연구보고서

유해물질 복합노출로 인한 건강영향에 대한 통계적 분석 방법

이지혜·예신희·이슬비·성정민·신무영

산업재해예방
안전보건공단
산업안전보건연구원

요 약 문

연구기간 2019년 1월 ~ 2019년 11월

핵심단어 복합노출, 톨루엔, 자일렌, 디메틸포름아미드, 특수건강진단

연구과제명 유해물질 복합노출로 인한 건강영향에 대한 통계적 분석 방법

1. 연구배경

최근 환경성 질환 역학여구에서는 수태부터 현재까지의 모든 노출의 총합을 뜻하는 엑스포솜(exposome), 태내에서부터 출생 후 현재까지 환경 노출을 반복적으로 측정하는 출생코호트 연구 등이 활발히 진행되어, 발전된 통계기법들을 활용하여 유해물질 복합노출로 인한 건강영향을 평가하는 연구가 수행되고 있다. 한편, 국내 작업환경에서 유해물질 노출량은 줄어들고 있고 복합 노출이 빈번히 이루어지는 상황에서 복합노출의 건강영향에 대한 연구가 부족한 실정이다. 따라서 직업성 질환 역학연구에서 유해물질 복합노출의 건강영향을 확인하기 위한 적절한 통계기법에 대한 검토가 필요하다.

2. 주요 연구내용

기존 문헌 검토를 통하여 환경성 질환에서 유해물질 복합 노출에 사용되는 통계기법과 특징, 장단점, 직업성 질환 역학연구의 적용 가능성을 고찰하였다. 선택된 통계적 방법에 대한 시범 연구로 2013년부터 2017년 특수건강진단 자료를 분석하였다. 노출 변수로 요 중 마뇨산, 요 중 메틸마뇨산, 요 중 디메틸포름아미드를 사용하였고, 건강

영향평가를 위해서 간기능 검사 수치를 사용하였다. 주성분분석(Principle Component Analysis, PCA), 부분 최소제곱(partial least square, PLS) 회귀, 베이지안커널머신 회귀(Bayesian kernel machine regression, BKMR)를 적용하여 유해물질 복합 노출과 건강영향을 평가하였다.

2013년부터 2017년 특수건강진단 자료를 사용하여 주성분분석과 부분최고제곱 회귀, 베이지안커널머신 회귀를 적용해보았을 때, 2013년 자료에서만 톨루엔, 자일렌, 디메틸포름아미드의 복합노출이 증가할 때 혈중 아스파테이트아미노전이효소가 상승하는 결과를 부분최소재곱 회귀와 베이지안커널머신 회귀 분석을 통해 확인할 수 있었다.

3. 연구 활용방안

직업성 질환 역학연구에서도 유해물질의 상호작용이나 비선형성 등을 고려한 분석 방법에 대한 고찰이 필요한 상황에서, 본 연구를 통해서 각 상황에 맞는 통계방법 선택에 가이드라인으로 활용할 수 있다.

4. 연락처

- 산업안전보건연구원 직업건강연구실 역학조사부 중부권역학조사팀 이지혜
- ☎ 032) 510. 0752
- E-mail jhlee01@kosha.or.kr

차 례

Ⅰ. 서론 ····· 1

1. 연구의 필요성 ····· 1
2. 연구목적 ····· 3

Ⅱ. 연구내용 및 방법 ····· 4

1. 연구내용 및 범위 ····· 4
2. 연구방법 ····· 5
3. 연구윤리 ····· 8

Ⅲ. 연구결과 ····· 9

1. 유해물질 복합 노출을 평가한 통계적 방법 검토 ····· 9
2. 역학 연구 문헌고찰 ····· 20
3. 국내 산업보건 이차자료 특성 검토 ····· 42
4. 복합노출 건강영향 평가의 목표 설정 ····· 43
5. 유해물질 복합노출 연구의 목표에 맞는 복합노출 통계방법 검토 ····· 44
6. 연구진 및 관계자 회의 진행 ····· 47
7. 복합노출의 건강영향을 추정하기 위한 통계분석 방법의 적용 48

1) 연구대상자 및 자료의 특성 ····· 48
2) 단일 생물학적 노출지표와 간기능검사 결과 간의 관련성 분석 ····· 60
3) 주성분분석(Principal component analysis, PCA)을 활용한 복합노출과 간기능검사 결과 간의 관련성 분석 ····· 70
4) 부분최소제곱 회귀(Partial least square regression)를 활용한 복합노출과 간기능검사 결과 간의 관련성 분석 ····· 70
5) 베이지안커널머신 회귀(Bayesian kernal machine regression, BKMR)를 활용한 복합노출과 간기능검사 결과 간의 관련성 분석139

Ⅳ. 고찰 및 결론 ····· 219

Ⅴ. 참고문헌 ····· 221

Ⅵ. Abstract ····· 228

Ⅶ. 부록 ····· 230

표 차 례

<표 2-1> 복합노출 건강영향 평가 통계방법의 예 ····· 5
<표 3-1> 통계분석 모델에 따른 특성 ····· 12
<표 3-2> 각 연구별 연구방법과 통계분석모델 및 특성 ····· 21
<표 3-3> 국내 산업보건 이차자료 특성 ····· 42
<표 3-4> 연구 목표에 따른 통계방법 평가 ····· 45
<표 3-5> 연구 목표 외 통계방법의 장단점 평가 ····· 46
<표 3-6> 연도 별 연구대상자의 특성 (1) ····· 50
<표 3-7> 연도 별 연구대상자의 특성 (2) ····· 51
<표 3-8> 연도 별 노출 및 결과 자료의 특성 ····· 52
<표 3-9> 2013년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터 ····· 71
<표 3-10> 2014년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터 ····· 74
<표 3-11> 2015년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터 ····· 77
<표 3-12> 2016년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터 ····· 80
<표 3-13> 2017년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터 ····· 83
<표 3-14> 2013-2017년 자료(최근 자료 중복제거)에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터 ····· 86
<표 3-15> 2013-2017년 자료(과거 자료 중복제거)에서 세 가지 노출변수

의 주성분분석 후 산출된 구성성분의 로딩벡터 ····· 90
<표 3-16> 2013년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ····· 96
<표 3-17> 2013년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ···· 98
<표 3-18> 2013년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ···· 100
<표 3-19> 2014년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ····· 102
<표 3-20> 2014년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ·· 104
<표 3-21> 2014년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ···· 106
<표 3-22> 2015년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ····· 108
<표 3-23> 2015년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ·· 110
<표 3-24> 2015년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 112
<표 3-25> 2016년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ····· 114

<표 3-26> 2016년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ··116
<표 3-27> 2016년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ····118
<표 3-28> 2017년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ····120
<표 3-29> 2017년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ··122
<표 3-30> 2017년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ··124
<표 3-31> 2013-2017년 자료(최근자료 중복제거)에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ····126
<표 3-32> 2013-2017년 자료(최근자료 중복제거)에서 세 가지 노출변수와 혈중 알라닌아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ····128
<표 3-33> 2013-2017년 자료(최근자료 중복제거)에서 세 가지 노출변수와 혈중 감마글루타밀전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ····130
<표 3-34> 2013-2017년 자료(과거자료 중복제거)에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ····132
<표 3-35> 2013-2017년 자료(과거자료 중복제거)에서 세 가지 노출변수와 혈중 알라닌아미노전이효소의 부분최소제곱법 회귀 후 산출된 첫 번

째 성분의 로딩벡터 ····· 134
<표 3-36> 2013-2017년 자료(최근자료 중복제거)에서 세 가지 노출변수와 혈중 감마글루타밀전이효소의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터 ····· 136
<표 3-37> 베이지안커널머신 회귀 분석 속도 ····· 216

그 림 차 례

[그림 3-1] 복합 노출 통계 세미나 ····· 47
[그림 3-2] 2013년 자료에서 세 가지 노출변수의 상관분석 결과 ····· 53
[그림 3-3] 2014년 자료에서 세 가지 노출변수의 상관분석 결과 ····· 54
[그림 3-4] 2015년 자료에서 세 가지 노출변수의 상관분석 결과 ····· 55
[그림 3-5] 2016년 자료에서 세 가지 노출변수의 상관분석 결과 ····· 56
[그림 3-6] 2017년 자료에서 세 가지 노출변수의 상관분석 결과 ····· 57
[그림 3-7] 2013-2017년 자료(최근 자료 중복제거)에서 세 가지 노출변수의 상관분석 결과 ····· 58
[그림 3-8] 2013-2017년 자료(과거 자료 중복제거)에서 세 가지 노출변수의 상관분석 결과 ····· 59
[그림 3-9] 다중 선형회귀분석을 이용한 효 중 마뇨산과 혈중 아스파테이트아미노전이효소 간의 관련성 ····· 60
[그림 3-10] 다중 선형회귀분석을 이용한 효 중 마뇨산과 혈중 알라닌아미노전이효소 간의 관련성 ····· 61
[그림 3-11] 다중 선형회귀분석을 이용한 효 중 마뇨산과 혈중 감마글루타밀전이효소 간의 관련성 ····· 62
[그림 3-12] 다중 선형회귀분석을 이용한 효 중 메틸마뇨산과 혈중 아스파테이트아미노전이효소 간의 관련성 ····· 63
[그림 3-13] 다중 선형회귀분석을 이용한 효 중 메틸마뇨산과 혈중 알라닌아미노전이효소 간의 관련성 ····· 64
[그림 3-14] 다중 선형회귀분석을 이용한 효 중 메틸마뇨산과 혈중 감마글루타밀전이효소 간의 관련성 ····· 65
[그림 3-15] 다중 선형회귀분석을 이용한 효 중 N,N-디메틸포름아미드

와 혈중 아스파테이트아미노전이효소 간의 관련성 ····· 66
[그림 3-16] 다중 선형회귀분석을 이용한 효 중 N,N-디메틸포름아미드와 혈중 알라닌아미노전이효소 간의 관련성 ····· 67
[그림 3-17] 다중 선형회귀분석을 이용한 효 중 N,N-디메틸포름아미드와 혈중 감마글루타밀전이효소 간의 관련성 ····· 68
[그림 3-18] 2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ···· 71
[그림 3-19] 2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소 간의 관련성 ····· 72
[그림 3-20] 2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 73
[그림 3-21] 2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ···· 74
[그림 3-22] 2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소 간의 관련성 ····· 75
[그림 3-23] 2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 76
[그림 3-24] 2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ···· 77
[그림 3-25] 2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소 간의 관련성 ····· 78
[그림 3-26] 2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 79
[그림 3-27] 2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ···· 80
[그림 3-28] 2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소 간의 관련성 ····· 81
[그림 3-29] 2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및

두 번째 주성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 82
[그림 3-30] 2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ···· 83
[그림 3-31] 2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소 간의 관련성 ····· 84
[그림 3-32] 2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 85
[그림 3-33] 2013-2017년 자료(최근 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ····· 87
[그림 3-34] 2013-2017년 자료(최근 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소 간의 관련성 ····· 88
[그림 3-35] 2013-2017년 자료(최근 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 89
[그림 3-36] 2013-2017년 자료(과거 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ····· 91
[그림 3-37] 2013-2017년 자료(과거 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소 간의 관련성 ····· 92
[그림 3-38] 2013-2017년 자료(과거 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 93
[그림 3-39] 2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ····· 97
[그림 3-40] 2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분

과 혈중 알라닌아미노전이효소 간의 관련성 ····· 99
[그림 3-41] 2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 101
[그림 3-42] 2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ····· 103
[그림 3-43] 2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소 간의 관련성 ····· 105
[그림 3-44] 2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 107
[그림 3-45] 2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ····· 109
[그림 3-46] 2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아라닌아미노전이효소 간의 관련성 ····· 111
[그림 3-47] 2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 113
[그림 3-48] 2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ····· 115
[그림 3-49] 2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소 간의 관련성 ····· 117
[그림 3-50] 2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 119
[그림 3-51] 2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ····· 121
[그림 3-52] 2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소 간의 관련성 ····· 123
[그림 3-53] 2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 125
[그림 3-54] 2013-2017년 자료(최근 자료 중복제거)에서 다중 선형회귀분

석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ····· 127
[그림 3-55] 2013-2017년 자료(최근 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소 간의 관련성 ····· 129
[그림 3-56] 2013-2017년 자료(최근 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 131
[그림 3-57] 2013-2017년 자료(과거 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소 간의 관련성 ····· 133
[그림 3-58] 2013-2017년 자료(과거 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소 간의 관련성 ····· 135
[그림 3-59] 2013-2017년 자료(과거 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소 간의 관련성 ····· 137
[그림 3-60] 2013년 자료에서 세 가지 노출물질과 혈중 아스파테이트아미노전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 140
[그림 3-61] 2013년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소 간의 비선형관계 ····· 141
[그림 3-62] 2013년 자료에서 세 가지 노출물질과 혈중 아스파테이트아미노전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 ····· 142
[그림 3-63] 2013년 자료에서 세 가지 노출물질이 혈중 아스파테이트아미노전이효소에 미치는 복합노출 영향 ····· 143
[그림 3-64] 2013년 자료에서 각 노출물질이 혈중 아스파테이트아미노전

이효소에 개별적으로 미치는 영향 ····· 144
[그림 3-65] 2013년 자료에서 세 가지 노출물질과 혈중 알라닌아미노전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 145
[그림 3-66] 2013년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소 간의 비선형관계 ····· 146
[그림 3-67] 2013년 자료에서 세 가지 노출물질과 혈중 알라닌아미노전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 ····· 147
[그림 3-68] 2013년 자료에서 세 가지 노출물질이 혈중 알라닌아미노전이효소애 미치는 복합노출 영향 ····· 148
[그림 3-69] 2013년 자료에서 각 노출물질이 혈중 알라닌아미노전이효소에 개별적으로 미치는 영향 ····· 149
[그림 3-70] 2013년 자료에서 세 가지 노출물질과 혈중 감마글루타밀전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 150
[그림 3-71] 2013년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소 간의 비선형관계 ····· 151
[그림 3-72] 2013년 자료에서 세 가지 노출물질과 혈중 감마글루타밀전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 ····· 152
[그림 3-73] 2013년 자료에서 세 가지 노출물질이 혈중 감마글루타밀전이효소애 미치는 복합노출 영향 ····· 153
[그림 3-74] 2013년 자료에서 각 노출물질이 혈중 감마글루타밀전이효소에 개별적으로 미치는 영향 ····· 154
[그림 3-75] 2014년 자료에서 세 가지 노출물질과 혈중 아스파테이트아미노전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 155

[그림 3-76] 2014년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소 간의 비선형관계 ····· 156
[그림 3-77] 2014년 자료에서 세 가지 노출물질과 혈중 아스파테이트아미노전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 ····· 157
[그림 3-78] 2014년 자료에서 세 가지 노출물질이 혈중 아스파테이트아미노전이효소에 미치는 복합노출 영향 ····· 158
[그림 3-79] 2014년 자료에서 각 노출물질이 혈중 아스파테이트아미노전이효소에 개별적으로 미치는 영향 ····· 159
[그림 3-80] 2014년 자료에서 세 가지 노출물질과 혈중 알라닌아미노전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 160
[그림 3-81] 2014년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소 간의 비선형관계 ····· 161
[그림 3-82] 2014년 자료에서 세 가지 노출물질과 혈중 알라닌아미노전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 ····· 162
[그림 3-83] 2014년 자료에서 세 가지 노출물질이 혈중 알라닌아미노전이효소에 미치는 복합노출 영향 ····· 163
[그림 3-84] 2014년 자료에서 각 노출물질이 혈중 알라닌아미노전이효소에 개별적으로 미치는 영향 ····· 164
[그림 3-85] 2014년 자료에서 세 가지 노출물질과 혈중 감마글루타밀전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 165
[그림 3-86] 2014년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소 간의 비선형관계 ····· 166
[그림 3-87] 2014년 자료에서 세 가지 노출물질과 혈중 감마글루타밀전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용

..... 167
[그림 3-88] 2014년 자료에서 세 가지 노출물질이 혈중 감마글루타밀전이효소에 미치는 복합노출 영향 168
[그림 3-89] 2014년 자료에서 각 노출물질이 혈중 감마글루타밀전이효소에 개별적으로 미치는 영향 169
[그림 3-90] 2015년 자료에서 세 가지 노출물질과 혈중 아스파테이트아미노전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 170
[그림 3-91] 2015년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소 간의 비선형관계 171
[그림 3-92] 2015년 자료에서 세 가지 노출물질과 혈중 아스파테이트아미노전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 172
[그림 3-93] 2015년 자료에서 세 가지 노출물질이 혈중 아스파테이트아미노전이효소에 미치는 복합노출 영향 173
[그림 3-94] 2015년 자료에서 각 노출물질이 혈중 아스파테이트아미노전이효소에 개별적으로 미치는 영향 174
[그림 3-95] 2015년 자료에서 세 가지 노출물질과 혈중 알라닌아미노전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 175
[그림 3-96] 2015년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소 간의 비선형관계 176
[그림 3-97] 2015년 자료에서 세 가지 노출물질과 혈중 알라닌아미노전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 177
[그림 3-98] 2015년 자료에서 세 가지 노출물질이 혈중 알라닌아미노전이효소에 미치는 복합노출 영향 178
[그림 3-99] 2015년 자료에서 각 노출물질이 혈중 알라닌아미노전이효소

