학교 이하는 86,3%인 4,739명, 대학교 이상은 13.7%인 754명이었다. 남녀로 비교했을 때 남성의 22.4%인 589명이 대학교 이상이라고 응답했으며, 여성의 대학교 이상 비율은 5.8%인 165명으로 남성에 비해 여성의 대학교 이상 비율이 더 낮게 나타났다.

교육 연한은 전체 평균은 9.3년이었으며, 남성은 10.6년, 여성은 7.5년이었다. 남성의 교육 연한 평균이 여성보다 높게 나타났다.

심혈관계질환은 과거에 진단받고 현재 치료 중인 사람을 대상으로 하였고, 전체 집단의 2.4%인 130명이 심혈관계질환 현재 치료 중이었으며 남녀의 비율은 각각 2.6%, 2.2%로 비슷했다. 주요 심혈관계질환 위험요인으로는 흡연, 음주, 고혈압, 당뇨, 고지혈증 등이 있으며, 건강 행태 위험요인인 흡연과 음주는 남자가 여성에 비해 비율이 더 높았다. 남성의 46.7%인 1,222명이 현재 흡연 중이라고 응답했으며 여성의 2.6%인 75명만 현재 흡연 중이라고 응답했다.

표 6. 지역사회 기반 코호트 대상자의 일반적 특성

	전체		남성		여성	
	n	(%)	n	(%)	n	(%)
계	5,493	(100)	2,616	(47.6)	2,877	(52.4)
연령	51.5±0.2		51.1±0.3		52.0±0.3	
연령군						
40-49세	2,748	(50.0)	1,378	(52.7)	1,370	(47.6)
50-59세	1,496	(27.2)	700	(26.8)	796	(27.7)
60-69세	1,249	(22.8)	538	(20.5)	711	(24.7)
교육						
교육 연한 (평균, 표준편차)	9.3	(3.4)	10.6	(3.4)	7.5	(2.4)
초등학교 이하	2,301	(41.9)	758	(24.5)	1,543	(64.3)
중학교	1,239	(22.6)	707	(22.9)	532	(22.2)
고등학교	1,295	(23.6)	1,104	(32.8)	281	(11.7)
대학 이상	658	(12.0)	614	(19.9)	44	(1.8)
심혈관계질환						
아니오	5,363	(97.6)	2,549	(97.4)	2,814	(97.8)
예	130	(2.4)	67	(2.6)	63	(2.2)
주요 심혈관계질환 위험요인						
고혈압						
아니오	4,909	(89.4)	2,383	(91.1)	2,526	(87.8)
예	584	(10.6)	233	(8.9)	351	(12.2)
당뇨						
아니오	5,269	(95.9)	2,494	(95.3)	2,775	(96.5)
예	224	(4.1)	122	(4.7)	102	(3.5)
고지혈증						
아니오	5,465	(99.5)	2.602	(99.5)	2,863	(99.5)
예	28	(0.5)	13	(0.5)	14	(0.5)

흡연						
아니오	4,196	(76.4)	1,394	(53.3)	2,802	(97.4)
예	1,297	(23.6)	1,222	(46.7)	75	(2.6)
비만(BMI 기준 25)						
BMI 25 미만	3,159	(57.5)	1,487	(56.8)	1,672	(58.2)
BMI 25 이상	2,334	(42.5)	1,133	(43.2)	1,201	(41.8)
비만(BMI 기준 30)						
BMI 30 미만	5,221	(95.1)	2,490	(95.0)	2,731	(95.1)
BMI 30 이상	272	(5.0)	130	(5.0)	142	(4.9)

농촌 코호트의 대상자의 일반적 특성은 [표 7]에 제시하였다. 농촌 코호트 기반 시점의 참여자는 약 28,000명이었으나 한국인칩 유전 정보 제공에 동의한 대상자는 8,105명이었다. 남성은 37.3%인 3,059명, 여성은 62.3%인 5,046명이으로 여성이 차지하는 비율이 더 높았다. 연령군은 10세 단위로 산출하였으며 50-59세와 60-69세 집단의 비율이 높았다.

교육 수준을 살펴보면 전체 8,105명 중 고등학교 이하는 90.1%인 7,305명, 대학교 이상은 9.9%인 800명이었다. 남녀로 비교했을 때 남성의 16.4%인 494명이 대학교 이상이라고 응답했으며, 여성의 대학교 이상 비율은 6.1%인 306명으로 남성에 비해 여성의 대학교 이상 비율이 더 낮게 나타났다.

교육 연한에서 전체 교육 연한의 평균은 7.8년이었고 남성은 9.3년, 여성은 6.9년으로 남성의 교육 연한 평균이 여성보다 높게 나타났다.

심혈관계질환은 과거에 진단받고 현재 치료 중인 사람을 대상으로 하였고, 전체 집단의 3.7%인 299명이 심혈관계질환 현재 치료 중이었으며 남녀의 비율은 각각 4.0%, 3.5%로 비슷했다. 주요 심혈관계질환 위험요인으로는 흡연, 음주, 고혈압, 당뇨, 고지혈증 등이 있으며, 건강 행태 위험요인인 흡연과 음주는 남자가 여성에 비해 비율이 더 높았다. 남성의 23.1%인 707명이 현재 흡연 중이라고 응답했으며 여성의 1.5%인 74명만 현재 흡연 중이라고 응답했다.

표 7. 농촌 코호트 대상자의 일반적 특성

	전체		남성		여성	
	n	(%)	n	(%)	n	(%)
계	8,105	(100)	3,059	(37.7)	5,046	(62.3)
연령	58.4±0.2		59.7±0.3		57.7±0.2	
연령군						
40-49세	1,456	(18.0)	431	(14.1)	1,025	(20.3)
50-59세	2,852	(35.2)	1,008	(33.0)	1,844	(36.6)
60-69세	2,923	(36.0)	1,242	(40.6)	1,681	(33.3)
70세 이상	874	(10.8)	378	(12.3)	496	(9.8)
교육						
교육 연한 (평균, 표준 편차)	7.8	(4.4)	9.3	(4.2)	6.9	(4.3)
초등학교 이하	4,272	(52.7)	1,198	(39.2)	3,074	(60.9)
중학교	1,455	(18.0)	642	(21.0)	813	(16.1)
고등학교	1,578	(19.5)	725	(23.7)	853	(16.9)
대학 이상	800	(9.9)	494	(16.2)	306	(6.1)
심혈관계질환						
아니오	7,806	(96.3)	2,935	(96.0)	4,871	(96.5)
예	299	(3.7)	124	(4.0)	175	(3.5)
주요 심혈관계질환 위험요인						
고혈압						
아니오	6,195	(76.4)	2,352	(76.9)	3,843	(76.2)
예	1,910	(23.6)	707	(23.1)	1,203	(23.8)
당뇨						
아니오	7,547	(93.1)	2,825	(92.4)	4,722	(93.6)
예	558	(6.9)	234	(7.6)	324	(6.4)
고지혈증						
아니오	7,791	(96.1)	2,960	(96.8)	4,831	(95.7)

예	314	(3.9)	99	(3.2)	215	(4.3)
흡연						
아니오	7,324	(90.4)	2,352	(76.9)	4,972	(98.5)
예	781	(9.6)	707	(23.1)	74	(1.5)
비만(BMI 기준 25)						
BMI 25 미만	4,737	(58.5)	1,822	(59.6)	2,915	(57.8)
BMI 25 이상	3,368	(41.6)	1,237	(40.4)	2,131	(42.2)
비만(BMI 기준 30)						
BMI 30 미만	7,760	(95.7)	2,988	(97.7)	4,772	(94.9)
BMI 30 이상	345	(4.3)	71	(2.3)	274	(5.1)

도시 코호트의 대상자의 일반적 특성은 [표 8]에 제시하였다. 도시 코호트 기반 시점의 참여자는 약 170,000명이었으나 한국인칩 유전정보 제공에 동의한 대상자는 58,696명이었다. 남성은 34.6%인 20,293명, 여성은 65.4%인 38.403명이으로 농촌 코호트와 마찬가지로 여성이 차지하는 비율이 더 높았다. 연령군은 10세 단위로 산출하였으며 50-59세 집단의 비율이 가장 높았다.

교육 수준을 살펴보면 전체 58,696명 중 고등학교 이하는 68.8%인 40,357명, 대학교 이상은 31.2%인 18,339명이었다. 남녀로 비교했을 때 남성의 43.9%인 8,909명이 대학교 이상이라고 응답했으며, 여성의 대학교 이상 비율은 24.6%인 9,430명으로 남성에 비해 여성의 대학교 이상 비율이 더 낮게 나타났다. 하지만 도시 코호트의 대학교 이상 대상자의 비율이 지역사회 기반 코호트, 농촌 코호트보다 높게 나타났다.

교육 연한에서 전체 평균은 11.7년 이었으며, 남성은 12.8년, 여성은 11.2년으로 나타났다. 남성의 교육 연한 평균이 여성보다 높게 나타났으며, 도시 코호트의 교육 연한이 나머지 두 코호트보다 높게 나타났다.

심혈관계질환은 과거에 진단받고 현재 치료 중인 사람을 대상으로 하였고, 전체 집단의 2.8%인 1,622명이 심혈관계질환 현재 치료 중이었으며 남녀의 비율은 각각 4.4%, 1.9%로 남성에서 더 높았다. 주요 심혈관계질환 위험요인으로는 흡연, 음주, 고혈압, 당뇨, 고지혈증

등이 있으며, 건강 행태 위험요인인 흡연과 음주는 남자가 여성에 비해 비율이 더 높았다. 남성의 8.5%인 1,722명이 현재 흡연 중이라고 응답했으며 여성의 0.5%인 191명만 현재 흡연 중이라고 응답했다.

표 8. 도시 코호트 대상자의 일반적 특성

	전체		남성		여성	
	n	(%)	n	(%)	n	(%)
계	58,696	(100)	20,293	(34.6)	38,403	(65.4)
연령	53.8±0.1		55.2±0.1		53.1±0.1	
연령군						
40-49세	18,389	(31.3)	5,466	(26.9)	12,923	(33.7)
50-59세	24,653	(42.0)	7,789	(38.4)	16,864	(43.9)
60-69세	14,838	(25.3)	6,568	(32.4)	8,270	(21.5)
70세 이상	816	(1.4)	470	(2.3)	346	(0.9)
교육						
교육 연한(평균, 표준편차)	11.7	(3.7)	12.8	(3.6)	11.2	(3.7)
초등학교 이하	9,056	(15.4)	1,950	(9.6)	7,106	(18.5)
중학교	9,318	(15.9)	2,493	(12.3)	6,825	(17.8)
고등학교	21,983	(37.5)	6,941	(34.2)	15,042	(39.2)
대학 이상	18,339	(31.2)	8,909	(43.9)	9,430	(24.6)
심혈관계질환						
아니오	57,074	(97.2)	19,395	(95.6)	37,679	(98.1)
예	1,622	(2.8)	898	(4.4)	724	(1.9)
주요 심혈관계질환 위험요인						
고혈압						
아니오	47,782	(81.4)	15,695	(77.3)	32,087	(83.6)
예	10.914	(18.6)	4,598	(22.7)	6,316	(16.4)
당뇨						
아니오	55,148	(94.0)	18,521	(91.3)	36,627	(95.4)
예	3,548	(6.0)	1,772	(8.7)	1,776	(4.6)
고지혈증						
아니오	52,433	(89.3)	18,061	(89.0)	34,372	(89.5)

예	6,263	(10.7)	2,232	(11.0)	4,031	(10.5)
흡연						
아니오	56,783	(96.7)	18,571	(91.5)	38,212	(99.5)
예	1,913	(3.3)	1,722	(8.5)	191	(0.5)
비만(BMI 기준 25)						
BMI 25 미만	39,812	(67.8)	12,240	(60.3)	27,572	(71.8)
BMI 25 이상	18,884	(32.2)	8,053	(39.7)	10,831	(28.2)
비만(BMI 기준 30)						
BMI 30 미만	57,096	(97.3)	19,767	(97.4)	37,329	(97.2)
BMI 30 이상	1,600	(2.7)	526	(2.6)	1,074	(2.8)

1.2. KoGES 코호트에서 계산한 심혈관계질환 발생 위험도 (로지스틱 분석)

코호트 자료는 일정 기간 동안 연구대상자를 추적관찰하기 때문에, 질병 위험도를 계산할 수 있다. 환자-대조군 모형으로 가정하여 교육수준과 심혈관계질환에 대해 로지스틱 회귀분석을 실시해 오즈비를 계산하였으며, 그 결과는 [표 9]에 제시하였다. 로지스틱 분석에 포함된 심혈관계질환 대상자는 기저 조사 시 이미 해당 질환을 진단받은 사람이며, 새로 진단 받은 사람은 제외하였다. 지역사회 코호트와 농촌 코호트는 표본 수가 적어 오즈비의 95% 신뢰구간이 유의하지 않았다. 도시 코호트는 표본 수가 충분하여 95% 신뢰구간과 p-value가 안정적으로 산출되었다.

지역사회 코호트에서 대학 이상인 집단의 심혈관계질환 오즈비는 0.85, 농촌 코호트에서 대학 이상인 집단의 심혈관계질환 오즈비는 0.73, 도시 코호트에서는 0.79였다.

표 9. KoGES 코호트에서 교육 수준에 따른 심혈관계질환 발생 위험도(로지스틱 회귀분석)

Cohort	Education	OR	(95% CI)	p-value	Adjusted OR	(95% CI)	p-value
지역사회	고등학교 이하	1			1		
	대학 이상	0.90	(0.33 - 2.42)	0.83	0.85	(0.32 - 2.28)	0.75
농촌	고등학교 이하	1			1		
	대학 이상	0.65	(0.38 - 1.13)	0.13	0.73	(0.43 - 1.26)	0.26
도시	고등학교 이하	1			1		
	대학 이상	0.70	(0.62 - 0.79)	<0.001	0.79	(0.72 - 0.89)	<0.001

a. 고혈압, 당뇨병, 흡연, 고지혈증 등 심혈관계질환 위험요인 보정

2. GWAS 결과 - 교육 수준

데이터 질 관리 과정을 거친 후 교육 수준과 연관성이 있는 SNP들을 찾아 [표 13]-[표 18]에 제시하였다. 도시 코호트에서 추출한 결과와 지역사회, 농촌 코호트 통합자료에서 추출한 결과를 모두 표로 제시했다. 표에 제시한 지표는 염색체 번호, 염색체 내에서의 위치, SNP 이름, 기준 대립 유전자, 효과 대립 유전자, 효과 대립 유전자 빈도, 효과 크기, 표준오차, p-value, 표본수이다. 도시 코호트에서는 p-value가 5e-5일 때 126개의 SNP를 구했다. 해당 SNP의 위치 정보, 효과 크기, 표준오차 등을 함께 제시했다.

지역사회 기반 코호트, 농촌 코호트, 통합 자료에서 교육 수준을 예측하는 SNP 역시 앞선 과정을 거쳐 추출하였다. 2개 코호트 통합 자료에서는 교육 수준을 예측하는 SNP는 163개가 산출되었다. 도시 코호트의 126개와 지역사회, 농촌 코호트의 163개 중 중복되는 SNP는 55개였으며, 이 SNP들을 멘델 무작위 분석에 이용하였다. 55개 SNP의 p-value 한계점은 5e-5였으며, p-value가 5e-6일 때 SNP는 33개, 5e-7일 때 SNP는 13개가 추출되었다. 각 p-value 한계점에 대하여 멘델 무작위 분석을 실시하였으며, p-value 5e-7의 경우에 SNP의 숫자가 적어 결과변수인 심혈관계질환에 대해서만 멘델 무작위 분석을 실시했다. 심혈관계질환 위험요인에 대한 분석은 SNP의 연결에서 적절한 SNP가 산출되지 않아 5e-7의 SNP 13개는 분석에

서 제외하였다.

표 10. 도시 코호트에서 GWAS 분석을 통하여 교육 수준을 예측한 SNP 목록

Chromosome	position	REF	ALT	EAF	Beta	se	p-value
1	85410061	C	T	2.28%	-0.0870	0.0206	2.51E-05
1	245768933	A	G	9.72%	-0.0216	0.0479	6.34E-06
1	234545928	T	C	11.02%	0.1324	0.0317	3.10E-05
2	61539291	T	G	2.31%	0.2712	0.0662	4.28E-05
2	161905001	C	T	49.46%	0.0818	0.0199	4.01E-05
2	161977647	A	G	48.73%	0.0898	0.0199	6.42E-06
2	162101261	G	A	48.86%	-0.0831	0.0199	3.23E-05
2	162214414	CT	C	41.44%	0.0862	0.0202	2.13E-05
2	162258433	T	C	35.30%	0.0949	0.0208	5.51E-06
2	228256328	G	A	12.23%	0.1380	0.0304	5.59E-06

3	18203815	C	T	1.70%	-0.3219	0.0768	2.81E-05
3	140942232	G	A	48.96%	-0.0452	0.0925	9.91E-07
3	144026997	A	G	13.24%	-0.0416	0.0873	1.88E-06
3	153585628	T	G	1.01%	-0.0340	0.0815	3.01E-05
3	140824608	T	C	1.51%	-0.3359	0.0813	3.63E-05
4	9785134	T	C	36.03%	0.0918	0.0207	9.30E-06
4	97157004	T	C	7.18%	0.1577	0.0384	4.15E-05
4	92431580	G	A	1.04%	0.0137	0.0332	3.53E-05
4	121613198	G	C	14.82%	-0.0374	0.0919	4.80E-05
5	32918833	A	G	2.04%	-0.2910	0.0702	3.45E-05
5	90860370	A	G	49.18%	-0.0822	0.0199	3.84E-05
5	90919651	A	G	49.72%	-0.0811	0.0199	4.83E-05
5	90977625	TA	T	49.21%	0.0815	0.0199	4.40E-05

5	54852576	C	T	1.29%	0.0830	0.02	3.22E-05
5	114210843	C	G	46.40%	0.0361	0.0888	4.89E-05
6	27974067	GT	G	43.56%	0.0858	0.02	1.80E-05
6	28100640	C	T	38.99%	0.0858	0.0203	2.50E-05
6	31322706	C	G	5.39%	0.0858	0.0204	2.56E-05
6	67132090	T	C	1.85%	0.0916	0.0204	6.98E-06
6	35499601	A	AAAT	41.58%	-0.0827996	0.0202	4.35E-05
6	77664001	CT	C	20.27%	0.1068	0.0248	1.68E-05
6	78812964	C	T	1.88%	0.3042	0.0735	3.52E-05
6	110186638	T	A	46.11%	-0.0864	0.0199	1.44E-05
7	41147	G	A	1.97%	0.0307	0.0723	2.22E-05
7	41170	G	A	3.44%	-0.0386	0.0927	3.13E-05

7	81580	C	A	3.50%	-0.0243	0.0520	3.02E-06
7	77256407	A	G	2.48%	-0.2839	0.0641	9.59E-06
7	111119191	T	G	1.95%	0.2990	0.0720	3.28E-05
7	112379807	C	G	1.51%	-0.3311	0.0813	4.71E-05
7	127907054	C	T	11.79%	0.1450	0.0309	2.76E-06
7	130049901	G	T	2.38%	0.3033	0.0654	3.61E-06
7	156833233	T	C	1.55%	-0.3628	0.0807	6.96E-06
8	132729215	G	A	1.04%	0.4693	0.0979	1.67E-06
8	225168	C	T	4.04%	0.0292	0.0409	4.75E-05
8	234623	T	G	3.52%	0.0392	0.0675	5.61E-06
8	239111	T	C	3.07%	0.0134	0.0204	5.12E-05
8	750906	C	T	5.21%	0.0272	0.0259	2.94E-06
8	761499	C	T	4.11%	0.0065	0.0557	9.06E-05

8	772139	C	T	2.81%	-0.0918	0.0909	3.13E-06
8	888744	C	T	3.39%	-0.0944	0.0365	9.66E-05
8	1135049	G	A	3.83%	0.0569	0.0926	5.39E-05
8	1140691	C	G	4.98%	0.0295	0.0424	4.86E-06
8	1151923	C	G	2.60%	-0.0620	0.0373	9.61E-06
8	1189658	T	C	3.17%	0.0591	0.0642	3.57E-06
8	1192943	A	G	4.75%	0.0829	0.0913	3.64E-06
8	1326979	C	T	4.65%	0.0236	0.0202	2.43E-06
8	1333258	A	G	31.66%	-0.1005	0.0404	1.29E-06
8	1360171	T	C	3.47%	0.0174	0.0267	5.14E-06
8	1472228	G	A	3.00%	-0.0806	0.0732	2.71E-07
8	1483124	C	A	2.57%	0.0294	0.0199	1.40E-07

8	1485960	G	T	2.36%	0.0386	0.0416	3.53E-06
9	35194447	C	T	6.42%	0.1899	0.0405	2.80E-06
10	30170236	A	T	2.10%	0.2827	0.0694	4.65E-05
10	36331720	T	A	2.00%	-0.3064	0.0711	1.65E-05
11	71616877	G	T	2.18%	-0.2906	0.0682	2.08E-05
11	104594658	C	T	8.09%	-0.1509	0.0364	3.40E-05
12	110297784	C	T	5.89%	0.1755	0.0424	3.55E-05
12	110342598	G	A	9.20%	0.1538	0.0346	8.73E-06
12	110390979	T	C	7.44%	0.0185	0.0381	1.17E-06
12	111337494	CT	C	18.75%	0.1102	0.0256	1.65E-05
12	111414461	T	G	14.11%	0.0248	0.0287	4.85E-18
12	111684461	G	A	49.99%	0.0826	0.0199	3.49E-05
12	111718231	C	A	18.23%	0.0220	0.0259	2.31E-17

12	111902051	GGT	G	46.36%	-0.0902	0.0201	7.09E-06
12	111994852	A	C	48.98%	0.1082	0.0201	6.88E-08
12	112168009	G	A	15.74%	0.0267	0.0275	3.08E-22
12	112230019	G	C	15.80%	0.0267	0.0275	2.58E-22
12	112325766	A	C	48.82%	0.0846	0.02	2.31E-05
12	112433843	CAA	C	44.80%	-0.0907	0.0201	6.15E-06
12	112468206	C	T	15.77%	0.0261	0.0275	1.97E-21
12	112552274	C	T	23.52%	0.0167	0.0237	1.55E-12
12	112627350	G	A	16.48%	0.0255	0.0270	4.52E-21
12	112678697	G	A	48.31%	0.0865	0.02	1.54E-05
12	112736118	A	G	15.85%	0.0257	0.0275	7.37E-21
12	112829103	CAT	C	42.25%	-0.0894	0.0203	1.12E-05

12	112923393	G	A	44.80%	0.0922	0.0201	4.53E-06
12	112982775	A	G	40.27%	-0.0863	0.0204	2.29E-05
12	113117897	G	T	15.32%	0.0229	0.0278	1.52E-16
12	113153857	C	T	37.99%	0.0909	0.0206	1.05E-05
12	113164710	C	A	37.20%	0.0905	0.0206	1.18E-05
12	113196733	G	A	39.95%	0.0865	0.0204	2.28E-05
12	113588466	C	T	10.20%	0.1371	0.0330	3.31E-05
12	113713062	A	C	8.00%	0.1808	0.0367	8.80E-07
12	114299584	G	C	3.13%	0.2428	0.0571	2.16E-05
13	28473690	T	C	3.34%	-0.2351	0.0554	2.24E-05
13	58382574	AT	A	14.78%	-0.1142	0.0280	4.77E-05
13	58408628	GT	G	14.62%	-0.1151	0.0282	4.51E-05
13	58525494	A	G	17.45%	-0.1094	0.0263	3.34E-05

13	58593657	A	G	17.88%	-0.1105	0.0261	2.30E-05
13	58611635	T	C	3.95%	-0.2377	0.0512	3.48E-06
13	83212401	A	G	32.40%	0.0902	0.0212	2.21E-05
13	98747553	G	A	10.74%	0.1313	0.0321	4.52E-05
13	110045774	CT	C	38.32%	-0.0859	0.0205	3.37E-05
14	24455996	A	C	2.46%	0.2726	0.0637	1.91E-05
14	60004577	G	T	20.98%	0.0994	0.0244	4.88E-05
14	79372796	A	G	1.26%	-0.3662	0.0887	3.73E-05
15	89630650	G	A	1.57%	0.32856	0.0799	3.94E-05
16	19380214	C	T	17.13%	-0.0126	0.0265	2.01E-06
16	56729619	G	A	1.38%	0.0355	0.0855	3.26E-05
16	89653173	T	C	47.48%	0.0854	0.0199	1.84E-05

17	14690676	C	T	1.19%	-0.3783	0.0915	3.59E-05
17	43149988	T	C	49.91%	0.0838	0.0199	2.72E-05
17	43161353	C	T	48.74%	0.0854	0.0199	1.92E-05
17	46502682	T	C	1.70%	-0.0165	0.0368	7.04E-06
17	72917513	C	T	4.58%	0.0815	0.02	4.46E-05
18	35613906	G	A	1.17%	-0.3936	0.0924	2.07E-05
18	54939057	A	T	2.00%	-0.2885	0.0710	4.89E-05
18	55511115	A	T	12.56%	0.1368	0.03	5.11E-06
18	61248488	C	A	1.02%	0.4046	0.0989	4.33E-05
19	5831672	G	C	49.96%	-0.0818	0.0199	4.11E-05
21	15947361	C	T	1.81%	-0.3297	0.0746	1.01E-05
21	23803358	GAT	G	1.36%	0.3500	0.0861	4.86E-05
21	37638293	G	T	3.23%	-0.26892	0.0563	1.81E-06

21	43963436	G	A	8.59%	-0.1633	0.0356	4.51E-06
22	44253964	C	A	49.86%	-0.0818	0.0198	3.77E-05
22	50985023	C	A	3.91%	0.2091	0.0512	4.57E-05

표 11. 지역사회/농촌 코호트에서 GWAS 분석을 통하여 교육 수준을 예측한 SNP 목록

Chromosome	position	REF	ALT	EAF	Beta	se	p-value
1	85410061	C	T	2.27%	0.0653	0.0153	2.04E-05
1	245768933	A	G	9.72%	-0.0313	0.0077	4.90E-05
3	140942232	G	A	48.95%	0.0186	0.0045	4.57E-05
3	144026997	A	G	13.24%	-0.0276	0.0068	4.99E-05
3	153585628	T	G	1.00%	0.0988	0.0230	1.85E-05
3	153657202	G	A	1.09%	0.0954	0.022	1.64E-05
4	14556767	G	A	3.17%	0.0554	0.0131	2.45E-05
4	92431580	G	A	1.03%	0.0922	0.0225	4.31E-05
4	121613198	G	C	14.81%	-0.0267	0.0064	3.46E-05
5	10919826	A	G	1.71%	-0.07605	0.0175	1.44E-05

5	54852576	C	T	1.29%	0.0847	0.0204	3.51E-05
5	114210843	C	G	46.40%	-0.0201	0.0046	1.19E-05
6	31322706	C	G	5.38%	0.0413	0.0101	4.82E-05
6	67132090	T	C	1.84%	0.0761	0.0170	7.96E-06
7	41147	G	A	1.97%	-0.0691	0.0162	1.97E-05
7	41170	G	A	3.44%	-0.0538	0.0123	1.37E-05
7	81580	C	A	3.50%	-0.0605	0.0123	9.14E-07
7	105732	C	T	4.12%	0.0480	0.0113	2.36E-05
8	166593	A	G	3.36%	-0.0594	0.1256	2.21E-06
8	225168	C	T	4.04%	0.0640	0.0114	2.20E-08
8	234623	T	G	3.52%	0.0508	0.0123	4.07E-05
8	239111	T	C	3.06%	-0.0615	0.0130	2.39E-06
8	250459	G	A	6.46%	-0.0376	0.0091	4.05E-05

8	295470	G	A	2.78%	-0.0609	0.0137	9.59E-06
8	366533	A	C	3.44%	0.0532	0.0124	1.80E-05
8	368599	C	T	3.08%	-0.0560	0.0130	1.83E-05
8	407191	C	T	15.18%	-0.0257	0.0062	4.31E-05
8	417368	A	G	4.14%	-0.0481	0.0113	2.07E-05
8	420826	C	T	2.68%	-0.0601	0.0138	1.49E-05
8	468504	T	C	3.97%	0.0546	0.0115	2.35E-06
8	471160	T	C	5.02%	0.0525	0.0103	3.76E-07
8	479339	A	G	6.70%	0.0377	0.0090	2.93E-05
8	480106	G	A	5.57%	0.0457	0.0098	3.17E-06
8	527569	G	C	4.05%	-0.0490	0.0114	1.83E-05
8	528906	G	A	1.59%	0.0790	0.0182	1.42E-05

8	593066	G	A	2.33%	-0.0671	0.0149	7.42E-06
8	649786	C	T	5.19%	0.0487	0.0101	1.63E-06
8	678090	C	T	4.33%	0.0525	0.0110	2.04E-06
8	692992	A	G	3.95%	0.0470	0.0115	4.78E-05
8	707414	C	G	10.14%	-0.0312	0.0074	2.80E-05
8	718031	A	C	3.75%	0.0586	0.0119	9.26E-07
8	744601	G	A	2.05%	-0.0655	0.0158	3.41E-05
8	750906	C	T	5.20%	-0.0555	0.0101	4.52E-08
8	761499	C	T	4.11%	-0.0542	0.0113	1.68E-06
8	772139	C	T	2.81%	-0.0698	0.0136	3.37E-07
8	778394	A	G	3.33%	0.0610	0.0125	1.29E-06
8	841243	G	A	3.18%	0.0646	0.0129	5.58E-07
8	849012	C	T	1.46%	0.0774	0.0190	4.89E-05

8	852928	GAAC	G	5.06	0.0521	0.0102	3.83E-07
8	883445	G	A	2.61	0.0628	0.0143	1.13E-05
8	888744	C	T	3.39	0.0789	0.0124	2.42E-10
8	921967	C	A	5.82	0.0395	0.0096	4.10E-05
8	927728	C	T	19.15	0.0261	0.0058	6.97E-06
8	986268	T	G	3.49	-0.0543	0.0123	1.03E-05
8	995388	T	C	5.12	0.0530	0.0102	2.22E-07
8	1025432	G	T	4.42	-0.0470	0.0109	1.73E-05
8	1026110	A	C	3.57	0.0630	0.0121	2.25E-07
8	1040139	C	T	3.59	-0.0609	0.0121	5.18E-07
8	1042397	G	A	6.97	0.0402	0.0088	6.12E-06
8	1090105	T	C	3.50	-0.0625	0.0123	3.83E-07

8	1108090	C	T	2.06	-0.0826	0.0159	2.33E-07
8	1123729	C	A	5.97	0.0420	0.0095	1.05E-05
8	1135049	G	A	3.83	-0.0733	0.0117	4.79E-10
8	1140691	C	G	4.98	0.0492	0.0103	2.19E-06
8	1151923	C	G	2.60	0.0606	0.0142	2.02E-05
8	1189658	T	C	3.17	0.0579	0.0129	7.28E-06
8	1192943	A	G	4.75	0.0645	0.0105	1.07E-09
8	1200217	C	CT	4.73	0.0458	0.0105	1.51E-05
8	1202088	G	T	1.55	0.0837	0.0182	4.61E-06
8	1204545	C	G	4.23	0.0466	0.0112	3.54E-05
8	1211021	G	C	3.54	0.0511	0.0123	3.23E-05
8	1218539	A	T	3.46	0.0615	0.0125	8.77E-07
8	1229413	G	A	3.98	-0.0482	0.0114	2.70E-05

8	1234652	A	G	5.25	0.0410	0.01	4.77E-05
8	1236401	C	A	1.91	0.0738	0.0167	1.02E-05
8	1237978	A	G	2.76	-0.0686	0.0138	7.79E-07
8	1238077	C	G	4.58	-0.0487	0.0107	6.22E-06
8	1240252	A	G	3.33	-0.0533	0.0125	2.15E-05
8	1259891	A	G	4.53	0.0556	0.0108	2.73E-07
8	1270475	C	T	22.14	-0.0223	0.0054	3.87E-05
8	1273160	C	T	5.44	0.0429	0.0099	1.63E-05
8	1292021	T	C	1.77	0.0796	0.0172	4.06E-06
8	1292118	C	T	3.36	0.0632	0.0125	4.71E-07
8	1295127	T	C	4.53	-0.0457	0.0108	2.42E-05
8	1304830	G	C	22.38	-0.0239	0.0055	1.39E-05

8	1326979	C	T	4.65	-0.0526	0.0107	8.64E-07
8	1333258	A	G	31.66	-0.0262	0.0048	8.47E-08
8	1360171	T	C	3.47	0.0603	0.0122	9.05E-07
8	1420305	C	A	2.44	0.0630	0.0146	1.66E-05
8	1426509	C	G	2.88	-0.0616	0.0135	5.05E-06
8	1429538	C	G	3.83	0.0507	0.0117	1.60E-05
8	1432889	T	C	6.69	0.0373	0.0091	4.07E-05
8	1444557	G	T	1.28	0.0825	0.0202	4.75E-05
8	1472228	G	A	2.99	0.0728	0.0133	4.84E-08
8	1483124	C	A	2.57	-0.0816	0.0143	1.21E-08
8	1485960	G	T	2.35	0.0906	0.0149	1.39E-09
11	29320154	G	T	2.59	0.0654	0.0145	6.86E-06
11	35870877	A	T	2.81	0.0596	0.0139	2.00E-05

11	103171277	GC	G	1.99	0.0686	0.0163	2.78E-05
11	126083026	AT	A	36.03	-0.0229	0.0047	1.71E-06
11	126184045	T	G	35.2	-0.0220	0.0048	4.89E-06
11	126200536	T	C	30.74	-0.0226	0.0049	5.70E-06
12	69185487	C	G	1.82	-0.0723	0.0172	2.69E-05
12	71792975	G	C	2.55	0.0593	0.0145	4.50E-05
12	110390979	T	C	7.44	0.0624	0.0061	3.11E-06
12	111337494	CT	C	18.74	-0.0178	0.0092	8.47E-06
12	111414461	T	G	14.11	-0.0550	0.0277	4.75E-06
12	111684461	G	A	49.98	-0.0857	0.0093	3.60E-07
12	111718231	C	A	18.23	-0.0077	0.0008	9.31E-06
12	111902051	GGT	G	46.36	0.0302	0.0066	6.49E-05

12	111994852	A	C	48.97	-0.08	0.0091	3.79E-05
12	112168009	G	A	15.73	-0.0288	0.0027	2.98E-06
12	112230019	G	C	15.79	0.0233	0.0023	3.27E-06
12	112468206	C	T	15.77	0.0136	0.0047	7.73E-05
12	112552274	C	T	23.51	0.0987	0.0079	2.13E-05
12	112627350	G	A	16.48	-0.0019	0.0020	9.23E-06
12	112678697	G	A	48.30	0.0192	0.0090	8.31E-05
12	112736118	A	G	15.84	-0.0404	0.0087	6.43E-05
12	112829103	CAT	C	42.25	-0.0064	0.0088	9.41E-06
12	112923393	G	A	44.79	-0.0074	0.0020	7.13E-07
12	112982775	A	G	40.26	0.0237	0.0019	2.32E-06
12	113117897	G	T	15.31	-0.0237	0.0028	4.06E-06
13	67087326	GT	G	44.87	-0.19294	0.0462	3.08E-05

14	51100617	C	T	1.23%	-0.0871	0.0209	3.21E-05
14	51176530	G	GA	1.33	-0.0821	0.02	4.25E-05
14	63002960	A	G	17.19	0.0251	0.0060	3.25E-05
15	56159100	C	T	1.20	0.0884	0.0210	2.62E-05
15	56220023	T	C	1.18	0.0928	0.0211	1.20E-05
15	56242878	G	A	1.16	0.0935	0.0213	1.23E-05
15	66871840	A	G	48.86	0.1949	0.0462	2.51E-05
16	1821810	G	GCC CC	2.59	0.0631	0.0145	1.35E-05
16	5004328	C	T	1.72	0.0725	0.0174	3.19E-05
16	5008164	G	T	2.85	-0.0584	0.0135	1.72E-05
16	5033749	G	A	3.02	-0.0671	0.0131	3.42E-07
16	5062968	C	A	3.66	-0.0553	0.0120	4.19E-06

16	5081516	T	A	5.66	-0.0504	0.0097	2.38E-07
16	5113679	C	T	4.40	-0.0479	0.0110	1.35E-05
16	5116466	T	C	13.37	0.0322	0.0066	1.20E-06
16	5117771	T	G	3.06	-0.0539	0.0131	4.13E-05
16	5186073	C	G	28.16	0.0211	0.0050	2.56E-05
16	5187039	C	A	4.96	0.0491	0.0103	2.07E-06
16	19380214	C	T	17.12	0.0654	0.0106	7.97E-10
16	56729619	G	A	1.37	0.0580	0.0105	3.33E-08
16	89653173	T	C	47.48	0.0478	0.0105	5.73E-06
16	5808379	C	T	5.54	-0.0574	0.0099	6.68E-09
16	76295833	T	G	6.81	0.0557	0.0089	4.90E-10
16	76349723	A	G	5.07	-0.0419	0.0103	4.62E-05
16	76365253	T	G	3.71	-0.0545	0.0119	5.30E-06

16	76480494	A	G	15.18	0.0367	0.0063	8.35E-09
16	76516286	AG	A	5.00	-0.0486	0.0103	2.57E-06
16	76520873	C	G	4.10	-0.0734	0.0113	9.41E-11
16	76542071	T	A	1.82	0.0723	0.0171	2.47E-05
16	76552955	G	C	36.08	0.0272	0.0047	7.60E-09
16	76554810	C	G	1.54	-0.0811	0.0183	9.43E-06
16	81479959	T	A	3.51	-0.0592	0.0122	1.37E-06
16	81517658	G	A	5.51	-0.0447	0.0099	6.23E-06
16	81700305	C	T	7.29	-0.0379	0.0087	1.51E-05
16	85546975	C	T	7.49	0.0475	0.0086	3.46E-08
16	85753810	G	A	2.28	-0.0725	0.0151	1.63E-06
16	85909923	C	T	2.95	-0.0611	0.0132	4.36E-06

16	85971515	CA	C	7.68	0.0372	0.0084	1.09E-05
17	46502682	T	C	1.70	0.0747	0.0177	2.65E-05
17	72917513	C	T	4.57	0.0465	0.0109	2.04E-05

표 12. 도시 코호트와 지역사회/농촌 코호트에서 중복되는 SNP 목록(p-value <5e-5)

Chr	position	REF	ALT	Nearest Gene	Functional consequence	EAF	Beta	se	p-value
1	85410061	C	T	MCOLN2	intron	2.28%	-0.0870	0.0206	2.51E-05
1	24576893	A	G	NCMAP	intron	9.72%	-0.0216	0.0479	6.34E-06
3	140942232	G	A	PXYLP1	intergenic	48.96%	-0.0452	0.0925	9.91E-07
3	144026997	A	G	LOC105374140	intron	13.24%	-0.0416	0.0873	1.88E-06
3	153585628	T	G	SH3BP5	intron	1.01%	-0.0340	0.0815	3.01E-05
4	92431580	G	A	CCSER1	intron	1.04%	0.0137	0.0332	3.53E-05
4	121613198	G	C	PRDM5	intron	14.82%	-0.0374	0.0919	4.80E-05
5	54852576	C	T	AK4P2	intergenic	1.29%	0.0830	0.02	3.22E-05
5	114210843	C	G	CTNND2	intron	46.40%	0.0361	0.0888	4.89E-05
6	31322706	C	G	HLA-B	intron	5.39%	0.0858	0.0204	2.56E-05

6	67132090	T	C	NUFIP1P1	intergenic	1.85%	0.0916	0.0204	6.98E-06
7	41147	G	A	LINC03015	intergenic	1.97%	0.0307	0.0723	2.22E-05
7	41170	G	A	LINC03015	intergenic	3.44%	-0.0386	0.0927	3.13E-05
7	81580	C	A	LOC101929756	intron	3.50%	-0.0243	0.0520	3.02E-06
8	225168	C	T	LOC101927566	upstream	4.04%	0.0292	0.0409	4.76E-05
8	234623	T	G	FAM87A	intergenic	3.52%	0.0392	0.0675	5.61E-06
8	239111	T	C	FAM87A	intergenic	3.07%	0.0134	0.0204	5.13E-05
8	750906	C	T	DLGAP2	intron	5.21%	0.0272	0.0259	2.95E-06
8	761499	C	T	DLGAP2	intron	4.11%	0.0065	0.0557	9.06E-05
8	772139	C	T	DLGAP2	non-coding	2.81%	-0.0918	0.0909	3.13E-06
8	888744	C	T	DLGAP2	intron	3.39%	-0.0944	0.0365	9.66E-05
8	1135049	G	A	DLGAP2	intron	3.83%	0.0569	0.0926	5.39E-05
8	1140691	C	G	DLGAP2	intron	4.98%	0.0295	0.0424	4.86E-06

8	1151923	C	G	DLGAP2	intron	2.60%	-0.0620	0.0373	9.61E-06
8	1189658	T	C	DLGAP2	intron	3.17%	0.0591	0.0642	3.57E-06
8	1192943	A	G	DLGAP2	intron	4.75%	0.0829	0.0913	3.64E-06
8	1326979	C	T	DLGAP2	intron	4.65%	0.0236	0.0202	2.43E-06
8	1333258	A	G	DLGAP2	intron	31.66%	-0.1005	0.0404	1.29E-06
8	1360171	T	C	DLGAP2	intron	3.47%	0.0174	0.0267	5.14E-06
8	1472228	G	A	DLGAP2	intron	3.00%	-0.0806	0.0732	2.71E-07
8	1483124	C	A	DLGAP2	intron	2.57%	0.0294	0.0199	1.40E-07
8	1485960	G	T	DLGAP2	intron	2.36%	0.0386	0.0416	3.53E-06
12	110390979	T	C	GIT2	missence	7.44%	0.0185	0.0381	1.17E-06
12	111337494	CT	C	CCDC63	intron	18.75%	0.1102	0.0256	1.65E-05
12	111414461	T	G	CUX2	intergenic	14.11%	0.0248	0.0287	4.85E-18

12	111684461	G	A	CUX2	intron	49.99%	0.0826	0.0199	3.49E-05
12	111718231	C	A	CUX2	intron	18.23%	0.0220	0.0259	2.31E-17
12	111902051	GGT	G	ATXN2	intron	46.36%	-0.0902	0.0201	7.09E-06
12	111994852	A	C	ATXN2	intron	48.98%	0.1082	0.0201	6.88E-08
12	112168009	G	A	ACAD10	intron	15.74%	0.0267	0.0275	3.08E-22
12	112230019	G	C	ALDH2	intron	15.80%	0.0267	0.0275	2.58E-22
12	112468206	C	T	NAA25	intron	15.77%	0.0261	0.0275	1.97E-21
12	112552274	C	T	NAA25	intergenic	23.52%	0.0167	0.0237	1.55E-12
12	112627350	G	A	HECTD4	intron	16.48%	0.0255	0.0270	4.52E-21
12	112678697	G	A	HECTD4	intron	48.31%	0.0865	0.02	1.54E-05
12	112736118	A	G	HECTD4	intron	15.85%	0.0257	0.0275	7.37E-21
12	112829103	CAT	C	HECTD4	intergenic	42.25%	-0.0894	0.0203	1.12E-05
12	112923393	G	A	PTPN11	intron	44.80%	0.0922	0.0201	4.53E-06

12	112982775	A	G	PTPN11	intergenic	40.27%	-0.0863	0.0204	2.29E-05
12	113117897	G	T	RPH3A	intron	15.32%	0.0229	0.0278	1.52E-16
16	19380214	C	T	LOC10537 1114	intron	17.13%	-0.0126	0.0265	2.01E-06
16	56729619	G	A	NUP93-D T	intron	1.38%	0.0355	0.0855	3.26E-05
16	89653173	T	C	CPNE7	intron	47.48%	0.0854	0.0199	1.84E-05
17	46502682	T	C	SKAP1	intron	1.70%	-0.0165	0.0368	7.04E-06
17	72917513	C	T	USH1G	intron	4.58%	0.0815	0.02	4.46E-05

표 13. 도시 코호트와 지역사회, 농촌 코호트에서 중복되는 SNP 목록(p-value <5e-6)

Chr	position	REF	ALT	Nearest Gene	Functional consequence	EAF	Beta	se	p-value
1	24576893	A	G	NCMAP	intron	9.72%	-0.0216	0.0479	6.34E-06
3	140942232	G	A	PXYLP1	intergenic	48.96%	-0.0452	0.0925	9.91E-07
3	144026997	A	G	LOC105374140	intron	13.24%	-0.0416	0.0873	1.88E-06
6	67132090	T	C	NUFIP1P1	intergenic	1.85%	0.0916	0.0204	6.98E-06
7	81580	C	A	LOC101929756	intron	3.50%	-0.0243	0.0520	3.02E-06
8	234623	T	G	FAM87A	intergenic	3.52%	0.0392	0.0675	5.61E-06
8	750906	C	T	DLGAP2	intron	5.21%	0.0272	0.0259	2.95E-06
8	772139	C	T	DLGAP2	non-coding	2.81%	-0.0918	0.0909	3.13E-06
8	1140691	C	G	DLGAP2	intron	4.98%	0.0295	0.0424	4.86E-06
8	1151923	C	G	DLGAP2	intron	2.60%	-0.0620	0.0373	9.61E-06

8	1189658	T	C	DLGAP2	intron	3.17%	0.0591	0.0642	3.57E-06
8	1192943	A	G	DLGAP2	intron	4.75%	0.0829	0.0913	3.64E-06
8	1326979	C	T	DLGAP2	intron	4.65%	0.0236	0.0202	2.43E-06
8	1333258	A	G	DLGAP2	intron	31.66%	-0.1005	0.0404	1.29E-06
8	1360171	T	C	DLGAP2	intron	3.47%	0.0174	0.0267	5.14E-06
8	1472228	G	A	DLGAP2	intron	3.00%	-0.0806	0.0732	2.71E-07
8	1483124	C	A	DLGAP2	intron	2.57%	0.0294	0.0199	1.40E-07
8	1485960	G	T	DLGAP2	intron	2.36%	0.0386	0.0416	3.53E-06
12	110390979	T	C	GIT2	missence	7.44%	0.0185	0.0381	1.17E-06
12	111414461	T	G	CUX2	intergenic	14.11%	0.0248	0.0287	4.85E-18
12	111718231	C	A	CUX2	intron	18.23%	0.0220	0.0259	2.31E-17
12	111902051	GGT	G	ATXN2	intron	46.36%	-0.0902	0.0201	7.09E-06
12	111994852	A	C	ATXN2	intron	48.98%	0.1082	0.0201	6.88E-08

12	112168009	G	A	ACAD10	intron	15.74%	0.0267	0.0275	3.08E-22
12	112230019	G	C	ALDH2	intron	15.80%	0.0267	0.0275	2.58E-22
12	112468206	C	T	NAA25	intron	15.77%	0.0261	0.0275	1.97E-21
12	112552274	C	T	NAA25	intergenic	23.52%	0.0167	0.0237	1.55E-12
12	112627350	G	A	HECTD4	intron	16.48%	0.0255	0.0270	4.52E-21
12	112736118	A	G	HECTD4	intron	15.85%	0.0257	0.0275	7.37E-21
12	112923393	G	A	PTPN11	intron	44.80%	0.0922	0.0201	4.53E-06
12	113117897	G	T	RPH3A	intron	15.32%	0.0229	0.0278	1.52E-16
16	19380214	C	T	LOC105371114	intron	17.13%	-0.0126	0.0265	2.01E-06
17	46502682	T	C	SKAP1	intron	1.70%	-0.0165	0.0368	7.04E-06

표 14. 도시 코호트와 지역사회, 농촌 코호트에서 중복되는 SNP 목록(p-value <5e-7)

Chr	position	REF	ALT	Nearest Gene	Functional consequence	EAF	Beta	se	p-value
3	140942232	G	A	PXYLP1	intergenic	48.96%	-0.0452	0.0925	9.91E-07
8	1472228	G	A	DLGAP2	intron	3.00%	-0.0806	0.0732	2.71E-07
8	1483124	C	A	DLGAP2	intron	2.57%	0.0294	0.0199	1.40E-07
12	111414461	T	G	CUX2	intergenic	14.11%	0.0248	0.0287	4.85E-18
12	111718231	C	A	CUX2	intron	18.23%	0.0220	0.0259	2.31E-17
12	111994852	A	C	ATXN2	intron	48.98%	0.1082	0.0201	6.88E-08
12	112168009	G	A	ACAD10	intron	15.74%	0.0267	0.0275	3.08E-22
12	112230019	G	C	ALDH2	intron	15.80%	0.0267	0.0275	2.58E-22
12	112468206	C	T	NAA25	intron	15.77%	0.0261	0.0275	1.97E-21
12	112552274	C	T	NAA25	intergenic	23.52%	0.0167	0.0237	1.55E-12

12	112627350	G	A	HECTD4	intron	16.48%	0.0255	0.0270	4.52E-21
12	112736118	A	G	HECTD4	intron	15.85%	0.0257	0.0275	7.37E-21
12	113117897	G	T	RPH3A	intron	15.32%	0.0229	0.0278	1.52E-16

3. 멘델 무작위 분석 결과 - 심혈관계질환

3.1. 1-표본 멘델 무작위 분석 - KoGES 코호트 통합

이 연구에서 1-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 다형성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 15. 이질성 검정 결과(교육 수준 - 심혈관계질환)

		Q	p-value
교육 연한	MR-Egger	20.3	0.93
	Inverse variance weighted	20.4	0.94

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 다형성 잔차항이 들어있는 것으로 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 16. 다형성 검정 결과(교육 수준 - 심혈관계질환)

		intercept	se	p-value
교육 연한	MR-Egger	0.005	0.02	0.73

멘델 무작위 분석과 민감도 분석을 [표 17]에 제시했다. p값에 따라 멘델 무작위 분석 결과를 계산하였다. IVW로 계산된 OR은 0.76이었으며 MR-Egger와 Weighted median은 0.79, 0.76의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 심혈관계질환의 위험성이 24% 감소한다는 것을 의미한다.

표 17. 1-표본 멘델 무작위 분석 결과(교육 수준 - 심혈관계질환)

		OR	95% 신뢰구간
1 표준편차의 교육연한 (p-value 5e-5)	IVW	0.76	(0.66-0.85)
	MR-Egger	0.79	(0.69-0.89)
	Weighted median	0.76	(0.67-0.86)
1 표준편차의 교육연한 (p-value 5e-6)	IVW	0.74	(0.66-0.82)
	MR-Egger	0.75	(0.65-0.88)
	Weighted median	0.76	(0.61-0.95)
1 표준편차의 교육연한 (p-value 5e-7)	IVW	0.73	(0.65-0.83)
	MR-Egger	0.74	(0.68-0.80)
	Weighted median	0.73	(0.64-0.83)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 9]에 위 1-표본 멘델 무작위 분석 결과에 대한 그림을 함께 제시했다. [그림 9]에 따르면 가로축은 교육에 대한 SNP의 효과이고 세로축은 심혈관계질환에 대한 SNP의 효과이다. 교육 수준이 높을수록 심혈관계질환의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

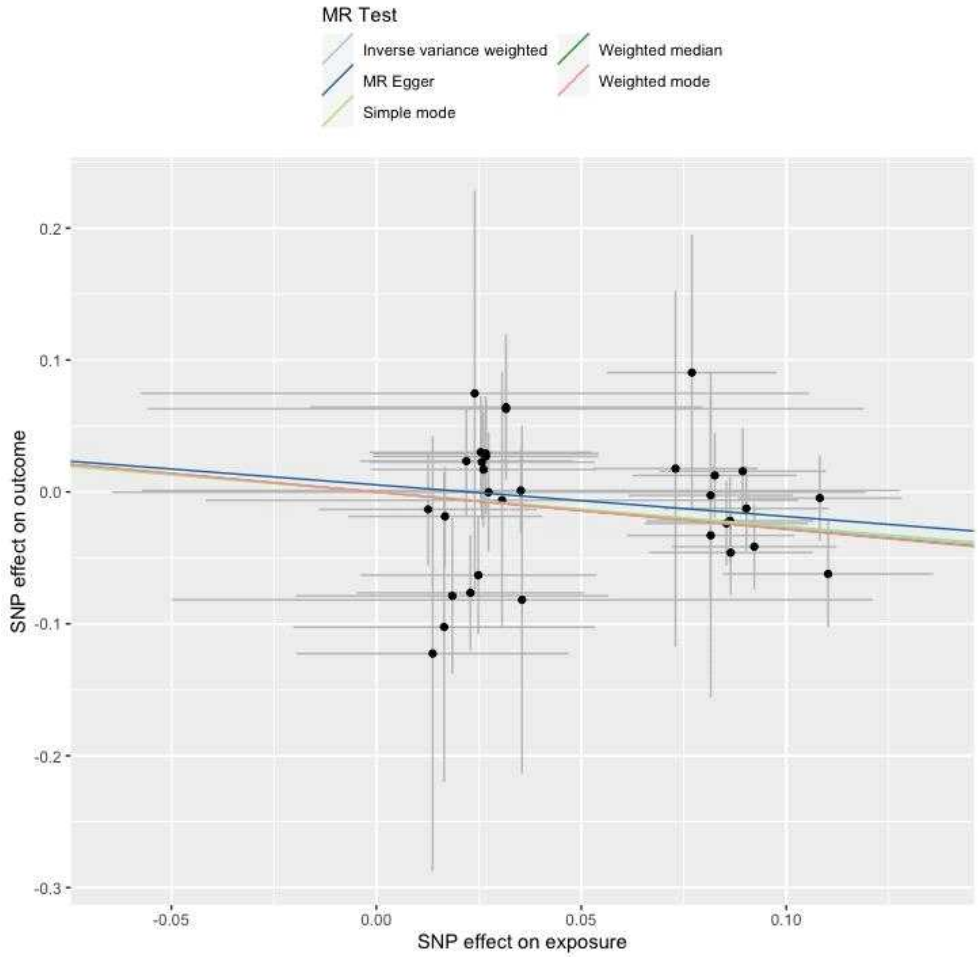

그림 9. 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 심혈관계질환)

표 18. 교육 수준과 심혈관계질환에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	-0.03	-0.06	0.04	0.05	6.34E-07	0.24
1	85410061	T	C	0.02	0.02	-0.07	-0.09	0.02	0.01	5.10E-06	0.38
3	140942232	A	G	0.48	0.49	-0.03	-0.001	0.09	0.03	2.99E-06	0.97
3	144026997	G	A	0.13	0.13	-0.03	-0.06	0.08	0.04	3.88E-06	0.19
3	153585628	G	T	0.01	0.01	-0.02	-0.07	0.08	0.01	2.12E-06	0.62
4	121613198	C	G	0.14	0.15	-0.02	0.002	0.09	0.04	2.80E-05	0.99
4	92431580	A	G	0.01	0.01	0.01	-0.12	0.03	0.01	1.53E-05	0.45
5	114210843	G	C	0.46	0.46	0.04	0.002	0.08	0.03	4.89E-06	0.93
5	54852576	T	C	0.01	0.01	0.07	0.017	0.01	0.01	2.22E-05	0.19
6	67132090	C	T	0.01	0.01	0.08	-0.03	0.02	0.01	3.98E-06	0.78
7	41147	A	G	0.01	0.02	0.03	-0.006	0.07	0.09	3.22E-05	0.94
12	110390979	C	T	0.07	0.07	0.01	-0.07	0.03	0.05	2.12E-06	0.18

12	111337494	C	CT	0.18	0.18	0.11	−0.06	0.02	0.04	3.17E−05	0.12
12	111414461	G	T	0.14	0.14	0.02	−0.06	0.02	0.04	2.85E−18	0.16
12	111684461	A	G	0.49	0.49	0.08	0.01	0.01	0.03	4.49E−05	0.69
12	111718231	A	C	0.18	0.18	0.02	0.02	0.02	0.04	2.31E−15	0.57
12	111902051	G	GGT	0.46	0.46	−0.09	0.01	0.02	0.03	3.21E−06	0.69
12	111994852	C	A	0.48	0.48	0.10	−0.004	0.02	0.03	2.29E−08	0.88
12	112168009	A	G	0.15	0.15	0.02	0.02	0.02	0.04	4.08E−20	0.50
12	112230019	C	G	0.15	0.15	0.02	0.02	0.02	0.04	3.58E−21	0.53
12	112468206	T	C	0.15	0.15	0.02	0.01	0.02	0.04	4.97E−20	0.69
12	112552274	T	C	0.23	0.23	0.01	−0.01	0.02	0.03	1.25E−11	0.62
12	112627350	A	G	0.16	0.16	0.02	0.03	0.02	0.04	6.32E−20	0.48
12	112678697	A	G	0.48	0.48	0.08	−0.04	0.01	0.03	3.75E−05	0.15
12	112736118	G	A	0.15	0.15	0.02	0.02	0.02	0.04	2.27E−21	0.60
12	112829103	C	CAT	0.42	0.42	−0.08	−0.01	0.02	0.03	3.22E−05	0.63
12	112923393	A	G	0.44	0.44	0.09	−0.04	0.02	0.03	4.53E−06	0.20

12	112982775	G	A	0.40	0.40	−0.08	0.02	0.02	0.03	3.99E−05	0.50
12	113117897	T	G	0.15	0.15	0.02	−0.07	0.02	0.04	8.52E−16	0.07
16	19380214	T	C	0.17	0.17	−0.01	0.01	0.02	0.04	1.41E−06	0.75
16	56729619	A	G	0.01	0.01	0.03	−0.08	0.08	0.01	6.46E−06	0.53
16	89653173	C	T	0.47	0.47	0.08	−0.02	0.01	0.03	3.84E−05	0.45
17	46502682	C	T	0.01	0.01	−0.01	0.10	0.03	0.011	5.60E−06	0.38
17	72917513	T	C	0.04	0.04	0.08	−0.002	0.01	0.07	2.46E−05	0.97

a.*ALT = effect allele, EAF = effect allele frequency

표 19. 교육 수준과 심혈관계질환에 대한 멘델 무작위 분석에서 confounder 검정 결과

Chromosome	position	ALT	REF	p-value					
				Exposure	Outcome	Smoking	Hypertension	Diabetes	Hyperlipidemia
1	245768933	G	A	6.34E-07	0.24	0.76	0.48	0.19	0.07
1	85410061	T	C	5.10E-06	0.38	0.38	0.27	0.78	0.94
3	140942232	A	G	2.99E-06	0.97	0.80	0.93	0.65	0.72
3	144026997	G	A	3.88E-06	0.19	0.75	0.16	0.52	0.85
3	153585628	G	T	2.12E-06	0.62	0.29	0.99	0.02	0.44
4	121613198	C	G	2.80E-05	0.99	0.76	0.14	0.23	0.93
4	92431580	A	G	1.53E-05	0.45	0.57	0.55	0.43	0.45
5	114210843	G	C	4.89E-06	0.93	0.36	0.11	0.66	0.83
5	54852576	T	C	2.22E-05	0.19	0.21	0.75	0.76	0.80
6	67132090	C	T	3.98E-06	0.78	0.75	0.79	0.22	0.50
7	41147	A	G	3.22E-05	0.94	0.22	0.60	0.47	0.34
12	110390979	C	T	2.12E-06	0.18	0.89	0.17	0.13	0.88

12	111337494	C	CT	3.17E-05	0.12	0.71	0.21	0.23	0.76
12	111414461	G	T	2.85E-18	0.16	0.71	0.23	0.93	0.20
12	111684461	A	G	4.49E-05	0.69	0.50	0.17	0.13	0.52
12	111718231	A	C	2.31E-15	0.57	0.67	0.13	0.22	0.16
12	111902051	G	GGT	3.21E-06	0.69	0.59	0.21	0.12	0.59
12	111994852	C	A	2.29E-08	0.88	0.59	0.56	0.38	0.85
12	112168009	A	G	4.08E-20	0.50	0.76	0.17	0.63	0.72
12	112230019	C	G	3.58E-21	0.53	0.73	0.16	0.49	0.80
12	112468206	T	C	4.97E-20	0.69	0.80	0.57	0.82	0.76
12	112552274	T	C	1.25E-11	0.62	0.52	0.13	0.56	0.60
12	112627350	A	G	6.32E-20	0.48	0.74	0.65	0.23	0.67
12	112678697	A	G	3.75E-05	0.15	0.37	0.51	0.64	0.30
12	112736118	G	A	2.27E-21	0.60	0.69	0.23	0.10	0.74
12	112829103	C	CAT	3.22E-05	0.63	0.97	0.13	0.13	0.06
12	112923393	A	G	4.53E-06	0.20	0.42	0.17	0.40	0.50
12	112982775	G	A	3.99E-05	0.50	0.91	0.50	0.15	0.07

12	113117897	T	G	8.52E-16	0.07	0.39	0.64	0.46	0.13
16	19380214	T	C	1.41E-06	0.75	0.67	0.49	0.54	0.13
16	56729619	A	G	6.46E-06	0.53	0.12	0.88	0.79	0.59
16	89653173	C	T	3.84E-05	0.45	0.25	0.63	0.13	0.15
17	46502682	C	T	5.60E-06	0.38	0.27	0.52	0.46	0.76
17	72917513	T	C	2.46E-05	0.97	0.56	0.92	0.26	0.37

*ALT = effect allele, EAF = effect allele frequency

3.2. 2-표본 멘델 무작위 분석 - 도시 코호트, 지역사회·농촌 코호트

이 연구에서 2-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 20. 이질성 검정 결과(교육 수준 - 심혈관계질환)

		Q	p-value
교육 연한	MR-Egger	9.1	0.99
	Inverse variance weighted	11.1	0.99

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 다형성 잔차항이 들어있는 것으로 잔차는 0.01로 나타났지만 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 21. 다형성 검정 결과(교육 수준 – 심혈관계질환)

		intercept	se	p-value
교육 연한	MR-Egger	-0.05	0.03	0.16

MR 분석과 민감도 분석을 [표 21]에 제시했다. p값에 따라 멘델 무작위 분석 결과를 계산하였다. IVW로 계산된 OR은 0.72이었으며 MR-Egger와 Weighted median은 0.76, 0.65의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 심혈관계질환의 위험성이 28% 감소한다는 것을 의미한다.

표 22. 2-표본 멘델 무작위 분석 결과(교육 수준 – 심혈관계질환)

		OR	95% 신뢰구간
1 표준편차의 교육 연한 (p-value 5e-5)	IVW	0.72	(0.68-0.75)
	MR-Egger	0.76	(0.69-0.84)
	Weighted median	0.65	(0.60-0.69)
1 표준편차의 교육 연한 (p-value 5e-6)	IVW	0.68	(0.62-0.75)
	MR-Egger	0.68	(0.57-0.81)
	Weighted median	0.68	(0.62-0.75)
1 표준편차의 교육 연한 (p-value 5e-7)	IVW	0.69	(0.61-0.79)
	MR-Egger	0.71	(0.61-0.84)
	Weighted median	0.74	(0.65-0.85)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 10]에 위 2-표본 멘델 무작위 분석 결과에 대한 그림을 함께 제시했다. [그림 10]에 따르면 가로축은 교육에 대한 SNP의 효과이고 세로축은 심혈관계질환에 대한 SNP의 효과이다. 교육 수준이 높을수록 심혈관계질환의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

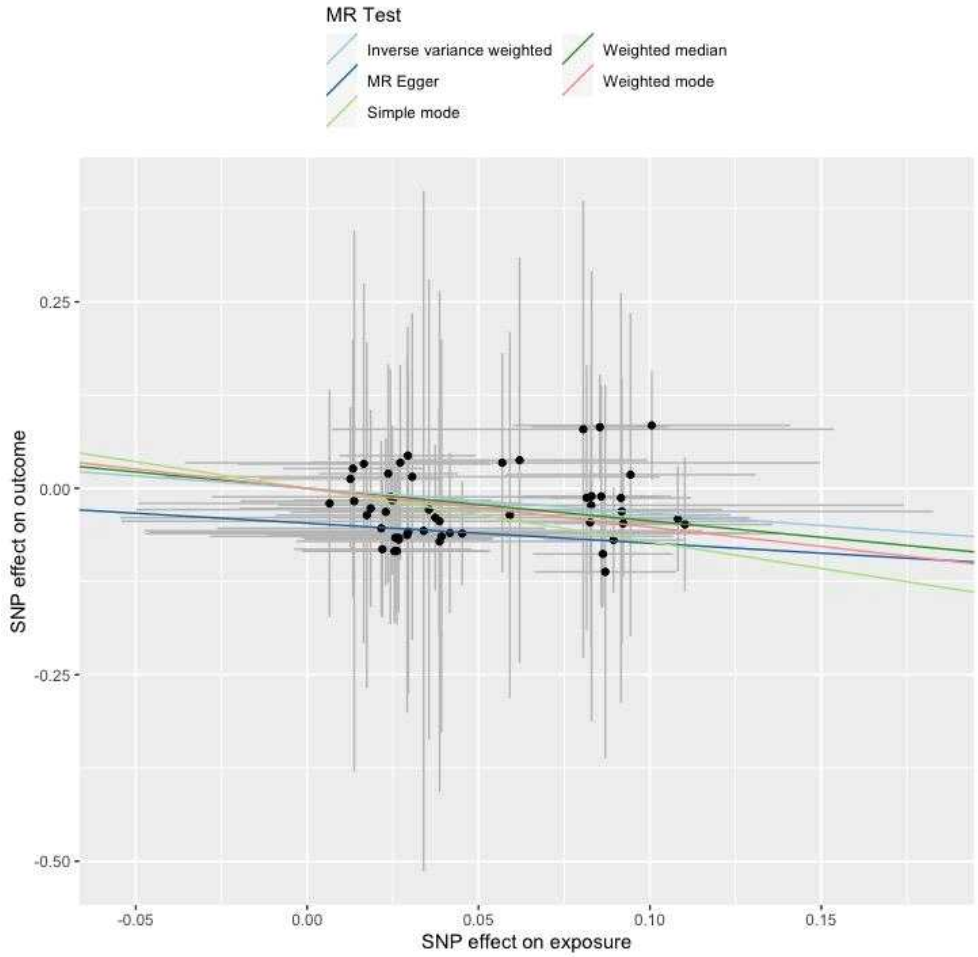

그림 10. 2-표본 멘델 무작위 분석의 산점도 (교육 수준 – 심혈관계질환)

표 23. 교육 수준과 심혈관계질환에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	0.02	−0.05	0.04	0.01	6.34E−06	0.64
1	85410061	T	C	0.02	0.02	0.08	−0.11	0.02	0.02	2.51E−05	0.65
3	140942232	A	G	0.48	0.48	−0.04	0.06	0.09	0.06	9.91E−07	0.38
3	144026997	G	A	0.13	0.13	−0.04	0.05	0.08	0.01	1.88E−06	0.35
3	153585628	G	T	0.01	0.01	−0.03	0.05	0.08	0.04	3.01E−05	0.21
4	121613198	C	G	0.14	0.14	−0.03	0.03	0.09	0.09	4.80E−05	0.68
4	92431580	A	G	0.01	0.01	0.01	−0.01	0.03	0.03	3.53E−05	0.63
5	114210843	G	C	0.46	0.46	0.03	0.009	0.08	0.06	4.89E−05	0.89
5	54852576	T	C	0.01	0.01	0.08	−0.01	0.01	0.03	3.22E−05	0.72
6	31322706	G	C	0.05	0.05	0.08	−0.01	0.02	0.01	2.56E−05	0.46
6	67132090	C	T	0.01	0.01	0.09	−0.01	0.02	0.02	6.98E−06	0.63
7	41147	A	G	0.01	0.01	0.03	0.01	0.07	0.02	2.22E−05	0.48

7	41170	A	G	0.03	0.03	−0.03	0.04	0.09	0.01	3.13E−05	0.34
7	81580	A	C	0.03	0.03	−0.02	0.01	0.05	0.01	3.02E−06	0.50
8	1135049	A	G	0.03	0.03	0.05	0.03	0.09	0.01	5.39E−05	0.19
8	1140691	G	C	0.04	0.04	0.02	−0.05	0.04	0.02	4.86E−06	0.70
8	1151923	G	C	0.02	0.02	−0.06	−0.03	0.03	0.02	9.61E−06	0.16
8	1189658	C	T	0.03	0.03	0.05	−0.03	0.06	0.02	3.57E−06	0.14
8	1192943	G	A	0.04	0.04	0.08	−0.02	0.09	0.01	3.64E−06	0.23
8	1326979	T	C	0.04	0.04	0.02	0.01	0.02	0.01	2.43E−06	0.17
8	1333258	G	A	0.31	0.31	−0.10	−0.08	0.04	0.07	1.29E−06	0.24
8	1360171	C	T	0.03	0.03	0.01	−0.03	0.02	0.02	5.14E−06	0.11
8	1472228	A	G	0.02	0.02	−0.08	−0.07	0.07	0.03	2.71E−07	0.98
8	1483124	A	C	0.02	0.02	0.02	0.04	0.01	0.01	1.40E−07	0.11
8	1485960	T	G	0.02	0.02	0.03	−0.07	0.04	0.03	3.53E−06	0.34
8	225168	T	C	0.04	0.04	0.02	−0.06	0.04	0.02	4.76E−05	0.87
8	234623	G	T	0.03	0.03	0.03	−0.06	0.06	0.02	5.61E−06	0.14
8	239111	C	T	0.03	0.03	0.01	0.02	0.02	0.01	5.13E−05	0.12

8	750906	T	C	0.05	0.05	0.02	0.03	0.02	0.01	2.95E-06	0.88
8	761499	T	C	0.04	0.04	0.006	-0.02	0.05	0.01	9.06E-05	0.18
8	772139	T	C	0.02	0.02	-0.09	0.03	0.09	0.01	3.13E-06	0.83
8	888744	T	C	0.03	0.03	-0.09	-0.01	0.03	0.02	9.66E-05	0.39
12	110390979	C	T	0.07	0.07	0.01	-0.02	0.03	0.01	1.17E-06	0.83
12	111337494	C	CT	0.18	0.18	0.11	-0.04	0.02	0.08	1.65E-05	0.95
12	111414461	G	T	0.14	0.14	0.02	-0.01	0.02	0.01	4.85E-18	0.87
12	111684461	A	G	0.49	0.49	0.08	-0.04	0.01	0.006	3.49E-05	0.51
12	111718231	A	C	0.18	0.18	0.02	-0.08	0.02	0.09	2.31E-17	0.37
12	111994852	C	A	0.48	0.48	0.10	-0.04	0.02	0.07	6.88E-08	0.56
12	112168009	A	G	0.15	0.15	0.02	-0.06	0.02	0.09	3.08E-22	0.49
12	112230019	C	G	0.15	0.15	0.02	-0.06	0.02	0.09	2.58E-22	0.48
12	112468206	T	C	0.15	0.15	0.02	-0.08	0.02	0.09	1.97E-21	0.39
12	112627350	A	G	0.16	0.16	0.02	-0.08	0.02	0.09	4.52E-21	0.38
12	112736118	G	A	0.15	0.15	0.02	-0.06	0.02	0.09	7.37E-21	0.49
12	112829103	C	CA	0.42	0.42	0.08	-0.06	0.02	0.07	1.12E-05	0.32

			T								
12	112923393	A	G	0.44	0.44	0.09	−0.04	0.02	0.07	4.53E−06	0.50
12	112982775	G	A	0.40	0.40	0.08	−0.08	0.02	0.07	2.29E−05	0.21
12	113117897	T	G	0.15	0.15	0.02	−0.03	0.02	0.09	1.52E−16	0.74
16	19380214	T	C	0.17	0.17	−0.01	−0.01	0.02	0.09	2.01E−06	0.18
16	56729619	A	G	0.01	0.01	0.03	−0.02	0.08	0.03	3.26E−05	0.92
16	89653173	C	T	0.47	0.47	0.08	0.08	0.01	0.07	1.84E−05	0.55
17	46502682	C	T	0.01	0.01	−0.01	−0.03	0.03	0.02	7.04E−06	0.17
17	72917513	T	C	0.04	0.04	0.08	−0.01	0.01	0.01	4.46E−05	0.47

4. 멘델 무작위 분석 결과 - 심혈관계질환 위험요인

4.1. 1-표본 멘델 무작위 분석

4.1.1. 고혈압

이 연구에서 1-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 24. 이질성 검정 결과(교육 수준 - 고혈압)

		Q	p-value
교육 연한	MR-Egger	33.8	0.33
	Inverse variance weighted	34.3	0.36

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 잔차항이 들어있는 것으로 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 25. 다형성 검정 결과(교육 수준 - 고혈압)

		intercept	se	p-value
교육 연한	MR-Egger	0.005	0.01	0.50

MR 분석과 민감도 분석을 [표 26]에 제시했다. IVW로 계산된 OR은 0.72이었으며 MR-Egger와 Weighted median은 0.75, 0.75의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 심혈관계질환의 위험성이 28% 감소한다는 것을 의미한다.