에 개별적으로 미치는 영향 ····· 179
[그림 3-100] 2015년 자료에서 세 가지 노출물질과 혈중 감마글루타밀전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 180
[그림 3-101] 2015년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소 간의 비선형관계 ····· 181
[그림 3-102] 2015년 자료에서 세 가지 노출물질과 혈중 감마글루타밀전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 ····· 182
[그림 3-103] 2015년 자료에서 세 가지 노출물질이 혈중 감마글루타밀전이효소애 미치는 복합노출 영향 ····· 183
[그림 3-104] 2015년 자료에서 각 노출물질이 혈중 감마글루타밀전이효소에 개별적으로 미치는 영향 ····· 184
[그림 3-105] 2016년 자료에서 세 가지 노출물질과 혈중 아스파테이트아미노전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 185
[그림 3-106] 2016년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소 간의 비선형관계 ····· 186
[그림 3-107] 2016년 자료에서 세 가지 노출물질과 혈중 아스파테이트아미노전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 ····· 187
[그림 3-108] 2016년 자료에서 세 가지 노출물질이 혈중 아스파테이트아미노전이효소애 미치는 복합노출 영향 ····· 188
[그림 3-109] 2016년 자료에서 각 노출물질이 혈중 아스파테이트아미노전이효소에 개별적으로 미치는 영향 ····· 189
[그림 3-110] 2016년 자료에서 세 가지 노출물질과 혈중 알라닌아미노전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 190

[그림 3-111] 2016년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소 간의 비선형관계 ····· 191
[그림 3-112] 2016년 자료에서 세 가지 노출물질과 혈중 알라닌아미노전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 ····· 192
[그림 3-113] 2016년 자료에서 세 가지 노출물질이 혈중 알라닌아미노전이효소에 미치는 복합노출 영향 ····· 193
[그림 3-114] 2016년 자료에서 각 노출물질이 혈중 알라닌아미노전이효소에 개별적으로 미치는 영향 ····· 194
[그림 3-115] 2016년 자료에서 세 가지 노출물질과 혈중 감마글루타밀전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 195
[그림 3-116] 2016년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소 간의 비선형관계 ····· 196
[그림 3-117] 2016년 자료에서 세 가지 노출물질과 혈중 감마글루타밀전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 ····· 197
[그림 3-118] 2016년 자료에서 세 가지 노출물질이 혈중 감마글루타밀전이효소에 미치는 복합노출 영향 ····· 198
[그림 3-119] 2016년 자료에서 각 노출물질이 혈중 감마글루타밀전이효소에 개별적으로 미치는 영향 ····· 199
[그림 3-120] 2017년 자료에서 세 가지 노출물질과 혈중 아스파테이트아미노전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 200
[그림 3-121] 2017년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소 간의 비선형관계 ····· 201
[그림 3-122] 2017년 자료에서 세 가지 노출물질과 혈중 아스파테이트아미노전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호

작용 ····· 202
[그림 3-123] 2017년 자료에서 세 가지 노출물질이 혈중 아스파테이트아미노전이효소애 미치는 복합노출 영향 ····· 203
[그림 3-124] 2017년 자료에서 각 노출물질이 혈중 아스파테이트아미노전이효소에 개별적으로 미치는 영향 ····· 204
[그림 3-125] 2017년 자료에서 세 가지 노출물질과 혈중 알라닌아미노전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 205
[그림 3-126] 2017년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소 간의 비선형관계 ····· 206
[그림 3-127] 2017년 자료에서 세 가지 노출물질과 혈중 알라닌아미노전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 ····· 207
[그림 3-128] 2017년 자료에서 세 가지 노출물질이 혈중 알라닌아미노전이효소애 미치는 복합노출 영향 ····· 208
[그림 3-129] 2017년 자료에서 각 노출물질이 혈중 알라닌아미노전이효소에 개별적으로 미치는 영향 ····· 209
[그림 3-130] 2017년 자료에서 세 가지 노출물질과 혈중 감마글루타밀전이효소의 베이지안커널머신 회귀 분석 후 각 노출물질의 사후포함확률 ····· 210
[그림 3-131] 2017년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소 간의 비선형관계 ····· 211
[그림 3-132] 2017년 자료에서 세 가지 노출물질과 혈중 감마글루타밀전이효소의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용 ····· 212
[그림 3-133] 2017년 자료에서 세 가지 노출물질이 혈중 감마글루타밀전이효소애 미치는 복합노출 영향 ····· 213
[그림 3-134] 2017년 자료에서 각 노출물질이 혈중 감마글루타밀전이효

소에 개별적으로 미치는 영향 214

Ⅰ. 서 론

1. 연구의 필요성

혼합물 독성학의 일반적 원칙으로서 화학물질의 병용 시 관찰되는 작용은 기본적으로 세 가지로 정의되는데, 유사작용(용량/농도 부가작용)(similar action(dose/concentration addition)), 비유사작용(독립적 작용)(dissimilar action(independent action)), 상호작용(interactions)이다(이미주, 2016).

이 중 정량화할 수 있는 반응은 용량/농도 부가 작용과 독립적 작용으로 두 가지 원칙 모두 혼합물 내의 화학물질들이 서로의 독성에 영향을 미치지 않는다는 추정에 기반을 두고 있다. 하지만 실제로 화학물질의 혼합물들이 유사하거나 독립적으로 작용하는 경우는 거의 존재하지 않고 있다. 상호작용은 둘 이상의 화학물질의 영향이 기대했던 것 보다 더 심해지거나(synergistic, potentiating, supra-additive) 더 약해지는(antagonistic, inhibitive, subadditive, infra-additive) 현상을 설명하며, 상대적인 용량 수준, 경로, 시기 선택, 노출 기간 및 생물학적 표적 이에 따라 다양하게 나타날 수 있다(이미주, 2016).

현재 유럽 연합(European Union, EU)에서는 혼합 화합물의 위험성 평가를 수행하기 위한 방법론적인 접근을 일반적으로 수용하지 않고 있으며, 사례별 접근은 리뷰를 통해 혼합물 종류에 따라 수행하고 있는데, 이 같은 실정은 우리나라도 마찬가지이다(이미주, 2016).

하지만, 많은 연구에서 기준치 이하의 복합물질에 노출된 경우에도

유의한 건강 영향이 나타날 수 있다고 보고하고있다(Cavieres et al., 2002; Rajapakse et al., 2002; Welshons et al., 2003). Woodruff et al. (2008)이 분석한 사례 연구에서 환경 화학 물질에 대한 기존의 또는 지속적인 노출뿐만 아니라, 명백한 질병의 위험에 독립적으로 기여하는 기존의 생물학적 또는 질병 감수성을 고려하는 것이 중요하다는 것을 보여 주었다(Woodruff et al., 2008). 예를 들어, 갑상선 호르몬 저장과 순환 반감기의 차이로 인해 영아에서 갑상선의 요오드 흡수에 대한 화학물질의 간섭의 영향은 성인보다 실질적으로 더 크다.

최근 환경성 질환 역학연구에서는 수태부터 현재까지의 모든 노출의 총합을 뜻하는 엑스포솜(exposome)[1], 태내에서부터 출생 후 현재까지 환경 노출을 반복적으로 측정하는 출생코호트 연구 등이 활발히 진행되어 유해물질 복합노출 평가 방법에 대한 요구가 높아졌고, 이로 인해 감독된 주요 구성요소 분석(supervised principal component analysis, SPCA), 최소 절대 제곱 축소 및 선택 연산자(least absolute square shirinkage and selection operator, LASSO), 부분 최소-제곱 회귀(partial least-squares regression, PLSR), 베이지안 모형 평균(bayesian model averaging, BMA) 등 발전된 통계기법들을 활용하여 유해물질 복합노출로 인한 건강영향을 평가하는 연구가 수행되고 있다.

특수건강검진데이터, 건강보험공단데이터, 통계청 사망자료 등 직업성 질환 연구에 대용량 자료의 활용도가 높아지고 있는 현재, 직업성 질환 역학연구에서도 유해물질 복합노출 평가에 적절한 통계기법을 탐색하고 선택하여 유해물질 복합노출로 인한 건강영향을 확인할 필요가 있다.

---

1) 평생 환경 노출의 건강영향에 대한 척도. 우리가 가지고 있는 모든 유전정보의 총합을 유전체(게놈, genome)라고 함. 반대로 우리가 노출되는 모든 환경, 즉 식습관, 음주, 흡연 등을 포함하는 생활습관, 외부환경, 스트레스, 경제능력, 장내세균, 운동량 등을 모두 포함하여 엑스포솜(exposome)이라고 함. 이 엑스포솜이 유전자 발현에 영향을 미치는 것으로 알려져 있음.

2. 연구목적

본 연구는 특수건강검진 유해인자 항목에서 유해물질 복합노출로 인한 근로자의 건강영향을 확인하기 위해, 적절한 복합노출 평가 방법을 선택하기 위한 기획연구이다. 구체적인 목표는 다음과 같다.

첫째, 기존 역학연구 및 통계 방법론에 대한 문헌고찰을 통해 직업적 복합노출로 인한 건강영향을 평가하는데 사용될 수 있는 통계 방법의 장단점 및 적용범위를 고찰한다.

둘째, 국내 노동자의 건강영향을 평가한 이차 자료의 특성을 검토하여 분석에 적합한 이차자료를 선정한다.

셋째, 예비 연구(pilot study)를 수행하여 특수건강검진 유해물질 복합노출과 건강영향의 관련성을 분석하기 위한 적절한 통계방법을 제언한다.

Ⅱ. 연구내용 및 방법

1. 연구내용 및 범위

1) 문헌 검색 및 검토

○ 유해물질 복합노출과 건강영향 간의 관련성 평가에 사용되고 있는 통계방법을 고찰한다.

○ 먼저 유해물질 복합노출 분석에 적합한 통계방법을 정리한 논문을 검토하여 고찰할 통계 방법을 선정한다.

2) 이차 자료 검토

○ 노동자의 건강영향을 평가한 이차 자료의 특성을 검토한다.

○ 복합노출의 건강영향 평가 분석에 적합한 이차자료를 선정한다.

3) 복합노출 건강영향 평가 목표 설정 및 통계방법 선택

○ 복합노출 건강영향 평가 통계방법과 국내 노동자의 이차 자료의 특성을 고려하여 본 연구의 복합노출 건강영향 평가 목표를 설정한다.

○ 목표를 기준을 적절한 통계 방법을 선택한다.

4) 예비연구 수행

○ 예비 연구로 유해물질 복합노출과 간 기능 간의 관련성을 확인한다.

○ 특수건강검진 자료에 적용가능한 통계적 방법을 제언한다.

2. 연구방법

1) 문헌 검색 및 검토

○ Google scholar와 PubMed에서 문헌검색을 통해 복합노출과 건강영향 간의 관련성을 발전된 통계방법으로 평가한 역학연구와 리뷰연구를 찾아 검토한 후 복합노출 평가 통계방법의 장점과 단점을 비교한다. 복합노출의 건강영향을 평가할 수 있는 통계방법의 예는 다음과 같다 <표 2-1>.

<표 2-1> 복합노출 건강영향 평가 통계방법의 예

- 베이지안 모델 평균(Bayesian model averaging, BMA)
- 나무 기반 모델(tree based models)
- 다변량 적응 회귀 스플라인(multivariate adaptive regression splines, MARS)
- 탄성 순 회귀(elastic net regression, ENR)
- 베이지안 계층적 회귀(Bayesian hierarchical regression, BHR)
- 반-베이즈 계층적 회귀(semi-Bayes hierarchical regression, semi-BHR)
- 가중 분위수 합 회귀(weighted quantile sum regression, WQSR)
-주 성분 분석(Principal component analysis, PCA)
- 부분 최소 제곱 회귀(partial least squares regression, PLSR)
- 구조 방정식 모델링(structural equation modelling, SEM)
-베이지안 커널 기계 회귀(Bayesian kernel machine regression, BKMR)

○ 문헌검색 시 ‘multipollutant’, ‘multi-pollutant’, ‘copollutant’, ‘co-pollutants’, ‘coexposures’, ‘co-exposures’, ‘concomitant exposure’, ‘concomitant pollutant’, ‘mixture’, ‘Bayesian model averaging’, ‘classification and regression tree’, ‘random forest’, ‘multivariate

adaptive regression splines', 'elastic net regression', 'Bayesian hierarchical regression', 'semi-Bayes hierarchical regression', 'weighted quantile sum regression', 'principal component analysis', 'partial least squares regression', 'structural equation modelling', 'Bayesian kernel machine regression'과 같은 검색어를 조합하여 사용하였다.

○ 문헌 검색을 통해 찾은 리뷰 문헌 및 역학연구 문헌을 검토하여 각 통계방법의 장단점을 분석한다.

○ 직업성 질환 유해물질 복합노출에 적용 가능성 및 범위를 평가한다.

2) 국내 노동자의 건강영향을 평가한 이차 자료의 특성 검토

○ 국내 노동자의 건강자료를 포함한 대용량 이차자료의 특성을 노출변수, 결과 변수, 대상자 수, 자료의 형태, 대표성을 고려하여 검토한다.

○ 검토 대상 자료: 특수건강진단 자료, 건강보험공단 자료, 통계청 사망자료

3) 복합노출 건강영향 평가의 목표를 설정

○ 복합노출로 인한 건강영향을 평가할 수 있는 통계방법의 적용사례를 기반으로 활용 가능한 범위를 검토한다.

○ 국내 노동자의 자료를 사용하여 복합노출로 인한 건강영향을 평가하는 목적과 목표를 설정한다.

4) 국내 노동자의 건강영향 평가방법을 검토

○ 활용 목표를 기반으로 국내 노동자의 이차 자료를 활용하여 유해

물질의 복합노출로 인한 건강영향을 평가하기에 적절한 통계방법을 검토한다.

○ 활용 목표를 기반으로 적절한 통계 방법을 선택한다.

5) 연구진 및 관계자 회의 시행

○ 문헌고찰 시 매주 연구진 내부 회의를 진행하여 최신 통계방법의 활용방안에 대해 토의한다.

○ 전문가 회의를 통해 복합노출로 인한 건강영향을 평가하기 위한 통계방법을 R 패키지를 활용하여 실제 분석해보는 실습을 진행한다.

6) 특수건강검진 자료를 활용한 시범 연구

○ 문헌고찰과 연구진 및 관계자 회의를 통해 선택한 복합노출의 건강영향을 평가할 수 있는 통계방법(주성분분석(principle component analysis, PCA), 부분최소제곱 회귀(partial least square regression, PLS), 베이지안커널머신 회귀(bayesian kernal machine regression, BKMR))을 사용하여 저 농도 간독성 유해물질들의 복합 노출 시 혈중 간기능 검사 결과에 미치는 영향을 평가한다.

○ 2013년부터 2017년의 특수건강검진 자료를 사용하였으며, 요 중 마뇨산, 요 중 메틸마뇨산, 요 중 디메틸포름아미드를 모두 검사하고, 간기능검사 결과와 보정변수(나이, 성별, 체질량지수, 흡연상태, 음주 습관, 중등도 운동 습관)에 결측 값이 없는 사람들만 연구대상으로 분석하였다.

○ 저 농도 복합노출의 건강영향을 확인하기 위하여 실무지침 상 노출 기준 이상의 요 중 마뇨산(노출 기준: 2.5 g/g crea.), 요 중 메틸마뇨산(노출 기준: 1.5 g/g crea), 요 중 N,N-디메틸포름아미드(30mg/g

crea.) 결과 값을 가진 사람들은 연구대상자에서 제외하였다.

○ 요 중 마뇨산, 요 중 메틸마뇨산, 요 중 N,N-디메틸포름아미드의 검사 결과 값이 0이거나 검출한계(limit of detection, LOD) 미만인 사람은 연구대상자에서 제외하였다.

○ 설문지에서 B형 간염 보균에 대한 문항에서 B형 간염 보균자가 아니라고 대답한 사람만 연구대상자로 포함하였고, 이 문항에서 '예' 또는 '모름'이라고 대답한 사람과 결측 값인 사람은 연구대상자에서 제외하였다.

○ 모든 통계분석 전 왼쪽으로 치우쳐져 있는 노출 자료의 특성을 고려하여 세 가지 노출 변수를 모두 로그 변환하였고, 복합노출의 건강영향 평가 시 각 노출 변수의 단위와 분산이 다름으로 인해 각 노출 변수의 기여도가 잘못 평가될 가능성을 고려하여 모든 노출 변수는 평균 0, 표본편차 1로 표준화되었다.

3. 연구윤리

본 조사를 위해서 2019년 산업안전보건연구원 생명윤리심의위원회의 심의(institutional review board, IRB)를 통과하였다(생명윤리심의번호: 2019-IRB-3).

Ⅲ. 연구결과

1. 유해물질 복합 노출을 평가한 통계적 방법 검토

1) 유해물질 복합노출 분석을 위한 통계 방법

○ 유해물질 복합노출 분석에 적합한 통계방법을 정리한 리뷰 논문 네 편 Lazarevic 등(2019), Sun 등(2013), Tayler 등(2016), 이슬비(2019)을 선정하였다.

○ 상기 논문을 종합하면 유해물질 복합노출 분석을 위한 통계 방법을 다음 네 가지로 분류할 수 있다.