표 26. 1-표본 멘델 무작위 분석 결과(교육 수준 - 고혈압)

		OR	95% 신뢰구간
1 표준편차의 교육 연한 (p-value <5e-5)	IVW	0.72	(0.65-0.80)
	MR-Egger	0.75	(0.66-0.85)
	Weighted median	0.75	(0.63-0.90)
1 표준편차의 교육 연한 (p-value <5e-6)	IVW	0.76	(0.68-0.83)
	MR-Egger	0.78	(0.69-0.84)
	Weighted median	0.75	(0.68-0.84)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 11]에 위 멘델 무작위 분석 결과에 대한 그림을 함께 제시했다. [그림 11]에 따르면 가로축은 교육에 대한 SNP의 효과이고 세로축은 고혈압에 대한 SNP의 효과이다. 교육 수준이 높을수록 고혈압의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

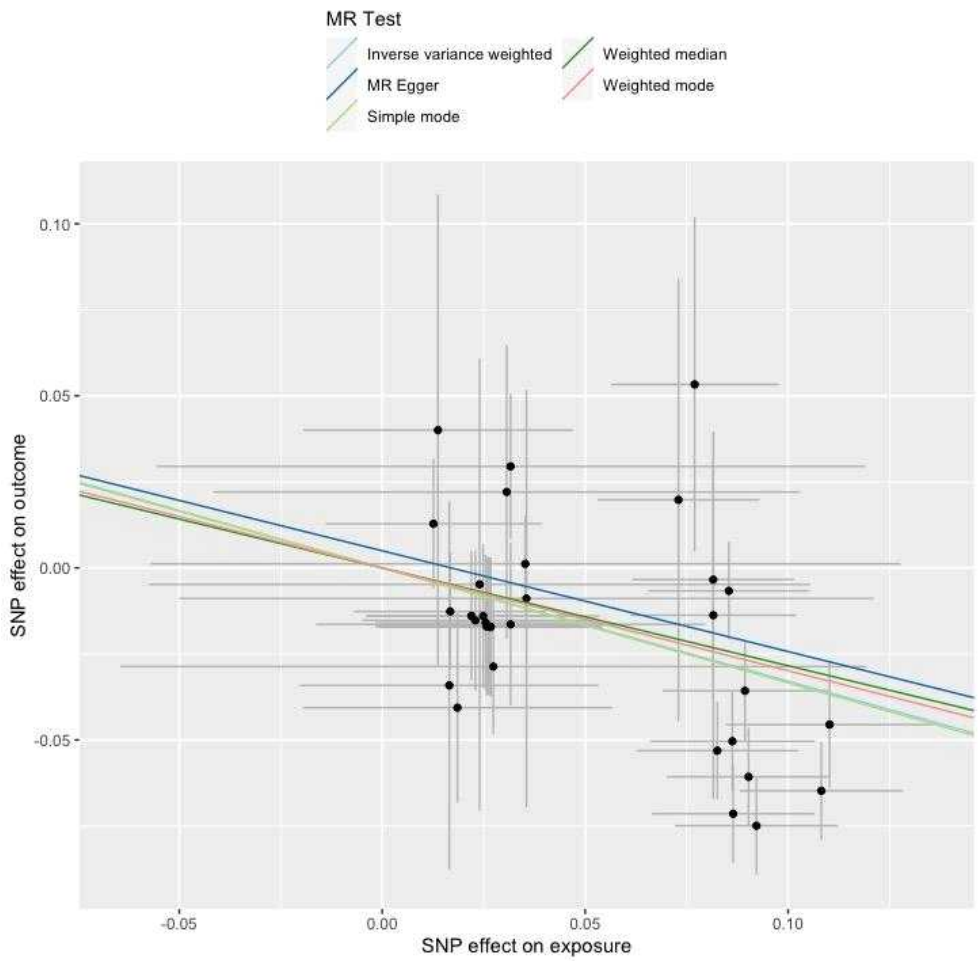

그림 11. 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 고혈압)

표 27. 교육 수준과 고혈압에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	−0.03	0.01	0.04	0.02	6.34E−07	0.48
1	85410061	T	C	0.02	0.02	−0.07	−0.05	0.02	0.04	5.10E−06	0.27
3	140942232	A	G	0.48	0.49	−0.03	−0.001	0.09	0.01	2.99E−06	0.93
3	144026997	G	A	0.13	0.13	−0.03	−0.02	0.08	0.02	3.88E−06	0.16
3	153585628	G	T	0.01	0.01	−0.02	0.004	0.08	0.06	2.12E−06	0.99
4	121613198	C	G	0.14	0.15	−0.02	0.02	0.09	0.01	2.80E−05	0.14
4	92431580	A	G	0.01	0.01	0.01	0.04	0.03	0.06	1.53E−05	0.55
5	114210843	G	C	0.46	0.46	0.04	0.05	0.08	0.01	4.89E−06	0.11
5	54852576	T	C	0.01	0.01	0.07	0.01	0.01	0.06	2.22E−05	0.75
6	67132090	C	T	0.01	0.01	0.08	−0.01	0.02	0.05	3.98E−06	0.79
7	41147	A	G	0.01	0.02	0.03	0.02	0.07	0.04	3.22E−05	0.60
12	110390979	C	T	0.07	0.07	0.01	−0.04	0.03	0.02	2.12E−06	0.17

12	111337494	C	CT	0.18	0.18	0.11	−0.04	0.02	0.01	3.17E−05	0.21
12	111414461	G	T	0.14	0.14	0.02	−0.01	0.02	0.02	2.85E−18	0.23
12	111684461	A	G	0.49	0.49	0.08	−0.05	0.01	0.01	4.49E−05	0.17
12	111718231	A	C	0.18	0.18	0.02	−0.01	0.02	0.01	2.31E−15	0.13
12	111902051	G	GGT	0.46	0.46	−0.09	0.06	0.02	0.01	3.21E−06	0.21
12	111994852	C	A	0.48	0.48	0.10	−0.06	0.02	0.01	2.29E−08	0.56
12	112168009	A	G	0.15	0.15	0.02	−0.01	0.02	0.02	4.08E−20	0.17
12	112230019	C	G	0.15	0.15	0.02	−0.01	0.02	0.02	3.58E−21	0.16
12	112468206	T	C	0.15	0.15	0.02	−0.01	0.02	0.02	4.97E−20	0.57
12	112552274	T	C	0.23	0.23	0.01	−0.01	0.02	0.01	1.25E−11	0.13
12	112627350	A	G	0.16	0.16	0.02	−0.01	0.02	0.01	6.32E−20	0.65
12	112678697	A	G	0.48	0.48	0.08	−0.07	0.01	0.01	3.75E−05	0.51
12	112736118	G	A	0.15	0.15	0.02	−0.01	0.02	0.02	2.27E−21	0.23
12	112829103	C	CAT	0.42	0.42	−0.08	0.03	0.02	0.01	3.22E−05	0.13
12	112923393	A	G	0.44	0.44	0.09	−0.07	0.02	0.01	4.53E−06	0.17

12	112982775	G	A	0.40	0.40	−0.08	0.05	0.02	0.01	3.99E−05	0.50
12	113117897	T	G	0.15	0.15	0.02	−0.01	0.02	0.02	8.52E−16	0.64
16	19380214	T	C	0.17	0.17	−0.01	−0.01	0.02	0.01	1.41E−06	0.49
16	56729619	A	G	0.01	0.01	0.03	−0.008	0.08	0.06	6.46E−06	0.88
16	89653173	C	T	0.47	0.47	0.08	−0.006	0.01	0.01	3.84E−05	0.63
17	46502682	C	T	0.01	0.01	−0.01	0.03	0.03	0.05	5.60E−06	0.52
17	72917513	T	C	0.04	0.04	0.08	−0.003	0.01	0.03	2.46E−05	0.92

표 28. 교육 수준과 고혈압에 대한 멘델 무작위 분석에서 confounder 검정 결과

Chromosome	position	ALT	REF	p-value			
				Exposure	Outcome	Smoking	Obesity
1	245768933	G	A	6.34E-07	0.48	0.76	0.41
1	85410061	T	C	5.10E-06	0.27	0.38	0.44
3	140942232	A	G	2.99E-06	0.93	0.80	0.51
3	144026997	G	A	3.88E-06	0.16	0.75	0.18
3	153585628	G	T	2.12E-06	0.99	0.29	0.65
4	121613198	C	G	2.80E-05	0.14	0.76	0.42
4	92431580	A	G	1.53E-05	0.55	0.57	0.69
5	114210843	G	C	4.89E-06	0.11	0.36	0.80
5	54852576	T	C	2.22E-05	0.75	0.21	0.14
6	67132090	C	T	3.98E-06	0.79	0.75	0.85
7	41147	A	G	3.22E-05	0.60	0.22	0.85
12	110390979	C	T	2.12E-06	0.17	0.89	0.74

12	111337494	C	CT	3.17E-05	0.21	0.71	0.34
12	111414461	G	T	2.85E-18	0.23	0.71	0.53
12	111684461	A	G	4.49E-05	0.17	0.50	0.14
12	111718231	A	C	2.31E-15	0.13	0.67	0.20
12	111902051	G	GGT	3.21E-06	0.21	0.59	0.08
12	111994852	C	A	2.29E-08	0.56	0.59	0.13
12	112168009	A	G	4.08E-20	0.17	0.76	0.69
12	112230019	C	G	3.58E-21	0.16	0.73	0.55
12	112468206	T	C	4.97E-20	0.57	0.80	0.23
12	112552274	T	C	1.25E-11	0.13	0.52	0.86
12	112627350	A	G	6.32E-20	0.65	0.74	0.41
12	112678697	A	G	3.75E-05	0.51	0.37	0.30
12	112736118	G	A	2.27E-21	0.23	0.69	0.35
12	112829103	C	CAT	3.22E-05	0.13	0.97	0.74
12	112923393	A	G	4.53E-06	0.17	0.42	0.34
12	112982775	G	A	3.99E-05	0.50	0.91	0.48

12	113117897	T	G	8.52E-16	0.64	0.39	0.10
16	19380214	T	C	1.41E-06	0.49	0.67	0.93
16	56729619	A	G	6.46E-06	0.88	0.12	0.81
16	89653173	C	T	3.84E-05	0.63	0.25	0.58
17	46502682	C	T	5.60E-06	0.52	0.27	0.37
17	72917513	T	C	2.46E-05	0.92	0.56	0.18

4.1.2. 당뇨병

이 연구에서 1-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 29. 이질성 검정 결과(교육 수준 - 당뇨병)

		Q	p-value
교육 연한	MR-Egger	18.1	0.97
	Inverse variance weighted	19.2	0.96

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 잔차항이 들어있는 것으로 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 30. 다형성 검정 결과(교육 수준 - 당뇨병)

		intercept	se	p-value
교육 연한	MR-Egger	0.01	0.01	0.30

멘델 무작위 분석과 민감도 분석을 [표 31]에 제시했다. IVW로 계산된 OR은 0.69이었으며 MR-Egger와 Weighted median은 0.72, 0.73의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 심혈관계질환의 위험성이 31% 감소한다는 것을 의미한다.

표 31. 1-표본 멘델 무작위 분석 결과(교육 수준 - 당뇨병)

		OR	95% 신뢰구간
1 표준편차의 교육 연한 (p-value <5e-5)	IVW	0.69	(0.58-0.80)
	MR-Egger	0.72	(0.64-0.82)
	Weighted median	0.73	(0.65-0.83)
1 표준편차의 교육 연한 (p-value <5e-6)	IVW	0.72	(0.62-0.83)
	MR-Egger	0.74	(0.63-0.85)
	Weighted median	0.71	(0.62-0.81)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 12]에 위 MR 결과에 대한 그림을 함께 제시했다. [그림 12]에 따르면 가로축은 교육에 대한 SNP의 효과이고 세로축은 당뇨병에 대한 SNP의 효과이다. 교육 수준이 높을수록 당뇨병의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

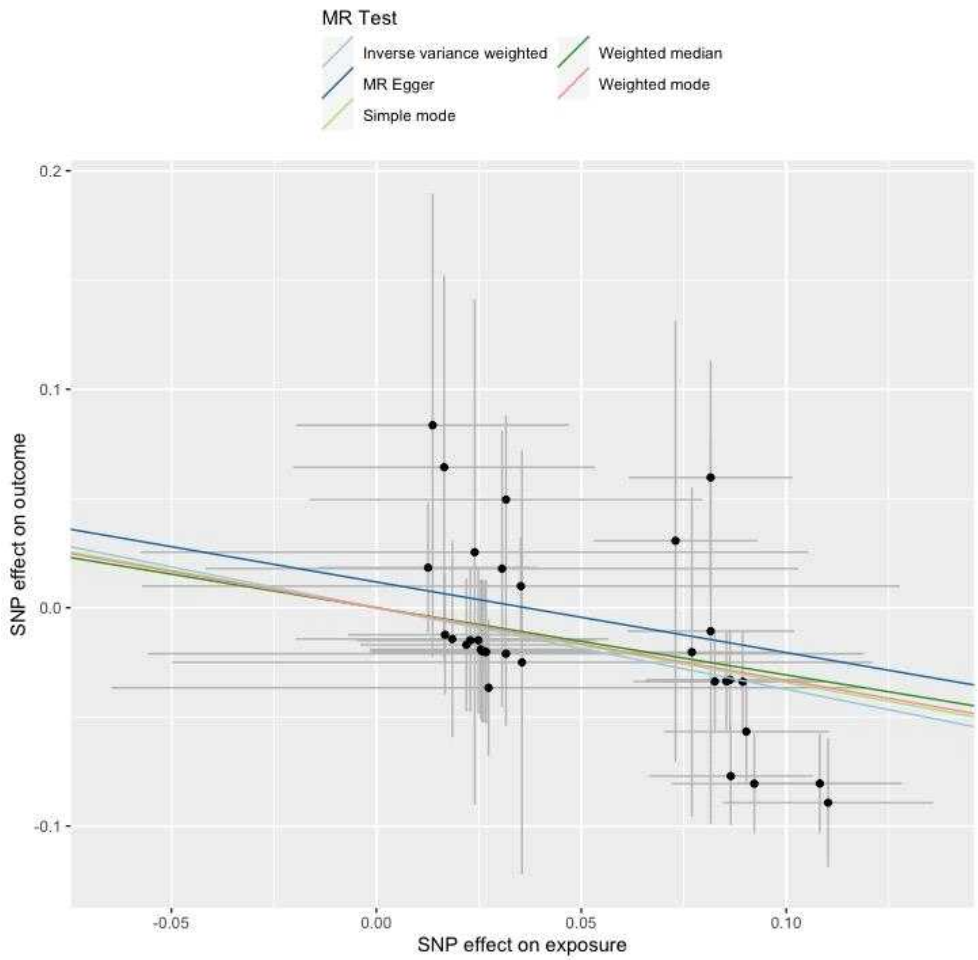

그림 12. 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 당뇨병)

표 32. 교육 수준과 당뇨병에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	−0.03	−0.04	0.04	0.03	6.34E−07	0.19
1	85410061	T	C	0.02	0.02	−0.07	0.02	0.02	0.07	5.10E−06	0.78
3	140942232	A	G	0.48	0.49	−0.03	−0.009	0.09	0.02	2.99E−06	0.65
3	144026997	G	A	0.13	0.13	−0.03	0.02	0.08	0.03	3.88E−06	0.52
3	153585628	G	T	0.01	0.01	−0.02	−0.02	0.08	0.01	2.12E−06	0.02
4	121613198	C	G	0.14	0.15	−0.02	0.03	0.09	0.03	2.80E−05	0.23
4	92431580	A	G	0.01	0.01	0.01	0.08	0.03	0.01	1.53E−05	0.43
5	114210843	G	C	0.46	0.46	0.04	−0.009	0.08	0.02	4.89E−06	0.66
5	54852576	T	C	0.01	0.01	0.07	0.03	0.01	0.01	2.22E−05	0.76
6	67132090	C	T	0.01	0.01	0.08	−0.01	0.02	0.08	3.98E−06	0.22
7	41147	A	G	0.01	0.02	0.03	0.01	0.07	0.06	3.22E−05	0.47
12	110390979	C	T	0.07	0.07	0.01	−0.01	0.03	0.04	2.12E−06	0.13

12	111337494	C	CT	0.18	0.18	0.11	−0.08	0.02	0.02	3.17E−05	0.23
12	111414461	G	T	0.14	0.14	0.02	−0.01	0.02	0.03	2.85E−18	0.93
12	111684461	A	G	0.49	0.49	0.08	−0.03	0.01	0.02	4.49E−05	0.13
12	111718231	A	C	0.18	0.18	0.02	−0.01	0.02	0.03	2.31E−15	0.22
12	111902051	G	G GT	0.46	0.46	−0.09	0.05	0.02	0.02	3.21E−06	0.12
12	111994852	C	A	0.48	0.48	0.10	−0.08	0.02	0.02	2.29E−08	0.38
12	112168009	A	G	0.15	0.15	0.02	−0.02	0.02	0.03	4.08E−20	0.63
12	112230019	C	G	0.15	0.15	0.02	−0.02	0.02	0.03	3.58E−21	0.49
12	112468206	T	C	0.15	0.15	0.02	−0.02	0.02	0.03	4.97E−20	0.82
12	112552274	T	C	0.23	0.23	0.01	−0.01	0.02	0.02	1.25E−11	0.56
12	112627350	A	G	0.16	0.16	0.02	−0.01	0.02	0.03	6.32E−20	0.23
12	112678697	A	G	0.48	0.48	0.08	−0.07	0.01	0.02	3.75E−05	0.64
12	112736118	G	A	0.15	0.15	0.02	−0.01	0.02	0.03	2.27E−21	0.10
12	112829103	C	CA T	0.42	0.42	−0.08	0.03	0.02	0.02	3.22E−05	0.13
12	112923393	A	G	0.44	0.44	0.09	−0.08	0.02	0.02	4.53E−06	0.40

12	112982775	G	A	0.40	0.40	−0.08	0.03	0.02	0.02	3.99E−05	0.15
12	113117897	T	G	0.15	0.15	0.02	−0.01	0.02	0.03	8.52E−16	0.46
16	19380214	T	C	0.17	0.17	−0.01	−0.01	0.02	0.02	1.41E−06	0.54
16	56729619	A	G	0.01	0.01	0.03	−0.02	0.08	0.09	6.46E−06	0.79
16	89653173	C	T	0.47	0.47	0.08	−0.03	0.01	0.02	3.84E−05	0.13
17	46502682	C	T	0.01	0.01	−0.01	−0.06	0.03	0.08	5.60E−06	0.46
17	72917513	T	C	0.04	0.04	0.08	0.05	0.01	0.05	2.46E−05	0.26

표 33. 교육 수준과 당뇨병에 대한 멘델 무작위 분석에서 confounder 검정 결과

Chromosome	position	ALT	REF	p-value			
				Exposure	Outcome	Smoking	Obesity
1	245768933	G	A	6.34E-07	0.19	0.76	0.41
1	85410061	T	C	5.10E-06	0.78	0.38	0.44
3	140942232	A	G	2.99E-06	0.65	0.80	0.51
3	144026997	G	A	3.88E-06	0.52	0.75	0.18
3	153585628	G	T	2.12E-06	0.02	0.29	0.65
4	121613198	C	G	2.80E-05	0.23	0.76	0.42
4	92431580	A	G	1.53E-05	0.43	0.57	0.69
5	114210843	G	C	4.89E-06	0.66	0.36	0.80
5	54852576	T	C	2.22E-05	0.76	0.21	0.14
6	67132090	C	T	3.98E-06	0.22	0.75	0.85
7	41147	A	G	3.22E-05	0.47	0.22	0.85
12	110390979	C	T	2.12E-06	0.13	0.89	0.74

12	111337494	C	CT	3.17E-05	0.23	0.71	0.34
12	111414461	G	T	2.85E-18	0.93	0.71	0.53
12	111684461	A	G	4.49E-05	0.13	0.50	0.14
12	111718231	A	C	2.31E-15	0.22	0.67	0.20
12	111902051	G	GGT	3.21E-06	0.12	0.59	0.08
12	111994852	C	A	2.29E-08	0.38	0.59	0.13
12	112168009	A	G	4.08E-20	0.63	0.76	0.69
12	112230019	C	G	3.58E-21	0.49	0.73	0.55
12	112468206	T	C	4.97E-20	0.82	0.80	0.23
12	112552274	T	C	1.25E-11	0.56	0.52	0.86
12	112627350	A	G	6.32E-20	0.23	0.74	0.41
12	112678697	A	G	3.75E-05	0.64	0.37	0.30
12	112736118	G	A	2.27E-21	0.10	0.69	0.35
12	112829103	C	CAT	3.22E-05	0.13	0.97	0.74
12	112923393	A	G	4.53E-06	0.40	0.42	0.34
12	112982775	G	A	3.99E-05	0.15	0.91	0.48

12	113117897	T	G	8.52E-16	0.46	0.39	0.10
16	19380214	T	C	1.41E-06	0.54	0.67	0.93
16	56729619	A	G	6.46E-06	0.79	0.12	0.81
16	89653173	C	T	3.84E-05	0.13	0.25	0.58
17	46502682	C	T	5.60E-06	0.46	0.27	0.37
17	72917513	T	C	2.46E-05	0.26	0.56	0.18

4.1.3. 고지혈증

이 연구에서 1-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 34. 이질성 검정 결과(교육 수준 - 고지혈증)

		Q	p-value
교육 연한	MR-Egger	34.8	0.29
	Inverse variance weighted	38.9	0.19

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 잔차항이 들어있는 것으로 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 35. 이질성 검정 결과(교육 수준 – 고지혈증)

		intercept	se	p-value
교육 연한	MR-Egger	0.01	0.01	0.07

멘델 무작위 분석과 민감도 분석을 [표 36]에 제시했다. IVW로 계산된 OR은 0.89이었으며 MR-Egger와 Weighted median은 0.85, 0.93의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 고지혈증의 위험성이 11% 감소한다는 것을 의미한다.

표 36. 1-표본 멘델 무작위 분석 결과(교육 수준 – 고지혈증)

		OR	95% 신뢰구간
1 표준편차의 교육 연한 (p-value <5e-5)	IVW	0.89	(0.81-0.98)
	MR-Egger	0.85	(0.79-0.90)
	Weighted median	0.93	(0.87-0.99)
1 표준편차의 교육 연한 (p-value <5e-6)	IVW	0.89	(0.80-0.97)
	MR-Egger	0.86	(0.78-0.91)
	Weighted median	0.92	(0.86-0.98)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 13]에 위 MR 결과에 대한 그림을 함께 제시했다. [그림 13]에 따르면 가로축은 교육에 대한 SNP의 크기이고 세로축은 고지혈증에 대한 SNP의 크기이다. 교육 수준이 높을수록 고지혈증의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

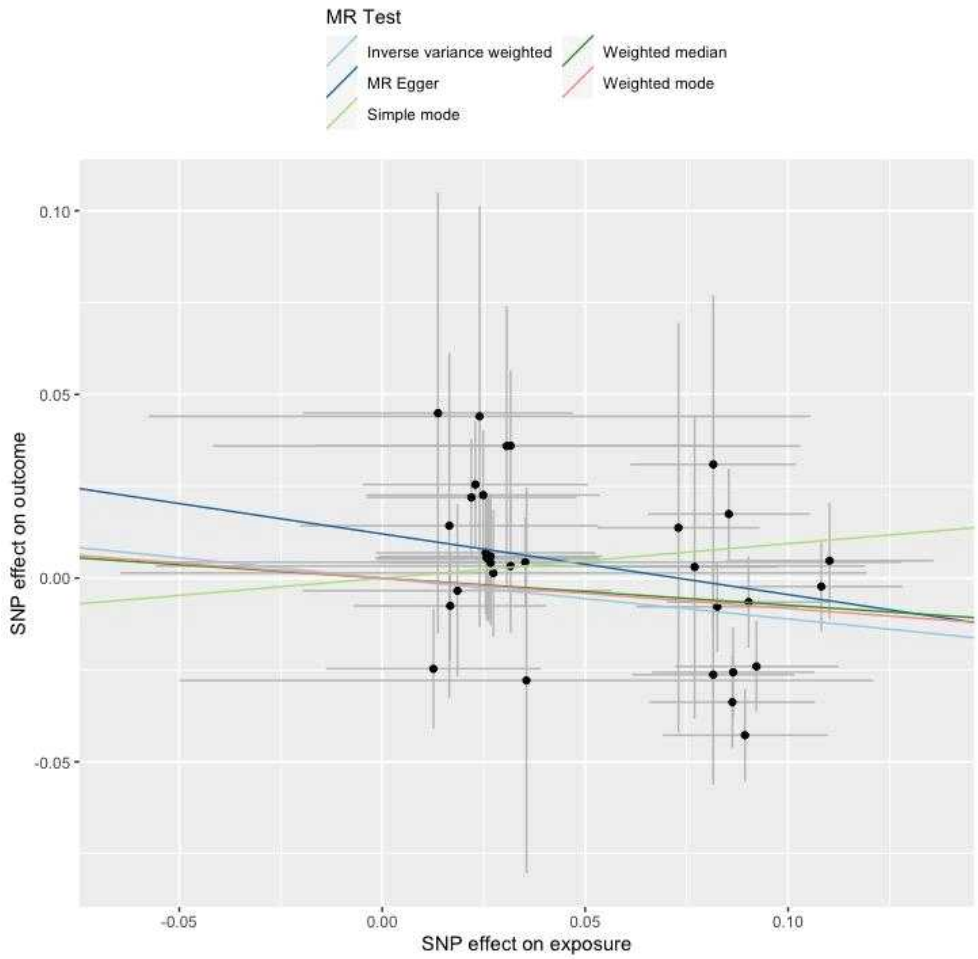

그림 13. 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 고지혈증)

표 37. 교육 수준과 고지혈증에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	-0.03	-0.03	0.04	0.02	6.34E-07	0.07
1	85410061	T	C	0.02	0.02	-0.07	-0.003	0.02	0.04	5.10E-06	0.94
3	140942232	A	G	0.48	0.49	-0.03	-0.004	0.09	0.01	2.99E-06	0.72
3	144026997	G	A	0.13	0.13	-0.03	-0.003	0.08	0.01	3.88E-06	0.85
3	153585628	G	T	0.01	0.01	-0.02	-0.04	0.08	0.05	2.12E-06	0.44
4	121613198	C	G	0.14	0.15	-0.02	-0.001	0.09	0.01	2.80E-05	0.93
4	92431580	A	G	0.01	0.01	0.01	0.04	0.03	0.06	1.53E-05	0.45
5	114210843	G	C	0.46	0.46	0.04	0.002	0.08	0.01	4.89E-06	0.83
5	54852576	T	C	0.01	0.01	0.07	0.01	0.01	0.05	2.22E-05	0.80
6	67132090	C	T	0.01	0.01	0.08	0.03	0.02	0.04	3.98E-06	0.50
7	41147	A	G	0.01	0.02	0.03	0.03	0.07	0.03	3.22E-05	0.34
12	110390979	C	T	0.07	0.07	0.01	-0.003	0.03	0.02	2.12E-06	0.88

12	111337494	C	CT	0.18	0.18	0.11	0.004	0.02	0.01	3.17E-05	0.76
12	111414461	G	T	0.14	0.14	0.02	0.02	0.02	0.01	2.85E-18	0.20
12	111684461	A	G	0.49	0.49	0.08	-0.007	0.01	0.01	4.49E-05	0.52
12	111718231	A	C	0.18	0.18	0.02	0.02	0.02	0.01	2.31E-15	0.16
12	111902051	G	GGT	0.46	0.46	-0.09	0.006	0.02	0.01	3.21E-06	0.59
12	111994852	C	A	0.48	0.48	0.10	-0.002	0.02	0.01	2.29E-08	0.85
12	112168009	A	G	0.15	0.15	0.02	0.005	0.02	0.01	4.08E-20	0.72
12	112230019	C	G	0.15	0.15	0.02	0.004	0.02	0.01	3.58E-21	0.80
12	112468206	T	C	0.15	0.15	0.02	0.005	0.02	0.01	4.97E-20	0.76
12	112552274	T	C	0.23	0.23	0.01	-0.007	0.02	0.01	1.25E-11	0.60
12	112627350	A	G	0.16	0.16	0.02	0.006	0.02	0.01	6.32E-20	0.67
12	112678697	A	G	0.48	0.48	0.08	-0.02	0.01	0.01	3.75E-05	0.30
12	112736118	G	A	0.15	0.15	0.02	0.005	0.02	0.01	2.27E-21	0.74
12	112829103	C	CAT	0.42	0.42	-0.08	0.042765	0.02	0.012546	3.22E-05	0.06
12	112923393	A	G	0.44	0.44	0.09	-0.02404	0.02	0.012425	4.53E-06	0.50

12	112982775	G	A	0.40	0.40	−0.08	0.033764	0.02	0.012584	3.99E−05	0.07
12	113117897	T	G	0.15	0.15	0.02	0.025459	0.02	0.017157	8.52E−16	0.13
16	19380214	T	C	0.17	0.17	−0.01	0.02469	0.02	0.016359	1.41E−06	0.13
16	56729619	A	G	0.01	0.01	0.03	−0.02785	0.08	0.05242	6.46E−06	0.59
16	89653173	C	T	0.47	0.47	0.08	0.01744	0.01	0.012321	3.84E−05	0.15
17	46502682	C	T	0.01	0.01	−0.01	−0.01427	0.03	0.046959	5.60E−06	0.76
17	72917513	T	C	0.04	0.04	0.08	−0.02629	0.01	0.029871	2.46E−05	0.37

표 38. 교육 수준과 고지혈증에 대한 멘델 무작위 분석에서 confounder 검정 결과

Chromosome	position	ALT	REF	p-value			
				Exposure	Outcome	Smoking	Obesity
1	245768933	G	A	6.34E-07	0.07	0.76	0.41
1	85410061	T	C	5.10E-06	0.94	0.38	0.44
3	140942232	A	G	2.99E-06	0.72	0.80	0.51
3	144026997	G	A	3.88E-06	0.85	0.75	0.18
3	153585628	G	T	2.12E-06	0.44	0.29	0.65
4	121613198	C	G	2.80E-05	0.93	0.76	0.42
4	92431580	A	G	1.53E-05	0.45	0.57	0.69
5	114210843	G	C	4.89E-06	0.83	0.36	0.80
5	54852576	T	C	2.22E-05	0.80	0.21	0.14
6	67132090	C	T	3.98E-06	0.50	0.75	0.85
7	41147	A	G	3.22E-05	0.34	0.22	0.85
12	110390979	C	T	2.12E-06	0.88	0.89	0.74

12	111337494	C	CT	3.17E-05	0.76	0.71	0.34
12	111414461	G	T	2.85E-18	0.20	0.71	0.53
12	111684461	A	G	4.49E-05	0.52	0.50	0.14
12	111718231	A	C	2.31E-15	0.16	0.67	0.20
12	111902051	G	GGT	3.21E-06	0.59	0.59	0.08
12	111994852	C	A	2.29E-08	0.85	0.59	0.13
12	112168009	A	G	4.08E-20	0.72	0.76	0.69
12	112230019	C	G	3.58E-21	0.80	0.73	0.55
12	112468206	T	C	4.97E-20	0.76	0.80	0.23
12	112552274	T	C	1.25E-11	0.60	0.52	0.86
12	112627350	A	G	6.32E-20	0.67	0.74	0.41
12	112678697	A	G	3.75E-05	0.30	0.37	0.30
12	112736118	G	A	2.27E-21	0.74	0.69	0.35
12	112829103	C	CAT	3.22E-05	0.06	0.97	0.74
12	112923393	A	G	4.53E-06	0.50	0.42	0.34
12	112982775	G	A	3.99E-05	0.07	0.91	0.48

12	113117897	T	G	8.52E-16	0.13	0.39	0.10
16	19380214	T	C	1.41E-06	0.13	0.67	0.93
16	56729619	A	G	6.46E-06	0.59	0.12	0.81
16	89653173	C	T	3.84E-05	0.15	0.25	0.58
17	46502682	C	T	5.60E-06	0.76	0.27	0.37
17	72917513	T	C	2.46E-05	0.37	0.56	0.18

4.1.4. 비만 - 체질량지수 25 kg/m^2 이상

이 연구에서 1-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 39. 이질성 검정 결과(교육 수준 - 비만, 체질량지수 25kg/m^2 이상)

		Q	p-value
교육 연한	MR-Egger	36.4	0.40
	Inverse variance weighted	21.5	0.68

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 잔차항이 들어있는 것으로 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 40. 다형성 검정 결과(교육 수준 – 비만, 체질량지수 25kg/m^2 이상)

		intercept	se	p-value
교육 연한	MR-Egger	-0.03	0.01	0.44

멘델 무작위 분석과 민감도 분석을 [표 41]에 제시했다. IVW로 계산된 OR은 0.90이었으며 MR-Egger와 Weighted median은 0.90, 0.90의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 비만의 위험성이 10% 감소한다는 것을 의미한다.

표 41. 1-표본 멘델 무작위 분석 결과(교육 수준 – 비만, 체질량지수 25kg/m^2 이상)

		OR	95% 신뢰구간
	IVW	0.90	(0.85-0.95)
1 표준편차의 교육 연한	MR-Egger	0.90	(0.84-0.97)
(p-value <5e-5)	Weighted median	0.90	(0.84-0.97)
	IVW	0.89	(0.84-0.94)
1 표준편차의 교육 연한	MR-Egger	0.91	(0.86-0.98)
(p-value <5e-6)	Weighted median	0.91	(0.85-0.97)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 14]에 위 1-표본 멘델 무작위 분석 결과에 대한 그림을 함께 제시했다. [그림 14]에 따르면 가로축은 교육에 대한 SNP의 크기이

고 세로축은 비만에 대한 SNP의 크기이다. 교육 수준이 높을수록 비만의 위험성이 감소하기 때문에 그래프의 기울기는 음이다. IVW의 직선 기울기보다 MR-Egger 기울기가 매우 완만하게 나타났다.

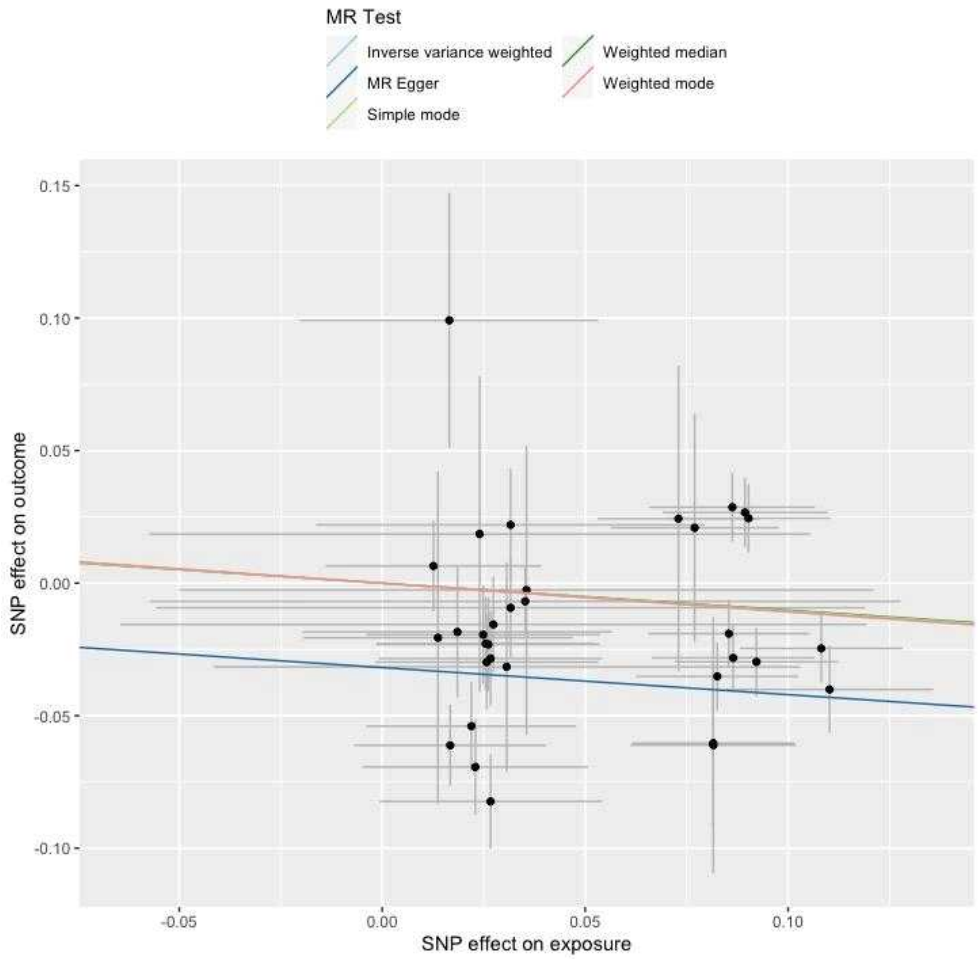

그림 14. 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 비만, 체질량지수 25kg/m^2 이상)

표 42. 교육 수준과 비만에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	-0.03	-0.02	0.04	0.02	6.34E-07	0.30
1	85410061	T	C	0.02	0.02	-0.07	-0.02	0.02	0.04	5.10E-06	0.62
3	140942232	A	G	0.48	0.49	-0.03	0.006	0.09	0.01	2.99E-06	0.58
3	144026997	G	A	0.13	0.13	-0.03	0.009	0.08	0.01	3.88E-06	0.96
3	153585628	G	T	0.01	0.01	-0.02	-0.01	0.08	0.05	2.12E-06	0.75
4	121613198	C	G	0.14	0.15	-0.02	0.01	0.09	0.01	2.80E-05	0.38
4	92431580	A	G	0.01	0.01	0.01	-0.02	0.03	0.06	1.53E-05	0.74
5	114210843	G	C	0.46	0.46	0.04	-0.01	0.08	0.01	4.89E-06	0.88
5	54852576	T	C	0.01	0.01	0.07	0.02	0.01	0.05	2.22E-05	0.67
6	67132090	C	T	0.01	0.01	0.08	-0.06	0.02	0.04	3.98E-06	0.20
7	41147	A	G	0.01	0.02	0.03	-0.03	0.07	0.03	3.22E-05	0.42
12	110390979	C	T	0.07	0.07	0.01	-0.01	0.03	0.02	2.12E-06	0.71

12	111337494	C	CT	0.18	0.18	0.11	-0.04	0.02	0.01	3.17E-05	0.14
12	111414461	G	T	0.14	0.14	0.02	-0.01	0.02	0.01	2.85E-18	0.19
12	111684461	A	G	0.49	0.49	0.08	-0.03	0.01	0.01	4.49E-05	0.60
12	111718231	A	C	0.18	0.18	0.02	-0.05	0.02	0.01	2.31E-15	0.12
12	111902051	G	GGT	0.46	0.46	-0.09	-0.02	0.02	0.01	3.21E-06	0.58
12	111994852	C	A	0.48	0.48	0.10	-0.02	0.02	0.01	2.29E-08	0.55
12	112168009	A	G	0.15	0.15	0.02	-0.02	0.02	0.01	4.08E-20	0.10
12	112230019	C	G	0.15	0.15	0.02	-0.08	0.02	0.01	3.58E-21	0.37
12	112468206	T	C	0.15	0.15	0.02	-0.02	0.02	0.01	4.97E-20	0.31
12	112552274	T	C	0.23	0.23	0.01	-0.06	0.02	0.01	1.25E-11	0.60
12	112627350	A	G	0.16	0.16	0.02	-0.02	0.02	0.01	6.32E-20	0.31
12	112678697	A	G	0.48	0.48	0.08	-0.02	0.01	0.01	3.75E-05	0.28
12	112736118	G	A	0.15	0.15	0.02	-0.02	0.02	0.01	2.27E-21	0.73
12	112829103	C	CAT	0.42	0.42	-0.08	-0.02	0.02	0.01	3.22E-05	0.40
12	112923393	A	G	0.44	0.44	0.09	-0.02	0.02	0.01	4.53E-06	0.21
12	112982775	G	A	0.40	0.40	-0.08	-0.02	0.02	0.01	3.99E-05	0.28

12	113117897	T	G	0.15	0.15	0.02	−0.06	0.02	0.01	8.52E−16	0.11
16	19380214	T	C	0.17	0.17	−0.01	−0.006	0.02	0.01	1.41E−06	0.96
16	56729619	A	G	0.01	0.01	0.03	−0.002	0.08	0.05	6.46E−06	0.96
16	89653173	C	T	0.47	0.47	0.08	−0.01	0.01	0.01	3.84E−05	0.13
17	46502682	C	T	0.01	0.01	−0.01	−0.09	0.03	0.04	5.60E−06	0.39
17	72917513	T	C	0.04	0.04	0.08	−0.06	0.01	0.03	2.46E−05	0.53

표 43. 교육 수준과 비만에 대한 멘델 무작위 분석에서 confounder 검정 결과

Chromosome	position	ALT	REF	p-value					
				Exposure	Outcome	Smoking	Diabetes	Hypertension	Hyperlipidemia
1	245768933	G	A	6.34E-07	0.30	0.76	0.19	0.48	0.07
1	85410061	T	C	5.10E-06	0.62	0.38	0.78	0.27	0.94
3	140942232	A	G	2.99E-06	0.58	0.80	0.65	0.93	0.72
3	144026997	G	A	3.88E-06	0.96	0.75	0.52	0.16	0.85
3	153585628	G	T	2.12E-06	0.75	0.29	0.02	0.99	0.44
4	121613198	C	G	2.80E-05	0.38	0.76	0.23	0.14	0.93
4	92431580	A	G	1.53E-05	0.74	0.57	0.43	0.55	0.45
5	114210843	G	C	4.89E-06	0.88	0.36	0.66	0.11	0.83
5	54852576	T	C	2.22E-05	0.67	0.21	0.76	0.75	0.80
6	67132090	C	T	3.98E-06	0.20	0.75	0.22	0.79	0.50
7	41147	A	G	3.22E-05	0.42	0.22	0.47	0.60	0.34
12	110390979	C	T	2.12E-06	0.71	0.89	0.13	0.17	0.88

12	111337494	C	CT	3.17E-05	0.14	0.71	0.23	0.21	0.76
12	111414461	G	T	2.85E-18	0.19	0.71	0.93	0.23	0.20
12	111684461	A	G	4.49E-05	0.60	0.50	0.13	0.17	0.52
12	111718231	A	C	2.31E-15	0.12	0.67	0.22	0.13	0.16
12	111902051	G	GGT	3.21E-06	0.58	0.59	0.12	0.21	0.59
12	111994852	C	A	2.29E-08	0.55	0.59	0.38	0.56	0.85
12	112168009	A	G	4.08E-20	0.10	0.76	0.63	0.17	0.72
12	112230019	C	G	3.58E-21	0.37	0.73	0.49	0.16	0.80
12	112468206	T	C	4.97E-20	0.31	0.80	0.82	0.57	0.76
12	112552274	T	C	1.25E-11	0.60	0.52	0.56	0.13	0.60
12	112627350	A	G	6.32E-20	0.31	0.74	0.23	0.65	0.67
12	112678697	A	G	3.75E-05	0.28	0.37	0.64	0.51	0.30
12	112736118	G	A	2.27E-21	0.73	0.69	0.10	0.23	0.74
12	112829103	C	CAT	3.22E-05	0.40	0.97	0.13	0.13	0.06
12	112923393	A	G	4.53E-06	0.21	0.42	0.40	0.17	0.50

12	112982775	G	A	3.99E-05	0.28	0.91	0.15	0.50	0.07
12	113117897	T	G	8.52E-16	0.11	0.39	0.46	0.64	0.13
16	19380214	T	C	1.41E-06	0.96	0.67	0.54	0.49	0.13
16	56729619	A	G	6.46E-06	0.96	0.12	0.79	0.88	0.59
16	89653173	C	T	3.84E-05	0.13	0.25	0.13	0.63	0.15
17	46502682	C	T	5.60E-06	0.39	0.27	0.46	0.52	0.76
17	72917513	T	C	2.46E-05	0.53	0.56	0.26	0.92	0.37

4.1.5. 비만 - 체질량지수 30 kg/m^2 이상

이 연구에서 1-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 44. 이질성 검정 결과(교육 수준 - 비만, 체질량지수 30kg/m^2 이상)

		Q	p-value
교육 연한	MR-Egger	22.6	0.86
	Inverse variance weighted	23.5	0.86

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 잔차항이 들어있는 것으로 잔차의 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 45. 다형성 검정 결과(교육 수준 – 비만, 체질량지수 30kg/m^2 이상)

		intercept	se	p-value
교육 연한	MR-Egger	-0.01	0.01	0.34

멘델 무작위 분석과 민감도 분석을 [표 46]에 제시했다. IVW로 계산된 OR은 0.85이었으며 MR-Egger와 Weighted median은 0.88, 0.86의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 비만의 위험성이 15% 감소한다는 것을 의미한다.

표 46. 1-표본 멘델 무작위 분석 결과(교육 수준 – 비만, 체질량지수 30kg/m^2 이상)

		OR	95% 신뢰구간
1 표준편차의 교육 연한 (p-value <5e-5)	IVW	0.85	(0.83-0.87)
	MR-Egger	0.88	(0.84-0.91)
	Weighted median	0.86	(0.83-0.89)
1 표준편차의 교육 연한 (p-value <5e-6)	IVW	0.87	(0.84-0.90)
	MR-Egger	0.86	(0.82-0.91)
	Weighted median	0.87	(0.83-0.91)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 15]에 위 1-표본 멘델 무작위 분석 결과에 대한 그림을 함께 제시했다. [그림 15]에 따르면 가로축은 교육에 대한 SNP의 효과이고 세로축은 비만에 대한 SNP의 효과이다. 교육 수준이 높을수록 비만의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

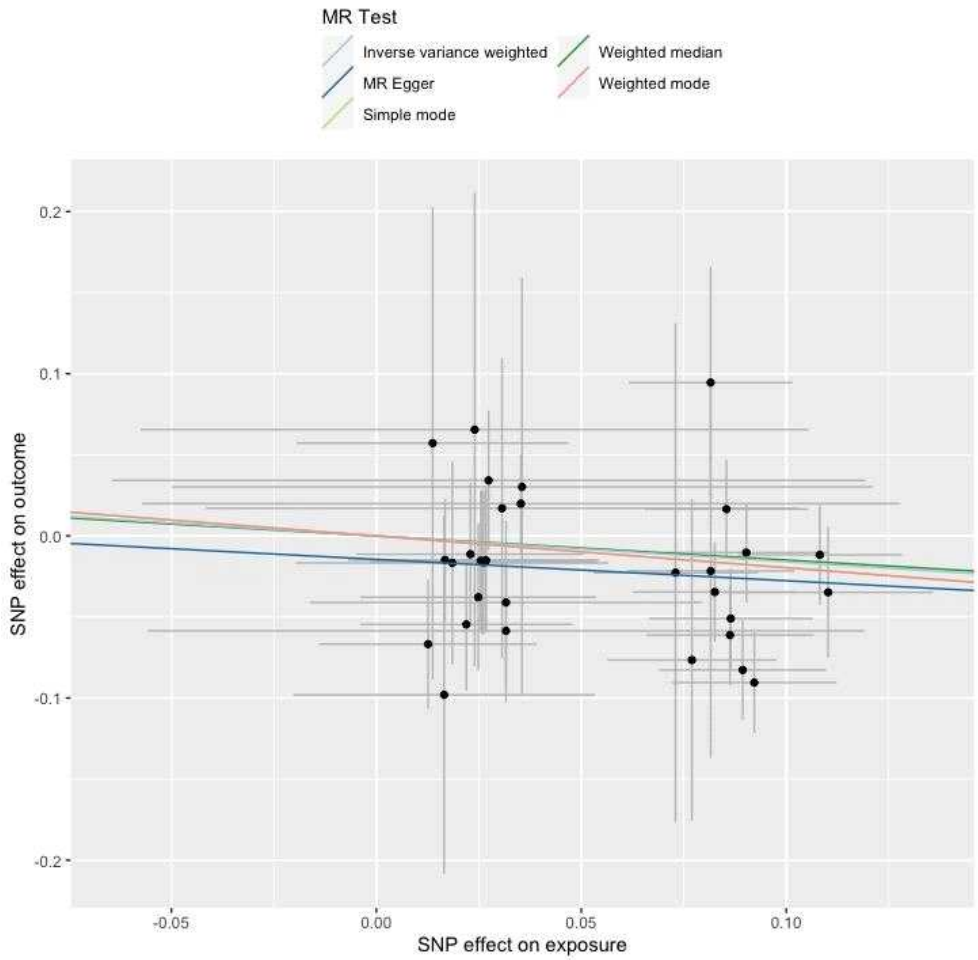

그림 15. 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 비만, 체질량지수 30kg/m^2 이상)

표 47. 교육 수준과 비만에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	−0.03	0.04	0.04	0.05	6.34E−07	0.41
1	85410061	T	C	0.02	0.02	−0.07	0.07	0.02	0.009	5.10E−06	0.44
3	140942232	A	G	0.48	0.49	−0.03	−0.01	0.09	0.03	2.99E−06	0.51
3	144026997	G	A	0.13	0.13	−0.03	0.05	0.08	0.04	3.88E−06	0.18
3	153585628	G	T	0.01	0.01	−0.02	−0.06	0.08	0.01	2.12E−06	0.65
4	121613198	C	G	0.14	0.15	−0.02	−0.03	0.09	0.04	2.80E−05	0.42
4	92431580	A	G	0.01	0.01	0.01	0.05	0.03	0.01	1.53E−05	0.69
5	114210843	G	C	0.46	0.46	0.04	−0.007	0.08	0.03	4.89E−06	0.80
5	54852576	T	C	0.01	0.01	0.07	−0.022	0.01	0.01	2.22E−05	0.14
6	67132090	C	T	0.01	0.01	0.08	−0.02	0.02	0.01	3.98E−06	0.85
7	41147	A	G	0.01	0.02	0.03	0.01	0.07	0.009	3.22E−05	0.85
12	110390979	C	T	0.07	0.07	0.01	−0.01	0.03	0.06	2.12E−06	0.74

12	111337494	C	CT	0.18	0.18	0.11	-0.03	0.02	0.04	3.17E-05	0.34
12	111414461	G	T	0.14	0.14	0.02	-0.03	0.02	0.04	2.85E-18	0.53
12	111684461	A	G	0.49	0.49	0.08	-0.03	0.01	0.03	4.49E-05	0.14
12	111718231	A	C	0.18	0.18	0.02	-0.05	0.02	0.04	2.31E-15	0.20
12	111902051	G	GGT	0.46	0.46	-0.09	0.01	0.02	0.03	3.21E-06	0.08
12	111994852	C	A	0.48	0.48	0.10	-0.01	0.02	0.03	2.29E-08	0.13
12	112168009	A	G	0.15	0.15	0.02	-0.01	0.02	0.04	4.08E-20	0.69
12	112230019	C	G	0.15	0.15	0.02	-0.01	0.02	0.04	3.58E-21	0.55
12	112468206	T	C	0.15	0.15	0.02	-0.01	0.02	0.04	4.97E-20	0.23
12	112552274	T	C	0.23	0.23	0.01	-0.01	0.02	0.03	1.25E-11	0.86
12	112627350	A	G	0.16	0.16	0.02	-0.01	0.02	0.04	6.32E-20	0.41
12	112678697	A	G	0.48	0.48	0.08	-0.05	0.01	0.03	3.75E-05	0.30
12	112736118	G	A	0.15	0.15	0.02	-0.01	0.02	0.04	2.27E-21	0.35
12	112829103	C	CAT	0.42	0.42	-0.08	0.08	0.02	0.03	3.22E-05	0.74
12	112923393	A	G	0.44	0.44	0.09	-0.09	0.02	0.03	4.53E-06	0.34

12	112982775	G	A	0.40	0.40	−0.08	0.06	0.02	0.03	3.99E−05	0.48
12	113117897	T	G	0.15	0.15	0.02	−0.01	0.02	0.04	8.52E−16	0.10
16	19380214	T	C	0.17	0.17	−0.01	0.06	0.02	0.03	1.41E−06	0.93
16	56729619	A	G	0.01	0.01	0.03	0.03	0.08	0.01	6.46E−06	0.81
16	89653173	C	T	0.47	0.47	0.08	0.01	0.01	0.03	3.84E−05	0.58
17	46502682	C	T	0.01	0.01	−0.01	0.09	0.03	0.01	5.60E−06	0.37
17	72917513	T	C	0.04	0.04	0.08	0.09	0.01	0.07	2.46E−05	0.18

표 48. 교육 수준과 비만에 대한 멘델 무작위 분석에서 confounder 검정 결과

Chromosome	position	ALT	REF	p-value					
				Exposure	Outcome	Smoking	Diabetes	Hypertension	Hyperlipidemia
1	245768933	G	A	6.34E-07	0.41	0.76	0.19	0.48	0.07
1	85410061	T	C	5.10E-06	0.44	0.38	0.78	0.27	0.94
3	140942232	A	G	2.99E-06	0.51	0.80	0.65	0.93	0.72
3	144026997	G	A	3.88E-06	0.18	0.75	0.52	0.16	0.85
3	153585628	G	T	2.12E-06	0.65	0.29	0.02	0.99	0.44
4	121613198	C	G	2.80E-05	0.42	0.76	0.23	0.14	0.93
4	92431580	A	G	1.53E-05	0.69	0.57	0.43	0.55	0.45
5	114210843	G	C	4.89E-06	0.80	0.36	0.66	0.11	0.83
5	54852576	T	C	2.22E-05	0.14	0.21	0.76	0.75	0.80
6	67132090	C	T	3.98E-06	0.85	0.75	0.22	0.79	0.50
7	41147	A	G	3.22E-05	0.85	0.22	0.47	0.60	0.34

12	110390979	C	T	2.12E-06	0.74	0.89	0.13	0.17	0.88
12	111337494	C	CT	3.17E-05	0.34	0.71	0.23	0.21	0.76
12	111414461	G	T	2.85E-18	0.53	0.71	0.93	0.23	0.20
12	111684461	A	G	4.49E-05	0.14	0.50	0.13	0.17	0.52
12	111718231	A	C	2.31E-15	0.20	0.67	0.22	0.13	0.16
12	111902051	G	GGT	3.21E-06	0.08	0.59	0.12	0.21	0.59
12	111994852	C	A	2.29E-08	0.13	0.59	0.38	0.56	0.85
12	112168009	A	G	4.08E-20	0.69	0.76	0.63	0.17	0.72
12	112230019	C	G	3.58E-21	0.55	0.73	0.49	0.16	0.80
12	112468206	T	C	4.97E-20	0.23	0.80	0.82	0.57	0.76
12	112552274	T	C	1.25E-11	0.86	0.52	0.56	0.13	0.60
12	112627350	A	G	6.32E-20	0.41	0.74	0.23	0.65	0.67
12	112678697	A	G	3.75E-05	0.30	0.37	0.64	0.51	0.30
12	112736118	G	A	2.27E-21	0.35	0.69	0.10	0.23	0.74
12	112829103	C	CAT	3.22E-05	0.74	0.97	0.13	0.13	0.06

12	112923393	A	G	4.53E-06	0.34	0.42	0.40	0.17	0.50
12	112982775	G	A	3.99E-05	0.48	0.91	0.15	0.50	0.07
12	113117897	T	G	8.52E-16	0.10	0.39	0.46	0.64	0.13
16	19380214	T	C	1.41E-06	0.93	0.67	0.54	0.49	0.13
16	56729619	A	G	6.46E-06	0.81	0.12	0.79	0.88	0.59
16	89653173	C	T	3.84E-05	0.58	0.25	0.13	0.63	0.15
17	46502682	C	T	5.60E-06	0.37	0.27	0.46	0.52	0.76
17	72917513	T	C	2.46E-05	0.18	0.56	0.26	0.92	0.37

4.1.6. 흡연

이 연구에서 1-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 49. 이질성 검정 결과(교육 수준 - 흡연)

		Q	p-value
교육 연한	MR-Egger	10.8	0.99
	Inverse variance weighted	10.9	0.99

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 잔차항이 들어있는 것으로 잔차의 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 50. 다형성 검정 결과(교육 수준 – 흡연)

		intercept	se	p-value
교육 연한	MR-Egger	-0.003	0.01	0.79

멘델 무작위 분석과 민감도 분석을 [표 51]에 제시했다. IVW로 계산된 OR은 0.69이었으며 MR-Egger와 Weighted median은 0.71, 0.73의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 비만의 위험성이 31% 감소한다는 것을 의미한다.

표 51. 1-표본 멘델 무작위 분석 결과(교육 수준 – 흡연)

		OR	95% 신뢰구간
1 표준편차의 교육 연한 (p-value <5e-5)	IVW	0.69	(0.59-0.82)
	MR-Egger	0.71	(0.62-0.81)
	Weighted median	0.73	(0.58-0.92)
1 표준편차의 교육 연한 (p-value <5e-6)	IVW	0.68	(0.58-0.81)
	MR-Egger	0.70	(0.63-0.80)
	Weighted median	0.69	(0.59-0.88)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 16]에 위 1-표본 멘델 무작위 분석 결과에 대한 그림을 함께 제시했다. [그림 16]에 따르면 가로축은 교육에 대한 SNP의 효과이고 세로축은 흡연에 대한 SNP의 효과이다. 교육 수준이 높을수록 흡연의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

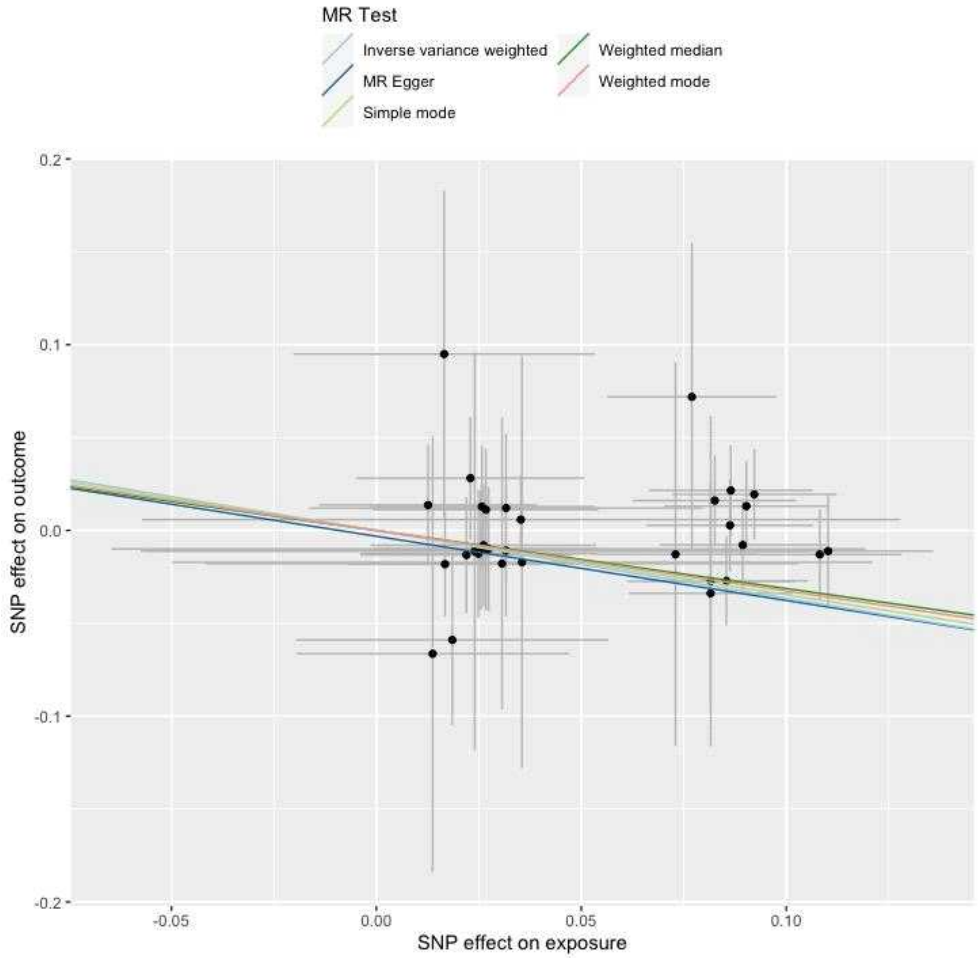

그림 16. 1-표본 멘델 무작위 분석의 산점도 (교육 수준 - 흡연)

표 52. 교육 수준과 흡연에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	−0.03	−0.01	0.04	0.04	6.34E−07	0.76
1	85410061	T	C	0.02	0.02	−0.07	−0.07	0.02	0.08	5.10E−06	0.38
3	140942232	A	G	0.48	0.49	−0.03	−0.005	0.09	0.02	2.99E−06	0.80
3	144026997	G	A	0.13	0.13	−0.03	0.01	0.08	0.03	3.88E−06	0.75
3	153585628	G	T	0.01	0.01	−0.02	0.01	0.08	0.01	2.12E−06	0.29
4	121613198	C	G	0.14	0.15	−0.02	0.009	0.09	0.03	2.80E−05	0.76
4	92431580	A	G	0.01	0.01	0.01	−0.06	0.03	0.11	1.53E−05	0.57
5	114210843	G	C	0.46	0.46	0.04	−0.02	0.08	0.02	4.89E−06	0.36
5	54852576	T	C	0.01	0.01	0.07	−0.01	0.01	0.10	2.22E−05	0.21
6	67132090	C	T	0.01	0.01	0.08	−0.02	0.02	0.08	3.98E−06	0.75
7	41147	A	G	0.01	0.02	0.03	−0.01	0.07	0.07	3.22E−05	0.22
12	110390979	C	T	0.07	0.07	0.01	−0.05	0.03	0.04	2.12E−06	0.89

12	111337494	C	CT	0.18	0.18	0.11	−0.01	0.02	0.03	3.17E−05	0.71
12	111414461	G	T	0.14	0.14	0.02	−0.01	0.02	0.03	2.85E−18	0.71
12	111684461	A	G	0.49	0.49	0.08	0.01	0.01	0.02	4.49E−05	0.50
12	111718231	A	C	0.18	0.18	0.02	−0.01	0.02	0.03	2.31E−15	0.67
12	111902051	G	GGT	0.46	0.46	−0.09	−0.01	0.02	0.02	3.21E−06	0.59
12	111994852	C	A	0.48	0.48	0.10	−0.01	0.02	0.02	2.29E−08	0.59
12	112168009	A	G	0.15	0.15	0.02	−0.009	0.02	0.03	4.08E−20	0.76
12	112230019	C	G	0.15	0.15	0.02	0.01	0.02	0.03	3.58E−21	0.73
12	112468206	T	C	0.15	0.15	0.02	−0.007	0.02	0.03	4.97E−20	0.80
12	112552274	T	C	0.23	0.23	0.01	−0.01	0.02	0.02	1.25E−11	0.52
12	112627350	A	G	0.16	0.16	0.02	−0.01	0.02	0.03	6.32E−20	0.74
12	112678697	A	G	0.48	0.48	0.08	0.02	0.01	0.02	3.75E−05	0.37
12	112736118	G	A	0.15	0.15	0.02	0.01	0.02	0.03	2.27E−21	0.69
12	112829103	C	CAT	0.42	0.42	−0.08	0.007	0.02	0.02	3.22E−05	0.97
12	112923393	A	G	0.44	0.44	0.09	0.01	0.02	0.02	4.53E−06	0.42

12	112982775	G	A	0.40	0.40	−0.08	−0.002	0.02	0.02	3.99E−05	0.91
12	113117897	T	G	0.15	0.15	0.02	0.02	0.02	0.03	8.52E−16	0.39
16	19380214	T	C	0.17	0.17	−0.01	−0.01	0.02	0.03	1.41E−06	0.67
16	56729619	A	G	0.01	0.01	0.03	−0.01	0.08	0.01	6.46E−06	0.12
16	89653173	C	T	0.47	0.47	0.08	−0.02	0.01	0.02	3.84E−05	0.25
17	46502682	C	T	0.01	0.01	−0.01	−0.09	0.03	0.08	5.60E−06	0.27
17	72917513	T	C	0.04	0.04	0.08	−0.03	0.01	0.05	2.46E−05	0.56

표 53. 교육 수준과 흡연에 대한 멘델 무작위 분석에서 confounder 검정 결과

Chromosome	position	ALT	REF	p-value					
				Exposure	Outcome	Diabetes	Hypertension	Hyperlipidemia	Obesity
1	245768933	G	A	6.34E-07	0.76	0.19	0.48	0.07	0.41
1	85410061	T	C	5.10E-06	0.38	0.78	0.27	0.94	0.44
3	140942232	A	G	2.99E-06	0.80	0.65	0.93	0.72	0.51
3	144026997	G	A	3.88E-06	0.75	0.52	0.16	0.85	0.18
3	153585628	G	T	2.12E-06	0.29	0.02	0.99	0.44	0.65
4	121613198	C	G	2.80E-05	0.76	0.23	0.14	0.93	0.42
4	92431580	A	G	1.53E-05	0.57	0.43	0.55	0.45	0.69
5	114210843	G	C	4.89E-06	0.36	0.66	0.11	0.83	0.80
5	54852576	T	C	2.22E-05	0.21	0.76	0.75	0.80	0.14
6	67132090	C	T	3.98E-06	0.75	0.22	0.79	0.50	0.85
7	41147	A	G	3.22E-05	0.22	0.47	0.60	0.34	0.85
12	110390979	C	T	2.12E-06	0.89	0.13	0.17	0.88	0.74

12	111337494	C	CT	3.17E-05	0.71	0.23	0.21	0.76	0.34
12	111414461	G	T	2.85E-18	0.71	0.93	0.23	0.20	0.53
12	111684461	A	G	4.49E-05	0.50	0.13	0.17	0.52	0.14
12	111718231	A	C	2.31E-15	0.67	0.22	0.13	0.16	0.20
12	111902051	G	GGT	3.21E-06	0.59	0.12	0.21	0.59	0.08
12	111994852	C	A	2.29E-08	0.59	0.38	0.56	0.85	0.13
12	112168009	A	G	4.08E-20	0.76	0.63	0.17	0.72	0.69
12	112230019	C	G	3.58E-21	0.73	0.49	0.16	0.80	0.55
12	112468206	T	C	4.97E-20	0.80	0.82	0.57	0.76	0.23
12	112552274	T	C	1.25E-11	0.52	0.56	0.13	0.60	0.86
12	112627350	A	G	6.32E-20	0.74	0.23	0.65	0.67	0.41
12	112678697	A	G	3.75E-05	0.37	0.64	0.51	0.30	0.30
12	112736118	G	A	2.27E-21	0.69	0.10	0.23	0.74	0.35
12	112829103	C	CAT	3.22E-05	0.97	0.13	0.13	0.06	0.74
12	112923393	A	G	4.53E-06	0.42	0.40	0.17	0.50	0.34

12	112982775	G	A	3.99E-05	0.91	0.15	0.50	0.07	0.48
12	113117897	T	G	8.52E-16	0.39	0.46	0.64	0.13	0.10
16	19380214	T	C	1.41E-06	0.67	0.54	0.49	0.13	0.93
16	56729619	A	G	6.46E-06	0.12	0.79	0.88	0.59	0.81
16	89653173	C	T	3.84E-05	0.25	0.13	0.63	0.15	0.58
17	46502682	C	T	5.60E-06	0.27	0.46	0.52	0.76	0.37
17	72917513	T	C	2.46E-05	0.56	0.26	0.92	0.37	0.18

4.2. 2-표본 멘델 무작위 분석 - 도시 코호트, 농촌·지역사회 코호트

4.2.1. 고혈압

이 연구에서 2-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 54. 이질성 검정 결과(교육 수준 - 고혈압)

		Q	p-value
교육 연한	MR-Egger	19.7	0.99
	Inverse variance weighted	20.1	0.99

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 잔차항이 들어있는 것으로 잔차의 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 55. 다형성 검정 결과(교육 수준 - 고혈압)

		intercept	se	p-value
교육 연한	MR-Egger	0.01	0.01	0.53

멘델 무작위 분석과 민감도 분석을 [표 56]에 제시했다. IVW로 계산된 OR은 0.76이었으며 MR-Egger와 Weighted median은 0.74, 0.72의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 고혈압의 위험성이 24% 감소한다는 것을 의미한다.

표 56. 2-표본 멘델 무작위 분석 결과(교육 수준 - 고혈압)

		OR	95% 신뢰구간
1 표준편차의 교육 연한 (p-value <5e-5)	IVW	0.76	(0.61-0.95)
	MR-Egger	0.74	(0.58-0.96)
	Weighted median	0.72	(0.58-0.90)
1 표준편차의 교육 연한 (p-value <5e-6)	IVW	0.77	(0.62-0.94)
	MR-Egger	0.78	(0.62-0.97)
	Weighted median	0.74	(0.60-0.91)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 17]에 위 2-표본 멘델 무작위 분석 결과에 대한 그림을 함께 제시했다. [그림 17]에 따르면 가로축은 교육에 대한 SNP의 크기이고 세로축은 고혈압에 대한 SNP의 크기이다. 교육 수준이 높을수록 고혈압의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

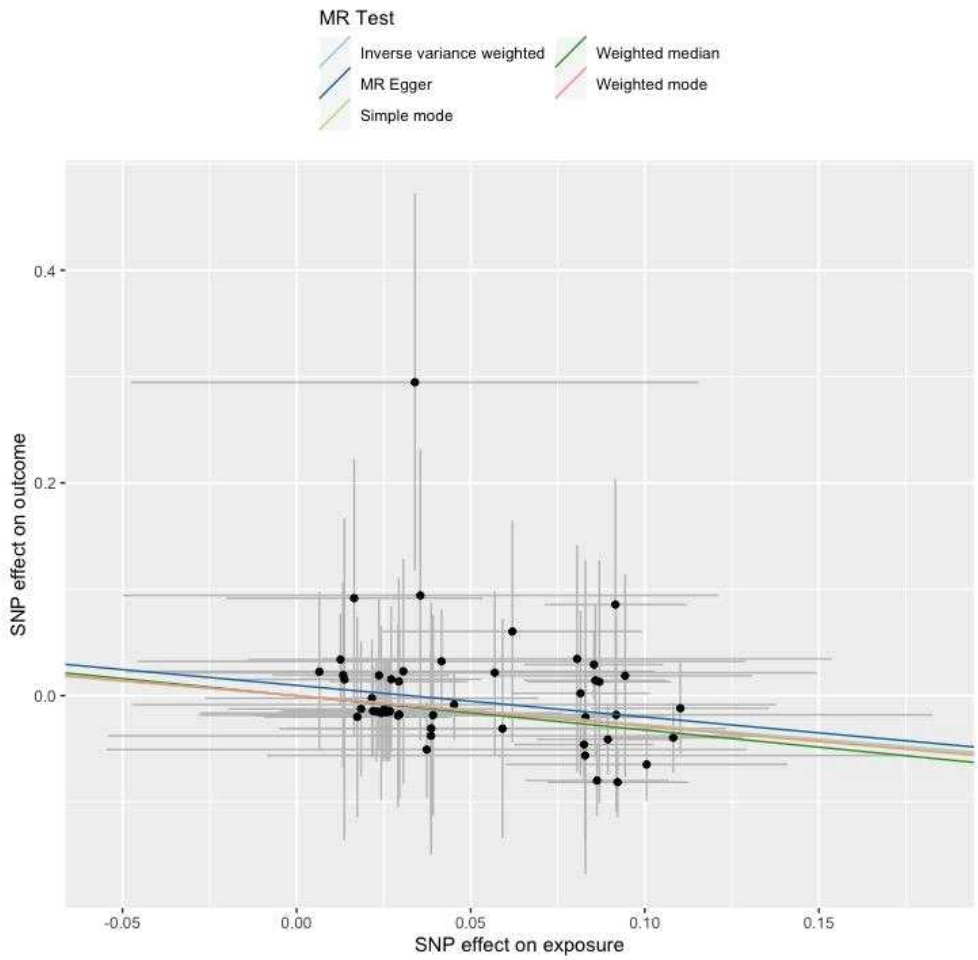

그림 17. 2-표본 멘델 무작위 분석의 산점도 (교육 수준 - 고혈압)

표 57. 교육 수준과 고혈압에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	0.02	0.002	0.04	0.05	6.34E-06	0.96
1	85410061	T	C	0.02	0.02	0.08	-0.01	0.02	0.01	2.51E-05	0.25
3	140942232	A	G	0.48	0.48	-0.04	0.008	0.09	0.03	9.91E-07	0.79
3	144026997	G	A	0.13	0.13	-0.04	-0.03	0.08	0.04	1.88E-06	0.50
3	153585628	G	T	0.01	0.01	-0.03	-0.02	0.08	0.01	3.01E-05	0.09
4	121613198	C	G	0.14	0.14	-0.03	0.05	0.09	0.04	4.80E-05	0.26
4	92431580	A	G	0.01	0.01	0.01	0.01	0.03	0.01	3.53E-05	0.31
5	114210843	G	C	0.46	0.46	0.03	0.05	0.08	0.03	4.89E-05	0.11
5	54852576	T	C	0.01	0.01	0.08	-0.02	0.01	0.01	3.22E-05	0.89
6	31322706	G	C	0.05	0.05	0.08	0.01	0.02	0.07	2.56E-05	0.84
6	67132090	C	T	0.01	0.01	0.09	0.08	0.02	0.01	6.98E-06	0.46
7	41147	A	G	0.01	0.01	0.03	0.02	0.07	0.01	2.22E-05	0.30

7	41170	A	G	0.03	0.03	-0.03	0.03	0.09	0.07	3.13E-05	0.17
7	81580	A	C	0.03	0.03	-0.02	0.01	0.05	0.08	3.02E-06	0.45
8	1135049	A	G	0.03	0.03	0.05	0.02	0.09	0.07	5.39E-05	0.50
8	1140691	G	C	0.04	0.04	0.02	-0.01	0.04	0.07	4.86E-06	0.25
8	1151923	G	C	0.02	0.02	-0.06	-0.06	0.03	0.01	9.61E-06	0.56
8	1189658	C	T	0.03	0.03	0.05	-0.03	0.06	0.01	3.57E-06	0.26
8	1192943	G	A	0.04	0.04	0.08	-0.05	0.09	0.07	3.64E-06	0.47
8	1326979	T	C	0.04	0.04	0.02	0.01	0.02	0.07	2.43E-06	0.77
8	1333258	G	A	0.31	0.31	-0.10	0.06	0.04	0.03	1.29E-06	0.59
8	1360171	C	T	0.03	0.03	0.01	-0.02	0.02	0.09	5.14E-06	0.33
8	1472228	A	G	0.02	0.02	-0.08	-0.03	0.07	0.01	2.71E-07	0.12
8	1483124	A	C	0.02	0.02	0.02	0.01	0.01	0.09	1.40E-07	0.17
8	1485960	T	G	0.02	0.02	0.03	-0.03	0.04	0.01	3.53E-06	0.94
8	225168	T	C	0.04	0.04	0.02	-0.01	0.04	0.08	4.76E-05	0.28
8	234623	G	T	0.03	0.03	0.03	-0.01	0.06	0.09	5.61E-06	0.51
8	239111	C	T	0.03	0.03	0.01	0.01	0.02	0.08	5.13E-05	0.26

8	750906	T	C	0.05	0.05	0.02	0.01	0.02	0.06	2.95E-06	0.24
8	761499	T	C	0.04	0.04	0.006	0.02	0.05	0.07	9.06E-05	0.25
8	772139	T	C	0.02	0.02	-0.09	0.01	0.09	0.09	3.13E-06	0.49
8	888744	T	C	0.03	0.03	-0.09	-0.01	0.03	0.09	9.66E-05	0.49
12	110390979	C	T	0.07	0.07	0.01	-0.01	0.03	0.06	1.17E-06	0.98
12	111337494	C	CT	0.18	0.18	0.11	-0.01	0.02	0.04	1.65E-05	0.54
12	111414461	G	T	0.14	0.14	0.02	-0.01	0.02	0.04	4.85E-18	0.91
12	111684461	A	G	0.49	0.49	0.08	-0.04	0.01	0.03	3.49E-05	0.15
12	111718231	A	C	0.18	0.18	0.02	-0.01	0.02	0.04	2.31E-17	0.78
12	111994852	C	A	0.48	0.48	0.10	-0.03	0.02	0.03	6.88E-08	0.22
12	112168009	A	G	0.15	0.15	0.02	-0.01	0.02	0.04	3.08E-22	0.19
12	112230019	C	G	0.15	0.15	0.02	-0.01	0.02	0.04	2.58E-22	0.11
12	112468206	T	C	0.15	0.15	0.02	-0.01	0.02	0.04	1.97E-21	0.10
12	112627350	A	G	0.16	0.16	0.02	-0.01	0.02	0.04	4.52E-21	0.25
12	112736118	G	A	0.15	0.15	0.02	-0.01	0.02	0.04	7.37E-21	0.73
12	112829103	C	CA	0.42	0.42	0.08	0.04	0.02	0.03	1.12E-05	0.20

			T								
12	112923393	A	G	0.44	0.44	0.09	−0.08	0.02	0.03	4.53E−06	0.13
12	112982775	G	A	0.40	0.40	0.08	0.07	0.02	0.03	2.29E−05	0.15
12	113117897	T	G	0.15	0.15	0.02	−0.01	0.02	0.04	1.52E−16	0.14
16	19380214	T	C	0.17	0.17	−0.01	−0.03	0.02	0.04	2.01E−06	0.43
16	56729619	A	G	0.01	0.01	0.03	0.09	0.08	0.01	3.26E−05	0.48
16	89653173	C	T	0.47	0.47	0.08	0.02	0.01	0.03	1.84E−05	0.37
17	46502682	C	T	0.01	0.01	−0.01	−0.09	0.03	0.01	7.04E−06	0.48
17	72917513	T	C	0.04	0.04	0.08	0.002	0.01	0.07	4.46E−05	0.97

4.2.2. 당뇨병

이 연구에서 2-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 58. 이질성 검정 결과(교육 수준 - 당뇨병)

		Q	p-value
교육 연한	MR-Egger	7.3	0.99
	Inverse variance weighted	7.3	0.99

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 잔차항이 들어있는 것으로 잔차의 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 59. 다형성 검정 결과(교육 수준 – 당뇨병)

		intercept	se	p-value
교육 연한	MR-Egger	-0.002	0.02	0.95

멘델 무작위 분석과 민감도 분석을 [표 60]에 제시했다. IVW로 계산된 OR은 0.76이었으며 MR-Egger와 Weighted median은 0.77, 0.77의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 당뇨병의 위험성이 24% 감소한다는 것을 의미한다.