가) 변수 선택만 가능한 통계방법 (variable selection, only)

나) 개별 효과를 평가할 수 있는 통계방법, 이 중 일부는 변수선택도 가능 (individual effect with/without variable selection)

다) 여러 노출 물질의 복합 효과를 평가할 수 있는 통계방법, 이 중 일부는 변수선택도 가능 (combined effect of various pollutants with/without variable selection)

라) 비선형성 및 상호작용 효과를 고려하여 복합 효과를 평가할 수 있는 통계방법 (combined effect considered non-linearity and interaction effect)

2) 통계분석 모델의 특성

○ 각 통계 분석 모델의 특성, 목적 및 장점과 단점을 <표 3-1>에 정리하였다.

○ 베이지안 모델 평균(Bayesian model averaging, BMA)는 변수

선택을 목적으로 수행되며 결과변수는 일반화 선형 모형(generalized linear model, GLM)로 표현이 가능하다. 상호작용 및 비선형성을 포함한다는 장점이 있지만, 노출 변수가 많은 수많은 모델을 고려할 때 계산이 비효율적이고 결과해석이 어려운 단점이 있다.

○ 나무기반 모델(tree based models)에는 분류 및 회귀나무(classification and regression tree, CART)와 무작위 숲(random forest, RF)이 있으며, 이를 통하여 변수를 선택할 수 있다. 이를 통하여 결과에 대하여 강력하게 예측할 수 있고 분류 및 회귀나무(CART)는 변수선택의 장점이 있으며 무작위숲(RF)은 중요도 점수를 제공한다. 하지만 해석에 어려움이 있으며, 무작위 숲은 변수 선택에 대한 임계값이 없어 설정에 어려움이 있다.

○ 최소 절대 수축 선택 연산(least absolute shrinkage selector operator, LASSO)와 탄성그물회귀(elastic net regression, ENR)는 축소의 방법을 통하여 변수를 선택하는 방법이다. 이를 통하여 회귀계수를 쉽게 해석할 수 있으며 잠재적 교란자를 통제할 수 있지만, 최소절대수축선택연산(LASSO)에서는 상관관계가 높은 노출의 화학물질 세트 중에서 하나만 선택할 수 있다는 단점이 있다.

○ 베이지안 계층적 회귀(Bayesian hierarchical regression, BHR), 특히 확률적 검색 변수를 사용한 방법은 축소를 통하여 변수를 선택할 수 있으며, 공동오염물질 및 잠재적 교란자를 통제할 수 있다. 비선형성도 고려가 가능하다. 하지만 계산이 까다로우며 분석 이전에 주제와 관련된 지식이 필요하다.

○ 가중 분위수 합 회귀(Weighted quantile sum regression, WQSR)는 각 노출 변수의 합 가중치를 추정하여 회귀모형에 사용하는 방법이다. 이 역시 변수선택 방법으로 이상치(outlier)의 영향을 줄일 수 있으며 다중공선성에 강하지만, 추정된 중량을 기준으로 중요한 오염물질을

식별하려면 특정 임계값 선택이 필요하다. 또한 상호작용 효과는 고려할 수 없다는 단점이 있다.

○ 주성분 분석(Principal component analysis, PCA)은 이전부터 사용된 전통적인 방법으로 여러 구성요소에 대한 노출 데이터를 줄이고 결과에 대해 회귀하는 방법이다. 이를 차원 축소라고 한다. 주성분 분석(PCA)에 의해 생성된 구성요소를 사용하여 회귀모델에서 혼란요인을 보정할 수 있지만, 이를 생물학적으로 해석하는 데 어려움이 있으며 노출변수가 연속형일 경우에만 사용이 가능하다.

○ 부분 최소 제곱 회귀 (Partial least squares regression, PLS)는 차원 축소 방법이며 연속형 변수와 범주형 변수에 모두 사용 가능하다. 부분최소제곱회귀(PLS)에 의해 생성된 구성요소를 사용하여 후속 회귀모델에서 혼란변수를 보정할 수 있는 가능성이 있다. 하지만 관련이 없는 변수의 포함으로 인한 영향이 있을 수 있다.

○ 구조 방정식 모델링(Structural equation modeling, SEM)도 전통적인 방법이다. 관측된 노출 변수가 잠재적인 실제노출변수로 보이는 경로모델링 접근방법으로, 하나 이상의 결과 변수에 영향을 미치는 것으로 가정한다. 이를 통하여 차원을 축소하며 여러 관련 결과 및 노출을 동시에 모델링 할 수 있는 장점이 있으나, 모델을 정확하게 만들기 위해 많은 사전 지식이 필요하고 실제로 검증하기가 어려울 수 있다.

○ 베이지안 커널 기계 회귀(Bayesian kernel machine regression, BKMR)는 다중 노출의 공동효과를 추정하고, 비선형성 및 상호작용을 허용하며, 구성 요소별 또는 계층적 변수 선택을 수행하는 방법이다. 관련성의 동인을 식별하면서 동시에 혼합물 건강 영향을 추정하고, 다중공선성에 견고 하며, 사후포함확률 추정은 각 노출의 중요도 지수로 사용가능한 장점이 있다. 하지만 현재는 매우 큰 자료에 대한 계산은 불가능한 실정이다.

<표 3-1> 통계분석모델에 따른 특성

1) 변수선택만 가능한 통계방법

방법	기술	목적	결과변수종류	장점	단점
베이지안 모델 평균 (Bayesian model averaging, BMA)	가능한 모든 모델의 부분집합과 최상의 부분집합에 대한 가중평균 추정치를 제공하는 베이지안 기법	변수선택	모든 일반화 선형 모형(general linear model, GLM), 링크 기능; 생존자료	- 모델 선택의 불확실성에 대한 고려 - 상호작용 및 비선형성 포함 지원	- 다중공선성에 강하지 않음 - 노출 변수가 많은 수많은 모델을 고려할 때 계산 비효율 - 결과 해석이 어려움 - 선형 또는 로지스틱 모델로 제한
나무기반모델 (Tree based models) -분류 및 회귀나무 (classification and regression tree, CART) - 무작위 숲 (random forest)	앙상블 방법 사용	변수선택	모두	- 비선형성과 상호작용을 고려할 수 있음 - 결과에 대한 강력한 예측 - CART : 변수선택가능 - 무작위 숲 : 가변 중요도 점수 제공	- 해석하기 어려움 - CART : 데이터 세트에 민감함 - 무작위 숲 : 변수 선택에 대한 임계값이 없음

2) 개별 효과를 평가할 수 있는 통계방법, 이 중 일부는 변수선택도 가능

방법	기술	목적	결과변수종류	장점	단점
일반화 선형 모델 (Generalized linear model, GLM)	한 개, 또는 여러개 단일 노출들의 일반화 선형모델	회귀	모든 일반화 선형 모형 링크 가능	- 회귀 계수의 쉬운 해석 - 잠재적인 교란자의 통제 - 상호작용 및 비선형성 식별	- 노출이 상관관계가 높은 경우 다준공선성이 문제 - 노출이 상관된 단일 오염 물질 모델의 높은 위양성률 - 변수 선택을 하지 않으므로 더 큰 노출 변수에는 적합하지 않음
- 최소 절대 수축 선택 연산 (Least absolute shrinkage selector operator, LASSO) - 릿지 (Ridge) - 탄성그물회귀(Elastic net regression, ENR)	축소(패널티 회귀)방법, LASSO 와 ENR은 변수 선택을 수행	축소 : 변수선택	모두	- 회귀 계수의 쉬운 해석 - 다중공선성에 견고함 - 최소 제곱 회귀보다 계수 변이가 낮음 - 잠재적인 교란자를 통제할 수 있음	- LASSO에서는 상관관계가 높은 노출의 화학물질 세트 중에서 하나만 선택할 수 있음 - 신뢰구간 생성을 포함하여 추론에 선택 수 통계 추론 도구가 필요함

2) 개별 효과를 평가할 수 있는 통계방법, 이 중 일부는 변수선택도 가능

방법	기술	목적	결과변수종류	장점	단점
- 베이지안 계층적 회귀(Bayesian hierarchical regression, BHR) - 확률적 검색 변수 선택한 BHR(BHR-SSVS)	모수가 될 수 있는 유연한 모델 종류	축소 : 변수선택	모두	- 공동 오염 물질 및 잠재적 교란자를 통제할 수 있음 - 수축 추정기로 인한 다중 공선성에 강함 - 고정 수축보다는 적응성; 분포를 기반으로 데이터가 수축량을 알려줌 - 비선형성과 상호작용을 고려할 수 있음	- 상세화에 접근가능하지 않은 주제와 관련한 지식이 필요할 수 있음 - 계산이 까다로움 - 베이지안 분석용 소프트웨어를 다루는 고급 기술이 필요함 (예. WinBUGS)

3) 여러 노출 물질의 복합 효과를 평가할 수 있는 통계방법, 이 중 일부는 변수선택도 가능

방법	기술	목적	결과변수종류	장점	단점
가중 분위수 합 회귀 (Weighted quantile sum regression, WQSR)	가중 노출 분위수 합계는 비선형 프로그래밍 및 부트 스트랩으로 추정 된 가중치 및 매개변수와 함께 회귀모델에 사용됨	변 수 선택; 회귀	모든 일반화 선형 모형 링크기능, 생존분석 자료	- 다중공선성에 강함 - 이상치(outlier)의 영향을 줄일수 있음 - 수축 방법보다 정확도, 비교감도, 특이도가 높을 수 있음	- 분위수를 사용함으로써 정보손실 - 추정 된 중량을 기준으로 중요한 오염물질을 식별하려면 특정 임계값 선택이 필요함 - 상호작용 효과는 고려할 수 없음

3) 여러 노출 물질의 복합 효과를 평가할 수 있는 통계방법, 이 중 일부는 변수선택도 가능

방법	기술	목적	결과변수종류	장점	단점
주성분 분석(Principal component analysis, PCA)	여러 구성 요소에 대한 노출 데이터를 줄이고 결과에 대해 회귀	차원 축소	모두	- 데이터를 상관되지 않은 여러 데이터로 줄여 다중공선성 제거 - PCA에 의해 생성된 구성요소를 사용하여 회귀모델에서 혼란요인을 보정할 수 있음	- 구성요소는 생물학적으로 해석하기 어려울 수 있음 - 노출변수가 연속형 일때만 사용가능
감독된 주성분 분석(Supervised principal component analysis, SPCA)	결과와 가장 관련있는 변수를 선택한 후 선택한 변수를 사용하여 구성요소로 노출 데이터를 줄임	차원축소 : 변수선택	연속형 변수, 생존분석자료	- 데이터를 상관되지 않은 여러 데이터로 줄여 다중공선성 제거 - 변수 선택을 수행하고 고정된 순서로 변수 중요도 점수를 제공함 - PCA에의해 생성된 구성요소를 사용하여 후속 회귀 모델에서 혼란요인을 보정할 수 있음	

3) 여러 노출 물질의 복합 효과를 평가할 수 있는 통계방법, 이 중 일부는 변수선택도 가능

방법	기술	목적	결과변수종류	장점	단점
- 부분 최소 제곱 회귀 (Partial least squares regression, PLS) - 부분 최소 제곱 회귀 판별 분석 (Partial least squares regression discriminant analysis, PLS-DA)	노출과 결과 사이의 공분산을 최대화하는 노출의 선형 조합을 형성하여 잠재변수를 구성하는 감독된 차원 축소 방법	차원축소	연속형 변수, 범주형 변수	- 직교 잠재 변수의 세트를 구성하여 다중공선성을 해결함 - PLS에 의해 생성된 구성요소를 사용하여 후속 회귀모델에서 혼란변수를 보정할 수 있는 가능성이 있음	- 관련이 없는 변수의 포함으로 인한 영향이 있을 수 있음
희박 부분 최소 제곱 회귀(Sparse partial least squares regression, SPLS, SPLS-DA)	PLS 방법론에 추가적으로 SPLS는 변수선택을 수행할 수있음	차원축소 : 변수선택	연속형 변수, 범주형 변수	- 변수선택을 수행하고 변수 중요도 점수 제공	- 상관관계가 높은 노출 변수 세트 중에서 하나의 노출을 임의로 선택할 수 있음

3) 여러 노출 물질의 복합 효과를 평가할 수 있는 통계방법, 이 중 일부는 변수선택도 가능

방법	기술	목적	결과변수종류	장점	단점
구조 방정식 모델링 (Structural equation modeling, SEM)	관측된 노출 변수가 잠재적인 실제노출변수의 징후로 보이는 경로모델링 접근방법으로 하나 이상의 결과 변수에 영향을 미치는 것으로 가정됨. 이 방법은 베이지안 설정에서 더 유연하게 사용될 수 있음	차원축소	연속형 변수, 범주형 변수	- 여러관련결과 및 노출을 동시에 모델링 할 수 있음 - 다중 관찰된 노출 변수를 나타내는 잠재 변수를 사용하여 다중 검정 문제가 줄어듬 - 각 결과-노출 쌍을 별도로 모델링 하는 것보다 더 큰 모델 특성을 가질수 있음	- 모델을 정확하게 만들기 위해 단일 결과-노출 모델보다 더 많은 지식이 필요함. - 실제로 검증하기 어려울 수 있으며 심각한 편견으로 이어질 수 있는 분석에 대한 여러 가지 강력한 가정을 전제로 함

4) 비선형성 및 상호작용 효과를 고려하여 복합 효과를 평가할 수 있는 통계방법

방법	기술	목적	결과변수종류	장점	단점
베이지안 커널 기계 회귀(Bayesian kernel machine regression, BKMR)	다중 노출의 공동효과를 추정하고 비선형성 및 상호작용을 허용하며 구성 요소별 또는 계층적 변수 선택을 수행함	스무딩 : 변수 선택; 수축	연속형 변수, 이분형 변수, 빈도변수	- 관련성의 동인을 식별하면서 동시에 혼합물 건강 영향을 추정 - 다준공선성에 견고 함 - 사후포함확률 추정 (각 노출의 중요도 지수로 사용가능) - 고차원 노출에서 비선형성 모델링 가능 - 노출 간 상호작용 효과를 모델링 할 수 있음	- 현재는 매우 큰 자료에 대한 계산이 불가능 - 사후포함확률(posterior inclusion probability)의 크기(상대순위는 아님)는 조정 매개변수 선택에 민감함 - 현재 중복 그룹을 허용하지 않음 - BKMR-HSV(계층적 변수선택)에서 그룹내에서 하나의 노출만 선택할 수 있음

2. 역학 연구 문헌고찰

앞서 고찰한 각 통계방법 중 새로운 방법을 중심으로 해당 분석방법을 적용한 논문을 검색하였다. 유해물질 복합 노출의 건강영향을 통계적으로 분석한 역학연구 총 22개를 고찰하였다. 이를 <표 3-2>로 정리하였다. 각 문헌의 세부 사항에 대해서는 [부록 1]에 정리하였다.