표 60. 2-표본 멘델 무작위 분석 결과(교육 수준 – 당뇨병)

		OR	95% 신뢰구간
1 표준편차의 교육 연한 (p-value <5e-5)	IVW	0.76	(0.70-0.81)
	MR-Egger	0.77	(0.72-0.83)
	Weighted median	0.77	(0.73-0.81)
1 표준편차의 교육 연한 (p-value <5e-6)	IVW	0.74	(0.69-0.81)
	MR-Egger	0.77	(0.70-0.84)
	Weighted median	0.76	(0.70-0.82)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 18]에 위 2-표본 멘델 무작위 분석 결과에 대한 그림을 함께 제시했다. [그림 18]에 따르면 가로축은 교육에 대한 SNP의 크기이고 세로축은 당뇨병에 대한 SNP의 크기이다. 교육 수준이 높을수록 당뇨병의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

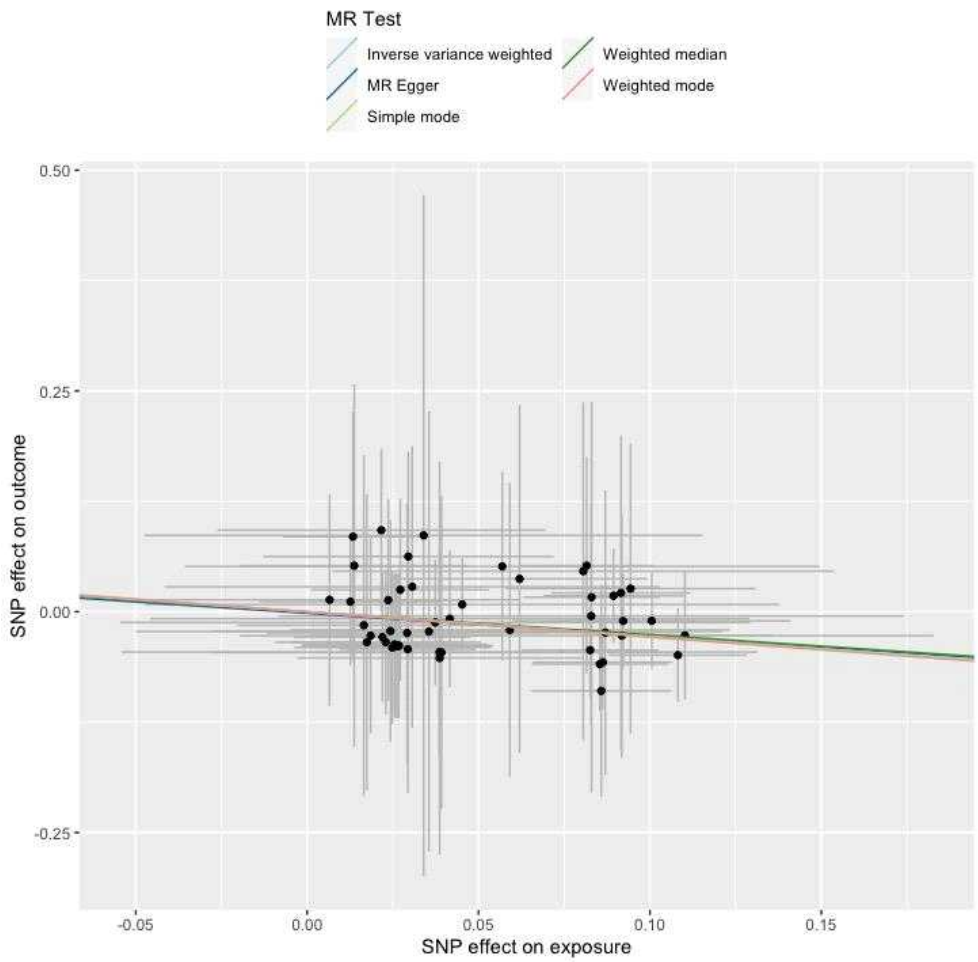

그림 18. 2-표본 멘델 무작위 분석의 산점도 (교육 수준 - 당뇨병)

표 61. 교육 수준과 당뇨병에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	0.02	−0.09	0.04	0.009	6.34E−06	0.31
1	85410061	T	C	0.02	0.02	0.08	0.02	0.02	0.01	2.51E−05	0.14
3	140942232	A	G	0.48	0.48	−0.04	−0.008	0.09	0.05	9.91E−07	0.87
3	144026997	G	A	0.13	0.13	−0.04	0.007	0.08	0.07	1.88E−06	0.91
3	153585628	G	T	0.01	0.01	−0.03	−0.08	0.08	0.03	3.01E−05	0.25
4	121613198	C	G	0.14	0.14	−0.03	0.01	0.09	0.07	4.80E−05	0.08
4	92431580	A	G	0.01	0.01	0.01	0.05	0.03	0.02	3.53E−05	0.11
5	114210843	G	C	0.46	0.46	0.03	0.03	0.08	0.05	4.89E−05	0.56
5	54852576	T	C	0.01	0.01	0.08	0.01	0.01	0.02	3.22E−05	0.45
6	31322706	G	C	0.05	0.05	0.08	−0.08	0.02	0.01	2.56E−05	0.45
6	67132090	C	T	0.01	0.01	0.09	0.02	0.02	0.01	6.98E−06	0.23
7	41147	A	G	0.01	0.01	0.03	0.02	0.07	0.01	2.22E−05	0.76

7	41170	A	G	0.03	0.03	-0.03	0.04	0.09	0.01	3.13E-05	0.76
7	81580	A	C	0.03	0.03	-0.02	0.02	0.05	0.01	3.02E-06	0.08
8	1135049	A	G	0.03	0.03	0.05	0.05	0.09	0.01	5.39E-05	0.15
8	1140691	G	C	0.04	0.04	0.02	0.06	0.04	0.01	4.86E-06	0.59
8	1151923	G	C	0.02	0.02	-0.06	-0.03	0.03	0.01	9.61E-06	0.58
8	1189658	C	T	0.03	0.03	0.05	-0.02	0.06	0.01	3.57E-06	0.21
8	1192943	G	A	0.04	0.04	0.08	-0.004	0.09	0.01	3.64E-06	0.96
8	1326979	T	C	0.04	0.04	0.02	0.01	0.02	0.01	2.43E-06	0.25
8	1333258	G	A	0.31	0.31	-0.10	0.01	0.04	0.05	1.29E-06	0.06
8	1360171	C	T	0.03	0.03	0.01	-0.03	0.02	0.01	5.14E-06	0.39
8	1472228	A	G	0.02	0.02	-0.08	-0.04	0.07	0.01	2.71E-07	0.16
8	1483124	A	C	0.02	0.02	0.02	-0.04	0.01	0.01	1.40E-07	0.79
8	1485960	T	G	0.02	0.02	0.03	-0.05	0.04	0.02	3.53E-06	0.18
8	225168	T	C	0.04	0.04	0.02	-0.02	0.04	0.01	4.76E-05	0.10
8	234623	G	T	0.03	0.03	0.03	-0.04	0.06	0.01	5.61E-06	0.90
8	239111	C	T	0.03	0.03	0.01	0.08	0.02	0.01	5.13E-05	0.54

8	750906	T	C	0.05	0.05	0.02	0.02	0.02	0.01	2.95E-06	0.15
8	761499	T	C	0.04	0.04	0.006	0.01	0.05	0.01	9.06E-05	0.25
8	772139	T	C	0.02	0.02	-0.09	0.02	0.09	0.01	3.13E-06	0.49
8	888744	T	C	0.03	0.03	-0.09	-0.02	0.03	0.01	9.66E-05	0.11
12	110390979	C	T	0.07	0.07	0.01	-0.02	0.03	0.01	1.17E-06	0.13
12	111337494	C	CT	0.18	0.18	0.11	-0.02	0.02	0.07	1.65E-05	0.22
12	111414461	G	T	0.14	0.14	0.02	-0.04	0.02	0.08	4.85E-18	0.27
12	111684461	A	G	0.49	0.49	0.08	-0.04	0.01	0.05	3.49E-05	0.40
12	111718231	A	C	0.18	0.18	0.02	-0.02	0.02	0.07	2.31E-17	0.91
12	111994852	C	A	0.48	0.48	0.10	-0.04	0.02	0.05	6.88E-08	0.35
12	112168009	A	G	0.15	0.15	0.02	-0.03	0.02	0.08	3.08E-22	0.26
12	112230019	C	G	0.15	0.15	0.02	-0.03	0.02	0.08	2.58E-22	0.21
12	112468206	T	C	0.15	0.15	0.02	-0.03	0.02	0.08	1.97E-21	0.20
12	112627350	A	G	0.16	0.16	0.02	-0.03	0.02	0.07	4.52E-21	0.33
12	112736118	G	A	0.15	0.15	0.02	-0.03	0.02	0.084	7.37E-21	0.21
12	112829103	C	CAT	0.42	0.42	0.08	-0.01	0.02	0.05	1.12E-05	0.97

12	112923393	A	G	0.44	0.44	0.09	−0.01	0.02	0.05	4.53E−06	0.49
12	112982775	G	A	0.40	0.40	0.08	0.05	0.02	0.05	2.29E−05	0.28
12	113117897	T	G	0.15	0.15	0.02	−0.03	0.02	0.08	1.52E−16	0.26
16	19380214	T	C	0.17	0.17	−0.01	−0.01	0.02	0.07	2.01E−06	0.11
16	56729619	A	G	0.01	0.01	0.03	−0.02	0.08	0.02	3.26E−05	0.36
16	89653173	C	T	0.47	0.47	0.08	−0.05	0.01	0.05	1.84E−05	0.91
17	46502682	C	T	0.01	0.01	−0.01	0.01	0.03	0.01	7.04E−06	0.42
17	72917513	T	C	0.04	0.04	0.08	0.05	0.01	0.01	4.46E−05	0.67

4.2.3. 고지혈증

이 연구에서 2-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 62. 이질성 검정 결과(교육 수준 - 고지혈증)

		Q	p-value
교육 연한	MR-Egger	52.3	0.34
	Inverse variance weighted	55.0	0.29

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 잔차항이 들어있는 것으로 잔차의 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 63. 다형성 검정 결과(교육 수준 – 고지혈증)

		intercept	se	p-value
교육 연한	MR-Egger	-0.06	0.04	0.12

멘델 무작위 분석과 민감도 분석을 [표 64]에 제시했다. IVW로 계산된 OR은 0.88이었으며 MR-Egger와 Weighted median은 0.91, 0.90의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 고지혈증의 위험성이 12% 감소한다는 것을 의미한다.

표 64. 2-표본 멘델 무작위 분석 결과(교육 수준 – 고지혈증)

		OR	95% 신뢰구간
1 표준편차의 교육 연한 (p-value <5e-5)	IVW	0.88	(0.83-0.93)
	MR-Egger	0.91	(0.81-1.01)
	Weighted median	0.90	(0.83-0.97)
1 표준편차의 교육 연한 (p-value <5e-6)	IVW	0.89	(0.84-0.92)
	MR-Egger	0.91	(0.80-1.02)
	Weighted median	0.90	(0.81-0.96)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 19]에 위 2-표본 멘델 무작위 분석 결과에 대한 그림을 함께 제시했다. [그림 19]에 따르면 가로축은 교육에 대한 SNP의 효과이고 세로축은 고지혈증에 대한 SNP의 효과이다. 교육 수준이 높을수록 고지혈증의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

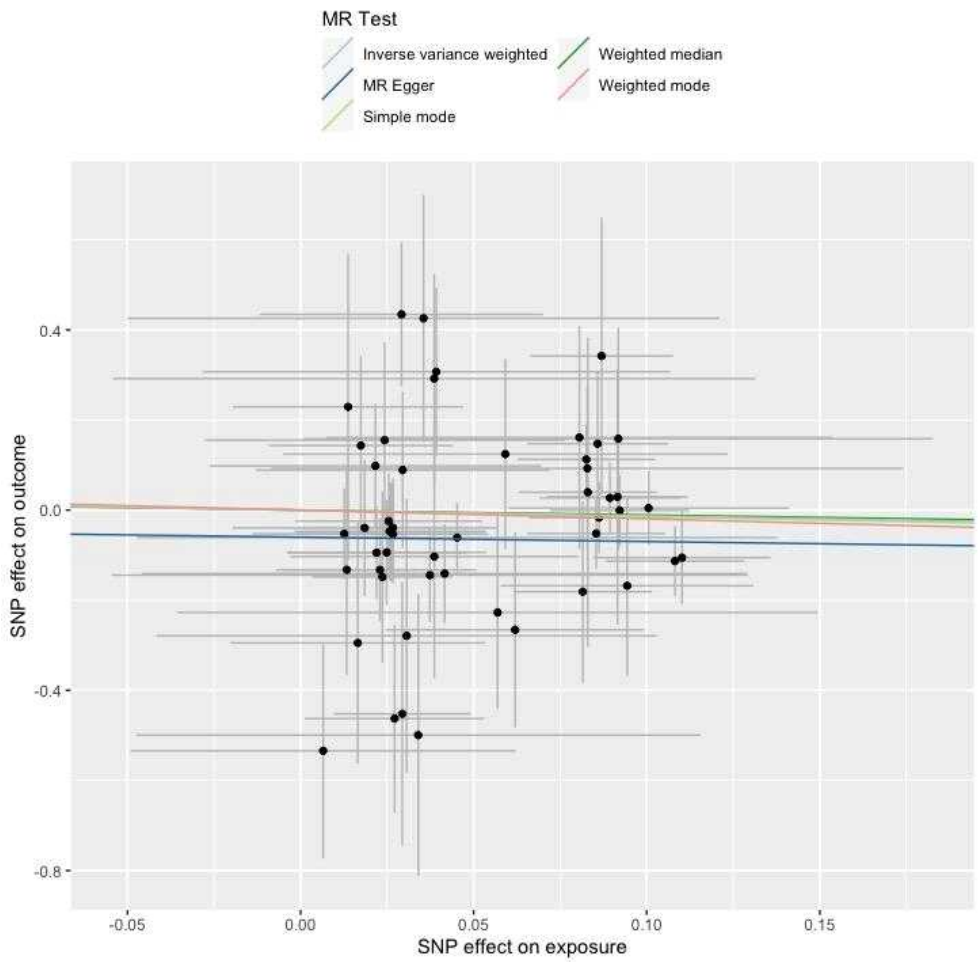

그림 19. 2-표본 멘델 무작위 분석의 산점도(교육 수준 - 고지혈증)

표 65. 교육 수준과 고지혈증에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	0.02	-0.09	0.04	0.01	6.34E-06	0.47
1	85410061	T	C	0.02	0.02	0.08	-0.03	0.02	0.03	2.51E-05	0.26
3	140942232	A	G	0.48	0.48	-0.04	0.06	0.09	0.07	9.91E-07	0.42
3	144026997	G	A	0.13	0.13	-0.04	0.01	0.08	0.01	1.88E-06	0.20
3	153585628	G	T	0.01	0.01	-0.03	0.04	0.08	0.03	3.01E-05	0.10
4	121613198	C	G	0.14	0.14	-0.03	0.01	0.09	0.01	4.80E-05	0.16
4	92431580	A	G	0.01	0.01	0.01	0.02	0.03	0.03	3.53E-05	0.50
5	114210843	G	C	0.46	0.46	0.03	-0.08	0.08	0.07	4.89E-05	0.28
5	54852576	T	C	0.01	0.01	0.08	0.03	0.01	0.03	3.22E-05	0.90
6	31322706	G	C	0.05	0.05	0.08	0.01	0.02	0.01	2.56E-05	0.36
6	67132090	C	T	0.01	0.01	0.09	0.0204	0.02	0.02	6.98E-06	0.91
7	41147	A	G	0.01	0.01	0.03	-0.02	0.07	0.03	2.22E-05	0.35

7	41170	A	G	0.03	0.03	-0.03	-0.02	0.09	0.02	3.13E-05	0.21
7	81580	A	C	0.03	0.03	-0.02	-0.01	0.05	0.02	3.02E-06	0.47
8	1135049	A	G	0.03	0.03	0.05	-0.02	0.09	0.02	5.39E-05	0.28
8	1140691	G	C	0.04	0.04	0.02	0.08	0.04	0.01	4.86E-06	0.60
8	1151923	G	C	0.02	0.02	-0.06	0.02	0.03	0.02	9.61E-06	0.22
8	1189658	C	T	0.03	0.03	0.05	0.01	0.06	0.02	3.57E-06	0.55
8	1192943	G	A	0.04	0.04	0.08	0.09	0.09	0.01	3.64E-06	0.60
8	1326979	T	C	0.04	0.04	0.02	-0.0	0.02	0.01	2.43E-06	0.43
8	1333258	G	A	0.31	0.31	-0.10	-0.004	0.04	0.008	1.29E-06	0.95
8	1360171	C	T	0.03	0.03	0.01	0.01	0.02	0.01	5.14E-06	0.46
8	1472228	A	G	0.02	0.02	-0.08	-0.01	0.07	0.02	2.71E-07	0.51
8	1483124	A	C	0.02	0.02	0.02	-0.04	0.01	0.02	1.40E-07	0.12
8	1485960	T	G	0.02	0.02	0.03	-0.01	0.04	0.02	3.53E-06	0.70
8	225168	T	C	0.04	0.04	0.02	0.04	0.04	0.01	4.76E-05	0.66
8	234623	G	T	0.03	0.03	0.03	0.03	0.06	0.01	5.61E-06	0.98
8	239111	C	T	0.03	0.03	0.01	-0.01	0.02	0.02	5.13E-05	0.56

8	750906	T	C	0.05	0.05	0.02	−0.04	0.02	0.02	2.95E−06	0.26
8	761499	T	C	0.04	0.04	0.006	−0.05	0.05	0.02	9.06E−05	0.25
8	772139	T	C	0.02	0.02	−0.09	−0.01	0.09	0.02	3.13E−06	0.52
8	888744	T	C	0.03	0.03	−0.09	0.01	0.03	0.01	9.66E−05	0.39
12	110390979	C	T	0.07	0.07	0.01	−0.03	0.03	0.01	1.17E−06	0.79
12	111337494	C	CT	0.18	0.18	0.11	−0.01	0.02	0.01	1.65E−05	0.30
12	111414461	G	T	0.14	0.14	0.02	−0.09	0.02	0.01	4.85E−18	0.41
12	111684461	A	G	0.49	0.49	0.08	0.01	0.01	0.007	3.49E−05	0.14
12	111718231	A	C	0.18	0.18	0.02	−0.09	0.02	0.01	2.31E−17	0.35
12	111994852	C	A	0.48	0.48	0.10	−0.01	0.02	0.007	6.88E−08	0.14
12	112168009	A	G	0.15	0.15	0.02	−0.03	0.02	0.01	3.08E−22	0.72
12	112230019	C	G	0.15	0.15	0.02	−0.05	0.02	0.01	2.58E−22	0.62
12	112468206	T	C	0.15	0.15	0.02	−0.04	0.02	0.01	1.97E−21	0.65
12	112627350	A	G	0.16	0.16	0.02	−0.02	0.02	0.01	4.52E−21	0.81
12	112736118	G	A	0.15	0.15	0.02	−0.04	0.02	0.01	7.37E−21	0.66
12	112829103	C	CAT	0.42	0.42	0.08	−0.02	0.02	0.007	1.12E−05	0.73

12	112923393	A	G	0.44	0.44	0.09	−0.007	0.02	0.007	4.53E−06	0.99
12	112982775	G	A	0.40	0.40	0.08	0.01	0.02	0.007	2.29E−05	0.83
12	113117897	T	G	0.15	0.15	0.02	−0.01	0.02	0.01	1.52E−16	0.24
16	19380214	T	C	0.17	0.17	−0.01	0.05	0.02	0.01	2.01E−06	0.60
16	56729619	A	G	0.01	0.01	0.03	0.04	0.08	0.02	3.26E−05	0.12
16	89653173	C	T	0.47	0.47	0.08	−0.05	0.01	0.007	1.84E−05	0.50
17	46502682	C	T	0.01	0.01	−0.01	0.02	0.03	0.02	7.04E−06	0.27
17	72917513	T	C	0.04	0.04	0.08	−0.01	0.01	0.02	4.46E−05	0.36

4.2.4. 비만 – 체질량지수 25 kg/m^2 이상

이 연구에서 2-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 66. 이질성 검정 결과(교육 수준 – 비만, 체질량지수 25kg/m^2 이상)

		Q	p-value
교육 연한	MR-Egger	10.6	0.99
	Inverse variance weighted	9.7	0.99

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 잔차항이 들어있는 것으로 잔차의 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 67. 다형성 검정 결과(교육 수준 – 비만, 체질량지수 25kg/m^2 이상)

		intercept	se	p-value
교육 연한	MR-Egger	-0.004	0.01	0.82

멘델 무작위 분석 분석과 민감도 분석을 [표 68]에 제시했다. IVW로 계산된 OR은 0.90이었으며 MR-Egger와 Weighted median은 0.91, 0.90의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 비만의 위험성이 10% 감소한다는 것을 의미한다.

표 68. 2-표본 멘델 무작위 분석 결과(교육 수준 – 비만, 체질량지수 25kg/m^2 이상)

		OR	95% 신뢰구간
	IVW	0.90	(0.85-0.95)
1 표준편차의 교육 연한	MR-Egger	0.91	(0.81-1.02)
(p-value <5e-5)	Weighted median	0.90	(0.84-0.96)
	IVW	0.91	(0.85-0.94)
1 표준편차의 교육 연한	MR-Egger	0.90	(0.81-1.00)
(p-value <5e-6)	Weighted median	0.91	(0.85-0.97)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 20]에 위 2-표본 멘델 무작위 분석 결과에 대한 그림을 함께 제시했다. [그림 20]에 따르면 가로축은 교육에 대한 SNP의 효과이고 세로축은 비만에 대한 SNP의 효과이다. 교육 수준이 높을수록 비만의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

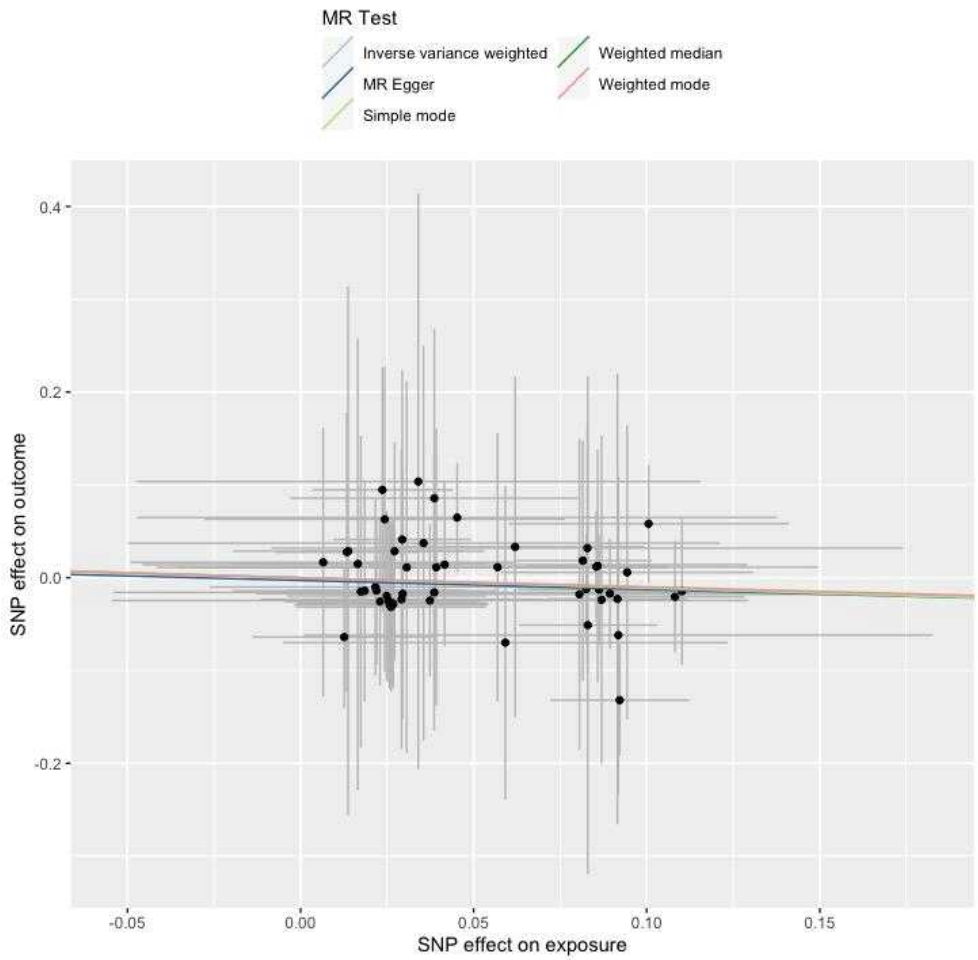

그림 20. 2-표본 멘델 무작위 분석의 산점도 (교육 수준 – 비만, 체질량지수 25kg/m^2 이상)

표 69. 교육 수준과 비만에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	0.02	0.01	0.04	0.009	6.34E-06	0.29
1	85410061	T	C	0.02	0.02	0.08	0.02	0.02	0.01	2.51E-05	0.17
3	140942232	A	G	0.48	0.48	-0.04	-0.01	0.09	0.01	9.91E-07	0.23
3	144026997	G	A	0.13	0.13	-0.04	-0.01	0.08	0.007	1.88E-06	0.06
3	153585628	G	T	0.01	0.01	-0.03	-0.01	0.08	0.009	3.01E-05	0.32
4	121613198	C	G	0.14	0.14	-0.03	-0.01	0.09	0.05	4.80E-05	0.39
4	92431580	A	G	0.01	0.01	0.01	-0.01	0.03	0.007	3.53E-05	0.08
5	114210843	G	C	0.46	0.46	0.03	-0.02	0.08	0.05	4.89E-05	0.57
5	54852576	T	C	0.01	0.01	0.08	-0.02	0.01	0.008	3.22E-05	0.21
6	31322706	G	C	0.05	0.05	0.08	-0.02	0.02	0.009	2.56E-05	0.13
6	67132090	C	T	0.01	0.01	0.09	-0.03	0.02	0.009	6.98E-06	0.56
7	41147	A	G	0.01	0.01	0.03	-0.02	0.07	0.008	2.22E-05	0.38

7	41170	A	G	0.03	0.03	-0.03	-0.02	0.09	0.009	3.13E-05	0.12
7	81580	A	C	0.03	0.03	-0.02	0.01	0.05	0.005	3.02E-06	0.40
8	1135049	A	G	0.03	0.03	0.05	-0.01	0.09	0.005	5.39E-05	0.27
8	1140691	G	C	0.04	0.04	0.02	0.01	0.04	0.005	4.86E-06	0.30
8	1151923	G	C	0.02	0.02	-0.06	-0.02	0.03	0.009	9.61E-06	0.44
8	1189658	C	T	0.03	0.03	0.05	0.06	0.06	0.007	3.57E-06	0.40
8	1192943	G	A	0.04	0.04	0.08	0.03	0.09	0.02	3.64E-06	0.07
8	1326979	T	C	0.04	0.04	0.02	0.01	0.02	0.005	2.43E-06	0.84
8	1333258	G	A	0.31	0.31	-0.10	-0.01	0.04	0.02	1.29E-06	0.54
8	1360171	C	T	0.03	0.03	0.01	0.01	0.02	0.01	5.14E-06	0.15
8	1472228	A	G	0.02	0.02	-0.08	-0.06	0.07	0.005	2.71E-07	0.27
8	1483124	A	C	0.02	0.02	0.02	-0.01	0.01	0.008	1.40E-07	0.87
8	1485960	T	G	0.02	0.02	0.03	-0.01	0.04	0.03	3.53E-06	0.73
8	225168	T	C	0.04	0.04	0.02	0.02	0.04	0.008	4.76E-05	0.76
8	234623	G	T	0.03	0.03	0.03	0.02	0.06	0.02	5.61E-06	0.91
8	239111	C	T	0.03	0.03	0.01	0.08	0.02	0.05	5.13E-05	0.15

8	750906	T	C	0.05	0.05	0.02	-0.05	0.02	0.02	2.95E-06	0.84
8	761499	T	C	0.04	0.04	0.006	0.01	0.05	0.01	9.06E-05	0.30
8	772139	T	C	0.02	0.02	-0.09	-0.02	0.09	0.02	3.13E-06	0.34
8	888744	T	C	0.03	0.03	-0.09	0.01	0.03	0.02	9.66E-05	0.57
12	110390979	C	T	0.07	0.07	0.01	0.01	0.03	0.01	1.17E-06	0.28
12	111337494	C	CT	0.18	0.18	0.11	-0.06	0.02	0.01	1.65E-05	0.70
12	111414461	G	T	0.14	0.14	0.02	0.01	0.02	0.01	4.85E-18	0.43
12	111684461	A	G	0.49	0.49	0.08	-0.01	0.01	0.01	3.49E-05	0.89
12	111718231	A	C	0.18	0.18	0.02	-0.03	0.02	0.01	2.31E-17	0.85
12	111994852	C	A	0.48	0.48	0.10	-0.06	0.02	0.01	6.88E-08	0.67
12	112168009	A	G	0.15	0.15	0.02	0.03	0.02	0.01	3.08E-22	0.81
12	112230019	C	G	0.15	0.15	0.02	0.09	0.02	0.01	2.58E-22	0.47
12	112468206	T	C	0.15	0.15	0.02	-0.05	0.02	0.006	1.97E-21	0.36
12	112627350	A	G	0.16	0.16	0.02	-0.01	0.02	0.01	4.52E-21	0.37
12	112736118	G	A	0.15	0.15	0.02	0.01	0.02	0.01	7.37E-21	0.91
12	112829103	C	CAT	0.42	0.42	0.08	0.04	0.02	0.01	1.12E-05	0.82

12	112923393	A	G	0.44	0.44	0.09	0.08	0.02	0.01	4.53E-06	0.63
12	112982775	G	A	0.40	0.40	0.08	-0.02	0.02	0.01	2.29E-05	0.14
12	113117897	T	G	0.15	0.15	0.02	0.01	0.02	0.01	1.52E-16	0.45
16	19380214	T	C	0.17	0.17	-0.01	0.02	0.02	0.01	2.01E-06	0.06
16	56729619	A	G	0.01	0.01	0.03	0.02	0.08	0.01	3.26E-05	0.15
16	89653173	C	T	0.47	0.47	0.08	0.01	0.01	0.01	1.84E-05	0.90
17	46502682	C	T	0.01	0.01	-0.01	0.06	0.03	0.01	7.04E-06	0.71
17	72917513	T	C	0.04	0.04	0.08	-0.005	0.01	0.01	4.46E-05	0.97

4.2.5. 비만 – 체질량지수 30 kg/m^2 이상

이 연구에서 2-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 70. 이질성 검정 결과(교육 수준 – 비만, 체질량지수 30kg/m^2 이상)

		Q	p-value
교육 연한	MR-Egger	10.0	0.99
	Inverse variance weighted	10.0	0.99

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 잔차항이 들어있는 것으로 잔차의 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 71. 다형성 검정 결과(교육 수준 – 비만, 체질량지수 30kg/m^2 이상)

		intercept	se	p-value
교육 연한	MR-Egger	-0.003	0.03	0.92

멘델 무작위 분석과 민감도 분석을 [표 72]에 제시했다. IVW로 계산된 OR은 0.85이었으며 MR-Egger와 Weighted median은 0.82, 0.87의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 비만의 위험성이 15% 감소한다는 것을 의미한다.

표 72. 2-표본 멘델 무작위 분석 결과(교육 수준 – 비만, 체질량지수 30kg/m^2 이상)

		OR	95% 신뢰구간
	IVW	0.85	(0.80-0.90)
1 표준편차의 교육 연한	MR-Egger	0.82	(0.73-0.92)
(p-value <5e-5)	Weighted median	0.87	(0.81-0.93)
	IVW	0.86	(0.80-0.91)
1 표준편차의 교육 연한	MR-Egger	0.84	(0.74-0.93)
(p-value <5e-6)	Weighted median	0.88	(0.81-0.94)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 21]에 위 2-표본 멘델 무작위 분석 결과에 대한 그림을 함께 제시했다. [그림 21]에 따르면 가로축은 교육에 대한 SNP의 효과이고 세로축은 비만에 대한 SNP의 효과이다. 교육 수준이 높을수록 비만의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

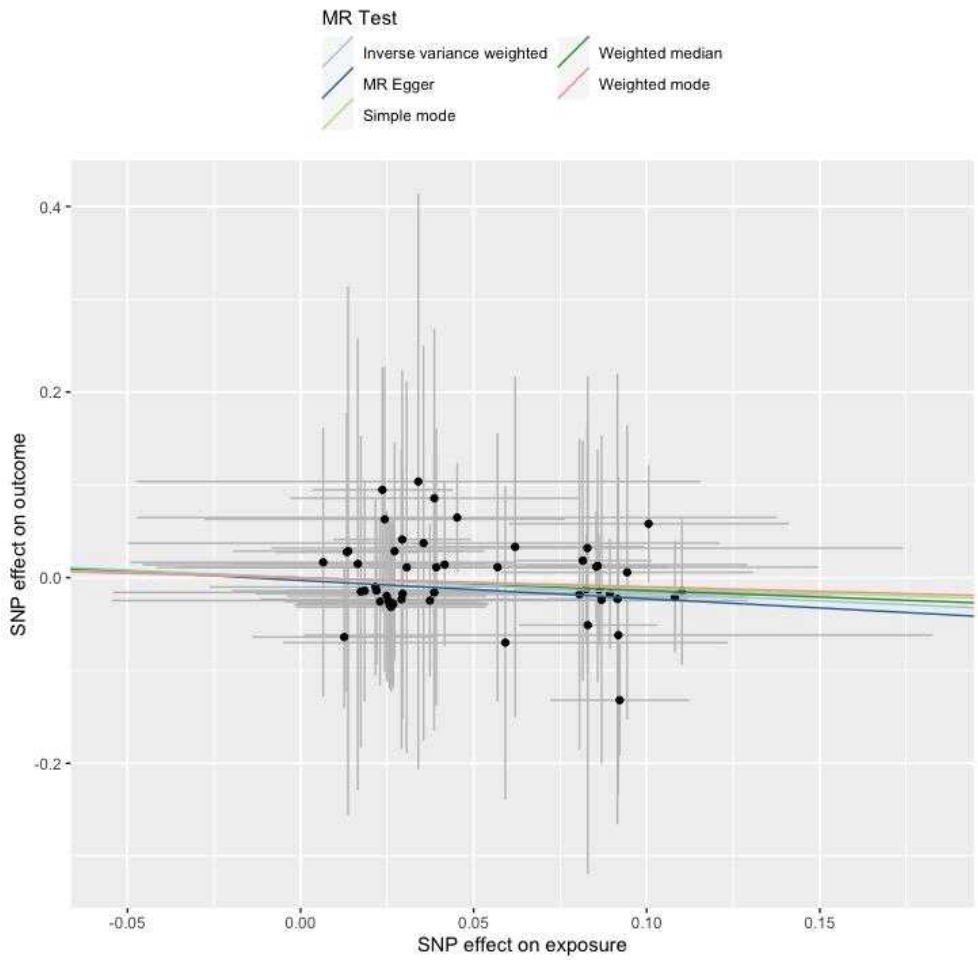

그림 21. 2-표본 멘델 무작위 분석의 산점도 (교육 수준 – 비만, 체질량지수 30kg/m^2 이상)

표 73. 교육 수준과 비만에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	0.02	0.01	0.04	0.009	6.34E-06	0.29
1	85410061	T	C	0.02	0.02	0.08	0.02	0.02	0.01	2.51E-05	0.17
3	140942232	A	G	0.48	0.48	-0.04	-0.06	0.09	0.05	9.91E-07	0.27
3	144026997	G	A	0.13	0.13	-0.04	-0.01	0.08	0.008	1.88E-06	0.87
3	153585628	G	T	0.01	0.01	-0.03	-0.01	0.08	0.03	3.01E-05	0.73
4	121613198	C	G	0.14	0.14	-0.03	0.02	0.09	0.008	4.80E-05	0.76
4	92431580	A	G	0.01	0.01	0.01	0.02	0.03	0.02	3.53E-05	0.91
5	114210843	G	C	0.46	0.46	0.03	0.08	0.08	0.005	4.89E-05	0.15
5	54852576	T	C	0.01	0.01	0.08	-0.05	0.01	0.02	3.22E-05	0.84
6	31322706	G	C	0.05	0.05	0.08	0.01	0.02	0.01	2.56E-05	0.30
6	67132090	C	T	0.01	0.01	0.09	-0.02	0.02	0.02	6.98E-06	0.34
7	41147	A	G	0.01	0.01	0.03	0.01	0.07	0.02	2.22E-05	0.57

7	41170	A	G	0.03	0.03	-0.03	0.01	0.09	0.01	3.13E-05	0.28
7	81580	A	C	0.03	0.03	-0.02	-0.06	0.05	0.01	3.02E-06	0.70
8	1135049	A	G	0.03	0.03	0.05	0.01	0.09	0.01	5.39E-05	0.43
8	1140691	G	C	0.04	0.04	0.02	-0.01	0.04	0.01	4.86E-06	0.89
8	1151923	G	C	0.02	0.02	-0.06	-0.03	0.03	0.01	9.61E-06	0.85
8	1189658	C	T	0.03	0.03	0.05	-0.06	0.06	0.01	3.57E-06	0.67
8	1192943	G	A	0.04	0.04	0.08	0.03	0.09	0.01	3.64E-06	0.81
8	1326979	T	C	0.04	0.04	0.02	0.009	0.02	0.01	2.43E-06	0.47
8	1333258	G	A	0.31	0.31	-0.10	-0.05	0.04	0.006	1.29E-06	0.36
8	1360171	C	T	0.03	0.03	0.01	-0.01	0.02	0.01	5.14E-06	0.37
8	1472228	A	G	0.02	0.02	-0.08	0.01	0.07	0.01	2.71E-07	0.91
8	1483124	A	C	0.02	0.02	0.02	0.04	0.01	0.01	1.40E-07	0.82
8	1485960	T	G	0.02	0.02	0.03	0.08	0.04	0.01	3.53E-06	0.63
8	225168	T	C	0.04	0.04	0.02	-0.02	0.04	0.01	4.76E-05	0.14
8	234623	G	T	0.03	0.03	0.03	0.01	0.06	0.01	5.61E-06	0.45
8	239111	C	T	0.03	0.03	0.01	0.02	0.02	0.01	5.13E-05	0.06

8	750906	T	C	0.05	0.05	0.02	0.02	0.02	0.01	2.95E-06	0.15
8	761499	T	C	0.04	0.04	0.006	0.01	0.05	0.01	9.06E-05	0.90
8	772139	T	C	0.02	0.02	-0.09	0.06	0.09	0.01	3.13E-06	0.71
8	888744	T	C	0.03	0.03	-0.09	-0.005	0.03	0.01	9.66E-05	0.97
12	110390979	C	T	0.07	0.07	0.01	-0.01	0.03	0.01	1.17E-06	0.23
12	111337494	C	CT	0.18	0.18	0.11	-0.01	0.02	0.007	1.65E-05	0.06
12	111414461	G	T	0.14	0.14	0.02	-0.01	0.02	0.009	4.85E-18	0.32
12	111684461	A	G	0.49	0.49	0.08	-0.01	0.01	0.05	3.49E-05	0.39
12	111718231	A	C	0.18	0.18	0.02	-0.01	0.02	0.007	2.31E-17	0.08
12	111994852	C	A	0.48	0.48	0.10	-0.02	0.02	0.005	6.88E-08	0.57
12	112168009	A	G	0.15	0.15	0.02	-0.02	0.02	0.008	3.08E-22	0.21
12	112230019	C	G	0.15	0.15	0.02	-0.02	0.02	0.009	2.58E-22	0.13
12	112468206	T	C	0.15	0.15	0.02	-0.03	0.02	0.009	1.97E-21	0.55
12	112627350	A	G	0.16	0.16	0.02	-0.02	0.02	0.008	4.52E-21	0.38
12	112736118	G	A	0.15	0.15	0.02	-0.02	0.02	0.009	7.37E-21	0.12
12	112829103	C	CAT	0.42	0.42	0.08	0.01	0.02	0.005	1.12E-05	0.40

12	112923393	A	G	0.44	0.44	0.09	-0.01	0.02	0.005	4.53E-06	0.27
12	112982775	G	A	0.40	0.40	0.08	0.01	0.02	0.005	2.29E-05	0.30
12	113117897	T	G	0.15	0.15	0.02	-0.02	0.02	0.009	1.52E-16	0.44
16	19380214	T	C	0.17	0.17	-0.01	0.06	0.02	0.007	2.01E-06	0.40
16	56729619	A	G	0.01	0.01	0.03	0.03	0.08	0.02	3.26E-05	0.07
16	89653173	C	T	0.47	0.47	0.08	0.01	0.01	0.005	1.84E-05	0.84
17	46502682	C	T	0.01	0.01	-0.01	-0.01	0.03	0.02	7.04E-06	0.54
17	72917513	T	C	0.04	0.04	0.08	0.01	0.01	0.01	4.46E-05	0.15

4.2.6. 흡연

이 연구에서 2-표본 멘델 무작위 분석을 위해 교육 수준에 대한 SNP를 지역사회 기반 코호트, 농촌 코호트 통합 자료와 도시 코호트에서 추출 후 중복되는 55개의 SNP을 이용하여 멘델 무작위 분석을 시행하였다. 추가로 이질성(heterogeneity)과 다형성(pleiotropy) 검정을 실시하였다.

이질성 검정은 Q 통계량을 구하게 되는데 Q가 작을수록 이질성으로부터 자유롭다. Q 검정 결과에서 p-value가 0.05 이상이므로 귀무가설을 기각해 이질성 검정을 통과했다.

표 74. 이질성 검정 결과(교육 수준 - 흡연)

		Q	p-value
교육 연한	MR-Egger	52.3	0.34
	Inverse variance weighted	53.7	0.33

다형성 검정은 IVW의 민감도 분석인 MR-Egger 회귀분석으로 시행한다. MR-Egger 회귀분석은 IVW의 회귀모형에 잔차항이 들어있는 것으로 잔차의 p-value가 0.05 이상이어서 다형성 검정을 통과했다.

표 75. 다형성 검정 결과(교육 수준 – 흡연)

		intercept	se	p-value
교육 연한	MR-Egger	0.02	0.01	0.27

멘델 무작위 분석과 민감도 분석을 [표 76]에 제시했다. IVW로 계산된 OR은 0.66이었으며 MR-Egger와 Weighted median은 0.64, 0.63의 OR을 산출했다. 이는 교육 연한이 1표준편차인 4.0년 증가하면 흡연의 위험성이 34% 감소한다는 것을 의미한다.