<표 3-2> 각 연구별 연구방법과 통계분석모델 및 특성

번호	저자	연구 방법	데이터 명	연구 대상자 수	노출변수	결과변수	층화 분석	변수 선택	혼합 작용	상호 작용	비선 형성	노출감 수성 기간	혼합물 분석 통계
1	Millar등 (2018)	시계열 연구	가나보건 서비스 이용자 명단	6～59개월 아동10,029명	말라리아 치료 유무, 환경, 인구, 사회경제적 요인	말라리아 유병률	기후	○	○	○	○	○	BMA
2	Menon 등 (2014)	코호트 내 환자 대조군 연구	Nashiville 출생코호트	여성 191명	출산 시 채취한 양수, 산모 혈청, 제태 혈청의 바이오마커	조산	평가하지않음	○	×	×	×	×	MARS
3	Fang 등 (2016)	시계열 연구	사망원인 등록자료	중국 영주권자 1,038만명, 호흡기계 사망자 9559명	대기질, 기상조건	일일 호흡기계 사망자 수	평가하지않음	○	○	○	○	×	GLMM, optimal GAMM, GAMM-BMA

번호	저자	연구 방법	데이터 명	연구 대상자 수	노출변수	결과변수	층화 분석	변수 선택	혼합 작용	상호 작용	비선 형성	노출감 수성 기간	혼합물 분석 통계
4	Forns 등 (2016)	전향적 출생 코호트	노르웨이 HUMIS[1)]	612쌍의 모자	모유내 24종의 pops물질의 농도.	영유아 증상 체크리스트 (ITSC)를 활용한 아동행동 문제 점수	산모교육수준	○	○	×	○	×	PCA, ENR, BMA, 선형회귀분석
5	Lenters 등 (2016)	전향적 출생 코호트	INUENDO[2)]	1250쌍의 모자	산모혈청내 DEHP의 대사산물, DiNP프탈레이트, 8PFAS, OCs 농도	출생 시 몸무게	아동성별, 임신 전 BMI, 혈청 중 코티닌 수치	○	○	×	○	×	ENR, 선형회귀분석, GAMs(일반화가법모델)
6	Swartz 등 (2015)	환자-대조군 연구	텍사스 출생 결손 등록자료	1,108명의 환자군, 4,132명의 대조군	주변의 유해한 공기오염물질 농도	이분척추	평가하지 않음	○	×	×	×	×	BHR, BHR-SSVS

번호	저자	연구 방법	데이터 명	연구 대상자 수	노출변수	결과변수	층화 분석	변수 선택	혼합 작용	상호 작용	비선 형성	노출감 수성 기간	혼합물 분석 통계
7	Braun 등 (2014)	전향적 출생코호트	오하이오주 신시내티의 HOME[3] 연구	175쌍의 모자	산모 혈액 및 소변의 52개의 EDC 물질	4-5세 아동의 사회반응척도 평가	아동의 성별	○	×	×	×	×	2단계 semi-베이지안 계층적 선형 혼합 모델
8	Powell 등 (2015)	후향적 코호트	MCAPS[4]	65세 이상 medicare 등록자20,000명 초과	PM10-2.5, PM2.5	심혈관질환과 호흡기질환으로 응급병원 입원 일 수	평가하지 않음	○	×	×	×	×	2단계 베이지안 계층적 모델과 포아송 로그-선형 회귀 모델
9	Horton 등 (2018)	전향적 출생 코호트	ELEMENT[5] 연구	133쌍의 모자	임신전후 치아 생체 표지자로 망간, 아연, 납의 농도	8-11세 아동의 행동 평가 시스템 결과	평가하지 않음	○	○	×	×	○	lagged WQSR

번호	저자	연구 방법	데이터 명	연구 대상자 수	노출변수	결과변수	층화 분석	변수 선택	혼합 작용	상호 작용	비선 형성	노출감 수성 기간	혼합물 분석 통계
10	Romano 등 (2018)	전향적 출생 코호트	오하이오 주 신시내티 의 HOME 자료	276쌍의 모자	9종의 프탈레이트 모노에스테르대사산물	갑상선 자극호르몬, 총 및 유리 티록신, 트리이오도사이로닌	평가하지 않음	○	○	×	×	○	WQSR
11	Lenters 등 (2015)	단면 연구	없음	그린란드 남성 199명, 폴란드 남성 197명, 우크라이나 남성208명	프탈레이트 대사산물과 과불화알킬산 (PFAAs)	22종의 생식 표지자	평가하지 않음	○	○	×	×	×	sPLS

번호	저자	연구 방법	데이터 명	연구 대상자 수	노출변수	결과변수	층화 분석	변수 선택	혼합 작용	상호 작용	비선 형성	노출감 수성 기간	혼합물 분석 통계
12	Porru등 (2014)	병원 기반 환자 대조군 연구	없음	199명 방광암 환자, 213명 대조군	인구통계학적 요인 및 직업 등 방향족 아민(AA)과 다환 방향성 탄화수소(PAHs)에 노출 가능성 있는 근거	말초 혈액 림프구 DNA 부가물, 방광암 위험	평가하지 않음	○	×	×	×	×	SEM
13	Arlinghaus 등 (2012)	단면 연구	없음	86,137명의 남녀	고용상태, 주당 근무시간, 수면시간, 인구학적 요인, 건강정보, 업무관련 변수	산업재해: 최근 3개월간 독극물 또는 상해관련 사건의 경험	평가하지않음	○	○	×	×	×	SEM

번호	저자	연구 방법	데이터 명	연구 대상자 수	노출변수	결과변수	층화 분석	변수 선택	혼합 작용	상호 작용	비선 형성	노출감 수성 기간	혼합물 분석 통계
14	Valeri등 (2017)	전향적 출생코호트	없음	825 쌍의 모자	재태혈관의 비소, 망간, 납	BSID-Ⅲ를 이용한 신경발달평가	클리닉	○	○	○	○	×	BKMR, GAMs
15	Lee(2019)	전향적 출생코호트	MOCEH연구[6]	302쌍의 모자	수은, 납, 카드뮴, 비스페놀A, 프탈레이트 대사체 3종, 산화질소, 대기중 입자상물질	6개월간 아토피 피부염	성별	○	○	○	○	○	ENR, PLS-DA, BKMR
16	Luo등 (2019)	전향적 출생코호트	나바우 출생코호트	500명의 산모	혈액 및 소변의 우라늄, 총 비소, 아비산, 디메틸아소닉산, 아연	8-이소-PGF2α와 프로스타글란딘 F2α 비	평가하지 않음	○	○	○	×	×	aLasso(adaptive lasso), CART, RF

번호	저자	연구 방법	데이터 명	연구 대상자 수	노출변수	결과변수	층화 분석	변수 선택	혼합 작용	상호 작용	비선 형성	노출감 수성 기간	혼합물 분석 통계
17	Casas등 (2016)	전향적 출생 코호트	INMA[7]	488쌍의 모자	소변 중 BPA, 8종의 프탈레이트 대사물질	태아 성장	아동 성별	○	○	×	○	×	GLM, 선형 혼합 모형 GAMs
18	Gascon 등 (2015)	전향적 출생 코호트	INMA	657쌍의 모자	소변 중 BPA, 프탈레이트 대사물질, 4종의 DEHP 대사물질	아동의 폐감염증, 기관지염, 천명, 습진, 아토피, 천식	아동 나이, 성별	○	○	×	○	○	GLM, GEE(일반화 추정 방정식), GAMs, 로지스틱 회귀분석
19	Maresca 등 (2016)	전향적 출생 코호트	CCCEH[8]	424쌍의 모자	산모와 아동 소변의 9종의 프탈레이트 대사물질	BMI, 체지방율, 허리둘레, 체지방지수	아동 성별	×	○	×	×	○	PCA, 선형 모델, GLM

번호	저자	연구 방법	데이터 명	연구 대상자 수	노출변수	결과변수	층화 분석	변수 선택	혼합 작용	상호 작용	비선 형성	노출감 수성 기간	혼합물 분석 통계
20	Talbott 등 (2015)	환자대조군연구	펜실베니아 인구기반 연구	217명의 환자군, 224명의 인터뷰 통한 대조군, 4971명의 출생 증명서를 통한 대조군	30종의 대기오염물질	자폐스펙트럼 장애위험도	평가하지 않음	○	○	×	×	×	GLM
21	Kim 등 (2010)	단면연구	없음	역학조사보고서 153건	진단 시 나이, 흡연력, 잠재기, 일한기간, 성별, 조직학적 분류, 산업분류, 노출물질	직업성 폐암 인정 유무	평가하지 않음	○	×	×	×	×	CART, 로지스틱 회귀분석

번호	저자	연구 방법	데이터 명	연구 대상자 수	노출변수	결과변수	층화 분석	변수 선택	혼합 작용	상호 작용	비선 형성	노출감 수성 기간	혼합물 분석 통계
22	Bucklye 등 (2016)	전향적 출생 코호트	뉴욕 Mt. Sinai 아동의 환경 건강연구	180쌍의 모자	산모 소변 중 9종의 프탈레이트 대사산물	아동 체지방량	아동 성별	×	×	×	×	×	semi-BHR, GAMs

1) HUMIS : Human Milk Study

2) INUENDO : Greenland, Poland, Ukraine cohort

3) HOME : The Health Outcomes and Measures of the Environment

4) MCAPS : Medicare Cohort Air Pollution Study

5) ELEMENT : Eearly Life Exposures In Mexico to Environmental Toxicants

6) MOCEH : Mothers and Children's Environmental Health

7) INMA : INfancia y Medio Ambiente; Childhood and Environment

8) CCCEH : Columbia Center for Children's Environmental Health

* BMA : Bayesian model averaging/ MARS : Multivariate adaptive regression spline/ GLMM : Generalized linear regression/ GAMM :generalized additive mixed model/ ENR : elastic net regression/ BHR : Bayesian hierarchical logistic regression/ BHR-SSVS : BHR-Stochastic Search Variable Selection/ WQSR : Weighted quantile sum regression/ SEM : Structural equation modeling/ BKMR : Bayesian kernel machine regression/ GAMs : Generalized additive models/ sPLS : Sparse partial least squares/ PLS-DA : PLS-discriminant analysis/ CART : Classification and regression trees/ RF : Radon forest/ GLM : Generalized linear model/ GEE : Generalized estimating equations

1) Millar 등(2018)

(1) 제목: 베이지안 모형 평균(BMA)을 이용한 가나(Ghana) 북부의 잔류 말라리아에 대한 국소 위험 요소 탐지.

(2) 요약: 시계열 연구로 6-59개월 소아 10,029명을 대상으로 하였다. 말라리아 치료, 환경/도시화, 인구/사회 경제적 변수에 따른 말라리아 유병률을 베이지안 모형 평균(Bayesian model averaging, BMA)을 사용하여 분석하였다. 계절적 차이와 비선형 연관성을 조사하였으며, 말라리아 유병률에 대한 인구통계학적, 사회경제적, 환경적, 말라리아 개입변수의 상대적 영향은 베이지안 모형 평균(BMA)을 사용하여 변수를 선택하였다. 말라리아 위험과 가장 가까운 도심 및 보건 시설과의 거리 사이에는 명백한 비선형적 연관성을 보였으며 말라리아 위험과 나이 및 일부 종족 간의 연관성은 우기와 건기에 유의미한 차이가 있었다. 또한 베이지안 모형 평균(BMA)은 계절에 따른 유효성 검사 접근법을 사용한 표본외 예측 능력에서 다른 회귀 접근법보다 뛰어났다.

2) Menon 등(2014)

(1) 제목: 조산의 바이오 마커를 예측하기 위한 다변량 적응 회귀 스플라인(MARS) 분석

(2) 요약: 코호트 내 환자 대조군 연구로 Nashville 출생 코호트 자료를 분석하였다. 출산시 양수, 산모 혈청, 제태 혈청을 채취하여 바이오마커를 얻었으며 결과변수는 조산으로 다변량 적응 회귀 스플라인(Multivariate adaptive regression spline, MARS)을 사용하여 분석하였다. 다변량 적응 회귀 스펄라인은 전체 및 인종 계층화 생물학적 지표 자료에 대한 모델을 생성하였고, 임상 및 인구통계학적 자료는 모델에 기여하지 않았다. 여러 생물학적 변수에 대하여 인종 전체, 흑인, 백인의 인종적 차이가 있었다.

3) Fang 등(2016)

(1) 제목: 대기 오염과 호흡기 사망률 간의 연관성을 평가하기 위한 베이지안 모형 평균(BMA) 방법 : 시계열 연구

(2) 요약: 사망원인등록자료를 이용한 시계열 연구로 기간동안 호흡기계 사망자는 9,559명이었다. 미세먼지(PM_{10}), 산화질소, 일산화탄소, 일평균 기온, 상대습도, 풍속과 기압을 노출변수로, 결과변수는 8개 지역의 일일 호흡기계 사망자 수였다. 분석방법은 일반화 선형 회귀(generalized linear regression, GLMM)와 일반화 부가 혼합 모형(generalized additive mixed model, GAMM), GAMM-BMA 모형을 비교하였다. GAMM-BMA 모형으로 인한 미세먼지로 인한 호흡기계 사망률은 다른 모형을 사용한 경우보다 약간 상승하였다. 연구결과 베이지안 모형 평균(BMA)는 시계열 연구에서 대기오염의 건강영향을 평가할 때 유용한 도구임이 확인되었다.

4) Forns 등(2016)

(1) 제목: 다중-독성물질 분석을 위한 통계적 방법을 새로 적용하여 DDT를 조기 아동 행동 문제의 위험 요소로 확인함

(2) 요약: 노르웨이에서 수행된 전향적 출생 코호트로 612쌍의 모자로 구성되어 있다. 모유 내 24종의 지속적 유기오염물질(persistent organic pollutents, POPs)을 측정했으며, 결과변수는 영유아 증상 체크리스트(ITSC)를 활용한 12개월과 24개월의 아동행동문제를 연속 변수로 분석하였다. 상관관계가 있는 변수를 적은 수의 비상관 주성분으로 변형하는 방법인 주성분분석(principle component analysis, PCA)을 시행하였고, 두 가지 변수 선택 방법인 탄성 그물 회귀(elastic net regression, ENR)와 베이지안 모형 평균(BMA)으로 분석 후 선택된 노출에 대해 선형회귀분석을 수행하였다. 전통적인 방법과 현대적인 방법을 함께 사용하였고, 아동의 행동 문제에 대한 위험요인으로 DDT가 밝혀졌다.

5) Lenters 등(2015)

(1) 제목: 출생 전 프탈레이트, 퍼플루오로 알킬산 및 유기염소 노출과 3 가지 출생 코호트에서의 출생 체중: 탄성 그물 회귀(ENR)에 기반한 다중 오염 모형

(2) 요약: 그린랜드, 우크라이나, 폴란드의 출생 코호트를 이용하여 1,250쌍의 모자가 모집되었다. 노출변수는 산모혈청내 DEHP의 대사산물, DiNP 프탈레이트, 8 PFAS, OCs의 농도는 연속형 변수로분석하였고, DEHP와 DiNP는 분자 합으로 분석하였으며 결과변수는 출생시 몸무게였다. 성별, 임신 전 산모 체질량 지수, 혈청 코티닌으로 보정하였으며 변수 선택에는 탄성 그물 회귀(ENR)를 하였으며 효과 측정에는 선형회귀분석을 수행하였다. 탄성그물회귀(ENR)은 가장 영향력이 적은 변수의 계수를 0으로 만들었고, 전통적인 단계적 회귀는 노출 상호간의 보정이 없어서 이론적으로 선택 정확도가 낮아졌다.

6) Swartz 등(2015)

(1) 제목: 여러 오염 물질과 이분척추증(spina bifida)을 평가하기 위한 베이지안 계층적 접근법(BHA)

(2) 요약: 미국에서 수행된 환자-대조군 연구로, 1,108건의 환자군과 4,132건의 대조군을 모집하였다. 노출은 주변 유해 공기오염물질 농도에 대한 인구통계학적 추정으로 평가하였으며, 결과변수는 이분척추였다. 확률적 탐색 변수 선택(stochastic search variable selection, SSVS)을 포함하는 베이지안 계층적 로지스틱 회귀분석(Bayesian hierarchical logistic regression, BHLR)을 사용하였다. 결과 퀴놀린과 트리클로로에틸렌이 이분 척추증의 위험에 중요한 위험요인이 있음을 발견하였다.

7) Braun 등(2014)

(1) 제목: 내분비계 교란 물질(EDC)의 임신 노출과 4세 및 5세 어린이의 상호사회적, 반복적 및 고정관념 행동: HOME 연구

(2) 요약: 미국 오하이오주 신티내티의 전향적 출생 코호트 내의 175쌍의 모자가 모집되었다. 산모 혈액 및 소변의 52개의 내분비계교란물질(endocrine disrupting chemicals, EDC)을 이분변수나 연속형 변수로 분석하였고, 결과변수는 4-5세 아동의 사회반응척도(social responsiveness scale, SRS)평가를 연속형 변수로 분석하였다. 아동의 성별을 매개효과로 분류하였으며, 반복 측정된 사회반응척도를 설명하기 위해 2 단계 반-베이지안 계층적 선형 혼합 모형을 사용하였다. 일부 내분비계 교란물질은 자폐행동과 관련있었다. PFOA, β-헥사클로로사이클로헥산, PCB-178, PBDE-28, PBDE-85 및 트랜스-노나클로는 어린 시절 자폐증 행동과 관련 될 수 있는 요인으로 추가 조사가 필요하다.

8) Powell 등(2015)

(1) 제목: 메디케어 코호트 대기오염 연구(1999-2010)를 이용한 거친 입자상 물질과 병원 입원.

(2) 요약: 미국의 65세 이상 메디케어 등록자를 대상으로 한 후향적 코호트 연구로 $PM_{10-2.5}$, $PM_{2.5}$에 따른 심혈관질환과 호흡기질환으로 인한 응급병원 입원 일 수를 분석하였다. 지역(미국 동부와 서부)과 $PM_{10-2.5}$와 카운티 수준의 입원율 사이의 국가적 연관성을 알아보기 위해 2 단계 베이지안 계층 모델(two-stage Bayesian hierarchical model)을 사용했다. $PM_{10-2.5}$와 입원 간의 카운티 차원의 연관성을 추정하기 위해과 분산 된 포와송 로그 선형 모델(Poisson log-linear model)을 사용했다. $PM_{10-2.5}$의 일중 변동은 65세 이상 Medicare 가입자의 심혈관

질환 응급 입원과 유의한 연관성이 있었다. 이 연관성은 $PM_{2.5}$ 농도 보정에도 견고했다. $PM_{10-2.5}$의 10-μg/m^3 증가는 호흡 관련 입원의 유의한 증가와 관련이 없었다.

9) Horton 등(2018)

(1) 제목: 망간, 아연, 납의 태아 및 유아기 노출의 상아질 바이오마커와 아동기 행동

(2) 요약: 멕시코에서 133쌍의 모자로 구성된 전향적 출생 코호트 자료이다. 임신 전후 치아 상아질 바이오마커로 망간, 아연, 납의 농도를 측정했으며 아동은 8-11세까지 추적관찰하여 아동행동평가체계(behavioral assessment system for children, 2nd edition, BASC-2)를 측정했다. 지연된 가중치 분위수 합 회귀(lagged weighted quantile sum regression, WQSR)를 사용하였다. 지연된 가중치 분위수 합 회귀를 통하여 각각의 개별 금속뿐만 아니라 금속을 측정 할 때 시간 창(window)에 따라 달라지는 금속의 조합에 대해서도 통계적으로 유의미한 관계가 있음을 나타냈다. 태아기(prenatal)의 치아 망간은 보호적일 수 있으며, 출생 후 초기의 과도한 망간은 부정적 행동에 대한 위험을 증가시킬 수 있었다. 망간, 아연, 납의 농도가 높을수록 행동에 나쁜 영향을 줄 수 있다.

10) Romano 등(2018)

(1) 제목: 임신 중 모성 소변 프탈레이트 대사체와 모성 및 제대 혈청의 갑상선 호르몬 농도: HOME 연구

(2) 요약: 오하이오 신시내티의 전향적 출생 코호트 자료를 이용하여 276쌍의 모자를 대상으로 9종의 프탈레이트 모노에스테르 대사산물을 측정하였다. 갑상선 관련 결과변수는 갑상선 자극 호르몬(thyroid stimulating hormone, TSH), 총 및 유리 티록신(total thyroxine, TT4

and free thyroxine, FT4), 트리이오도사이로닌(triiodothyronine TT3, FT3)을 측정하였다. 본 연구에서 가중치 분위수 합 회귀(WQS)를 사용하여 단일 프탈레이트 지수를 만들었는데 각 대사산물의 가중치를 추정한 값이다. 이 프탈레이트 지수를 보정후 다변수 선형 회귀모형(multivariate linear regression)으로 총체적인 영향을 추정하였다. 결과 여러 프탈레이트에 대한 공동 노출이 특정 갑상선 호르몬(임산부와 신생아의 TT4, 신생아의 TSH)과 반비례하였다.

11) Lenters 등(2015)

(1) 제목: 프탈레이트, 과불화알킬산, 금속 및 유기염소와 생식 기능: 그린란드, 폴란드, 우크라이나 남성의 다중오염 평가

(2) 요약: 본 연구는 단면연구로 그린란드 남성 199명, 폴란드 197명, 우크라이나 208명을 대상으로 프탈레이트 대사산물과 과불화알킬산을 측정하였다. 결과변수로는 22종의 생식표지자를 사용하였다. 노출 프로파일과 각 결과간의 연관성을 평가하기 위해 희소부분최소제곱(sparse partial least squares, sPLS) 회귀 모델을 사용하였다. 희소부분최소제곱(sPLS) 회귀 모델에서 테스트 된 300 개 이상의 노출-결과 연관 관계 중 8 개 결과를 포함하는 10 개의 연관 관계를 발견하였다. 다음의 몇몇 연관 관계는 세 연구 집단에서 방향성이 두드러지게 일관되었다: 수은과 inhibin B, 그리고 카드뮴과 테스토스테론 사이의 양방향 연관성; DiNP 대사 산물과 테스토스테론, 폴리염화바이페닐-153과 진행성 정자 운동성, 그리고 DEHP 대사 산물과 부고환 기능의 지표 인 중성 α-글루코시다제 사이의 역상관 관계. 이러한 환경오염물질의 혼합물에 대한 종합 평가는 일부 유기 염소 및 프탈레이트가 남성 생식 건강의 일부 지표에 유해한 영향을 미친다는 것을 나타낸다.

12) Porru 등(2014)

(1) 제목: 구조 방정식 모델링을 사용한 직업, 환경, DNA 부가물, 유전자 다형성 및 방광암 간의 관계

(2) 요약: 환자 대조군 연구로 199명의 방광암 환자와 213명 대조군을 대상으로 하였다. 노출변수는 인구통계학적 요인 및 직업 등 방향족 아민(AA)과 다환 방향성 탄화수소(PAHs)에 노출 가능성이고 결과변수는 말초 혈액 림프구 DNA 부가물과 방광암 위험이었다. 구조방정식 모델링을 사용하였으며, 결과 처음으로 DNA 부가물 및 방광암 위험과 방향족 아민에 대한 직업 누적 노출 간의 관련성에 대한 근거를 제공하였고, 방광암 발생에 있어서 방향족 아민이 중심적인 역할을 한 것으로 판명되었다.

13) Arlinghaus 등(2012)

(1) 제목: 산업 재해와 피로 관련 위험 요소에 대한 구조 방정식 모델링 접근법

(2) 요약: 86,137명의 남녀를 대상으로 한 단면연구로 고용상태, 주당 근무시간, 수면시간, 인구학적요인, 건강정보, 업무관련 변수와 산업재해의 연관성을 분석하였다. 긴 근무 시간과 짧은 수면 시간은 독립적으로 부상의 위험을 증가 시켰으며 추가적인 위험 요소로는 성별, 직업, 급여 유형 및 체질량지수가 있었다. 동시에, 긴 노동 시간, 높은 심리적 고통 및 높은 체질량지수는 수면 지속 기간을 경유하여 부상의 위험을 증가시킴을 발견하였다.

14) Valeri 등(2017)

(1) 제목: 20-40개월의 신경 발달에 대한 금속 혼합물의 태아 노출의 공동 효과: 농촌 방글라데시 결과

(2) 요약: 전형적 코호트 연구로 아동의 신경 발달에 대한 비소(arsenic), 망간(manganese) 및 납(lead)에 대한 자궁 노출의 공동 효과

를 평가하고자 하였다. 참가자는 방글라데시의 두 개의 농촌 보건 진료소에서 모집된 825 쌍의 모자이고, 노출변수는 재태혈관의 비소, 망간, 납이며 결과변수는 신경발달평가를 수행하였다. 베이지안 커널 기계 회귀(BKMR)를 사용한 방법으로 상호작용과 비선형효과를 고려하였다. 결과, 비소와 망간 사이의 잠재적인 상승 작용뿐만 아니라 혼합물의 신경 독성의 증거를 발견했다.

15) Lee (2019)

(1) 제목: 임신 중 복합 환경유해물질 노출이 6개월 영유아 아토피 피부염 발생에 미치는 영향

(2) 요약: 산모·영유아의 환경유해인자 노출 및 건강영향 코호트 연구에 참여한 산모아 출생아를 대상으로 6개월 시기 아토피 피부염 발생에 중요한 영향을 미치는 환경오염물질을 확인하였다. 노출평가로 환경오염물질은 중금속(납, 수은, 카드뮴), 비스페놀 A, 프탈레이트 대사체 3종(ME-HHP, MEOHP, MnBP), 대기오염물질(이산화질소, 미세먼지, 초미세먼지)의 10가지이다. 성별로 보정하였으며, 먼저 단일노출시에 유의한 요인을 조사하였다. 이 10가지 물질에 대하여 탄성 그물 회귀(elastic net regression, ENR) 분석을 하여 임신 초기에는 납, 수은, MnBP, 미세먼지 노출이 선택되었으며, 임신 말기에는 납, 수은, MEHHP, MEOHP, 미세먼지가 선택되었다. 부분최소제곱-판별 분석(partial least square-discriminant analysis, PLS-DA)를 통해 모든 환경유해물질의 누적 노출 점수를 산출하였고 임신 말기 누적 노출 점수가 증가할수록 아토피 피부염 발생이 유의하게 증가하였다. 베이지안 커널 기계회귀(Bayesian kernal machine regression)을 통해 비선형연관성 및 교호작용이 존재함을 확인하였으며 이를 반영하여도 임신 말기의 누적 노출과 아토피 피부염 발생 사이에 유의한 연관성이 나타났다.

16) Luo 등(2019)

(1) 제목: 환경 화학 혼합물의 건강 영향을 평가하기 위한 2단계 접근 방식: 시뮬레이션 된 데이터 세트 및 실제 데이터에 적용

(2) 요약: 나바호(Navajo)인디언을 대상으로 한 출생 코호트 연구로 14-45세 산모 500명을 모집했다. 등록 시 혈액, 혈청, 소변 이렇게 세 가지 샘플에서 측정된 금속 노출로 우라늄, 총 비소, 비소(arsenous III) 산(arsenite, AsIII), 디메틸아르시닉산(dimethylarsinic acid, DMA), 아연을 측정하였다. 결과변수는 산화 스트레스의 바이오마커이며, 8-iso-PGF2α와 prostaglandinF2α 비(ratio)를 사용하였다. 연령, 흡연, 고용, 체질량지수를 보정했으며 (1) 단일 단계 적응형(adaptive) 올가미(lasso); (2) 분류 및 회귀 나무(classification and regression tree, CART)를 이용하고 이후 적응형 lasso 접근법의 2 단계; (3) 무작위 숲(random forest)을 이용하고 이후 적응형 lasso 접근법의 2단계를 분석하였다. 2 단계 접근 방식은 변수 선택 및 적응형 수축 전략의 장점을 결합하여, 고-처리량 화학 혼합물 노출이 건강 결과에 미치는 영향을 분석하는 강력한 방법을 제공하였다.

17) Casas 등(2016)

(1) 제목: 임신 중 비스페놀 A와 프탈레이트 노출과 태아 성장의 초음파적 측정

(2) 요약: 스페인 출생 코호트의 모-자녀 488 쌍에 대하여 비스페놀 A 와 프탈레이트에 대한 태아기 노출이 태아성장과 관련이 있는지 평가하였다. 태아 초음파 검사는 12, 20, 34 주에 시행하였고, 대퇴골 길이(FL, femur length), 머리 둘레(HC, head circumference), 복부 둘레(AC, abdominal circumference), 양쪽마루뼈지름(BPD, biparietal diameter)를 매개변수로 측정하였다. 종적 성장 곡선은 선형 혼합 모형(linear mixed model)을 사용했고 다중오염 노출의 점수를 매겼다.

비스페놀 A 또는 프탈레이트 노출과 태아성장의 연관성에 대한 증거는 거의 없으나, 프탈레이트 대사물인 MBzP 와 MnBP 는 일부 태아성장 매개변수와 관련이 있었다.

18) Gascon 등(2015)

(1) 제목: 비스페놀 A 및 프탈레이트와 소아기 호흡기계 감염 및 알레르기

(2) 요약: 스페인에서 수행된 657 모자 쌍으로 구성된 전향적 코호트 연구로 종적 분석으로 6 개월부터 7 세까지의 호흡기계 및 면역학적 결과는 일반화 선형 모형(generalized linear models)을 사용하였으며, 7 살의 천식과 4 세의 아토피에 대해서는 로지스틱 회귀분석을 수행하였다. 산전 비스페놀 A 와 고분자 프탈레이트는 소아의 천식과 호흡기계 감염의 위험을 상승시켰다.

19) Maresca 등(2016)

(1) 제목: 산전 프탈레이트 노출과 소아기 신체 크기 부석

(2) 요약: 뉴욕에서 수행된 코호트 연구로 424명의 모자 쌍을 분석하였다. 모성 소변에서 9종의 프탈레이트 대사체 농도와 3세와 5세의 소아 소변을 측정하였다. 요소 점수를 얻기 위해서는 주요성분분석(princple component analysis, PCA)를 수행하였고, 요소 점수와 소아의 체질량지수 분석에는 일반화 선형 모형을 사용하였다. 주요성분분석(PCA)을 통해 DEHP와 non-DEHP를 확인하였고, 가설과는 반대로, 산전 non-DEHP 프탈레이트 노출은 유아기의 남아에서 낮은 체질량지수, 허리 둘레, 체지방지수와 관련이 있었음.

20) Talbott 등(2015)

(1) 제목: 공기 독성물질과 자폐 스펙트럼 장애: 펜실베니아 남서부에서 수행한 환자-대조군 연구

(2) 요약: 환자 217명이고, 대조군은 두 형태로 한 군은 224명의 인터뷰한 대조군이고 다른 한 군은 5,007건의 출생 등록 대조군이었다. 30종 공기 독성물질 노출은 출생 거주지의 센서스 단계로 평가하였다. 먼저 주요성분 분석을 수행하였고, 이를 통한 7개의 오염물질은 단계적 로지스틱 회귀분석을 수행하였다. 임신기간 동안 스티렌과 크로뮴의 농도가 높은 지역에서 사는 것은 자폐스펙트럼장애 위험 증가와 관련이 있고, 방향족 탄화수소(PAHs)와 염화메틸렌도 위험 증가의 경계선에 있었다.

21) Kim 등(2010)

(1) 제목: 분류 및 회귀 분석을 사용한 직업성 폐암의 결정 나무

(2) 요약: 153건의 환자를 대상으로 대상 변수는 사례가 직업성 폐암으로 승인되었는지 여부였으며, 독립 변수는 연령, 성별, 흡연 기간, 조직학적 유형, 산업 유형, 대기 시간, 작업 기간 및 작업장의 노출 물질이었다. 분류 및 회귀 나무(classification and regression tree, CART)모형은 직업성 폐암의 예측 변수를 검색하는 데 사용되었다. 분류 및 회귀나무(CART) 모형에서 가장 좋은 예측 변수는 알려진 폐 발암 물질에 대한 노출이었고, 두 번째로 가장 좋은 예측 변수는 8.6년 이상의 잠재기이며, 세 번째로 가장 좋은 예측 변수는 11.25갑년 미만의 흡연력이었다.

22) Bucklye 등(2016)

(1) 제목: 뉴욕시 코호트에서 태아기 프탈레이트 노출 및 아동기 지방량

(2) 요약: 시나이산 소아 환경 보건 연구(The Mount Sinai Children's Environmental Health Study)는 1998년부터 2002년 사이에 뉴욕에 있는 시나이산 진단 및 치료 센터(Mount Sinai Diagnostic and Treatment Center)와 인접한 2개의 개인 의원으로부터 단일아를 임신한 479명의 초산부를 등록하였다. 산모들은 25-40주(평균 31.5주) 사이의 소변 샘플을 제공하였고, 다음 물질을 조사하였다. Monoethyl phthalate(MEP), mono-n-butyl phthalate(MnBP), monoiso-butyl phthalate(MiBP), mono phthalate(MCPP), monobenzyl phthalate(MBzP), DEHP의 대사산물 4개(MEHP, MEHHP, MEOHP, MECPP). 소아에는 생체 전기 임피던스 분석을 통하여 체중과 체성분을 평가하였다. 변수 선택에는 반-베이즈 계층적 회귀(semi-Bayesian hierarchical regression, semi-BHR)를 사용하였으며, 선형 혼합 효과 회귀(linear mixed effect regression) 모형을 사용하여 지방 질량의 연관성을 추정했다. 모든 프탈레이트 대사산물은 3기 산모 소변 샘플의 90%이상에서 검출되었으며 4분위수 모델에서 3기 산모의 요중 DEHP 대사산물의 합의 농도가 가장 높은 삼분위수에 해당하는 아동은 가장 낮은 삼분위수에 해당하는 아동과 비교하여 체지방률이 더 낮았다.

3. 국내 산업보건 이차자료 특성 검토

○ <표 3-3>과 같이 국내 산업보건연구에 활용 가능한 이차자료의 특성을 검토해보았다. 특수건강진단 자료, 건강보험공단 자료, 통계청 사망자료 중 복합노출 통계방법 적용을 위해 필요한 생물학적 노출 지표를 가진 자료는 특수건강진단 자료만 해당 되었고, 모든 자료는 종적자료의 특성을 가졌다. 각 자료는 유해인자 179종 유해인자에 노출되는 노동자의 건강, 직장가입자의 건강, 전 국민 사망을 각각 대표하는 자료로 대용량 자료의 특성을 가지고 있었다.

<표 3-3> 국내 산업보건 이차자료 특성

	특수건강진단 자료	건강보험공단 자료	통계청 사망자료
노출 변수	혈중/요중 생물학적 지표	업종 분류	노출 변수는 없으나 자료 연계 가능
결과 변수	생물학적 지표, 검진 판정 결과, 질병 유병	국가건강검진 결과, 질병 발생, 질병 유병	사망일시, 사망원인 등
대상자 수	2017년 약 200만명 수검	2018년 직장가입자 약 17500만명	2018년 29만 8,900명 사망
자료 형태	종적자료	종적자료	종적자료
대표성	유해인자 179종 노출 노동자 특수건강검진 결과 원 자료	직장가입자의 국가건강검진결과 및 병원 진료기록 원 자료	전 국민 사망

4. 복합노출 건강영향 평가의 목표 설정

○ 복합노출로 인한 건강영향을 평가할 수 있는 통계방법의 적용사례를 기반으로 활용 가능한 범위를 검토하였고, 국내 노동자의 자료의 특성을 고려하여 복합노출로 인한 건강영향을 평가하는 통계분석의 목적을 연구진 회의를 통해 다음과 같이 설정하였다.

○ 첫째, 여러 노출물질의 복합적인 건강영향을 평가할 수 있어야 하고, 둘째로 종적자료에 적용 가능한 방법이어야 한다. 셋째, 대용량 자료를 분석할 수 있어야 한다. 넷째, 복합적인 영향 중 각 노출 물질이 건강영향에 기여하는 정도를 확인할 수 있어야 한다. 다섯째, 분석 결과를 생물학적으로 해석이 가능해야 한다.