표 76. 2-표본 멘델 무작위 분석 결과(교육 수준 – 흡연)

		OR	95% 신뢰구간
1 표준편차의 교육 연한 (p-value <5e-5)	IVW	0.66	(0.58-0.76)
	MR-Egger	0.64	(0.55-0.74)
	Weighted median	0.63	(0.53-0.75)
1 표준편차의 교육 연한 (p-value <5e-6)	IVW	0.67	(0.57-0.77)
	MR-Egger	0.66	(0.55-0.75)
	Weighted median	0.65	(0.54-0.76)

a. OR = Odds ratio, IVW = Inverse variance weighted
b. 1 표준편차는 4.0년에 해당

[그림 22]에 위 2-표본 멘델 무작위 분석 결과에 대한 그림을 함께 제시했다. [그림 22]에 따르면 가로축은 교육에 대한 SNP의 효과이고 세로축은 흡연에 대한 SNP의 효과이다. 교육 수준이 높을수록 흡연의 위험성이 감소하기 때문에 그래프의 기울기는 음이다.

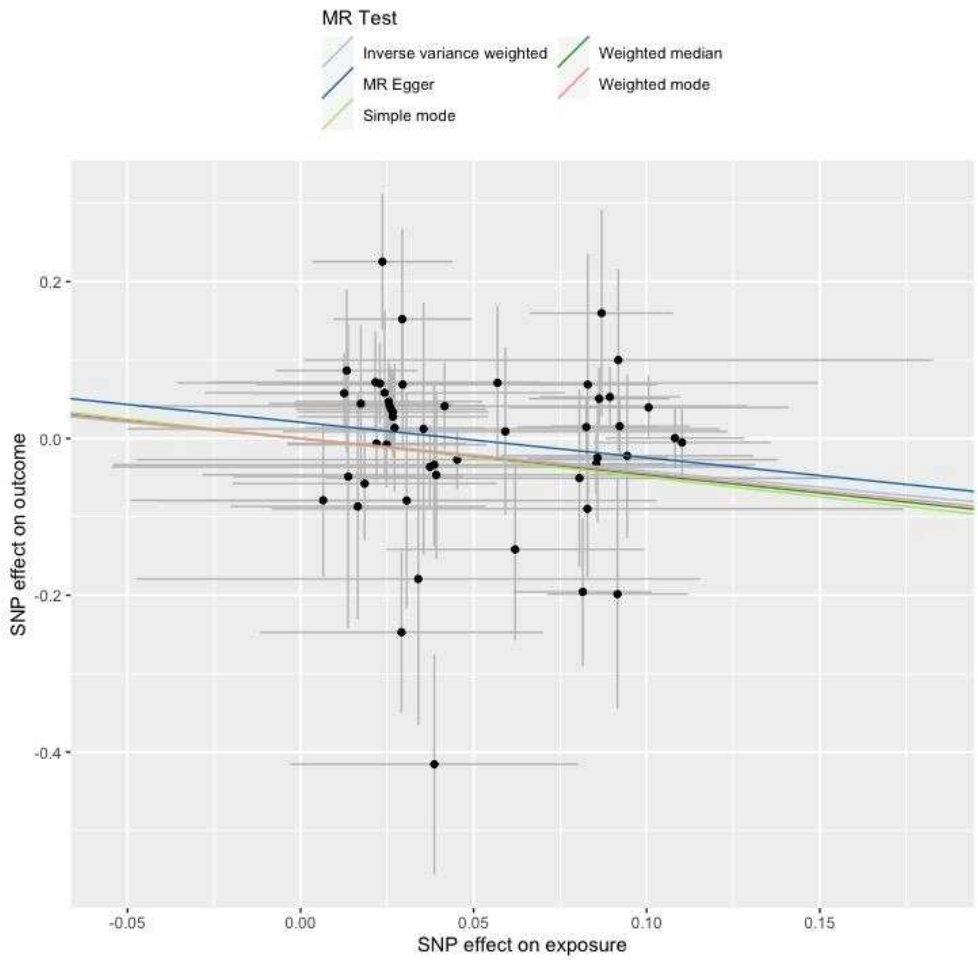

그림 22. 2-표본 멘델 무작위 분석의 산점도 (교육 수준 - 흡연)

표 77. 교육 수준과 흡연에 대한 멘델 무작위 분석에서 각 SNP의 연결 결과

Chromosome	position	ALT	REF	EAF		Beta		SE		P-value	
				Exposure	Outcome	Exposure	Outcome	Exposure	Outcome	Exposure	Outcome
1	245768933	G	A	0.09	0.09	0.02	-0.07	0.04	0.006	6.34E-06	0.26
1	85410061	T	C	0.02	0.02	0.08	-0.01	0.02	0.01	2.51E-05	0.22
3	140942232	A	G	0.48	0.48	-0.04	0.02	0.09	0.03	9.91E-07	0.47
3	144026997	G	A	0.13	0.13	-0.04	-0.04	0.08	0.005	1.88E-06	0.46
3	153585628	G	T	0.01	0.01	-0.03	0.01	0.08	0.01	3.01E-05	0.33
4	121613198	C	G	0.14	0.14	-0.03	0.03	0.09	0.005	4.80E-05	0.49
4	92431580	A	G	0.01	0.01	0.01	-0.04	0.03	0.01	3.53E-05	0.80
5	114210843	G	C	0.46	0.46	0.03	0.04	0.08	0.03	4.89E-05	0.28
5	54852576	T	C	0.01	0.01	0.08	0.06	0.01	0.01	3.22E-05	0.68
6	31322706	G	C	0.05	0.05	0.08	-0.02	0.02	0.008	2.56E-05	0.76
6	67132090	C	T	0.01	0.01	0.09	-0.01	0.02	0.01	6.98E-06	0.17
7	41147	A	G	0.01	0.01	0.03	-0.07	0.07	0.01	2.22E-05	0.56

7	41170	A	G	0.03	0.03	−0.03	0.03	0.09	0.01	3.13E−05	0.74
7	81580	A	C	0.03	0.03	−0.02	−0.05	0.05	0.01	3.02E−06	0.57
8	1135049	A	G	0.03	0.03	0.05	0.07	0.09	0.009	5.39E−05	0.47
8	1140691	G	C	0.04	0.04	0.02	0.06	0.04	0.008	4.86E−06	0.40
8	1151923	G	C	0.02	0.02	−0.06	0.01	0.03	0.01	9.61E−06	0.22
8	1189658	C	T	0.03	0.03	0.05	0.008	0.06	0.01	3.57E−06	0.93
8	1192943	G	A	0.04	0.04	0.08	−0.08	0.09	0.008	3.64E−06	0.30
8	1326979	T	C	0.04	0.04	0.02	0.02	0.02	0.008	2.43E−06	0.92
8	1333258	G	A	0.31	0.31	−0.10	−0.03	0.04	0.04	1.29E−06	0.33
8	1360171	C	T	0.03	0.03	0.01	0.04	0.02	0.01	5.14E−06	0.65
8	1472228	A	G	0.02	0.02	−0.08	0.05	0.07	0.01	2.71E−07	0.65
8	1483124	A	C	0.02	0.02	0.02	0.01	0.01	0.01	1.40E−07	0.18
8	1485960	T	G	0.02	0.02	0.03	−0.04	0.04	0.01	3.53E−06	0.29
8	225168	T	C	0.04	0.04	0.02	−0.02	0.04	0.01	4.76E−05	0.16
8	234623	G	T	0.03	0.03	0.03	−0.04	0.06	0.01	5.61E−06	0.66
8	239111	C	T	0.03	0.03	0.01	0.08	0.02	0.01	5.13E−05	0.40

8	750906	T	C	0.05	0.05	0.02	0.01	0.02	0.008	2.95E-06	0.87
8	761499	T	C	0.04	0.04	0.006	-0.07	0.05	0.009	9.06E-05	0.41
8	772139	T	C	0.02	0.02	-0.09	-0.09	0.09	0.01	3.13E-06	0.38
8	888744	T	C	0.03	0.03	-0.09	0.02	0.03	0.01	9.66E-05	0.83
12	110390979	C	T	0.07	0.07	0.01	-0.05	0.03	0.007	1.17E-06	0.42
12	111337494	C	CT	0.18	0.18	0.11	-0.05	0.02	0.04	1.65E-05	0.91
12	111414461	G	T	0.14	0.14	0.02	-0.07	0.02	0.005	4.85E-18	0.88
12	111684461	A	G	0.49	0.49	0.08	0.01	0.01	0.03	3.49E-05	0.70
12	111718231	A	C	0.18	0.18	0.02	-0.006	0.02	0.04	2.31E-17	0.89
12	111994852	C	A	0.48	0.48	0.10	0.04	0.02	0.03	6.88E-08	0.99
12	112168009	A	G	0.15	0.15	0.02	0.03	0.02	0.005	3.08E-22	0.51
12	112230019	C	G	0.15	0.15	0.02	0.02	0.02	0.05	2.58E-22	0.58
12	112468206	T	C	0.15	0.15	0.02	0.03	0.02	0.05	1.97E-21	0.46
12	112627350	A	G	0.16	0.16	0.02	0.04	0.02	0.05	4.52E-21	0.35
12	112736118	G	A	0.15	0.15	0.02	0.04	0.02	0.05	7.37E-21	0.41
12	112829103	C	CAT	0.42	0.42	0.08	-0.05	0.02	0.03	1.12E-05	0.17

12	112923393	A	G	0.44	0.44	0.09	0.01	0.02	0.03	4.53E-06	0.68
12	112982775	G	A	0.40	0.40	0.08	-0.05	0.02	0.03	2.29E-05	0.19
12	113117897	T	G	0.15	0.15	0.02	0.06	0.02	0.05	1.52E-16	0.17
16	19380214	T	C	0.17	0.17	-0.01	-0.05	0.02	0.05	2.01E-06	0.25
16	56729619	A	G	0.01	0.01	0.03	0.01	0.08	0.01	3.26E-05	0.93
16	89653173	C	T	0.47	0.47	0.08	-0.03	0.01	0.03	1.84E-05	0.39
17	46502682	C	T	0.01	0.01	-0.01	0.08	0.03	0.01	7.04E-06	0.54
17	72917513	T	C	0.04	0.04	0.08	-0.01	0.01	0.009	4.46E-05	0.37

V. 연구 결과의 요약

1.1. 연구 대상자의 일반적 특성

지역사회 기반 코호트, 농촌 코호트, 도시 코호트 대상자는 각각 5,493명, 8,105명, 58,696명이었다. 지역사회 기반 코호트는 남녀 비율이 비슷했지만 농촌 코호트와 도시 코호트는 여성이 차지하는 비율이 더 높았다. 남녀로 비교했을 때 도시 코호트에서 남성의 43.9%인 8,909명이 대학교 이상이라고 응답했으며, 여성의 대학교 이상 비율은 24.6%인 9,430명으로 남성에 비해 여성의 대학교 이상 비율이 더 낮게 나타났다. 하지만 도시 코호트의 대학교 이상 대상자의 비율이 지역사회 기반 코호트, 농촌 코호트보다 높게 나타났다.

심혈관계질환은 과거에 진단받고 현재 치료 중인 사람을 대상으로 하였고, 도시 코호트 기준 전체 집단의 2.8%인 1,622명이 심혈관계질환 현재 치료 중이었으며 남녀의 비율은 각각 4.4%, 1.9%로 남성에서 더 높았다. 주요 심혈관계질환 위험요인으로는 흡연, 음주, 고혈압, 당뇨, 고지혈증 등이 있으며, 건강 행태 위험요인인 흡연과 음주는 남자가 여성에 비해 비율이 더 높았다. 남성의 8.5%인 1,722명이 현재 흡연 중이라고 응답했으며 여성의 0.5%인 191명만 현재 흡연 중이라고 응답했다.

1.2. KoGES 코호트에서 계산한 심혈관계질환 위험도(로지스틱 회귀분석)

코호트 자료는 일정 기간 동안 연구대상자를 추적관찰하기 때문에, 질병 위험도를 계산할 수 있다. 환자-대조군 모형으로 가정하여 교육수준과 심혈관계질환에 대해 로지스틱 회귀분석을 실시해 오즈비를 계산하였으며, 그 결과는 표 9에 제시하였다. 지역사회 코호트와 농촌 코호트는 표본 수가 적어 오즈비의 95% 신뢰구간이 유의하지 않았다. 도시 코호트는 표본 수가 충분하여 95% 신뢰구간과 p-value가 안정적으로 산출되었다.

지역사회 코호트에서 대학 이상인 집단의 심혈관계질환 오즈비는 0.85, 농촌 코호트에서 대학 이상인 집단의 심혈관계질환 오즈비는 0.73, 도시 코호트에서는 0.79였다.

표본 수가 충분하지 않은 지역사회 코호트와 농촌 코호트의 오즈비는 대학교 이상인 집단에서 낮았지만 95% 신뢰구간이 유의하지 않았다. 표본 수가 충분한 도시 코호트에서 계산한 오즈비는 유의하였다. 따라서 도시 코호트에서 계산한 오즈비를 참조하는 것이 더 신뢰할 수 있는 결과이다.

1.3. 교육 수준에 대한 GWAS 결과

도시 코호트에서 58,696명이 대상자였으며, 126개의 SNP가 추출되었으며 모두 p-value가 5e-5 이하였다. 해당 SNP의 위치 정보, 효과 크기, 표준오차 등을 함께 제시했다.

지역사회 기반 코호트, 농촌 코호트 통합 자료에서 13,598명이 대상자였으며, 교육 수준을 예측하는 SNP 역시 앞선 과정을 거쳐 추출하였다. 2개 코호트 통합 자료에서는 교육 수준을 예측하는 SNP는 163개가 발견되었다.

126개 SNP과 163개 SNP에서 공통되는 SNP 55개를 최종적으로 선정하여 멘델 무작위 분석에 이용하였다.

1.4. 멘델 무작위 분석 결과

멘델 무작위 분석은 3단계로 이루어졌다. 첫째, KoGES 3개 코호트를 통합한 자료에서 1-표본 멘델 무작위 분석을 실시하였다. 둘째, KoGES 코호트 중 도시 코호트와 지역사회·농촌 코호트를 나누어 2-표본 멘델 무작위 분석을 실시하였다. MR 분석 결과 심혈관계질환과 심혈관계질환 위험요인에 대해 교육 수준이 인과적으로 영향을 미치는 것을 확인하였다.

멘델 무작위 분석 결과와 로지스틱 회귀분석을 이용한 OR을 비교하여 아래 [표 78]에 제시하였다.

1.4.1. 심혈관계질환

교육만 놓고 비교하였을 때 코호트 역학 자료를 이용하여 계산한 OR은 0.79이었다. MR 분석 결과, 1-표본 멘델 무작위 분석에서 산출한 OR은 0.59, 2-표본 멘델 무작위 분석에서 산출한 OR은 0.47이었다.

표 78. 심혈관계질환에 대한 관찰 연구 결과와 멘델 무작위 분석 연구 결과 요약

	Observational	Mendelian Randomization		
	OR (95% CI)	OR (95% CI)		
		IVW	MR- Egger	Weighted Median
지역사회 코호트	0.85 (0.32 - 2.28)			
농촌 코호트	0.73 (0.43 -1.26)			
도시 코호트	0.79 (0.72-0.89)			
One Sample MR				
p-value 5e-5		0.76 (0.66-0.85)	0.79 (0.69-0.89)	0.76 (0.67-0.86)
p-value 5e-6		0.74 (0.66-0.82)	0.75 (0.65-0.88)	0.76 (0.61-0.95)
p-value 5e-7		0.73 (0.65-0.83)	0.74 (0.68-0.80)	0.73 (0.64-0.83)

Two Sample MR			
p-value 5e-5	0.72 (0.68-0.75)	0.76 (0.69-0.84)	0.65 (0.60-0.69)
p-value 5e-6	0.68 (0.62-0.75)	0.68 (0.57-0.81)	0.68 (0.62-0.75)
p-value 5e-7	0.69 (0.61-0.79)	0.71 (0.61-0.84)	0.74 (0.65-0.85)

a. 1 표준편차는 4.0년에 해당

1.4.2. 심혈관계질환 위험요인

심혈관계질환 위험요인에 대한 결과 요약은 아래 [표 79, 80]에 제시하였다. 이분형 변수와 연속형 변수를 나누어 분석하였으며, 대학교 이상의 교육 수준에서 심혈관계질환 위험요인의 위험성이 더 낮았다.

표 79. 심혈관계질환 위험요인에 대한 멘델 무작위 분석 연구 결과 요약 (p-value <5e-5)

	OR (95% CI)		
	IVW	MR-Egger	Weighted Median
One Sample MR			
Binary variables			
고혈압	0.72 (0.65-0.80)	0.75 (0.66-0.85)	0.75 (0.63-0.90)
당뇨병	0.69 (0.58-0.80)	0.72 (0.64-0.82)	0.73 (0.65-0.83)
고지혈증	0.89 (0.81-0.98)	0.85 (0.79-0.90)	0.93 (0.87-0.99)
비만(BMI 25 kg/m^2이상)	0.90 (0.85-0.95)	0.90 (0.84-0.97)	0.90 (0.84-0.97)
비만(BMI 30 kg/m^2이상)	0.85 (0.83-0.87)	0.88 (0.84-0.91)	0.86 (0.83-0.89)
흡연	0.69 (0.59-0.82)	0.71 (0.62-0.81)	0.73 (0.58-0.92)
Two Sample MR			
Binary variables			

고혈압	0.76 (0.61-0.95)	0.74 (0.58-0.96)	0.72 (0.58-0.90)
당뇨병	0.76 (0.70-0.81)	0.77 (0.72-0.83)	0.77 (0.73-0.81)
고지혈증	0.88 (0.83-0.93)	0.91 (0.81-1.01)	0.90 (0.83-0.97)
비만(BMI 25 kg/m^2이상)	0.90 (0.85-0.95)	0.91 (0.81-1.02)	0.90 (0.84-0.96)
비만(BMI 30 kg/m^2이상)	0.85 (0.80-0.90)	0.82 (0.73-0.92)	0.87 (0.81-0.93)
흡연	0.66 (0.58-0.76)	0.64 (0.55-0.74)	0.63 (0.53-0.75)

a. 1 표준편차는 4.0년

표 80. 심혈관계질환 위험요인에 대한 멘델 무작위 분석 연구 결과 요약 (p-value <5e-6)

	OR (95% CI)		
	IVW	MR-Egger	Weighted Median
One Sample MR			
Binary variables			
고혈압	0.76 (0.68-0.83)	0.78 (0.69-0.84)	0.75 (0.68-0.84)
당뇨병	0.72 (0.62-0.83)	0.74 (0.63-0.85)	0.71 (0.62-0.81)
고지혈증	0.89 (0.80-0.97)	0.86 (0.78-0.91)	0.92 (0.86-0.98)
비만(BMI 25 kg/m^2이상)	0.89 (0.84-0.94)	0.91 (0.86-0.98)	0.91 (0.85-0.97)
비만(BMI 30 kg/m^2이상)	0.87 (0.84-0.90)	0.86 (0.82-0.91)	0.87 (0.83-0.91)
흡연	0.68 (0.58-0.81)	0.70 (0.63-0.80)	0.69 (0.59-0.88)
Two Sample MR			
Binary variables			
고혈압	0.77 (0.62-0.94)	0.78 (0.62-0.97)	0.74 (0.60-0.91)
당뇨병	0.74 (0.69-0.81)	0.77 (0.70-0.84)	0.76 (0.70-0.82)
고지혈증	0.89 (0.84-0.92)	0.91 (0.80-1.02)	0.90 (0.81-0.96)
비만(BMI 25 kg/m^2이상)	0.91 (0.85-0.94)	0.90 (0.81-1.00)	0.91 (0.85-0.97)
비만(BMI 30 kg/m^2이상)	0.86	0.84	0.88

	(0.80-0.91)	(0.74-0.93)	(0.81-0.94)
흡연	0.67 (0.57-0.77)	0.66 (0.55-0.75)	0.65 (0.55-0.76)

a. 1 표준편차는 4.0년

VI. 고찰

1. 교육 수준에 대한 GWAS 연구

1.1. 외국의 교육에 대한 SNP와 비교

GWAS 분석을 통해 교육을 예측하는 SNP는 최종적으로 55개를 선정했고, 이는 해외에서 수행된 연구 결과와 다소 차이를 보였다. 해외 연구에서 수행된 GWAS에서 교육에 대한 SNP는 1,271개(Lee, et al., 2018), 162개(Okbay, et al., 2016), 74개(Okbay, et al., 2016) 등등으로 연구마다 다소 차이가 있었다. GWAS를 통해 산출된 SNP 개수가 많고 적음에 따라 교육에 대한 설명력이 변하는 것은 아니다.

기존의 교육에 대한 SNP는 미국 자료에서 추출한 162개 SNP(Okbay, et al., 2016), 영국 자료에서 추출한 74개 SNP(Okbay, et al., 2016)가 주로 이용되었다. 162개 SNP와 74개 SNP 사이에 중복되는 SNP은 27개였다. KoGES 코호트에서 교육과 관련된 약 83만 개의 SNP가 추출되었으며, 기존 162개 SNP 중 4개의 SNP만 KoGES 코호트에 존재했다. 74개 SNP 중 2개의 SNP만 KoGES 코호트에 존재했으며 두 경우 모두 p값이 0.05 이상이었다. 162개 SNP과 74개 SNP에 대한 정보는 부록의 [표 81, 82, 83]에 제시하였다.

KoGES 코호트와 해외 자료 사이의 차이가 교육 수준에만 해당하

는 것인지 검증이 필요하다. Canela-Xandri 등이 수행한 UK Biobank 자료를 이용한 연구 결과와 Nam 등이 KoGES 코호트 자료를 이용한 연구 결과를 부록의 [표 84]에 비교하였다.

키, 체질량지수, 적혈구 수치, 혈소판 수치, 허리 둘레, 엉덩이 둘레, 키 등에서 UK biobank와 KoGES 결과가 다르게 나타났다. 같은 표현형이라 하더라도 KoGES SNP 개수가 더 적게 나타났으며 공통된 SNP는 0~6개 사이였다. 이는 UK Biobank와 KoGES 코호트를 구성하는 인종적 구성과 특성이 매우 다르다는 것을 의미한다. 연구집단의 구성과 인종적 특성이 다르면 GWAS 결과가 다르게 나타날 수 있으며, 각 자료의 GWAS 결과는 해당 연구집단의 특성을 반영한 것으로 판단 할 수 있다. 따라서 해외 자료에서 구한 SNP를 그대로 KoGES 코호트에 적용할 수 없다.

1.2. 교육을 예측하는 유전 변이에 대한 기전

교육을 받는 일은 사회적 현상이라고 할 수 있다. 교육은 짧은 시간이 아니라 긴 시간에 걸쳐 완성되기 때문에 다양한 요인들의 영향을 받게 된다. 개인의 교육 수준을 예측하는 유전 변이에 대한 기전은 크게 직접적인 경로와 간접적인 경로가 있다(Lee, et al., 2018; Okbay, et al., 2019).

기억, 주의, 학습과 같은 인지 기능이 교육과 직접적인 연관이 있기 때문에, 인지 기능과 관련된 유전 변이를 찾으면 그 유전 변이가 교육 수준을 예측한다고 볼 수 있다(Lee, et al., 2018). 인지 기능은 태아기 신경 구조 발달과 관련성이 높다. 신경전달물질, 시냅스, 대뇌피질 등을 발현시키는 유전자와 연관성이 있는 유전 변이를 가지고 있다면 더 높은 교육 수준을 가질 것으로 예측하는 것이다. 부록의 [표 85]은 여러 신경전달물질과 신경수용체에 관여하는 유전자를 보여준다.

성격, 성향, 행동 등과 같은 개인적 특성이 교육에 영향을 주는 것이 간접적인 경로다(Lee, et al., 2018). 인내심과 자기통제력 등이 높을수록 교육 성취도가 높으므로, 이러한 특성과 연관이 있는 유전 변이를 찾을 수 있다면 교육 수준을 예측하는 SNP으로 사용할 수 있다. 하지만 간접적인 경로는 직접적인 경로보다 외부 환경에 영향을 더 많이 받고, 개인의 성격, 행동 등을 측정할 수 있는 방법이 쉽지

않다.

교육 수준을 예측하는 유전 변이에 대한 기전은 몇 가지 한계점을 지닌다. 유전 변이가 위치한 영역은 주로 신경전달물질이나 신경수용체 발현에 관여하는 유전자와 가깝기 때문에 인지 기능과 연관성이 있는 것이지, 유전 변이가 직접적으로 교육과 관련된 유전자의 발현을 일으키는 경우는 드물다. 또한, GWAS 연구에서 사용되는 SNP은 집단 내에서 1% 이상의 빈도를 가지는 변이를 이용하기 때문에 표현형에 영향을 주는 모든 유전 변이를 찾는 것은 어렵다.

1.2. KoGES 결과에 대한 교육을 예측하는 유전 변이에 대한 기전

본 연구에서 교육에 대한 SNP을 추출할 때 재현가능성을 위해 도시 코호트에서 추출한 결과와 지역사회, 농촌 코호트에서 추출한 결과를 비교하였다. 두 결과에서 최종적으로 55개 SNP을 선정했으며, 해당 SNP에 대한 정보는 [표 86]에 제시하였다. 표에는 각각 SNP들이 어디에 위치하였는지를 표시하였다. 그리고 KoGES에서 사용한 SNP들이 위치한 유전자의 기능을 [표 87]에 나타내었다.

앞서 설명하였듯이 유전 변이가 교육을 예측하는 기전은 주로 직접적인 경로에 의해서다(Lee et al, 2018). 해당 유전 변이가 위치하고 있는 유전자의 기능을 살펴보았다. 55개 SNP의 위치는 다양한 유전자에 존재하였다. SNP들은 주로 인트론(intron) 변이에 해당하였다. 유전자는 엑손(exon)과 인트론(intron)으로 구성되어 있는데, 엑손은 단백질을 생성하는 데 필요한 부분으로, mRNA로 복사되어 직접 단백질 합성에 관여한다. 인트론은 직접 단백질 합성에 관여하지 않지만, 인트론 변이가 생길 경우 유전자의 기능에 영향을 미칠 수 있다.

SNP이 위치한 유전자가 신경전달물질, 신경 발달, 시냅스 구성, 에너지 대사 등에 관련되어 있는 경우 SNP가 해당 유전자의 영향을 받을 수 있다.

1번 염색체에 위치한 MCOLN2 유전자는 음이온 채널 활성에 관

여하는 유전자이며, 2번 염색체의 NCMAP 유전자는 마이엘린 관련 단백질과 연관되어 있다. 55개 염색체 중에서 8번 염색체와 12번 염색체에서 SNP들이 많이 분포했는데, 해당 위치에 있는 유전자는 DLGAP2, GIT2, CUX2, ATXN2, ACAD10, NAA25, HECTD4, PTPN11, RPH3A 등이다. 이 유전자들은 시냅스 구성, 신경세포 발달, 대사 관련 효소 등에 대해 기능을 갖고 있다. 특히 8번 염색체의 DLGAP2 유전자는 신경세포 전달과 시냅스 구성에서 중요한 역할을 하는 단백질과 관련이 있으며, CUX2 유전자는 신경세포 증식과 분화에 관련된 요인이다. 비록 이들 변이가 인트론에 위치하였지만, 인트론 변이가 있을 경우 해당 유전자의 영향을 받을 수 있다.

따라서 KoGES에서 추출한 SNP들이 위치한 유전자들의 기능을 고려했을 때, 이 SNP들이 신경전달물질, 신경 발달 등의 유전자 기능에 영향을 받는다고 볼 수 있다. 이 기능들은 교육에 대한 직접적인 경로에 해당하며, KoGES에서 산출한 SNP들이 교육을 예측하는 기전과 연관성이 있다고 해석할 수 있다.

GWAS는 표본 수가 많을수록 효과적이다. 충분한 표본 수가 확보되지 않으면 통계적 검정력이 떨어지고 결과의 신뢰성이 낮아질 수 있다. 본 연구에서는 72,294명의 표본 수가 있었지만, 해외 연구에서 사용한 표본 수는 10만 명 혹은 20만 명이 넘었다. 해외 연구에서 더 많은 표본 수를 사용했기 때문에 통계적 검정력은 우리 나라 결과에 비해 높을 수 있지만, KoGES 코호트 표본 수가 5만 명 이상으로 절대적으로 부족한 것은 아니다. 따라서 GWAS 분석에는 다양한 요인을 고려해야 한다.

교육을 예측하는 SNP는 교육은 복잡하고 다양한 요소들의 조합으로 결정된다(Okbay, et al., 2016). 유전 변이는 교육과 관련된 특성의 일부만을 설명할 수 있으며, 다른 비유전적 요인과 상호작용하는 관계를 반영하지 못할 가능성이 존재한다. 교육은 유전적인 영향 외에 환경적 영향도 받는데 GWAS 분석은 주로 유전적 변이에 초점을 맞춘다. 예를 들어, 교육에는 교육 환경, 가정 환경, 사회경제적 환경 등이 포함되는데 이러한 환경적 요인들은 GWAS에서 감지하기 어렵다. GWAS 분석을 통해 찾은 유전 변이는 해당 표현형과의 통계적 연관성을 의미하기 때문에 교육을 예측하는 SNP가 전체적인 설명력을 완벽하게 제공하지 못할 수 있다.

인구 구조와 유전적 다양성은 GWAS 결과에 영향을 미칠 수 있다. 특정 인구나 인종 집단에서 수행된 GWAS는 그 결과가 해당 집단에만 적용될 수 있으며, 다른 인구나 인종 집단에서는 일반화하기 어려

울 수 있다. 해외 GWAS 연구는 주로 유럽과 미국에서 수행되었으며, 해당 지역은 단일 인종이 아닌 다인종으로 인구 구성이 되어 있는 경우가 많다. 반면 KoGES 코호트에서는 한국인칩을 사용하여 한국인에 맞춘 유전 정보 분석이 가능하다. 본 연구를 통해 얻은 SNP와 해외 연구에서 사용한 SNP는 개수가 다른 것 외에 해당 SNP의 위치 역시 서로 다르다. 본 연구에서는 교육을 예측하는 SNP 10개 중에 9개가 12번 염색체에 집중적으로 위치하였다. 이는 한국인에 해당하는 인종적, 유전적 특성이라고 판단할 수 있다.

2. 멘델 무작위 분석 결과 해석

도시 코호트에서는 126개의 SNP가 산출되었고, 지역사회/농촌 코호트에서는 163개 SNP가 산출되었다. 두 경우에서 중복되는 SNP 55개를 본 연구의 멘델 무작위 분석에 이용하였다. P-value는 5e-5를 기준으로 했으며, p-value가 5e-6일 때는 33개의 SNP가 남았으며, 5e-7일 때는 13개의 SNP가 남았다.

멘델 무작위 분석은 p-value에 따라 계산하였다. 심혈관계질환에서 1-표본 멘델 무작위 분석을 통해 역분산가중치(inverse variance weighted, 이하 IVW)로 계산된 오즈비(odds ratio, 이하 OR)는 p-value 5e-5일 때 0.76, 5e-6일 때 0.74, 5e-7일 때 0.73이었다. 이는 1 표준편차의 교육 연한인 4.0년이 증가하면 심혈관계질환 위험이 24% 낮음을 뜻한다. 2-표본 멘델 무작위 분석에서 역분산가중치로 계산된 오즈비는 p-value 5e-5일 때 0.72, 5e-6일 때 0.68, 5e-7일 때 0.69이었다. 이는 1 표준편차의 교육 연한인 4.0년이 증가하면 심혈관계질환 위험이 28% 낮음을 의미한다.

P-value를 조정했을 때 SNP의 수가 달라졌지만 심혈관계질환과 위험요인의 오즈비는 1 이하로 나와 강건한 결과를 보여주었다. 본 연구의 멘델 무작위 분석을 통해 산출한 교육이 심혈관계질환에 미치는 영향은 교육이 일생에 걸쳐 나타내는 효과를 의미한다.

Bahk 등에 따르면, 우리나라에서 40년 동안 교육 수준의 증가와

사망률 감소를 통해 기대 수명이 증가하였다(Bahk, et al., 2017). 무학자, 초등학교 졸업자 등의 비율이 급격하게 줄어들고 고등학교 졸업 이상의 학력자의 비율이 증가하면서 인구집단의 교육 연한이 증가하였다. 사망률 감소의 원인에 교육 연한의 증가만 있는 것은 아니지만, 교육 수준의 증가가 기대 수명 증가에 기여했다고 볼 수 있다.

2.1. 기존 연구 결과와 비교

본 연구를 통해 교육 수준이 높을수록 심혈관계질환 위험을 낮추는데 인과적 영향이 있음을 발견했다. “자연의 무작위 배정”(Hingorani, et al., 2005)이라고 할 수 있는 유전 정보를 사용한 연구 결과는 관찰 연구 결과와 일치했으며, 연구 결과가 유전적 다형성 때문이라는 증거는 발견되지 않았다.

다양한 환경에 걸친 방대한 관찰 연구에서 교육과 심혈관계질환 사이의 연관성이 밝혀졌다. 반면 이러한 연관성의 인과 관계를 명시적으로 조사한 연구는 비교적 적게 이루어졌다. 인과성에 대한 기존 연구는 아래의 영역에서 이루어졌다. 첫째, 자연 실험을 통해 의무 교육법 변경 전후의 사망률을 비교한 연구로, 예를 들어 최소 교육을 강화하는 국가 법안이 도입되기 전후의 사망률을 살펴본 것이다. 네덜란드에서는 이러한 변화가 모든 원인에 의한 사망률 감소와 관련이

있었다(van Kippersluis, et al., 2009). 영국에서는 지금까지 가장 큰 규모의 연구에서 신체 활동, 체질량지수, 고혈압, 당뇨병, 관상동맥질환 및 모든 원인에 의한 사망률 감소에 대한 인과적 효과가 보고되었다(Davis et al., 2016). 스웨덴에서는 의무 교육을 연장하는 정책을 실시한 결과 40세 이후에 발생한 모든 원인에 의한 사망률이 낮아진 것으로 나타났다(Lager, et al., 2012).

둘째, 일란성 쌍둥이에 대한 연구가 이루어졌다. 일란성 쌍둥이는 동일한 유전자를 가지고 있으며, 쌍둥이 간 질병 결과의 차이는 유전적 영향에서 비롯될 수 없다. 만약 일란성 쌍둥이 간 차이가 있다면 환경적 요인(예: 한쪽이 더 교육을 오래 받은 경우)에 의한 차이일 것이다. 덴마크에서 수행된 쌍둥이 연구에서 교육과 사망률 및 관상동맥질환 발병률 간의 인과 관계에 대한 증거가 처음 발견되었다(Madsen, et al., 2010; Madsen, et al., 2014). 스웨덴에서 수행된 연구는 약 5만 명의 쌍둥이를 대상으로 하였으며 인과 관계에 대한 강력한 증거가 발견되었다(Lundborg, et al., 2016). 기존 인구 기반 연구와 쌍둥이 연구 간 결과를 비교했을 때 교육 수준과 사망률 사이의 연관성이 전혀 약화되지 않았다. 쌍둥이 연구보다 표본 수가 100배 이상 큰 비동일 형제 자매 자료를 사용한 연구에서 교육 수준과 사망률 간의 연관성이 약화되지 않는 것으로 나타났다(Naess, et al., 2012; Sondergaard, et al., 2012). 이는 쌍둥이 연구와 마찬가지로 형제 자매가 공유하는 환경적, 유전적 요인이 교육과 질병 사이의 연

관성을 교란할 가능성이 낮다는 것을 시사한다. 쌍둥이 연구와 형제자매 연구 모두 공유되지 않은 환경 요인으로 인한 교란 가능성이 있지만, 멘델 무작위 분석을 사용한 본 연구 결과를 종합하면, 교육 수준이 높을수록 심혈관계질환 위험이 감소할 수 있다는 인과적 해석이 가능하다.

해외와 본 연구에서 사회경제적 위치 지표를 교육 수준으로 이용했다 하더라도 결과를 해석할 때 고려해야 할 점이 있다. 교육 수준의 측정을 어떻게 했느냐에 따라서 결과의 비교 가능 여부가 결정된다. 예를 들어 해외와 본 연구에서 동일하게 높은 교육 수준의 집단의 심혈관계질환 위험도가 0.85가 나왔다고 하면, 이를 곧바로 해외와 본 연구 결과가 같다고 할 수 없다. 왜냐하면 교육 수준을 어떻게 측정하였느냐에 따라 결과의 해석도 달라지기 때문이다.

대부분 해외 연구에서 교육 수준은 교육 연한으로 사용했으며, 결과변수의 비교 단위는 1 표준편차에 해당하는 3.6년이었다. 본 연구는 교육 수준을 해외 연구와 비교를 위해 범주형 변수를 교육 연한으로 변환하였다. 무학의 교육 연한은 0년, 초등학교의 교육 연한은 6년, 중학교는 9년, 고등학교는 12년, 전문대 14년, 대학교 16년, 대학원 18년으로 설정하였다. 본 연구에서 전체 집단의 평균 교육 연한은 11.1세였으며 1 표준편차는 4.0년이었다.

해외에서 수행한 교육 수준과 심혈관계질환에 대한 멘델 무작위 분석 연구 결과의 요약을 [표 88]에 제시하였다. Tillmann 등이 수행한

연구(Tillmann, et al., 2017)에서, 교육 연한이 3.6년 증가하면 관상동맥질환의 OR은 0.67, 0.60이었다. Carter 등의 연구(Carter, et al., 2019)에 따르면, 교육 연한이 3.6년 증가할 때 관상동맥질환의 OR은 0.63, 뇌졸중의 OR은 0.71, 심근경색증의 OR은 0.61이었다. Gill 등의 연구(Gill, et al., 2019)에서 관상동맥질환의 OR은 0.65였고, Gao 등의 연구(Gao, et al., 2022)에서 허혈성 뇌졸중의 OR이 0.54로 나타났다. Cao 등의 연구(Cao, et al., 2020)에서 관상동맥질환의 OR은 0.64였다.

해외에서 진행된 멘델 무작위 분석 연구에서 교육 수준은 교육을 최종적으로 몇 년을 받았는지 연속형 변수로 주어졌다. 교육 수준을 범주형 변수로 나타낸 연구는 Wen 등이 수행한 사례(Wen, et al., 2020)가 있었으며, 허혈성 뇌졸중의 OR이 0.74였다.

본 연구에서 1표준편차에 해당하는 교육 연한은 4.0년이었고 1-표본 멘델 무작위 분석 결과 OR은 0.76, 2-표본 멘델 무작위 분석 결과 OR은 0.72이었다. 이는 교육 연한이 4.0년 증가하면 심혈관계질환 위험도가 24%, 28% 감소함을 의미한다. 해외 연구에서 사용된 3.6년의 교육 연한에 의한 효과는 본 연구 결과의 효과보다 크게 나타났다.