5. 유해물질 복합노출 연구의 목표에 맞는 복합노출 통계방법 검토

○ 복합노출의 영향을 평가 가능한 통계방법만 <표 3-4>에 포함하였으며, 그 외 연구 목표를 만족할 수 있는 방법인지를 평가하였다. 주성분분석의 경우에는 각 노출 물질이 건강영향에 기여하는 정도를 확인할 수 없었고, 노출 변수들의 공분산을 잘 설명해주는 주성분을 계산하는 방식으로 결과 변수와의 연관성을 고려하지 않은 분석으로 생물학적 해석이 가능하지 않다는 단점이 있었다. 베이지안커널머신 회귀분석의 경우에는 분석 시간이 오래 걸려 대용량 자료 분석은 가능하지 않다는 단점이 있었다.

○ 하지만, 본 연구의 목표 외에도 각 통계방법 마다 여러 가지 추가적인 장단점이 있어 내부 연구진 회의를 통해 <표 3-5>와 같이 추가적으로 통계방법의 장단점을 평가하였고, 그 결과 가중분위수합 회귀, 주성분분석, 부분최소제곱 회귀, 구조방정식 모델링, 베이지안커널머신 회귀가 연구 목적에 맞게 활용할 가치가 있다고 평가되었다.

○ 이 중 본 연구에서는 주성분분석(principle component anlaysis), 부분 최소 제곱 (partial least square, PLS), 베이지안 커널 기계 회귀 (Bayesian kernal machine regression, BKMR)을 실제로 활용해보고, 각 통계방법의 특성을 살펴보았다.

<표 3-4> 연구 목표에 따른 통계방법 평가

	복합노출의 영향 평가	종적 자료 적용	대용량 자료 분석	각 노출의 기여 정도 확인	생물학적 해석
가중분위수합 회귀	가능	가능	가능	가능	가능
주성분분석	가능	가능	가능	불가능	불가능
감독된 주성분분석	가능	가능	가능	불가능	불가능
부분최소제곱 회귀	가능	가능	가능	가능	가능
희박부분최소제곱 회귀	가능	가능	가능	가능	가능
구조방정식 모델링	가능	가능	가능	가능	가능
베이지안커널머신 회귀	가능	가능	불가능	가능	가능

<표 3-5> 연구 목표 외 통계방법의 장단점 평가

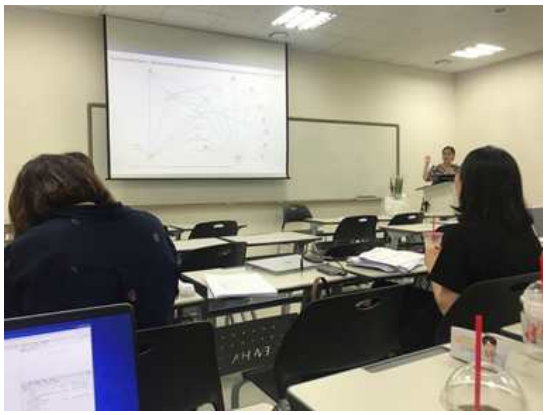
	평가 결과	목표 외 통계적 장단점 평가 내용
가중분위수합 회귀	△	가중분위수합 회귀는 측정된 가중치에 따라 중요한 노출을 알기 위해 구체적인 기준치가 있어야 하는데 이 기준치에 대한 기준을 설정하기 어려움
주성분분석	△	각 노출의 결과 변수에 미치는 영향의 기여 정도와 생물학적 해석에 대한 정보를 알기 어려우나, 전통적인 방법으로 시도해볼 필요는 있음
감독된 주성분분석	×	본 연구에서는 가설 설정 시 노출과 결과의 연관성을 미리 고려하므로, 통계적인 노출변수의 선택이 필요하지 않음
부분최소제곱 회귀	△	누적노출점수를 산출 가능할 때 로딩벡터를 통해 각 노출의 기여 정도를 알 수 있으나, 누적노출점수를 산출하는 동안 보정변수의 영향을 고려하지 못한 상태임
희박부분최소제곱 회귀	×	본 연구에서는 가설 설정 시 노출과 결과의 연관성을 미리 고려하므로, 통계적인 노출변수의 선택이 필요하지 않음
구조방정식 모델링	△	노출과 결과 변수들 간의 복잡한 관계에 대한 사전 지식이 풍부해야 함
베이지안커널머신 회귀	△	자료 용량이 큰 경우에는 분석이 불가능할 수 있으나, 노출변수간의 상호작용과 노출과 결과변수 간의 비선형성을 고려한 분석이 가능하다는 장점이 있음

6. 연구진 및 관계자 회의 진행

○ 매주 고찰한 문헌을 토대로 연구진 내부 회의를 진행하여 통계방법의 활용방안에 대해 토의하였다.

○ 2019년 7월 31일 (금) 복합노출 통계방법 세미나를 통해 복합노출로 인한 건강영향을 평가하기 위한 통계방법을 R 패키지를 활용하여 실제 분석해보는 실습을 다음 내용을 포함하여 진행하였다.

- ‘caret’, ‘glmnet’ 패키지를 활용한 탄성순회귀(Elastic net regression) 분석 실습
- ‘psych’ 패키지를 활용한 주성분분석(principal component analysis) 분석 실습
- ‘pls’ 패키지를 활용한 부분최소제곱 회귀(partial least squares regression) 분석 실습
- ‘bkmr’ 패키지를 활용한 베이지안커널머신 회귀(Bayesian kernel machine regression) 분석 실습
- ‘gWQS’ 패키지를 활용한 가중분위수합 회귀(Weighted quantile sum regression) 분석 실습



[그림 3-1] 복합노출 통계방법 세미나

7. 복합노출의 건강영향을 추정하기 위한 통계분석 방법의 적용

1) 연구대상자 및 자료의 특성

유해물질 복합노출의 건강영향을 평가하기 위해 노동자를 대상으로 한 이용가능한 이차자료에는 특수건강진단 자료, 건강보험공단 자료, 사망원인통계자료가 있다. 유해인자에 노출되는 노동자가 수검하는 특수건강진단자료에서는 혈중 및 요중 생물학적 지표를 통하여 노출 변수를 사용할 수 있으며, 혈액검사결과, 검진판정결과, 질병유병 등을 파악할 수 있다. 건강보험공단 자료에서는 직장가입자의 국가검진결과 및 질병명을 포함한 병원진료 기록을 파악할 수 있다. 사망원인통계자료에서는 사망자의 사망원인을 소개하고 있다.

특수건강진단 수검 전 사업장을 대상으로 작업환경측정을 수행하고 있는데, 노출기준을 크게 초과하는 유해물질은 현재 거의 없는 상태이다. 이러한 상황에서 유해물질 복합 노출로 인한 건강영향을 질병 발생이나 사망으로 평가하기 위해서는 장기간의 추적이 필요한 상황이다. 또한 질병발생이나 사망은 이분형 변수로 예와 아니오로 구분된다. 이는 유해물질 농도가 매우 높을 때 사용할 수 있지만, 저농도의 유해물질에 복합 노출된 경우에는 혈액검사 결과가 정상이어도 수치가 소폭 상승할 수 있다. 국민건강보험공단 자료와 사망원인통계에서는 이 미세한 차이를 볼 수가 없다. 하지만 특수건강진단에서는 혈액검사결과를 연속적인 수치로 분석할 수 있고, 국내 노동자의 건강을 대표할 수 있는 자료로 사료된다. 따라서 유해물질 복합노출에 대한 건강영향 분석에는 특수건강진단 자료를 사용하였다.

본 연구에서는 2013년부터 2017년 자료에서 각 연도 별로 자료를 분석하였고, 2013년부터 2017년 까지 특수건강검진을 받은 모든 대상자를 모아서 각 연구대상자의 가장 과거의 자료만 남겨서 중복 제거를 한 경우와 각 연구대상자의 가장 최근의 자료만 남겨서 중복 제거를 한 경우로 나누어

추가로 분석하였다. 특수건강진단 연구대상자와 자료의 특성은 <표 3-6> 및 <표 3-7>, <표 3-8>과 같다. 분석에 사용한 각 기간 별 자료에서 연구대상자의 특성을 살펴보면, 모든 자료에서 연구대상자의 특성과 노출 변수, 결과 변수의 특성이 비슷한 것을 확인할 수 있었다.

노출 변수는 왼쪽으로 기울어진 자료의 분포를 보였고, 세 가지 노출변수의 단위와 분산이 달랐다. 따라서, 모든 분석 전에 세 가지 노출변수를 로그 변환을 하였고, 복합노출의 건강영향을 평가하기 위한 통계분석에서 세 가지 노출물질의 기여도를 비교할 수 있도록 평균 0, 분산 1로 노출변수를 표준화한 뒤 상관분석, 다변량 선형회귀분석 및 주성분분석, 부분최소제곱 회귀, 베이지안커널머신 회귀 분석을 하였다.

세 가지 노출변수의 연관성을 확인하기 위해 각 연도 별로 상관분석을 하였을 때, 마뇨산과 메틸마뇨산은 일관되게 상관관계가 관찰되었으나 상관계수가 0.2 내외로 상관성이 크지는 않았다.

<표 3-6> 연도 별 연구대상자의 특성 (1)

	2013	2014	2015	2016	2017	2013-2017*	2013-2017**
	N=615	N=1220	N=1604	N=2351	N=2678	N=6330	N=6330
나이							
20대 이하	100 (16.3)	219 (18)	236 (14.7)	451 (19.2)	472 (17.6)	1189 (18.8)	1122 (17.7)
30-40대	369 (60)	705 (57.8)	852 (53.1)	1284 (54.6)	1401 (52.3)	3460 (54.7)	3429 (54.2)
50-60대	145 (23.6)	295 (24.2)	515 (32.1)	614 (26.1)	802 (30)	1675 (26.5)	1772 (28)
70대 이상	1 (0.2)	1 (0.1)	1 (0.1)	2 (0.1)	3 (0.1)	6 (0.1)	7 (0.1)
성별							
남	588 (95.6)	1125 (92.2)	1511 (94.2)	2115 (90)	2476 (92.5)	5761 (91)	5758 (91)
여	27 (4.4)	95 (7.8)	93 (5.8)	236 (10)	202 (7.5)	569 (9)	572 (9)
체질량지수(kg/m^2)							
<25	423 (68.8)	776 (63.6)	1007 (62.8)	1459 (62.1)	1637 (61.1)	3985 (63)	3954 (62.5)
≥25	192 (31.2)	444 (36.4)	597 (37.2)	892 (37.9)	1041 (38.9)	2345 (37.1)	2376 (37.5)
흡연상태							
비 흡연	173 (28.1)	378 (31)	472 (29.4)	780 (33.2)	988 (36.9)	2141 (33.8)	2153 (34)
과거 흡연	156 (25.4)	269 (22.1)	382 (23.8)	528 (22.5)	589 (22)	1276 (20.2)	1328 (21)
현재 흡연	286 (46.5)	573 (47)	750 (46.8)	1043 (44.4)	1101 (41.1)	2913 (46)	2849 (45)

*같은 사람에서 최근 자료를 중복 제거하여 가장 과거의 자료만 남김
**같은 사람에서 과거 자료를 중복 제거하여 가장 최근의 자료만 남김

<표 3-7> 연도 별 연구대상자의 특성 (2)

	2013	2014	2015	2016	2017	2013-2017*	2013-2017**
	N=615	N=1220	N=1604	N=2351	N=2678	N=6330	N=6330
음주습관(잔/주)							
0	176 (28.6)	336 (27.5)	438 (27.3)	614 (26.1)	867 (32.4)	1872 (29.6)	1906 (30.1)
1-7	117 (19)	166 (13.6)	229 (14.3)	400 (17)	382 (14.3)	980 (15.5)	940 (14.9)
≥7	322 (52.4)	718 (58.9)	937 (58.4)	1337 (56.9)	1429 (53.4)	3478 (54.9)	3484 (55)
중등도 이상 운동습관(회/주)							
<3	400 (65)	821 (67.3)	1060 (66.1)	1485 (63.2)	1660 (62)	4204 (66.4)	4110 (64.9)
≥3	215 (35)	399 (32.7)	544 (33.9)	866 (36.8)	1018 (38)	2126 (33.6)	2220 (35.1)

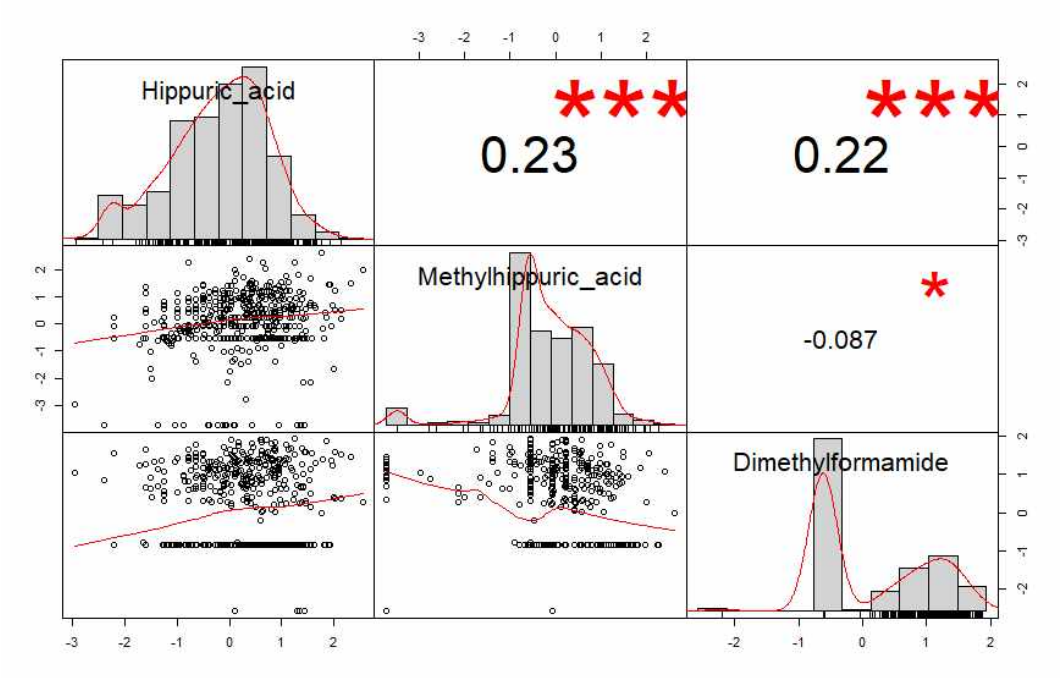
*같은 사람에서 최근 자료를 중복 제거하여 가장 과거의 자료만 남김
**같은 사람에서 과거 자료를 중복 제거하여 가장 최근의 자료만 남김

<표 3-8> 연도 별 노출 및 결과 자료의 특성

	2013	2014	2015	2016	2017	2013-2017*	2013-2017**
	N=615	N=1220	N=1604	N=2351	N=2678	N=6330	N=6330
	중위수 (최소값, 최대값)						
노출 변수							
마뇨산(g/g crea.)	0.14 (0.0042, 2.285)	0.146 (0.001, 2.293)	0.1575 (0.004, 2.45)	0.134 (0.006, 2.39)	0.15 (0.001, 2.33)	0.14 (0.001, 2.45)	0.148 (0.001, 2.45)
메틸마뇨산(g/g crea.)	0.02 (0.0001, 0.97)	0.02 (0.0001, 0.894)	0.01 (0.001, 0.916)	0.01 (0.001, 1.186)	0.013 (0.0002, 1.03)	0.01 (0.0001, 1.19)	0.011 (0.0001, 1.186)
디메틸포름아미드(mg/g crea.)	0.01 (0.0001, 14.55)	0.2495 (0.01, 14.82)	0.01 (0.001, 14.47)	0.01 (0.001, 14.50)	0.01 (0.001, 14.74)	0.01 (0.0001, 14.82)	0.01 (0.0001, 14.74)
결과 변수							
아스파테이트아미노전이효소(IU/L)	22 (9, 118)	23 (9, 339)	23 (9, 145)	23 (10, 427)	23 (2, 624)	23 (2, 427)	23 (2, 624)
알라닌아미노전이효소(IU/L)	20 (5, 159)	22 (6, 171)	22 (5, 199)	21 (5, 327)	21 (1.5, 229)	21 (1.5, 327)	21 (1.5, 327)
감마글루타밀전이효소(IU/L)	27 (8, 290)	28 (6, 897)	30 (6, 772)	29 (3, 762)	29 (0.8, 1613)	29 (0.8, 1613)	29 (0.8, 1613)

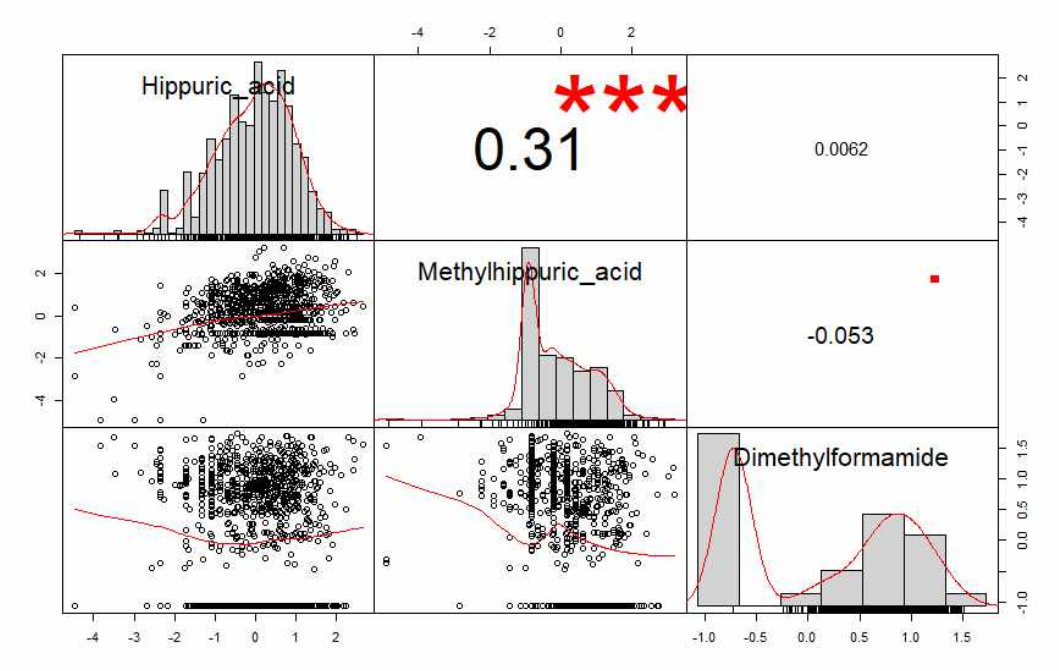
노출 평가 대상은 특수건강진단에서 현재 생물학적 노출 지표로 사용하고 있는 요중 마뇨산, 요중 메틸마뇨산과 요중 N,N-디메틸포름아미드이다. 위 세 가지 노출변수에 대하여 상관분석을 실시하였다.