주요 심혈관계질환 위험요인에 대한 해외 연구에서도 양상은 비슷하였다. 해외 연구에서 결과변수의 비교 단위는 1 표준편차에 해당하는 3.6년이었다. 주요 심혈관계질환 위험요인에 대한 해외의 멘델 무

작위 분석 연구 결과를 [표 89]에 제시하였다. Tillmann 등의 연구(Tillmann, et al., 2017)에서 흡연의 OR은 0.65, 당뇨병의 OR은 0.75이었다. 연구에 따라 심혈관계질환 위험요인의 효과 크기는 다르게 나타났다. 해외 연구에서 사용된 3.6년의 교육 연한에 의한 효과는 본 연구 결과의 효과가 작게 나타났다.

2.2. 관찰 연구 결과와 비교

본 연구는 코호트 자료를 이용해 로지스틱 회귀분석으로 심혈관계 질환의 OR을 분석했다. 3개 코호트 중 지역사회, 농촌 코호트는 표본 수가 부족해 결과의 신뢰구간이 유의미하지 않은 경우가 많았으며, 도시 코호트는 비교적 표본 수가 충분하여 유의미한 결과가 산출되었다. 심혈관계질환의 경우 로지스틱 회귀분석으로 계산한 OR은 0.79이었다. 이는 멘델 무작위 분석 결과보다 크기가 작게 계산되었다. 심혈관계질환 위험요인의 경우에도 비슷한 양상이었다. 관찰 연구 결과보다 멘델 무작위 분석 결과의 크기가 크게 나온 이유는 관찰 연구에서는 모든 교란변수를 보정할 수 없기 때문이다. 예를 들면 교육 수준 외에 소득, 직업, 거주 환경, 개인의 건강 행태 등의 다양한 교란변수가 존재한다.

2.3. 심혈관계질환 위험요인에 따른 OR 차이

본 연구 결과, 여러 심혈관계질환 위험요인의 OR이 각각 다르게 나타났다. 심혈관계질환 위험요인 중 흡연의 OR이 가장 크게 나타났으며 비만과 고지혈증의 OR이 비교적 작게 나타났다.

흡연, 고혈압, 당뇨병은 기존 관찰 연구에서 교육 수준, 소득 수준, 직업 수준에 따라 그 차이가 명확하게 나타났다. 본 연구에서 체질량지수 25 kg/m^2 이상의 비만에 대해서는 교육 수준의 효과가 크게 나타나지는 않았지만, 체질량지수 30 kg/m^2 이상의 비만에 대해서는 체질량지수 25 kg/m^2 이상의 비만보다 크게 산출되었다. 우리나라에서 비만의 기준은 체질량지수 25 kg/m^2 이상이지만 해외에서는 체질량지수 30 kg/m^2 이상으로 정의한다. 교육이 비만과 고지혈증에 직접적으로 미치는 영향 외에도 간접적으로 영향을 미치는 기전이 존재한다. 교육 수준이 높을수록 더 높은 소득을 가질 가능성이 있다. 높은 소득으로 인해 더 좋은 음식과 영양의 섭취를 할 수 있으며, 직업적으로는 육체노동직보다는 사무직을 가질 확률이 더 높다. 사무직 종사자의 경우 육체노동직 종사자에 비해 신체활동량이 부족하다.

비만에 영향을 주는 사회적 요인 중에 체형에 대한 사회적 압박이 존재한다. 외형적으로 날씬한 것을 사회적으로 인정받으려는 분위기가 있으며 이는 여성에서 두드러진다. 또한 높은 교육 수준을 가질수록 건강에 대한 지식과 정보 습득이 빠르며 주위 동료들에 의한 영향

도 크게 받는다. 30 kg/m^2 이상의 고도 비만에서는 교육 연한의 효과가 더 크게 나타났다고 볼 수 있다.

2.4. 내재적, 외재적 타당성

멘델 무작위 분석에 대한 초기 논의부터 연구자들은 인과 관계를 주장하기 위한 가정들이 많은 유전 변이에 대해 유효하지 않을 수 있음을 강조했다. 유전 변이가 결과에 직접적인 영향이 없거나 교란변수와 관련이 없다는 가정을 위반하게 되면 내재적 타당성을 위반하게 된다. 멘델 무작위 분석에서 내재적 타당성을 보장하기 위해 이질성과 다형성 검정을 실시한다.

멘델 무작위 분석에서 또 다른 문제는 외재적 타당성 문제다. 도구변수로 사용한 유전변이가 도구변수 가정을 만족하고 유효한 추정치를 얻는다면 이는 인과적 효과에 해당한다고 판단할 수 있다. 그러나 이러한 추정치를 실제 맥락에서 해석할 때 어떤 질문이 제기되는지가 중요하다. 유전 변이의 존재로 인한 원인의 변화에 따른 인과적 효과가 원인에 대한 개입(intervention)의 인과적 효과와 비슷한지가 외재적 타당성 문제다.

첫째, 개인의 유전자 변이 유무는 임신 시에 결정된다. 멘델 무작위 분석의 결과는 유전적 유무에 따른 노출군의 생애 효과를 의미한다

(Davey Smith, 2006). 많은 경우 임상 개입(clinical intervention)의 효과는 특정 위험요인에 대한 누적된 노출량에 따라 달라진다. 예를 들면, 임상 시험에서 스타틴(statin)을 1년 복용하면 관상동맥질환 위험은 10% 감소하고, 5년 동안 약 25%까지 감소한다(Ference, et al., 2012). 따라서 멘델 무작위 분석의 추정치는 노출에 대한 더 긴 변화의 영향을 나타낸다. 대부분의 임상 개입은 성숙한 개인을 대상으로 수행되며, 개인은 장기간에 걸친 노출에 대해 보상 기전을 가질 수 있다. 예를 들어 노출의 유전 변이가 자궁 내 단계에 영향을 미치는 경우, 유아기에 비타민D에 노출되는 것이 다발성 경화증의 위험을 줄일 수 있다는 연구가 있다(Gale, et al., 1995). 멘델 무작위 분석을 통해 비타민 D가 다발성 경화증의 인과적 위험요인이라는 증거를 제공함에도 불구하고(Morky, et al., 2015), 성인의 비타민 D 보충은 다발성 경화증의 위험을 감소시키지 못할 수 있다.

둘째, 멘델 무작위 분석에서 유전 변이는 노출의 평균 또는 보통 수준에 영향을 미칠 것으로 예상된다. 평생 동안의 노출에 영향을 미치는 장기적 무작위 배정 연구처럼 멘델 무작위 분석 연구는 일정한 장점이 존재한다. 하지만 멘델 무작위 분석은 개인의 급격한 생체표지자 수치의 증가에 따른 개인의 반응 행동에 대한 정보를 제공하지 못한다. 개인의 장기간 높은 평균의 노출은 건강에 영향을 미치지 않을 수 있지만, 노출의 급격한 반응은 건강에 영향을 미칠 수 있기 때문이다. 병리적 수준에서 단기간 표적 개입의 효과는 멘델 무작위 분

석으로 평가하기 어렵다.

셋째, 유전 변이에 의한 노출의 변화는 일반적으로 작다. 진단적으로 중요한 노출에 큰 변화를 일으키는 유전 변이를 찾기 어렵다. 멘델 무작위 분석에서 사용된 여러 유전 변이들은 노출 변화의 약 1~4%를 설명했다(Schatzkin, et al, 2009; Davey Smith, 2011). 모든 인구 집단에게 노출을 일정량으로 작게 낮추거나 높이는 개입이라면, 멘델 무작위 분석 연구는 개입의 효과에 관한 추정치를 제공할 수 있다.

넷째, 유전 변이와 임상적 개입은 노출에 대해 항상 동일한 작용기전을 가지지 않는다. 노출에 대한 유전적 변화는 다른 변수를 통해 인과적 경로로 작용할 수 있다. 예를 들어, FTO 유전자 변이는 포만감을 조절함으로써 체질량지수에 영향을 미친다(Wardle, et al, 2008). 따라서 음석 섭취를 줄이는 것에 기반하지 않은 체질량지수 개입의 효과는 FTO 유전자 변이를 도구변수로 사용한 멘델 무작위 분석 연구 추정치와 다를 수 있다.

다섯째, 유전 변이는 일반적으로 인구집단의 모든 개인에게 영향을 준다. 만약 개입이 전체 인구에 대해 이루어지는 경우, 인구 기반 코호트를 사용한 멘델 무작위 분석은 유효한 추정치를 제공할 수 있다. 하지만 개입이 특정 인구 집단에만 적용되도록 의도된 경우, 관련 추정치를 제공하는 멘델 무작위 분석에 적합한 코호트를 선택하기가 어렵다. 예를 들어, 혈압에 대한 개입은 임상적으로 진단된 고혈압 환자

에게만 적용될 수 있지만, 혈압과 관련된 유전 변이는 인구 집단 전체에 영향을 미칠 수 있다.

3. 연구의 정책적, 학문적 의의

3.1. 정책적 의의

교육 수준이 높을수록 더 좋은 건강 결과를 가진다는 사실이 많은 연구를 통해 밝혀졌지만, 교육의 증가는 비건강적 효과도 가져올 수 있다. 교육을 강화하는 정책은 경제적 생산성을 향상시키고, 삶의 만족도 향상과 같은 비건강적 혜택으로 이어질 가능성 높다(Botero, et al., 2013; Oreopoulos, et al., 2011). 또한, 교육의 확대가 결과적으로 건강이나 복지에 해를 끼칠 수 있다는 증거는 거의 존재하지 않는다. 교육의 증가가 건강에 미치는 영향에 대한 엄격한 과학적 논쟁이 계속되어야 하지만, 현재까지 대부분의 연구 결과에서 교육의 증가가 건강 결과를 개선시킨다는 것이 일반적이다.

본 연구는 교육의 증가가 심혈관계질환 위험도를 낮추는 인과적 효과를 규명하였다. 이 결과를 토대로 실용적이고, 수용 가능하며, 지속 가능한 방식으로 교육을 어떻게 확대할 것인가에 대한 논의가 필요하

다.

인구집단에서 교육을 향상시키는 정책을 시행했을 때 어느 수준까지 교육을 올려야 하며, 얼마만큼의 비용을 투입해야 하는지에 대해선 사회적 동의가 필요하다. 사회적으로 감당 가능한 한계비용(marginal cost)에 대한 논의가 필요하다. 효과적인 건강 정책의 도입을 위해서는 그 정책의 대상 질환이 인구집단에서 중요한 비중을 차지해야 한다. 심혈관계질환은 전체 인구집단의 사망률에 크게 기여하고 있는 중요한 질환이며, 심혈관계질환 위험요인인 고혈압, 당뇨병, 고지혈증, 비만, 흡연 등 역시 인구집단에서의 유병률과 질병부담이 높기 때문에 중요한 정책 대상이라고 할 수 있다. 따라서 본 연구의 결과는 인구집단에서 교육 수준을 향상시키는 정책의 도입에 대해 근거를 제공해 줄 수 있다.

정책 효과를 평가할 때 도구변수를 사용하여 분석할 수 있다. 도구변수 분석에서는 도구변수가 도구변수로서 적합한지가 가장 중요한 문제가 된다. 그리고 현실에서 적정한 도구변수를 찾는 것이 매우 어렵다. 반면, 멘델 무작위 분석은 유전변이를 도구변수로 이용한다는 차이점이 있다. 멘델 무작위 분석은 유전변이를 사용하기 때문에 정책 효과를 직접적으로 평가할 수는 없지만, 독립변수의 생애(life-time) 효과를 측정할 수 있다는 점에서 정책 도입의 근거를 제시할 수 있다는 장점이 있다.

건강불평등의 인과성을 이해하는 것은 중요하다. 원인과 결과를 이루는 기전을 이해할 수 있다면, 건강불평등을 어떻게 줄일 수 있는지에 대한 정보를 획득하고 건강불평등 현상의 윤리적 판단 기준을 제시할 수 있기 때문이다. 심혈관계질환은 세계적으로 중요한 건강 위험요인이다. 교육 수준의 증가가 심혈관계질환 위험을 감소시키는 것이 인과적으로 타당하다면, 정책적으로 건강불평등을 줄이고 국민의 건강 수준을 향상시키는 데 크게 기여할 수 있다.

3.2. 학문적 의의

본 연구는 쉽게 측정할 수 있는 사회경제적 위치 요인 중 하나인 교육과 심혈관계질환 사이의 인과적 관계를 분석했다. 멘델 무작위 분석을 이용하였으며, 이 연구 방법은 다른 연구 결과와 함께 교란으로 인한 bias를 줄임으로써 인과적 관계에 대한 이해를 향상시킬 수 있다. 본 연구는 KoGES 코호트 자료 중에서 지역사회 기반 코호트, 농촌 코호트, 도시 코호트를 사용했으며, 3개 코호트 참여자 중 유전 정보를 가지고 있는 대상자 수는 72,294명으로 기존의 한국에서 진행되었던 멘델 무작위 분석 연구보다 충분한 대상자 수를 확보하여 연구를 진행할 수 있었다. 지금까지 한국에서 교육 수준에 의한 주요 심혈관계질환 위험요인과 심혈관계질환에 대한 인과적 영향을 분석한 연구가 존재하지 않았다. 이 연구는 교육 수준을 예측하는 GWAS 분석을 우리나라에서 처음으로 수행하였고, 나아가 교육 수준이 주요 심혈관계질환 위험요인과 심혈관계질환에 미치는 인과적 영향을 멘델 무작위 분석으로 탐구했다. 일본의 유전 정보와 연계하여 2-표본 멘델 무작위 분석을 수행한 결과에서도 교육 수준이 심혈관계질환 위험을 감소시키는 것으로 나타났다.

사회경제적 요인에 노출된 후 일반적인 질병이 발생하기까지 긴 대기 기간이 있는 경우가 많다. 따라서 이러한 연구 분야는 인과 관

계에 대한 의문을 해결할 수 있는 무작위 대조 시험에 적합하지 않다. 하지만 본 연구는 무작위 대조 시험과 유사한 설계를 가진 멘델 무작위 분석을 이용하여 교육이 심혈관계질환에 미치는 인과적 영향을 규명하였다. 향후 연구의 주요 질문은 교육 수준이 심혈관계질환 위험을 감소시킨다면 그 인과성을 설명하는 기전은 무엇인지에 대한 것이다. 식습관과 같은 건강 행동이 중요한 매개체라면, 식습관에 대한 개입이 건강불평등을 줄이기 위한 정책의 초석이 될 수 있다.

교육과 건강을 연계한 연구의 신뢰성을 평가하려면 교육이 건강에 영향을 미칠 수 있는 기전을 이해하는 것이 도움이 된다(Berkman, et al., 2014). 첫째, 교육은 더 안전한 직업과 더 높은 소득으로 연결된다. 교육 수준이 건강에 직접 영향을 주기도 하지만, 직업과 소득을 매개로 하여 건강에 영향을 줄 수 있다. 둘째, 교육은 단순한 사실적 지식이 아니라 지속적인 인지적 또는 정서적 기술을 형성하는 데 도움을 준다. 문해력과 수리력은 개인이 건강한 결정을 내리는 데 도움을 준다. 높은 교육 수준을 가질수록 추상적으로 사고하고, 규칙을 준수하는 능력 등이 높을 수 있다. 셋째, 교육은 인적 관계 형성에 영향을 준다. 앞서 설명하였듯이 교육 수준이 높을수록 좋은 직업과 높은 소득을 얻을 가능성이 증가하기 때문에, 비슷한 수준의 학력, 직업, 소득을 가진 사회적 네트워크를 구성할 가능성이 높다.

최근 우리나라 상황을 비추어 보았을 때 교육 수준을 어떻게 측정할 것인지는 중요한 연구과제다. 30대 이하의 젊은 연령층에서 의무

교육 확대와 높은 대학 진학률로 인해 교육 수준이 급격히 높아졌다. 통계청에 의하면 2022년 대학 진학률이 73.3%에 이른다. 대부분 대학 졸업 이상의 학력을 갖게 되면 교육 수준이 사회경제적 위치 지표로서 갖는 변별력은 감소하게 된다. 대학 졸업자 내에서 지역별, 학교별 차이가 있는 점을 감안하여, 추후 더 세분화된 교육 수준의 측정 지표가 필요하다.

교육 외에 다른 독립변수를 이용하여 멘델 무작위 분석을 통해 인과적 영향을 분석하는 것도 가능하다. Do는 학원 교습 시간 제한 정책 도입을 도구변수로 하여 수면 시간이 비만에 미치는 인과적 영향을 탐구하였다(Do YK, 2019). 멘델 무작위 분석을 통해 수면을 예측하는 유전변이를 찾고, 비만에 대한 영향을 탐구한다면 기존의 연구 결과를 더 강건하게 비교해 볼 수 있다.

그간 우리나라에서 유전정보를 활용한 MR 분석 연구는 연구대상자 수가 부족하여 인과성 분석이 쉽지 않은 상황이었다. 유전정보를 이용하는 연구가 활발해지면서 예전에 비해 KoGES 코호트 자료의 대상자 수가 충분히 늘어남에 따라 MR 분석 연구가 많아질 것으로 기대된다. 앞으로 유전정보를 포함하는 대상자 수가 해외처럼 10만 명 이상 축적된다면 MR 분석 연구를 통한 인과성 평가는 더 강건해질 수 있다. 본 연구에서는 교육이 심혈관계질환에 인과적으로 영향을 분석했지만, 추후 다른 중요한 건강 위험요인에 대한 분석이 가능할 것이라 생각된다.

3.3. 연구의 제한점

이 연구는 자료원의 한계, 분석 방법 등에 있어 몇 가지 제한점을 가진다.

첫째, 자료의 표본 수가 해외 주요 연구 사례에 비해 아직 부족하다. 도시 코호트의 한국인칩 유전 정보가 공개되면서 많은 표본 수가 확보되었지만, 해외의 UK Biobank와 같이 10만 명이 넘는 표본 수에는 미치지 못했다. GWAS 분석은 참여 대상자가 많을수록 결과의 신뢰성이 증가하기 때문에 멘델 무작위 분석에서도 연구 대상자 수가 충분히 확보되어야 결과의 강건성이 보장될 수 있다.

둘째, 도구변수 가정을 완벽하게 만족시키지 못했다. 도구변수의 가정에는 3가지 조건이 있는데, 그 중 유전 변이가 다른 교란변수에 영향을 주지 않고 독립변수에만 연관성이 있어야 한다. 다른 표현형에도 영향을 준다면 그것은 다형성을 나타내게 된다. 본 연구에서 멘델 무작위 분석을 수행하고 IVW를 산출해 효과크기를 구했지만, 민감도 분석으로 사용한 MR-Egger 회귀분석 결과에서 IVW 결과와 완전히 일치하지는 않았다. MR-Egger 결과가 IVW가 완전히 일치한다면 다형성이 0이라는 의미가 되는데, 현실적으로 다형성을 100% 제거하는 것은 불가능하다. 기존 멘델 무작위 분석 연구에서도 MR-Egger 결과가 IVW와 완전히 일치하는 경우는 없었고, MR-Egger 결과가 IVW 결과와 반대로 나온다거나 너무 차이가 심

하게 벌어지는 경우가 아니라면 IVW 결과를 인과적 효과의 크기라고 판단할 수 있다.

셋째, 자녀의 건강 결과를 직접적으로 유발하는 부모의 행동과 연관될 때 가정이 깨지는 왕조 효과(dynastic effect)가 없다고 가정하였다. 왕조 효과는 여러 세대에 걸쳐 전달되는 특성으로 인해 멘델 무작위 분석의 인과 추정치에 교란이 발생하는 것을 의미한다. 예를 들면, 고등 교육을 선호하는 유전적 소인을 가진 부모는 자녀에게 더 좋은 식단을 제공하려고 할 수 있다. 하지만 부모의 교육 수준은 자녀의 일반적인 심혈관계질환 위험 요인을 제대로 예측하지 못한다고 나타났다(Kvaavik, et al., 2012).

넷째, 교육과 심혈관계질환 사이의 인과성을 중재할 수 있는 매개 경로의 효과를 파악하지 못했다. 교육이 직접적으로 심혈관계질환에 미치는 영향과 더불어 교육과 심혈관계질환 사이에 심혈관계질환 위험요인이 매개체로 존재할 수 있다. 본 연구는 교육이 심혈관계질환에 미치는 전체 효과를 살펴보았고, 각각의 매개 효과는 알지 못했다. 교육과 심혈관계질환에 대한 매개 분석은 후속 연구로 진행될 필요가 있다.

다섯째, 교육을 예측하는 SNP들의 기전은 주로 인지, 기억과 같은 지적 능력에 대한 유전자 기능으로 설명할 수 있는데, 이는 엄밀하게 증가한 교육 연한에 대한 기전은 아니다. 즉, 공부를 잘 할 수 있는 능력에 대한 기전과 오랜 기간 교육을 받는 기전은 다를 수 있다. 하

지만 인지, 학습 능력이 좋을수록 교육 수준이 높을 것이라고 가정한다면, 교육 연한의 증가에 대한 기전을 본 연구의 SNP들이 교육을 예측하는 기전으로 설명할 수 있다고 판단할 수 있다.

VII. 결론

이 연구는 교육 수준이 주요 심혈관계질환 위험요인과 심혈관계질환에 미치는 인과적 영향을 멘델 무작위 분석을 통해 밝혔다. 자료원으로 지역사회 기반 코호트, 농촌 코호트, 도시 코호트를 사용했으며, 멘델 무작위 분석을 위해 교육 수준과 심혈관계질환 위험요인, 심혈관계질환에 대한 GWAS 분석을 수행했다.

교육 수준은 교육 연한으로 변환하여 분석했다. 무학은 0년, 초등학교 중퇴는 3년, 초등학교 졸업 및 중학교 중퇴는 6년, 중학교 졸업 및 고등학교 중퇴는 9년, 고등학교 졸업은 12년, 전문대 졸업은 14년, 대학교 졸업은 16년, 대학원 이상은 18년의 연한으로 변환하였다. 전체 연구집단의 평균 교육 연한은 11.1년이었으며, 1 표준편차에 해당하는 교육 연한은 4.0년이었다.

GWAS 분석에 앞서 자료의 질 관리를 실시하였다. 전체 유전 정보는 국립자원인체자원은행에서 대체(imputation)을 실시하여 제공되어 유전 정보의 결측치는 없었다. 전체 SNP은 8,056,211개에서 마이너 대립유전자 빈도 1% 이하인 9,770개를 제외하였다. 하디-와인버그 검정을 실시해 멘델 유전법칙을 따르는 SNP를 남겼으며, 5e-6을 기준으로 72,098개 SNP을 제외하였다. 마지막으로 LD 전정(pruning)을 실시해 7,135,046개의 SNP을 제외해 총 839,297개의 SNP를 남

겼다.

위 과정을 통해 추출한 839,297개 SNP 중에서 교육 연한을 예측하는 SNP을 찾기 위해 도시 코호트와 지역사회/농촌 코호트 두 경우에서 추출하였다. 이는 교육 수준을 예측하는 SNP의 재현가능성(replication)을 위해 실시하였다. 도시 코호트에서는 126개의 SNP가 산출되었고, 지역사회/농촌 코호트에서는 163개 SNP가 산출되었다. 두 경우에서 중복되는 SNP 55개를 본 연구의 멘델 무작위 분석에 이용하였다. P-value는 5e-5를 기준으로 했으며, p-value가 5e-6일 때는 33개의 SNP가 남았으며, 5e-7일 때는 13개의 SNP가 남았다.

멘델 무작위 분석은 p-value에 따라 계산하였다. 심혈관계질환에서 1-표본 멘델 무작위 분석을 통해 역분산가중치(inverse variance weighted, 이하 IVW)로 계산된 오즈비(odds ratio, 이하 OR)는 p-value 5e-5일 때 0.76, 5e-6일 때 0.74, 5e-7일 때 0.73이었다. 이는 1 표준편차의 교육 연한인 4.0년이 증가하면 심혈관계질환 위험이 24% 낮음을 뜻한다. 2-표본 멘델 무작위 분석에서 역분산가중치로 계산된 오즈비는 p-value 5e-5일 때 0.72, 5e-6일 때 0.68, 5e-7일 때 0.69이었다. 이는 1 표준편차의 교육 연한인 4.0년이 증가하면 심혈관계질환 위험이 28% 낮음을 의미한다.

이 연구는 우리나라에서 교육 수준이 심혈관계질환 위험요인과 심혈관계질환에 미치는 인과적 영향을 처음으로 탐구하였다는 점에서 의의가 있다. 심혈관계질환은 중요한 사망원인과 질병부담 원인 중

하나인데, 이번 연구를 통해 교육 수준의 증가가 심혈관계질환 위험도를 감소시키는 크기가 인과적으로 측정되었다. 이 결과를 바탕으로 우리나라에서 사회경제적 불평등을 완화하기 위한 정책의 근거를 세우는 데 도움을 줄 수 있다.

이 연구에서 가지는 제한점이 몇 가지 있는데 교육 수준을 범주형 변수로 고려했고, 표본 수가 해외 연구 사례에 비해 아직은 부족했다는 점이다. 이러한 제한점들에 대해서는 후속 연구와 새로운 자료 구축을 통해 해결해야 할 것이다. 사회경제적 위치 지표를 교육 외에 다른 지표들을 사용할 수 있는지도 후속 연구가 진행해야할 과제다. 자료 구축이 지속적으로 이루어지면서 표본 수도 더 늘어날 것으로 생각되며 추후에 진행하는 멘델 무작위 분석 연구에도 도움이 될 것이다.

사회경제적 위치가 건강에 미치는 인과적 영향을 밝히는 것은 학문적으로도 중요한 의의가 있으며, 나아가 사회정책적으로도 건강불평등 개선을 위한 근거를 제시하는 토대가 된다.

부 록

표 81. 교육에 대한 해외의 162개 SNP 목록(p-value <5e-8)

SNP	Chr	Position	ALT	EAF	Effect size	P-value
rs301800	1	8490603	T	0.18	0.016	2.853E-08
rs56044892	1	41830086	T	0.20	-0.016	5.37E-09
rs12076635	1	44026656	C	0.79	0.018	3.107E-11
rs12410444	1	44188719	A	0.70	-0.017	6.012E-13
rs142328051	1	44371441	T	0.91	0.022	3.596E-08
rs12143094	1	72105376	C	0.06	0.029	2.729E-09
rs34305371	1	72733610	A	0.10	0.036	1.517E-20
rs2568955	1	72762169	T	0.25	-0.016	5.769E-10
rs12142680	1	73615892	A	0.09	0.026	8.967E-10
rs12145291	1	74161795	T	0.94	-0.029	2.209E-08
rs12754946	1	77989923	T	0.57	0.013	1.483E-08
rs1008078	1	91189731	T	0.40	-0.017	3.1E-14
rs12134151	1	96202443	C	0.50	-0.013	1.142E-08
rs4378243	1	98395881	T	0.83	0.018	1.041E-09
rs17372140	1	98572382	A	0.30	-0.014	9.191E-09
rs648163	1	199315998	T	0.26	0.014	1.384E-08
rs11588857	1	204587047	A	0.21	0.020	3.273E-13
rs35771425	1	211609768	T	0.79	0.018	2.71E-11
rs78365243	1	211737950	T	0.95	0.029	2.218E-08
rs2992632	1	243503764	A	0.72	0.016	3.245E-11

rs7590368	2	10961474	T	0.73	-0.014	2.722E-08
rs76076331	2	10977585	T	0.14	0.020	2.375E-10
rs7605827	2	15598740	A	0.29	0.016	4.855E-08
rs17504614	2	51080481	T	0.80	0.016	1.559E-08
rs1606974	2	51873599	A	0.12	0.022	1.821E-10
rs56158183	2	60632924	A	0.07	0.025	1.421E-08
rs7593947	2	60704933	A	0.51	0.015	2.394E-11
rs356992	2	60753593	C	0.30	0.017	4.028E-12
rs268134	2	65608363	A	0.25	0.014	3.528E-08
rs6715849	2	100306378	A	0.44	-0.015	1.648E-11
rs4851251	2	100753490	T	0.27	-0.015	1.357E-09
rs12987662	2	100821548	A	0.39	0.025	8.522E-28
rs71413877	2	100924822	A	0.04	0.035	1.91E-09
rs34106693	2	101151830	C	0.83	0.017	1.798E-08
rs77702819	2	101328728	T	0.09	0.022	2.929E-08
rs17824247	2	144152539	T	0.59	-0.016	2.405E-12
rs10178115	2	155451738	T	0.54	0.014	5.841E-10
rs10930008	2	161854736	A	0.73	-0.014	4.135E-08
rs16845580	2	161920884	T	0.63	0.016	4.135E-12
rs4500960	2	162818621	T	0.46	-0.014	2.56E-10
rs1596747	2	193802478	A	0.51	0.014	1.14E-09
rs4675248	2	202880230	A	0.40	-0.012	4.391E-08
rs12694681	2	226609241	T	0.69	0.014	1.805E-08
rs11687170	2	237058144	T	0.83	0.021	1.394E-09
rs7429990	3	47901803	A	0.27	-0.015	8.435E-09

rs62263033	3	48370424	T	0.96	0.037	5.595E-09
rs140711597	3	48469441	C	0.98	0.052	1.659E-08
rs34638686	3	48682658	T	0.10	0.023	1.509E-09
rs3172494	3	48731487	T	0.12	0.023	8.975E-11
rs55786114	3	48982335	T	0.07	-0.030	4.111E-11
rs113011189	3	49250007	T	0.09	-0.025	2.909E-08
rs13090388	3	49391082	T	0.31	0.026	2.577E-26
rs62262721	3	49768927	T	0.96	0.042	3.414E-09
rs11130222	3	49901060	A	0.59	0.025	3.676E-28
rs6800916	3	50052873	A	0.08	-0.024	1.70E-08
rs2624818	3	50056265	A	0.11	0.021	8.627E-09
rs112634398	3	50075494	A	0.95	0.038	2.744E-12
rs71326918	3	50174844	A	0.10	0.022	1.021E-08
rs35971989	3	51469248	A	0.84	0.018	2.946E-08
rs7610856	3	71579022	A	0.43	0.012	3.023E-08
rs62263923	3	85674790	A	0.64	-0.017	1.106E-13
rs56262138	3	86183716	A	0.30	0.014	2.293E-08
rs9755467	3	127143885	T	0.16	0.019	5.11E-10
rs7633857	3	160844942	C	0.52	-0.014	4.737E-08
rs12646808	4	3249828	T	0.66	0.015	3.792E-10
rs1967109	4	28720915	A	0.15	-0.017	4.404E-08
rs4308415	4	67821874	C	0.44	-0.013	2.52E-09
rs6839705	4	106144735	A	0.36	0.015	1.194E-10

rs4863692	4	140764124	T	0.32	0.017	4.606E-12
rs1912528	4	140945966	T	0.36	0.014	1.529E-09
rs12640626	4	176626272	A	0.58	0.013	1.658E-08
rs4493682	5	45188024	C	0.17	0.019	1.537E-10
rs1562242	5	57566494	T	0.48	-0.013	5.946E-09
rs61160187	5	60111579	A	0.61	-0.017	2.709E-14
rs113474297	5	60554934	T	0.13	-0.021	8.338E-10
rs10223052	5	60800336	A	0.36	0.016	3.559E-12
rs775326	5	62918416	A	0.32	-0.014	1.218E-08
rs12653396	5	87847273	A	0.56	-0.013	7.65E-09
rs6882046	5	87968864	A	0.74	-0.019	8.116E-14
rs700590	5	88106258	T	0.59	-0.013	2.839E-08
rs75090987	5	104042643	A	0.52	0.014	1.138E-09
rs152603	5	106774922	A	0.65	-0.013	2.012E-08
rs660001	5	113866598	A	0.21	-0.018	1.341E-10
rs62379838	5	120102028	T	0.69	0.013	4.063E-08
rs7776010	6	14723608	T	0.82	-0.021	2.606E-12
rs7772172	6	16662928	A	0.40	0.013	9.832E-09
rs6939294	6	16950631	T	0.23	0.016	2.895E-09
rs2179152	6	26325888	T	0.37	-0.013	9.301E-09
rs56231335	6	98187291	T	0.67	-0.017	7.165E-13
rs1338554	6	98346801	A	0.50	0.015	1.522E-11
rs9401593	6	98549801	A	0.52	-0.024	3.827E-28
rs56081191	6	98557732	A	0.07	0.028	3.672E-09

rs11756123	6	152218079	A	0.35	-0.015	6.432E-11
rs113779084	7	11871787	A	0.31	0.014	2.696E-08
rs12531458	7	39090698	A	0.52	0.012	3.812E-08
rs12702087	7	44812607	A	0.46	0.013	1.735E-09
rs756912	7	71741797	T	0.52	-0.014	1.141E-09
rs11976020	7	72247800	A	0.23	-0.015	4.434E-08
rs12534506	7	92662327	A	0.47	-0.014	3.173E-10
rs148490894	7	99531755	A	0.98	0.044	1.84E-08
rs2406253	7	100077273	A	0.81	0.016	4.639E-08
rs11771168	7	113904061	T	0.24	-0.015	2.557E-08
rs113520408	7	128402782	A	0.27	0.015	7.154E-09
rs17167170	7	133302345	A	0.80	0.019	1.785E-12
rs7791133	7	135237096	A	0.38	-0.014	2.333E-09
rs320700	7	137049477	A	0.65	0.014	3.905E-09
rs1106761	8	142619234	A	0.38	-0.016	1.366E-11
rs11774212	8	145686505	T	0.52	0.016	1.509E-12
rs11998763	9	1787687	A	0.54	0.017	4.614E-14
rs4741343	9	14075095	A	0.18	-0.016	2.318E-08
rs4741351	9	14222782	A	0.30	-0.015	2.978E-10
rs7029201	9	23358081	A	0.41	0.025	7.157E-27
rs7033137	9	72055158	C	0.76	0.015	1.774E-08
rs17425572	9	88006338	A	0.46	0.014	1.382E-09
rs10821136	9	96238731	T	0.34	0.013	3.584E-08
rs10818606	9	124618386	T	0.40	-0.014	5.672E-10

rs10761741	10	65066186	T	0.42	0.013	7.054E-09
rs7914680	10	67965010	T	0.71	-0.014	1.596E-08
rs1925576	10	68689083	A	0.54	-0.012	2.229E-08
rs149613931	10	103550281	T	0.06	-0.028	5.541E-09
rs73344830	10	103816828	A	0.42	0.015	9.929E-12
rs61874768	10	103880118	T	0.18	-0.016	3.852E-08
rs10786662	10	103989812	C	0.55	-0.017	4.629E-14
rs12761761	10	133775375	T	0.24	0.016	1.04E-08
rs7945718	11	12748819	A	0.62	0.014	1.263E-09
rs76878669	11	66092567	C	0.76	0.014	4.124E-08
rs7948975	11	90424638	T	0.64	0.014	1.143E-09
rs111321694	11	110950386	T	0.17	-0.016	4.329E-08
rs79925071	11	121998253	T	0.56	0.013	1.518E-08
rs1550973	11	131291728	A	0.35	-0.014	1.998E-09
rs10772644	12	13417617	C	0.88	0.020	1.649E-08
rs7964899	12	14595756	A	0.44	0.016	4.373E-13
rs2456973	12	56416928	A	0.67	-0.019	5.828E-16
rs1389473	12	92154270	A	0.38	-0.013	4.516E-09
rs10773002	12	123746961	A	0.25	0.022	7.742E-18
rs8002014	13	58358159	A	0.27	-0.024	3.797E-21
rs9556958	13	99100046	T	0.53	-0.015	1.208E-11
rs34344888	14	23387585	A	0.39	-0.016	8.867E-13
rs1115240	14	27090388	C	0.75	-0.016	7.053E-10
rs10483349	14	29629456	A	0.81	-0.017	7.108E-10

rs58694847	14	84916511	C	0.26	-0.018	4.981E-12
rs1378214	15	47579004	T	0.37	-0.015	1.846E-11
rs6493271	15	47613593	T	0.83	0.017	4.212E-09
rs281302	15	47686662	A	0.56	-0.013	2.879E-09
rs12900061	15	66009248	A	0.18	0.019	5.043E-11
rs4076457	15	78007213	T	0.25	0.015	8.849E-09
rs28420834	15	82513121	A	0.45	-0.014	2.668E-10
rs8049439	16	28837515	T	0.59	0.015	6.992E-11
rs11643654	16	51183728	A	0.60	0.013	1.997E-08
rs9914544	17	18787828	A	0.62	-0.013	4.663E-08
rs192818565	17	43991515	T	0.80	0.020	2.015E-12
rs9964724	18	35159124	T	0.68	0.018	2.389E-14
rs12956009	18	44768024	T	0.57	-0.013	3.75E-09
rs62100765	18	50735418	T	0.42	-0.015	1.078E-10
rs7241530	18	75907992	T	0.36	-0.013	2.281E-08
rs1382358	19	13171424	T	0.87	0.020	1.663E-08
rs111730030	19	13268826	T	0.06	-0.029	7.507E-09
rs12462428	19	16694610	T	0.81	0.016	3.31E-08
rs78387210	20	47823441	T	0.09	0.023	8.411E-09
rs6065080	20	59832791	T	0.36	-0.013	1.155E-08
rs35532491	22	34329603	A	0.90	-0.022	7.153E-09
rs7286601	22	51121416	T	0.54	-0.014	1.989E-09

표 82. 교육에 대한 해외의 74개 SNP 목록(p-value <5e-8)

SNP	Chr	Position	ALT	EAF	Effect size	P-value
rs301800	1	8490603	T	0.18	0.019	1.79E-08
rs11210860	1	43982527	A	0.37	0.017	2.36E-10
rs34305371	1	72733610	A	0.09	0.035	3.76E-14
rs2568955	1	72762169	T	0.24	-0.017	1.80E-08
rs1008078	1	91189731	T	0.41	-0.016	6.01E-10
rs11588857	1	204587047	A	0.21	0.020	5.27E-10
rs1777827	1	211613114	A	0.59	0.015	1.55E-08
rs2992632	1	243503764	A	0.72	0.017	8.23E-09
rs76076331	2	10977585	T	0.15	0.020	3.63E-08
rs11689269	2	15621917	C	0.33	0.016	1.28E-08
rs1606974	2	51873599	A	0.12	0.022	2.80E-08
rs11690172	2	57387094	A	0.59	0.015	1.99E-08
rs2457660	2	60757419	T	0.64	-0.017	7.11E-10
rs114598875	2	60976384	A	0.84	-0.020	2.41E-08
rs10496091	2	61482261	A	0.29	-0.018	5.62E-10
rs13402908	2	100333377	T	0.46	-0.018	1.70E-11
rs4851251	2	100753490	T	0.27	-0.017	1.91E-08
rs12987662	2	100821548	A	0.39	0.027	2.69E-24
rs17824247	2	144152539	T	0.59	-0.016	2.77E-09
rs16845580	2	161920884	T	0.63	0.016	2.65E-09

rs4500960	2	162818621	T	0.46	-0.016	3.75E-10
rs6739979	2	193731929	T	0.63	-0.015	4.70E-08
rs2245901	2	194296294	A	0.40	-0.016	4.54E-09
rs55830725	2	237056854	A	0.17	-0.022	5.37E-10
rs35761247	3	48623124	A	0.05	0.034	3.82E-08
rs62259535	3	48939052	A	0.96	0.048	2.63E-09
rs148734725	3	49406708	A	0.32	0.025	1.36E-18
rs11712056	3	49914397	T	0.57	0.024	3.30E-19
rs112634398	3	50075494	A	0.95	0.036	4.61E-08
rs62263923	3	85674790	A	0.64	-0.016	7.01E-09
rs6799130	3	160847801	C	0.52	-0.015	2.82E-08
rs12646808	4	3249828	T	0.66	0.016	4.00E-08
rs2610986	4	18037231	T	0.67	-0.016	2.01E-08
rs34072092	4	28801221	T	0.90	0.024	3.91E-08
rs3101246	4	42649935	T	0.60	-0.015	1.43E-08
rs4863692	4	140764124	T	0.31	0.018	1.56E-10
rs4493682	5	45188024	C	0.17	0.019	3.32E-08
rs2964197	5	57535206	T	0.50	0.015	3.02E-08
rs61160187	5	60111579	A	0.61	-0.017	3.49E-10
rs324886	5	87896602	T	0.39	-0.015	1.91E-08
rs10061788	5	87934707	A	0.18	0.021	2.46E-09
rs2431108	5	103947968	T	0.68	0.016	5.27E-09

rs1402025	5	113987898	T	0.78	0.017	3.42E-08
rs62379838	5	120102028	T	0.69	0.016	3.30E-08
rs56231335	6	98187291	T	0.67	-0.017	2.07E-09
rs9320913	6	98584733	A	0.48	0.024	2.46E-19
rs7767938	6	153367613	T	0.75	0.017	2.44E-08
rs2615691	7	23402104	A	0.04	-0.037	4.71E-08
rs12531458	7	39090698	A	0.51	0.014	3.11E-08
rs12671937	7	92654365	A	0.53	0.016	9.15E-10
rs113520408	7	128402782	A	0.27	0.017	1.97E-08
rs17167170	7	133302345	A	0.80	0.020	1.14E-09
rs11768238	7	135227513	A	0.34	-0.017	9.90E-10
rs12682297	8	145712860	A	0.46	-0.016	3.93E-09
rs1871109	9	1746016	T	0.55	-0.016	4.35E-10
rs13294439	9	23358875	A	0.59	-0.023	2.20E-17
rs895606	9	88003668	A	0.45	0.015	2.25E-08
rs7854982	9	124644562	T	0.46	-0.015	1.29E-08
rs11191193	10	103802408	A	0.66	0.018	5.44E-11
rs12772375	10	104082688	T	0.40	-0.015	1.56E-08
rs7945718	11	12748819	A	0.63	0.015	1.54E-08
rs7955289	12	14653667	A	0.61	0.017	4.49E-10
rs2456973	12	56416928	A	0.67	-0.020	1.06E-12
rs7131944	12	92159557	A	0.62	0.015	9.02E-09

rs572016	12	121279083	A	0.51	0.014	3.46E-08
rs7306755	12	123767929	A	0.21	0.023	1.26E-12
rs9537821	13	58402771	A	0.72	0.024	1.50E-16
rs1043209	14	23373986	A	0.61	0.018	1.82E-11
rs8005528	14	27098611	A	0.75	-0.018	7.19E-09
rs17119973	14	84913111	A	0.26	-0.019	3.55E-10
rs192818565	17	43991515	T	0.81	0.025	1.47E-12
rs12969294	18	35186122	A	0.34	-0.016	7.24E-09
rs2837992	21	42620520	T	0.39	0.015	3.80E-08
rs165633	22	29880773	A	0.74	-0.018	2.86E-09

표 83. 교육에 대한 해외의 SNP와 KoGES SNP 사이의 비교

	SNP	Chr	Position	ALT	REF	EAF	Beta	SE	p-value	KoGES	
										Beta	p-value
162 SNPs	rs4851251	2	100000000	C	T	0.7463	0.015	0.003	2.8E-08	0.08	0.71
	rs17824247	2	140000000	C	T	0.4198	0.018	0.003	5.3E-13	0.23	0.45
	rs2624818	3	50000000	A	G	0.1063	0.021	0.004	4.5E-07	0.43	0.14
	rs6882046	5	88000000	G	A	0.3134	0.021	0.003	7.9E-14	0.11	0.23
74 SNPs	rs4851251	2	100000000	C	T	0.7463	0.015	0.003	2.8E-08	0.08	0.71
	rs17824247	2	140000000	C	T	0.4198	0.018	0.003	5.3E-13	0.23	0.45

표 84. UK biobank SNP와 KoGES SNP 사이의 비교

	UK Biobank SNPs	KoGES SNPs	공통 SNPs
Weight	2,113	55	2
BMI	6,656	37	2
RBC	2,891	107	3
Platelet	4,534	133	3
Waist circumference	1,048	19	0
Hip circumference	1,448	35	0
Height	10,374	231	6

표 85. Priotized genes in neural communication

	Genes
Voltage-gated ion channels	SCN1A, SCN2A, KCNC2
Neurotransmitter	UNC13C, CPLX1, DNM1
Metabolic receptors	HTR2A, GRM1, CHRM1
Transmembrane transporters	ATP262, MCU, SLC12A5

자료원: Lee et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature Genetics. 2018:50:1112-1121에서 재구성하였음

표 86. KoGES에서 최종적으로 사용한 55개 SNP의 정보

Chr	position	REF	ALT	Nearest Gene	Functional consequence	EAF	Beta	se	p-value
1	85410061	C	T	MCOLN2	intron	2.28%	-0.0870	0.0206	2.51E-05
1	24576893	A	G	NCMAP	intron	9.72%	-0.0216	0.0479	6.34E-06
3	140942232	G	A	PXYLP1	intergenic	48.96%	-0.0452	0.0925	9.91E-07
3	144026997	A	G	LOC105374140	intron	13.24%	-0.0416	0.0873	1.88E-06
3	153585628	T	G	SH3BP5	intron	1.01%	-0.0340	0.0815	3.01E-05
4	92431580	G	A	CCSER1	intron	1.04%	0.0137	0.0332	3.53E-05
4	121613198	G	C	PRDM5	intron	14.82%	-0.0374	0.0919	4.80E-05
5	54852576	C	T	AK4P2	intergenic	1.29%	0.0830	0.02	3.22E-05
5	114210843	C	G	CTNND2	intron	46.40%	0.0361	0.0888	4.89E-05

6	31322706	C	G	HLA-B	intron	5.39%	0.0858	0.0204	2.56E-05
6	67132090	T	C	NUFIP1P1	intergenic	1.85%	0.0916	0.0204	6.98E-06
7	41147	G	A	LINC03015	intergenic	1.97%	0.0307	0.0723	2.22E-05
7	41170	G	A	LINC03015	intergenic	3.44%	-0.0386	0.0927	3.13E-05
7	81580	C	A	LOC10192 9756	intron	3.50%	-0.0243	0.0520	3.02E-06
8	225168	C	T	LOC10192 7566	upstream	4.04%	0.0292	0.0409	4.76E-05
8	234623	T	G	FAM87A	intergenic	3.52%	0.0392	0.0675	5.61E-06
8	239111	T	C	FAM87A	intergenic	3.07%	0.0134	0.0204	5.13E-05
8	750906	C	T	DLGAP2	intron	5.21%	0.0272	0.0259	2.95E-06
8	761499	C	T	DLGAP2	intron	4.11%	0.0065	0.0557	9.06E-05
8	772139	C	T	DLGAP2	non-coding	2.81%	-0.0918	0.0909	3.13E-06
8	888744	C	T	DLGAP2	intron	3.39%	-0.0944	0.0365	9.66E-05
8	1135049	G	A	DLGAP2	intron	3.83%	0.0569	0.0926	5.39E-05

8	1140691	C	G	DLGAP2	intron	4.98%	0.0295	0.0424	4.86E-06
8	1151923	C	G	DLGAP2	intron	2.60%	-0.0620	0.0373	9.61E-06
8	1189658	T	C	DLGAP2	intron	3.17%	0.0591	0.0642	3.57E-06
8	1192943	A	G	DLGAP2	intron	4.75%	0.0829	0.0913	3.64E-06
8	1326979	C	T	DLGAP2	intron	4.65%	0.0236	0.0202	2.43E-06
8	1333258	A	G	DLGAP2	intron	31.66%	-0.1005	0.0404	1.29E-06
8	1360171	T	C	DLGAP2	intron	3.47%	0.0174	0.0267	5.14E-06
8	1472228	G	A	DLGAP2	intron	3.00%	-0.0806	0.0732	2.71E-07
8	1483124	C	A	DLGAP2	intron	2.57%	0.0294	0.0199	1.40E-07
8	1485960	G	T	DLGAP2	intron	2.36%	0.0386	0.0416	3.53E-06
12	110390979	T	C	GIT2	missence	7.44%	0.0185	0.0381	1.17E-06
12	111337494	CT	C	CCDC63	intron	18.75%	0.1102	0.0256	1.65E-05

12	111414461	T	G	CUX2	intergenic	14.11%	0.0248	0.0287	4.85E-18
12	111684461	G	A	CUX2	intron	49.99%	0.0826	0.0199	3.49E-05
12	111718231	C	A	CUX2	intron	18.23%	0.0220	0.0259	2.31E-17
12	111902051	GGT	G	ATXN2	intron	46.36%	-0.0902	0.0201	7.09E-06
12	111994852	A	C	ATXN2	intron	48.98%	0.1082	0.0201	6.88E-08
12	112168009	G	A	ACAD10	intron	15.74%	0.0267	0.0275	3.08E-22
12	112230019	G	C	ALDH2	intron	15.80%	0.0267	0.0275	2.58E-22
12	112468206	C	T	NAA25	intron	15.77%	0.0261	0.0275	1.97E-21
12	112552274	C	T	NAA25	intergenic	23.52%	0.0167	0.0237	1.55E-12
12	112627350	G	A	HECTD4	intron	16.48%	0.0255	0.0270	4.52E-21
12	112678697	G	A	HECTD4	intron	48.31%	0.0865	0.02	1.54E-05
12	112736118	A	G	HECTD4	intron	15.85%	0.0257	0.0275	7.37E-21
12	112829103	CAT	C	HECTD4	intergenic	42.25%	-0.0894	0.0203	1.12E-05

12	112923393	G	A	PTPN11	intron	44.80%	0.0922	0.0201	4.53E-06
12	112982775	A	G	PTPN11	intergenic	40.27%	-0.0863	0.0204	2.29E-05
12	113117897	G	T	RPH3A	intron	15.32%	0.0229	0.0278	1.52E-16
16	19380214	C	T	LOC105371114	intron	17.13%	-0.0126	0.0265	2.01E-06
16	56729619	G	A	NUP93-DT	intron	1.38%	0.0355	0.0855	3.26E-05
16	89653173	T	C	CPNE7	intron	47.48%	0.0854	0.0199	1.84E-05
17	46502682	T	C	SKAP1	intron	1.70%	-0.0165	0.0368	7.04E-06
17	72917513	C	T	USH1G	intron	4.58%	0.0815	0.02	4.46E-05

표 87. 교육에 대한 SNP들이 위치한 Gene과 그 기능들

Gene	Functions
MCOLN2	Cation channel activity
NCMAP	Myelin associated protein
CCSER1	Serine-Rich Coiled-Coil Domain-Containing Protein. Diseases associated with CCSER1 include Hypomyelinating Leukoencephalopathy
PRDM5	PR Domain Zinc Finger Protein 5, cell differentiation
HLA-B	Protein coding gene, autoimmune disease
DLGAP2	a membrane-associated protein that may play a role in synapse organization and signalling in neuronal cells. This gene is biallelically expressed in the brain
GIT2	GIT proteins regulate cytoskeletal dynamics and participate in receptor internalization and membrane trafficking
CCDC63	spermiogenesis
CUX2	Transcription factor involved in the control of neuronal proliferation and differentiation in the brain. Dendrite development and branching, dendritic spine formation, and synaptogenesis in cortical layers.
ATXN2	Involved in EGFR trafficking, acting as negative regulator of endocytic EGFR internalization at the plasma membrane.
ACAD10	Acyl-CoA dehydrogenase related to metabolism.
ALDH2	Aldehyde dehydrogenase enzymes oxidize aldehydes
NAA25	This gene encodes the auxiliary subunit of the heteromeric N-terminal acetyltransferase B complex. Diseases associated with NAA25 include Microphthalmia
HECTD4	Predicted to enable ubiquitin-protein transferase activity. Involved in glucose homeostasis and glucose metabolic process. Predicted to be integral component of membrane. Diseases associated with HECTD4 include Neurodevelopmental

	Disorder With Seizures, Spasticity, And Complete Or Partial Agenesis Of The Corpus Callosum.
PTPN11	The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation.
RPH3A	The protein encoded by this gene is thought to be an effector for RAB3A, which is a small G protein that acts in the late stages of neurotransmitter exocytosis. The encoded protein may be involved in neurotransmitter release and synaptic vesicle traffic.
CPNE7	Calcium-dependent phospholipid-binding protein that may play a role in calcium-mediated intracellular processes.
SKAP1	Positively regulates T-cell receptor signaling by enhancing the MAP kinase pathway. Diseases associated with SKAP1 include Spinocerebellar Ataxia
USH1G	This protein plays a role in the development and maintenance of the auditory and visual systems

자료원: The natinoal library of medicine. https://ncbi.nlm.nih.gov accessed 2023.6.30.

표 88. 교육 수준과 심혈관계질환에 대한 멘델 무작위 분석 연구 결과(해외)

Study	Data Source	Participants	Education	Outcome	OR (IVW)	Per unit
Tillman et al.	SSGAC	349,306	Years of education (162 SNPs)	Coronary heart disease	0.67	3.6 years
Tillman et al.	SSGAC	349,306	Years of education (72 SNPs)	Coronary heart disease	0.60	3.6 years
Carter et al.	UK Biobank	503,317	Years of education (74 SNPs)	Coronary heart disease	0.63	3.6 years
Carter et al.	UK Biobank	503,317	Years of education (74 SNPs)	Stroke	0.71	3.6 years
Carter et al.	UK Biobank	503,317	Years of education (74 SNPs)	Myocardial infarction	0.61	3.6 years
Carter et al.	UK Biobank	503,317	Years of education (74 SNPs)	Cardiovascular disease	0.63	3.6 years
Gill et al.	23andMe	766,345	Years of education (625 SNPs)	Coronary heart disease	0.65	3.6 years
Xiuyun et al.	ARIC	11,509	Categorized (less than high school/ high school/ college or higher)	Ischemic stroke	0.76	
Gao et al.	SSGAC	293,723	Years of education (111 SNPs)	Ischemic stroke	0.54	3.6 years
Cao et al.	UK Biobank	111,349	Years of education (68 SNPs)	Coronay artery disease	0.64	1 year

표 89. 교육 수준과 심혈관계질환 위험요인에 대한 멘델 무작위 분석 연구 결과(해외)

Study	Cohort	Participants	Education	Outcome	Effect or OR (IVW)	Per unit
Tillman et al.	SSGAC	349,306	Years of education (162 SNPs)	Smoking status	0.65	3.6 years
Tillman et al.	SSGAC	349,306	Years of education (72 SNPs)	Type 2 Diabetes	0.75	3.6 years
Tillman et al.	SSGAC	349,306	Years of education (162 SNPs)	Systolic blood pressure	−1.36	3.6 years
Tillman et al.	SSGAC	349,306	Years of education (72 SNPs)	Glucose (mmol/L)	−0.02	3.6 years
Tillman et al.	SSGAC	349,306	Years of education (72 SNPs)	BMI (kg/m^2)	−0.17	3.6 years
Carter et al.	UK Biobank	503,317	Years of education (74 SNPs)	Systolic blood pressure	−0.15	3.6 years
Carter et al.	UK Biobank	503,317	Years of education (74 SNPs)	BMI (kg/m^2)	−0.22	3.6 years
Bockerman et al.	SSGAC	293,723	Years of education (74 SNPs)	BMI (kg/m^2)	−0.64	3.6 years

Zhou et al.	SSGAC	293,723	Years of education (73 SNPs)	Smoking status	0.63	3.6 years
Zhou et al.	SSGAC	293,723	Years of education (73 SNPs)	BMI (kg/m^2)	−0.28	3.6 years
Zhou et al.	SSGAC	293,723	Years of education (73 SNPs)	Triglyceride (mmol/L)	−0.24	3.6 years
Cao et al.	UK Biobank	111,349	Years of education (68 SNPs)	BMI (kg/m^2)	−0.17	1 year
Cao et al.	UK Biobank	111,349	Years of education (68 SNPs)	Type 2 Diabetes	0.61	1 year

참 고 문 헌

Anderson, EL, et al. Education, intelligence and Alzheimer's disease: evidence from a multivariable two-sample Mendelian randomization study. International Journal of Epidemiology. 2020;49(4):1163-1172.

Bahk J, Lynch JW, Khang YH. Forty years of economic growth and plummeting mortality: the mortality experience of the poorly educated in South Korea. Journal of Epidemiol Community Health. 2017;71(3):282-288.

Baicker K, Taubman SL, Allen HL, et al. The Oregon experiment--effects of Medicaid on clinical outcomes. New England Journal of Medicine. 2013;368(18):1713-22.

Berkman LF, Kawachi I, Glymour MM. Social epidemiology: Oxford University Press; 2014.

Bockerman P, Viinikainen J, Pulkki-Raback L, et al. Does higher education protect against obesity? Evidence using Mendelian

randomization. Preventive Medicine. 2017;101:195-8.

Botero J, Ponce A, et al. Education, Complaints, and Accountability. Journal of Law and Economics. 2013;56:959-96

Boyle PJ, Norman P, Popham F. Social mobility: evidence that it can widen health inequalities. Social Science and Medicine. 2009;68(10):1835-42.

Burgess S, Thompson SG. Mendelian randomization: methods for using genetic variants in causal estimation: CRC Press; 2015.

Bush WS, Moore JH. Lewitter F, Kann M. Genome-wide association studies. PLOS Computational Biology. 2012;8(12):e1002822.

Campbell F, Conti G, Heckman JJ, et al. Early Childhood Investments Substantially Boost Adult Health. Science. 2014;343(6178):1478-85.

Cao M, Cui B. Association of educational attainment with

adiposity, type 2 diabetes, and coronary artery diseases: a Mendelian randomization study. Frontiers in public health. 2020;8:112.

Carter AR, Gill D, Davies NM, et al. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. BMJ. 2019;365:l1855.

Canela-Xandri, O, Rawlik, K, Tenesa A. An atlas of genetic associations in UK Biobank. Nature genetics, 2018:50(11):1593-1599.

Card D. Estimating the return to schooling: progress on some persistent econometric problems. Econometrica. 2001;69(5):1127-60.

Case A, Fertig A, Paxson C. The lasting impact of childhood health and circumstance. Journal of Health Economics. 2005;24(2):365-89.

Cho Y, Shin SY, Won S, Relton CL, Davey Smith G, Shin MJ.

Alcohol intake and cardiovascular risk factors: A Mendelian randomisation study. Scientific Reports. 2015;5:18422.

Choi, Y, Lee SJ, Spiller W, Jung KJ, Lee JY, Kimm H, Back JH, Lee S, Jee SH. Causal associations between serum bilirubin levels and decreased stroke risk: a two-sample Mendelian randomization study. Arteriosclerosis, Thrombosis, and Vascular Biology. 2020;40(2):437-445.

Davey Smith G, Ebrahim S, 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?, International Journal of Epidemiology, 2003;32(1):1－22

Davis NM, Dickson M, et al. The causal effects of education on health, mortality, cognition, well-being, and income in the UK Biobank. bioRxiv. 2016

Davies NM, et al. Multivariable two-sample Mendelian randomization estimates of the effects of intelligence and education on health. Elife. 2019;8:e43990.

Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nature Methods. 2011;9,179-181,

Do YK. Causal Effect of Sleep Duration on Body Weight in Adolescents: A Population-based Study Using a Natural Experiment. Epidemiology. 2019;30(6):876-884.

Elo IT, Preston SH. Educational differentials in mortality: United States, 1979-1985. Social Science and Medicine. 1996;42(1):47-57.

Fawcett J, Blakely T. Cancer is overtaking cardiovascular disease as the main driver of socioeconomic inequalities in mortality: New Zealand (1981-99). Journal of Epidemiology and Community Health. 2007;61(1):59-66.

Gao L, Wang K, Ni QB, Fan H, Zhao L, Huang L, Yang M, Li H. Educational Attainment and Ischemic Stroke: A Mendelian Randomization Study. Frontiers in Genetics. 2022;12:794820.

Gill D, Efstathiadou A, Cawood K, Tzoulaki I, Dehghan A. Education protects against coronary heart disease and stroke independently of cognitive function: evidence from Mendelian randomization. International Journal of Epidemiology. 2019;48(5):1468-77.

Greenland S. An introduction to instrumental variables for epidemiologists. International Journal of Epidemiology. 2000;29(4):722-9.

Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. International Journal of Epidemiology. 2016;45:1717-1726.

Heckman JJ, Moon SH, Pinto R, Savelyev PA, Yavitz A. The Rate of Return to the High/Scope Perry Preschool Program. Journal of Public Economics. 2010;94(1-2):114-28.

Heckman JJ. Skill Formation and the Economics of Investing in

Disadvantaged Children. Science. 2006;312(5782):1900-2.

Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7.

Hingorani A, Humpgries S. Nature's randomised trials. Lancet. 2005;366:1906-8

Hong M, Park JW, Yang PS, et al. A mendelian randomization analysis: The causal association between serum uric acid and atrial fibrillation. European Journal of Clinical Investigation. 2020;50(10):e13300.

Howe LD, et al. Effects of body mass index on relationship status, social contact and socio-economic position: Mendelian randomization and within-sibling study in UK Biobank. International Journal of Epidemiology. 2020;49(4):1173-1184.

Jee YH, Lee SJ, Jung KJ, Jee SH. Alcohol Intake and Serum Glucose Levels from the Perspective of a Mendelian

Randomization Design: The KCPS-II Biobank. PLoS One. 2016;11(9):e0162930.

Jeon C, Lee JY, Lee SJ, Jung KJ, Kimm H, Jee SH. Bilirubin and risk of ischemic heart disease in Korea: a two-sample Mendelian randomization study. Epidemiology and Health. 2019;41.

Jung-Choi K, Khang YH, Cho HJ, Yun SC. Decomposition of educational differences in life expectancy by age and causes of death among South Korean adults. BMC Public Health. 2014;14:560.

Kari JT, et al. Education leads to a more physically active lifestyle: Evidence based on Mendelian randomization. Scandinavian Journal of Medicine & Science in Sports. 2020;30(7):1194-1204.

Kawachi I, Adler NE, Dow WH. Money, schooling, and health: Mechanisms and causal evidence. Annals of the New York Academy of Sciences. 2010;1186(1):56-68

Khang YH, Lynch JW, Kaplan GA. Health inequalities in Korea: age- and sex-specific educational differences in the 10 leading causes of death. International Journal of Epidemiology. 2004;33(2):299-308.

Khang YH, Kim HR. Explaining socioeconomic inequality in mortality among South Koreans: an examination of multiple pathways in a nationally representative longitudinal study. International Journal of Epidemiology. 2005;34(3):630-637.

Khang YH, Lynch J, Jung-Choi K, Cho HJ. Explaining Age Specific Inequalities in Mortality from All Causes, Cardiovascular Disease and Ischaemic Heart Disease Among South Korean Male Public Servants: Relative and Absolute Perspectives. Heart. 2008;94(1):75-82

Khang YH, Kim HR. Socioeconomic Inequality in mortality using 12-year follow-up data from nationally representative surveys in South Korea. International Journal for Equity in Health. 2016;15:51.

Kim H, Kim S, Han S, et al. Prevalence and incidence of atherosclerotic cardiovascular disease and its risk factors in Korea: a nationwide population-based study. BMC Public Health. 2019;19(1):1112.

Kitigawa EM, Hauser PM. Differential mortality in the United States: a study in socioeconomic epidemiology. Cambridge, MA: Harvard University Press; 1973.

Kivimäki M, Shipley MJ, Ferrie JE, et al. Best-practice interventions to reduce socioeconomic inequalities of coronary heart disease mortality in UK: a prospective occupational cohort study. Lancet. 2008;372(9650):1648-54.

Kvaavik E, Glymour M, et al. Parental education as a predictor of offspring behavioural and physiological cardiovascular disease risk factors. European Journal of Public Health. 2012;22:544-50

Lager AC, Torssander J. Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. Proceedings of the National Academy of Sciences. 2012;109:8461-6

Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine. 2008;27(8):1133-63.

Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges, International Journal of Epidemiology. International Journal of Epidemiology. 2016;45:908-915

Lee JJ, Wedow R et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature genetics. 2018;50(8):1112-1121

Lee SJ, Jee YH, Jung KJ, Hong S, Shin ES, Jee SH. Bilirubin and Stroke Risk Using a Mendelian Randomization Design. Stroke. 2017;48(5):1154-60.

Lee YH. Overview of Mendelian Randomization Analysis. Journal of Rheumatic Diseases. 2020;27:241-246.

Lee YS, Cho Y, Burgess S, et al. Serum gamma-glutamyl transferase and risk of type 2 diabetes in the general Korean population: a Mendelian randomization study. Human Molecular Genetics. 2016;25(17):3877-86.

Liao LZ, Zhuang XD, Zhang SZ, Liao XX, Li WD. Education and heart failure: New insights from the atherosclerosis risk in communities study and mendelian randomization study. International Journal of Cardiology. 2021;324:115-21.

Lim D, Kong KA, Lee HA, et al. The population attributable fraction of low education for mortality in South Korea with improvement in educational attainment and no improvement in mortality inequalities. BMC Public Health. 2015;15:313.

Lundborg P, Lyttkens CH, Nystedt P. The Effect of Schooling on Mortality: New Evidence From 50,000 Swedish Twins. Demography. 2016;53(4):1135-68.

Mackenbach JP. Health inequalities: Persistence and change in

modern welfare states: Oxford University Press, USA; 2019.

Mackenbach JP, Stirbu I, Roskam AJ, et al. Socioeconomic inequalities in health in 22 European countries New England Journal of Medicine. 2008;358(23):2468-2481.

Mackenbach JP, Kulhánová I, Artnik B, Bopp M, Borrell C, Clemens T, et al. Changes in mortality inequalities over two decades: register based study of European countries. BMJ. 2016;353:i1732.

Madsen M, Andersen AM, et al. Does educational status impact adult mortality in Denmark? A twin approach. American Journal of Epidemiology. 2010;177:225-34

Madsen M, Andersen PK, et al. Are the educational differences in incidence of cardiovascular disease explained by underlying familial factors? a twin study. Social Science and Medicine. 2014;118:182-90

Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new

multipoint method for genome-wide association studies by imputation of genotypes. Nature Genetics. 2007;39:906-913

Marioni RE, Ritchie SJ, et al. Genetic variants linked to education predict longevity. Proceedings of the National Academy of Sciences. 2016;113:13366-71

Moon SH, Kim YJ et al. The Korea Biobank Array: Design and Identification of coding variants associated with boold biochemical traits. Scientific Reports. 2019;9:1382

Morris TT, Heron J, Elenor CM, et al. Interpretation of Mendelian randomization using a single measure of an exposure that varies over time, International Journal of Epidemiology. 2022;51(6):1899-1909

Mountjoy E, Davies NM, Plotnikov D, et al. Education and myopia: assessing the direction of causality by mendelian randomisation. BMJ. 2018;361:k2022.

Muennig P, Robertson D, Johnson G, Campbell F, Pungello EP,

Neidell M. The Effect of an Early Education Program on Adult Health: The Carolina Abecedarian Project Randomized Controlled Trial. American Journal of Public Health. 2011;101(3):512-6.

Naess O, Hoff DA, et al. Education and adult cause-specific mortality - examining the impact of family factors shared by 871,367 Norwegian siblings. International Journal of Epidemiology. 2012;41:1683-91

Sondergaard G, Mortensen LH, et al. Does shared family background influence the impact of educational differences on early mortality? American Journal of Epidemiology. 2012;176:675-83

Naghavi M, Abajobir AA, Abbafati C, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1151-210.

Nam K, Kim J, Lee S. Genome-wide study on 72,298 individuals in Korean biobank data for 76 traits. Cell Genomics,

2022:2(10):100189.

Okbay A, Beauchamp JP, Fontana MA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533(7604):539-42.

Oreopoulos P, Salvanes KG, et al. Priceless: The Nonpecuniary benefits of schooling. Journal of Economic Perspectives. 2011;25:159-84

Pearl J. Causality: Cambridge university press; 2009.

Polimanti R, Ratanatharathorn A, Maihofer AX, et al. Association of Economic Status and Educational Attainment With Posttraumatic Stress Disorder: A Mendelian Randomization Study. JAMA Network Open. 2019;2(5):e193447.

Rubin DB. Causal Inference Using Potential Outcomes. Journal of the American Statistical Association. 2005; 100(469):322-31.

Sakaue S, Kanai M, et al. A cross-population atlas of genetic associations for 220 human phenotypes. Nature Genetics. 2021;53:1415-1424

Sanderson E, Davey Smith G, Bowden J, Munafo MR. Mendelian randomisation analysis of the effect of educational attainment and cognitive ability on smoking behaviour. Nature Communications. 2019;10(1):2949.

Sull JW, Lee TY, Jee SH. Effect of smoking on the association of HHEX (rs5015480) with diabetes among Korean women and heavy smoking men. BMC Medical Genetics. 2018;19(1):68.

Sull, JW, Kim G, Jee SH. Association of MC4R (rs17782313) with diabetes and cardiovascular disease in Korean men and women. BMC Medical Genetics. 2020;21(1):1-6.

Sull, JW, Kim S, Jee SH. Effects of Obesity and Family History of Diabetes on the Association of CETP rs6499861 with HDL-C Level in Korean Populations. Journal of Lipid and Atherosclerosis. 2019;8(2):252.

Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nature Reviews Genetics. 2019;20(8):467-84.

Tillmann T, Vaucher J, Okbay A, et al. Education and coronary heart disease: mendelian randomisation study. BMJ. 2017;358:j3542.

Uffelmann E, Huang QQ, Munung NS, De Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D. Genome-wide association studies. Nature Reviews Methods Primers. 2021;26(1):59

VanderWeele TJ. Explanation in causal inference: developments in mediation and interaction. International Journal of Epidemiology. 2016;45(6):1904-8.

van Kippersluis H, O'Donnell O, van Doorslaer E. Long run returns to education: does schooling lead to an extended old age? Journal of Human Resources. 2009;4:1-33

Viinikainen J, Bryson A, Bockerman P, et al. Does education protect against depression? Evidence from the Young Finns Study using Mendelian randomization. Preventive Medicine. 2018;115:134-9.

Wen X, et al. Education and stroke: evidence from epidemiology and Mendelian randomization study. Scientific Reports. 2020;10(1):1-11.

Yang SJ, Kwak SY, Jo G, Song TJ, Shin MJ. Serum metabolite profile associated with incident type 2 diabetes in Koreans: findings from the Korean Genome and Epidemiology Study. Scientific Reports. 2018;8(1):1-10.

Zhou H, Zhang Y, Liu J, et al. Education and lung cancer: a Mendelian randomization study. International Journal of Epidemiology. 2019;48(3):743-50.

Abstract

Causal Effect of Education on Major Cardiovascular Risk Factors and Cardiovascular Diseases: a Mendelian Randomization Analysis

Youngs Chang

Department of Health Policy and Management

The Graduate School

Seoul National University

The causal effect of socioeconomic factors on health is an important topic in health inequality research. Many observational studies have shown that higher socioeconomic status is associated with better health outcomes. While repeated observational studies

have shown a strong association between socioeconomic status and health outcomes, a research that explores the causal effects of socioeconomic status on health is necessary. Because association does not imply causation.

There are many factors that affect an individual's health. Age, gender, place of residence, genetics, health behaviors, and many other factors play a role. It is difficult to single out the specific factors we want to know about and determine their causal effect on health. The biggest obstacles to determining causality between a specific cause and effect are reverse causation and confounding.

The publicly well-known research method to control for reverse causation and confounding is randomized controlled trial (RCT), but RCT is time-consuming and expensive. All studies cannot be randomized due to ethical issues. Another way to investigate causality is to use instrumental variables. An instrumental variable is a variable that satisfies three following conditions. First, the instrumental variable is related to the independent variable. Second, the instrumental variable affects the outcome variable only through the independent variable. Third, the instrumental variable must not be associated with any confounders between the independent and outcome variables.

Mendelian randomization is a method of analyzing the causality between an independent variable and an outcome variable, with single nucleotide polymorphisms (SNPs) associated with the independent variable as the instrumental variable. Since genes are randomly inherited from parents with a 50% probability, genetic variants known to be associated with the independent variable can be set as instrumental variables. We can assume that the random inheritance of genes is analogous to random assignment. In a Mendelian randomization analysis, the genetic variants used as instrumental variables must satisfy the conditions of the instrumental variables.

This study utilized data from the Korean Genome and Epidemiology Study (KoGES) cohort provided by the National Biobank of Korea, part of Korea National Institute of Health, the Korea Disease Control and Prevention Agency. KoGES cohorts consist of community-based cohorts, rural cohorts, urban cohorts, and twin-family cohorts. Each cohort provides a combination of epidemiologic and genetic information, and this study used the data from the community-based cohort, rural cohort, and urban cohort.

There are two main analytical approaches in this study. First,

we use GWAS analysis to identify genetic variants that predict educational attainment. Second, we will conduct Mendelian randomization analysis of major cardiovascular risk factors and cardiovascular diseases as outcome variables, with genetic variants predicting educational attainment as instrumental variables. For the Mendelian randomization analysis, we conducted one sample Mendelian randomization using the combined KoGES cohort data and two sample Mendelian randomization by dividing the urban cohort and the community/rural cohort.

Education level was analyzed by converting it to years of education. No schooling was converted to 0 years, elementary school dropout to 3 years, elementary school completion and middle school dropout to 6 years, middle school completion and high school dropout to 9 years, high school completion to 12 years, community college completion to 14 years, university completion to 16 years, and graduate school or higher to 18 years. The mean number of years of education for the entire study population was 11.1 years, with one standard deviation of 4.0 years of education.

Data quality control was performed prior to GWAS analysis. Complete genetic information was provided by the National

Biobank of Korea, and all genetic data have no missing genetic information due to imputation prior to data delivery. We excluded 9,770 minor allele frequencies of 1% or less from 8,056,211 total SNPs. The Hardy-Weinberg test was performed to retain SNPs that followed Mendelian inheritance, and 72,098 SNPs were excluded based on p-value 5e-6. Finally, we performed LD pruning to remove 7,135,046 SNPs, leaving a total of 839,297 SNPs.

From the 839,297 SNPs extracted through the above process, we extracted SNPs from both urban and community/rural cohorts to find SNPs that predicted years of education. This was done to ensure replication of the SNPs predicting years of education. The urban cohort yielded 126 SNPs and the community/rural cohort yielded 163 SNPs. The 55 SNPs that overlapped in both cases were used in the Mendelian randomization analysis of this study. P-values were based on 5e-5, with 33 SNPs remaining at a p-value of 5e-6 and 13 SNPs remaining at a p-value of 5e-7.

Mendelian randomization analysis was carried out based on the p-value. In cardiovascular diseases, the odds ratio (OR) calculated by inverse variance weighted (IVW) using a one-sample Mendelian randomization was 0.76 for p-value 5e-5,

0.74 for 5e-6, and 0.73 for 5e-7. This means that a one standard deviation increase of 4.0 years of education is associated with a 24% lower risk of cardiovascular disease. In the two-sample Mendelian randomization, the odds ratios calculated with inverse variance weighting were 0.72 for p-value 5e-5, 0.68 for 5e-6, and 0.69 for 5e-7. This means that a one standard deviation increase of 4.0 years of education is associated with a 28% lower risk of cardiovascular disease.

In a one-sample Mendelian randomization analysis of cardiovascular disease risk factors, the odds ratio for hypertension was 0.72, the odds ratio for diabetes was 0.69, the odds ratio for hyperlipidemia was 0.89, the odds ratio for obesity ≥ 25 kg/m2 was 0.90, the odds ratio for obesity ≥ 30 kg/m2 was 0.85, and the odds ratio for smoking was 0.69. In a two-sample Mendelian randomization analysis, the odds ratio for hypertension was 0.76, the odds ratio for diabetes was 0.76, the odds ratio for hyperlipidemia was 0.88, the odds ratio for obesity ≥25 kg/m2 was 0.90, the odds ratio for obesity ≥30 kg/m2 was 0.85, and the odds ratio for smoking was 0.66.

We compared the SNPs related to the education attainment known from overseas studies with data from the KoGES cohort.

There were 162 and 74 SNPs identified in overseas studies. Upon reviewing the KoGES data, 4 out of the 162 SNPs and 2 out of the 74 SNPs were found in the KoGES data. In this study, we examined where the 55 SNPs related to the level of education are located within each chromosome. These 55 SNPs were located in various genes. The MCOLN2 gene located on chromosome 1 is involved in anion channel activation, and the NCMAP gene on chromosome 2 is associated with myelin-related proteins. Many SNPs were distributed on chromosomes 8 and 12 among the 55 chromosomes, and the genes located at these locations include DLGAP2, GIT2, CUX2, ATXN2, ACAD10, NAA25, HECTD4, PTPN11, RPH3A, and more. These genes have functions related to synaptic composition, neuronal development, and metabolic enzymes. Particularly, the DLGAP2 gene on chromosome 8 is related to a protein that plays an important role in neuronal transmission and synaptic composition, and the CUX2 gene is a factor related to neuronal proliferation and differentiation. Although these mutations are located in introns, the presence of intron mutations could influence the respective genes.

Therefore, considering the functions of the genes where the SNPs extracted from KoGES are located, it can be inferred that

these SNPs may be influenced by genetic functions such as neurotransmitter and neural development. These functions correspond to the direct pathways related to education, and the SNPs derived from KoGES can be interpreted as having relevance to the mechanisms predicting education.

This study holds significance as it is the first to investigate the causal influence of education level on cardiovascular risk factors and cardiovascular diseases in Korea. Cardiovascular disease is an important cause of mortality and burden of disease in Korea, and this study provides a causal measure of the magnitude by which an increase in educational attainment reduces the risk of cardiovascular disease. These findings can be used to inform policies to reduce socioeconomic inequalities in Korea.

There are several limitations of this study. The sample size was still insufficient compared to overseas studies. These limitations will need to be addressed through further research and new data collection. It is also a matter for further research to determine whether other indicators of socioeconomic status can be used in addition to education. As we continue to build our data set, we expect the sample size to increase, which will be helpful for future Mendelian randomization studies.

Identifying the causal effects of socioeconomic position on health is not only important academically, but also provides a foundation for proposing evidence for reducing health inequality in social policy.

keywords : education, cardiovascular disease risk factors, cardiovascular disease, GWAS, Mendelian randomization

Student Number : 2018-34360