2013년 자료에서 요중 마뇨산과 요중 메틸마뇨산의 상관계수는 0.23이었으며, 요중 마뇨산과 요중 N,N-디메틸포름아미드의 상관계수도 0.22로 비슷하였다. 다만 요중메틸마뇨산과 요중 N,N-디메틸포름아미드의 상관계수는 -0.087로 약한 음의 상관관계가 관찰되었다.



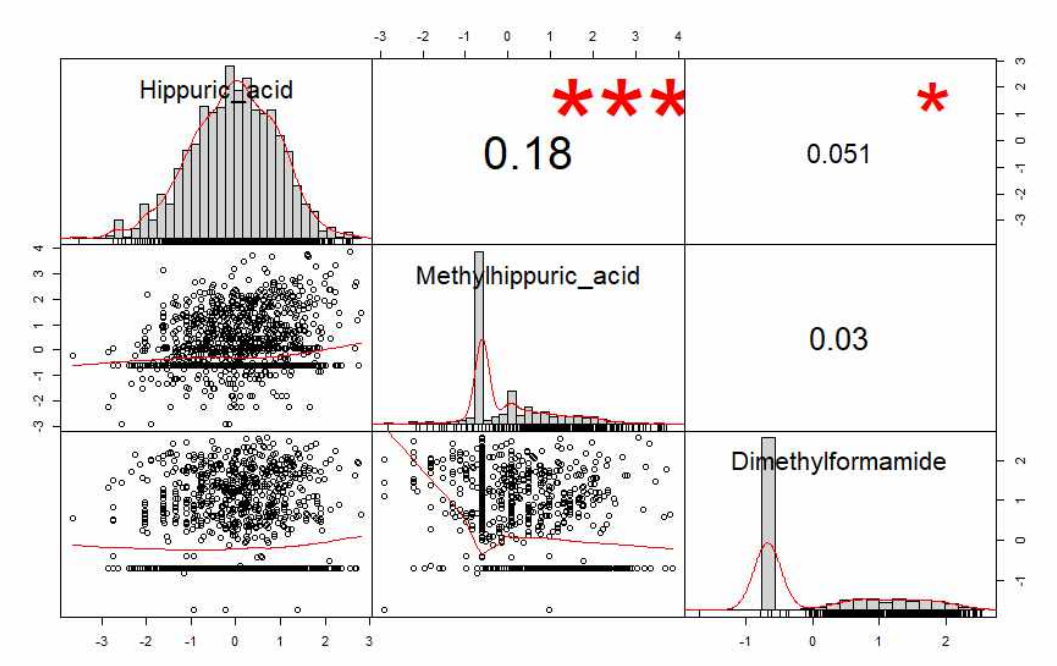
[그림 3-2] 2013년 자료에서 세 가지 노출변수의 상관분석 결과

2014년 자료에서 요중 마뇨산과 요중 메틸마뇨산의 상관계수는 0.31로 양의 상관관계가 관찰되었다.



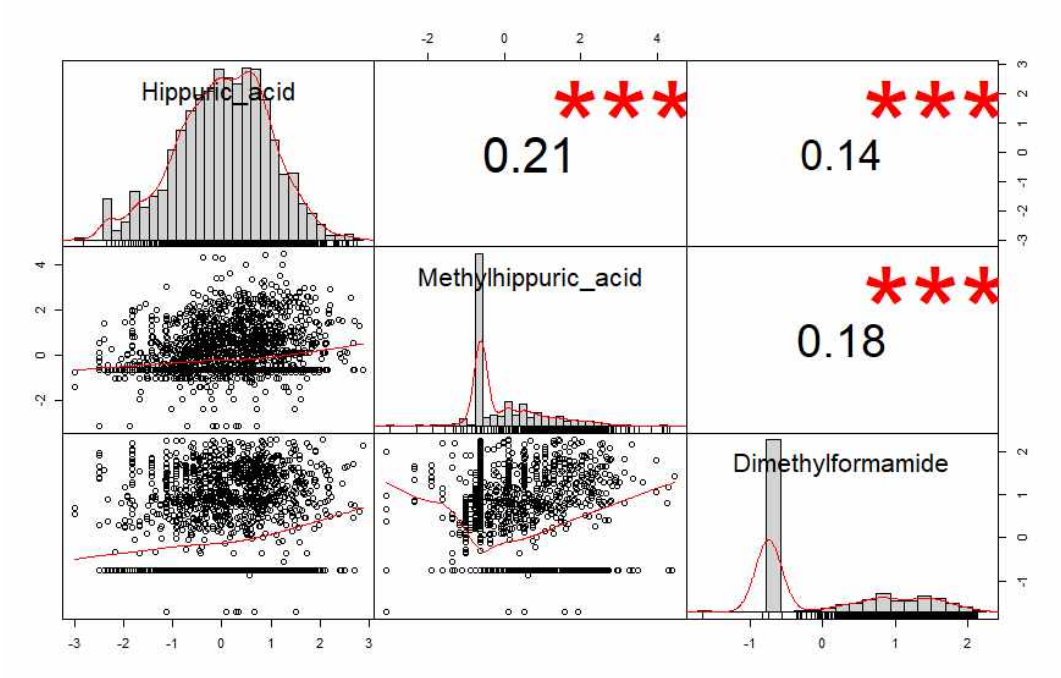
[그림 3-3] 2014년자료에서 세 가지 노출변수의 상관분석 결과

2015년 자료에서 요중 마뇨산과 요중 메틸마뇨산의 상관계수는 0.18이었으며, 요중 마뇨산과 요중 N,N-디메틸포름아미드의 상관계수는 0.051으로 약한 양의 상관관계가 관찰되었다.



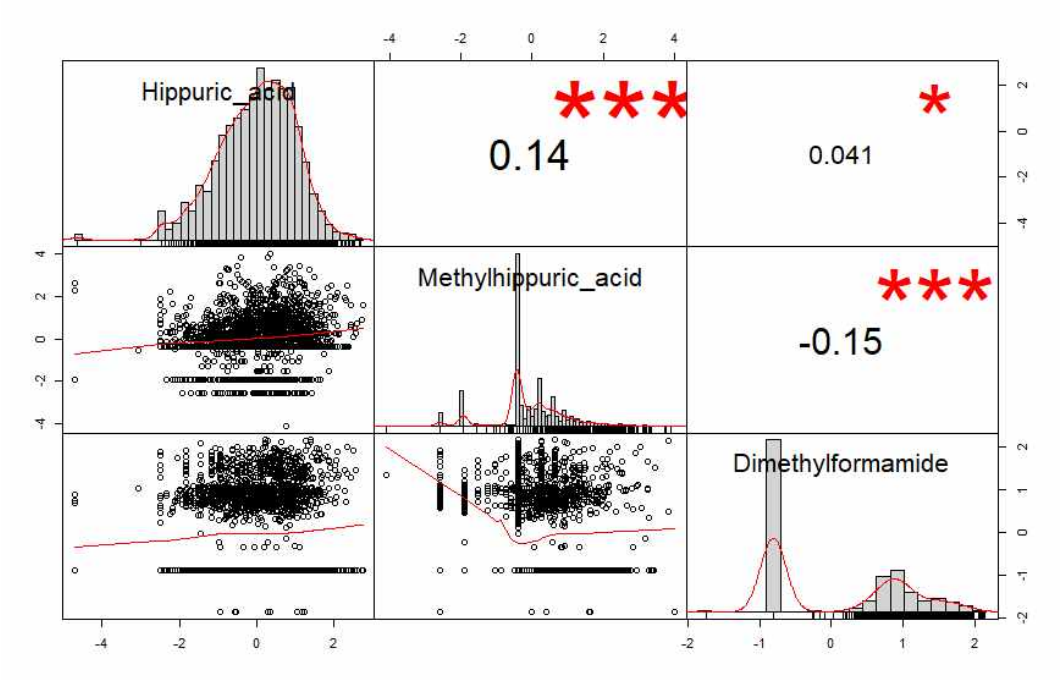
[그림 3-4] 2015년 자료에서 세 가지 노출변수의 상관분석 결과

2016년 자료에서 요중 마뇨산과 요중 메틸마뇨산의 상관계수는 0.21이었으며, 요중 마뇨산과 요중 N,N-디메틸포름아미드의 상관계수는 0.14이며, 요중 메틸마뇨산과 요중 N,N-디메틸포름아미드의 상관관계는 0.18로 관찰되었다.



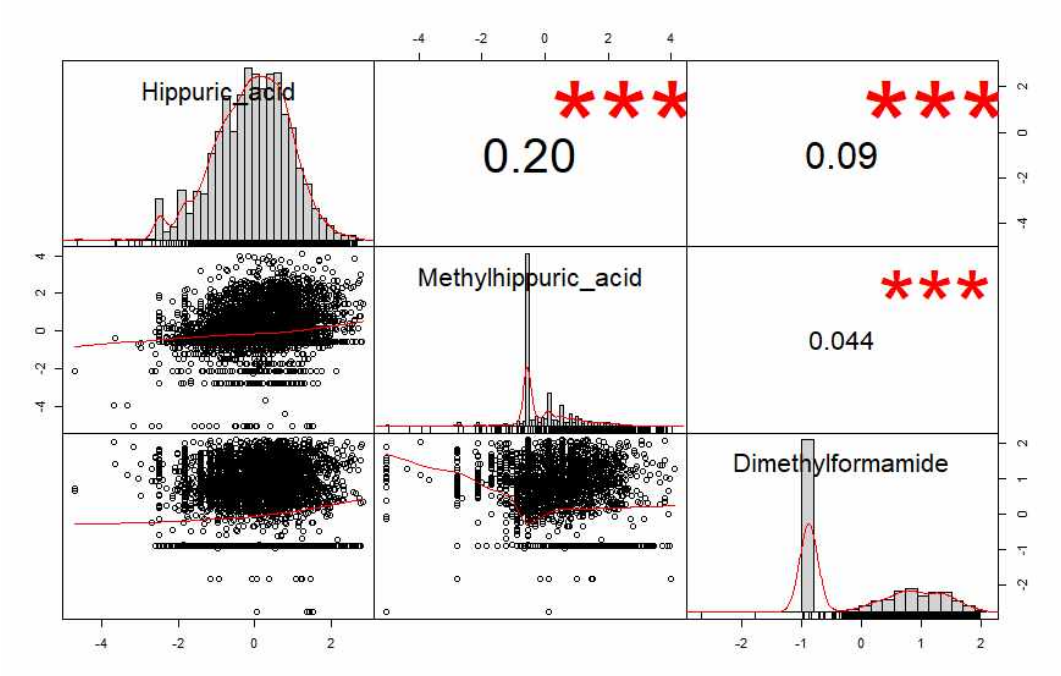
[그림 3-5] 2016년 자료에서 세 가지 노출변수의 상관분석 결과

2017년 자료에서 요중 마뇨산과 요중 메틸마뇨산의 상관관계는 0.14이었으며, 요중 마뇨산과 요중 N,N-디메틸포름아미드는 0.041이었으며, 요중 메틸마뇨산과 요중 N,N-디메틸포름아미드의 상관계수는 -0.15로 음의 상관관계가 관찰되었다.



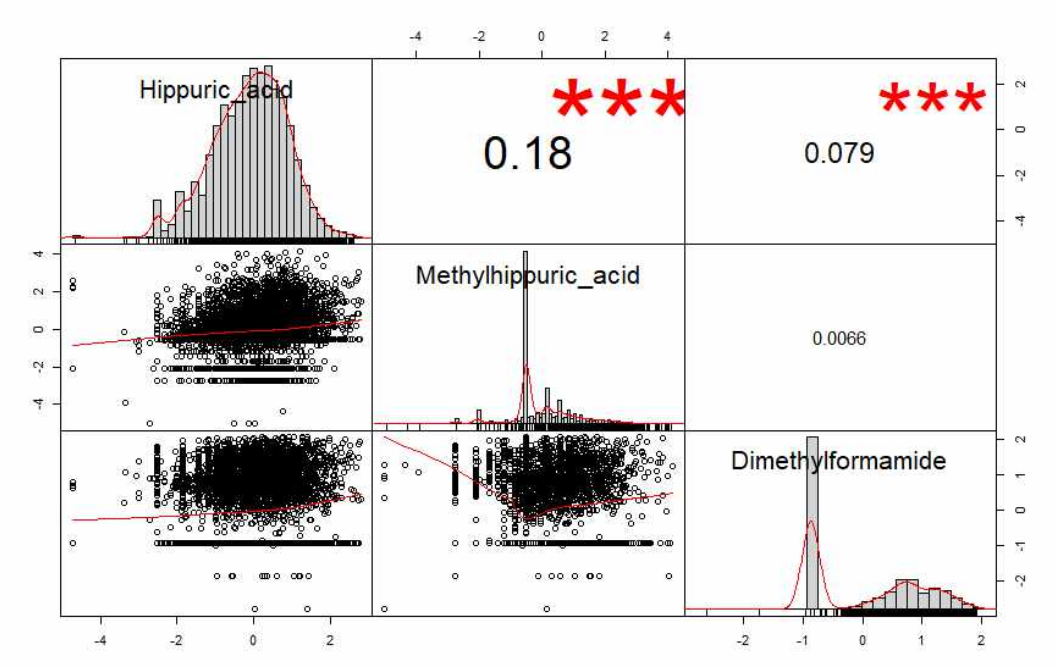
[그림 3-6] 2017년 자료에서 세 가지 노출변수의 상관분석 결과

2013년부터 2017년까지 자료를 통합하여 같은 사람의 특수건강검진 결과 중 최근 자료는 중복 제거하여 가장 과거의 자료만 남긴 자료에서 세 노출변수의 상관관계를 분석하면 요중 마뇨산과 요중 메틸마뇨산의 상관계수는 0.20이었으며, 요중 마뇨산과 요중 N,N-디메틸포름아미드의 상관계수는 0.09이며, 요중 메틸마뇨산과 요중 N,N-디메틸포름아미드의 상관계수는 0.044로 조사되었다.



[그림 3-7] 2013-2017년 자료(최근 자료 중복제거)에서 세 가지 노출변수의 상관분석 결과

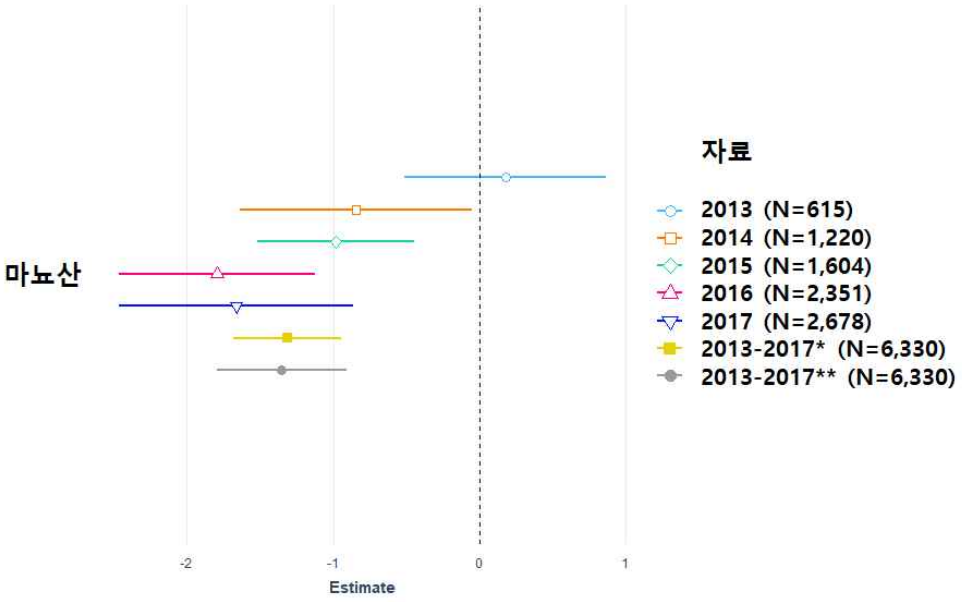
2013년부터 2017년까지 자료를 통합하여 같은 사람의 특수건강검진 결과 중 과거 자료는 중복 제거하여 가장 최근의 자료만 남긴 자료에서 세 노출변수의 상관관계를 분석하면 요중 마뇨산과 요중 메틸마뇨산의 상관계수는 0.18이었으며, 요중 마뇨산과 요중 N,N-디메틸포름아미드는 0.079이었다.



[그림 3-8] 2013-2017년 자료(과거 자료 중복제거)에서 세 가지 노출변수의 상관분석 결과

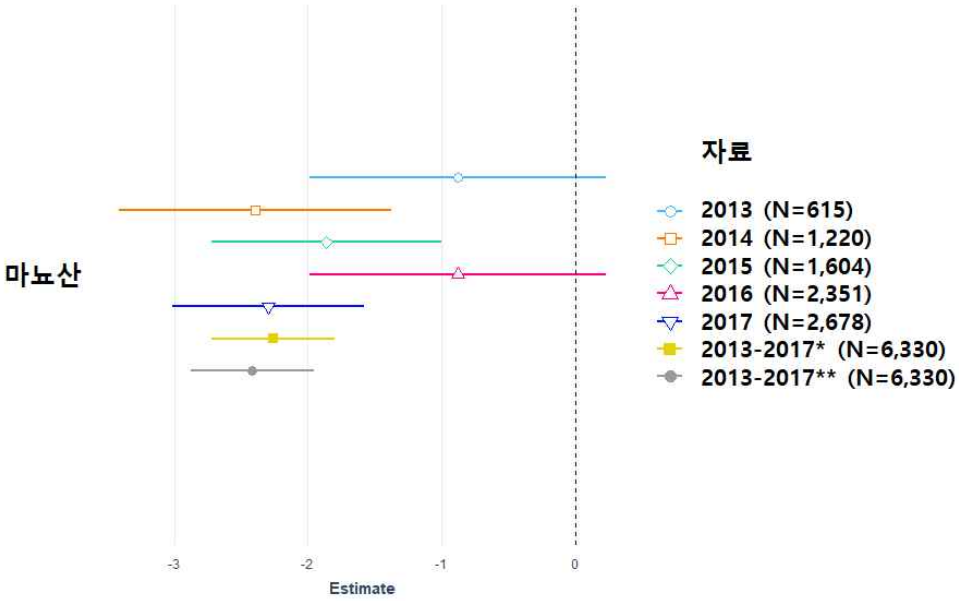
2) 단일 생물학적 노출지표와 간기능검사 결과 간의 관련성 분석

그림 3-9는 마뇨산 단일 노출과 아스파테이트아미노전이효소 간의 관련성을 다중 선형회귀분석을 이용하여 분석한 결과이다. 2013년부터 2017년까지 각 해의 자료를 분석하였고, 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 가장 과거의 자료를 남기는 경우와 같은 사람의 자료 중 과거 자료를 중복 제거하여 가장 최근의 자료를 남기는 경우로 자료를 가공하여 추가로 분석하였다. 나이, 체질량지수, 흡연 상태, 음주량, 중등도 이상 운동 유무를 보정하였을 때, 2013년 자료에서 마뇨산과 아스파테이트아미노전이효소의 관련성은 유의하지 않았으나, 그 외의 자료에서는 음의 연관성을 유의하게 보였다.



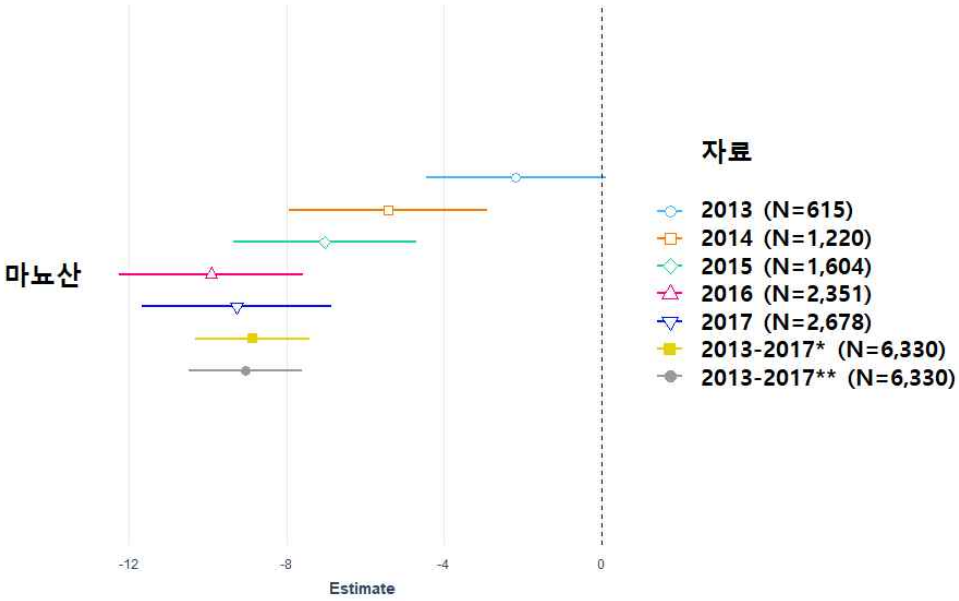
[그림 3-9] 다중 선형회귀분석을 이용한 요 중 마뇨산(hippuric acid)과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

그림 3-10은 마뇨산 단일 노출과 알라닌아미노전이효소 간의 관련성을 다중 선형회귀분석을 이용하여 분석한 결과이다. 2013년부터 2017년까지 각 해의 자료를 분석하였고, 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 가장 과거의 자료를 남기는 경우와 같은 사람의 자료 중 과거 자료를 중복 제거하여 가장 최근의 자료를 남기는 경우로 자료를 가공하여 추가로 분석하였다. 나이, 체질량지수, 흡연 상태, 음주량, 중등도 이상 운동 유무를 보정하였을 때, 2013년과 2016년 자료에서 마뇨산과 알라닌아미노전이효소의 관련성은 유의하지 않았으나, 그 외의 자료에서는 음의 연관성을 유의하게 보였다.



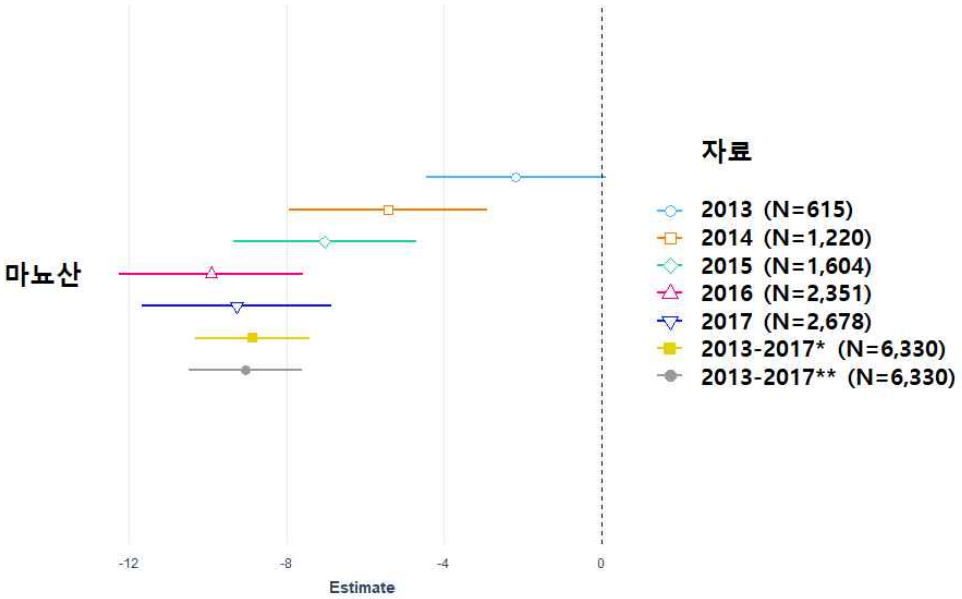
[그림 3-10] 다중 선형회귀분석을 이용한 요 중 마뇨산(hippuric acid)과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

그림 3-11은 마뇨산 단일 노출과 감마글루타밀전이효소 간의 관련성을 다중 선형회귀분석을 이용하여 분석한 결과이다. 2013년부터 2017년까지 각 해의 자료를 분석하였고, 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 가장 과거의 자료를 남기는 경우와 같은 사람의 자료 중 과거 자료를 중복 제거하여 가장 최근의 자료를 남기는 경우로 자료를 가공하여 추가로 분석하였다. 나이, 체질량지수, 흡연 상태, 음주량, 중등도 이상 운동 유무를 보정하였을 때, 2013년 자료에서 마뇨산과 감마글루타밀전이효소의 관련성은 유의하지 않았으나, 그 외의 자료에서는 음의 연관성을 유의하게 보였다.



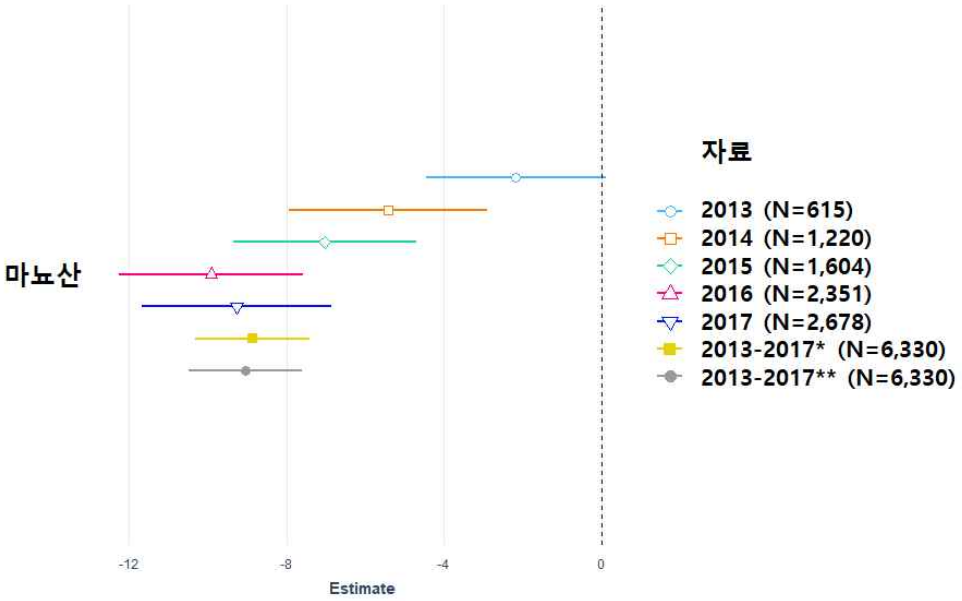
[그림 3-11] 다중 선형회귀분석을 이용한 요 중 마뇨산(hippuric acid)과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

그림 3-12는 메틸마뇨산 단일 노출과 아스파테이트아미노전이효소 간의 관련성을 다중 선형회귀분석을 이용하여 분석한 결과이다. 2013년부터 2017년까지 각 해의 자료를 분석하였고, 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 가장 과거의 자료를 남기는 경우와 같은 사람의 자료 중 과거 자료를 중복 제거하여 가장 최근의 자료를 남기는 경우로 자료를 가공하여 추가로 분석하였다. 나이, 체질량지수, 흡연 상태, 음주량, 중등도 이상 운동 유무를 보정하였을 때, 2013, 2014, 2015, 2017년 자료와 2013부터 2017년 자료 중 중복제거 후 가장 과거의 자료를 남긴 경우에 메틸마뇨산과 아스파테이트아미노전이효소의 관련성은 유의하지 않았으나, 그 외의 자료에서는 음의 연관성을 유의하게 보였다.



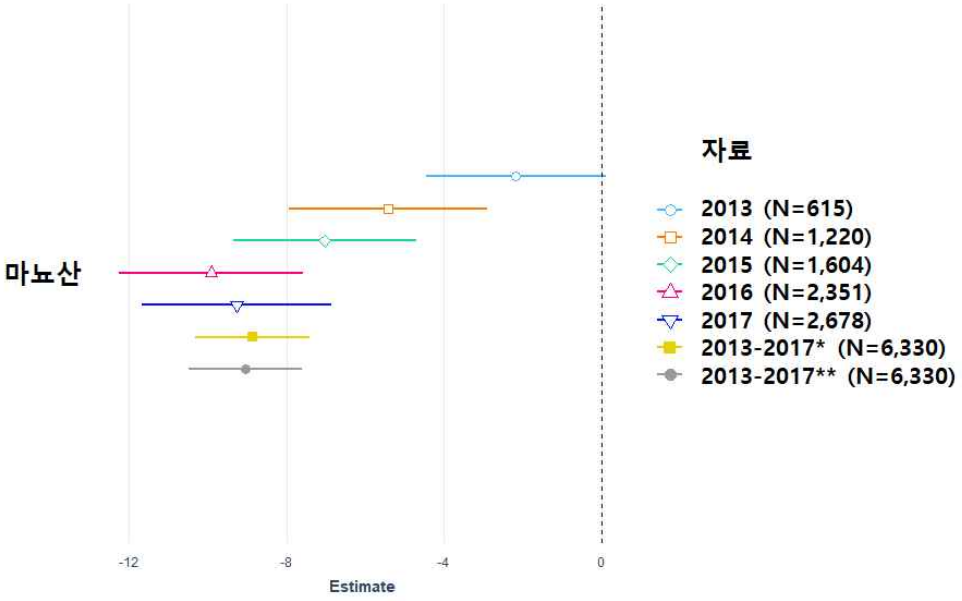
[그림 3-12] 다중 선형회귀분석을 이용한 요 중 메틸마뇨산(methyl hippuric acid)과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

그림 3-13은 메틸마뇨산 단일 노출과 알라닌아미노전이효소 간의 관련성을 다중 선형회귀분석을 이용하여 분석한 결과이다. 2013년부터 2017년까지 각 해의 자료를 분석하였고, 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 가장 과거의 자료를 남기는 경우와 같은 사람의 자료 중 과거 자료를 중복 제거하여 가장 최근의 자료를 남기는 경우로 자료를 가공하여 추가로 분석하였다. 나이, 체질량지수, 흡연 상태, 음주량, 중등도 이상 운동 유무를 보정하였을 때, 2017년 자료에서 메틸마뇨산과 알라닌아미노전이효소에서 유의한 음의 관련성을 관찰하였으나, 그 외의 자료에서는 유의한 관련성이 관찰되지 않았다.



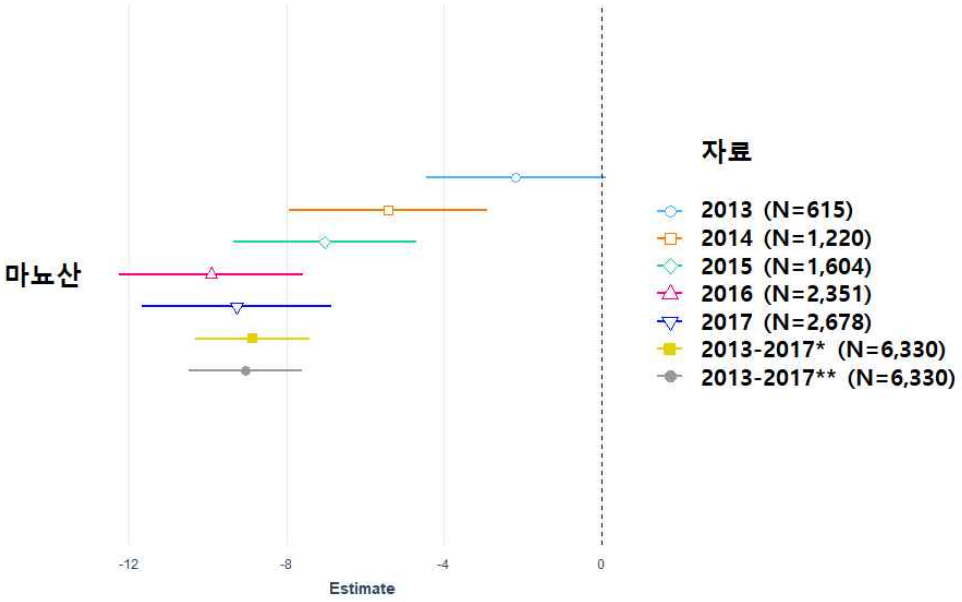
[그림 3-13] 다중 선형회귀분석을 이용한 요 중 메틸마뇨산(methyl hippuric acid)과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

그림 3-14는 메틸마뇨산 단일 노출과 감마글루타밀전이효소 간의 관련성을 다중 선형회귀분석을 이용하여 분석한 결과이다. 2013년부터 2017년까지 각 해의 자료를 분석하였고, 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 가장 과거의 자료를 남기는 경우와 같은 사람의 자료 중 과거 자료를 중복 제거하여 가장 최근의 자료를 남기는 경우로 자료를 가공하여 추가로 분석하였다. 나이, 체질량지수, 흡연 상태, 음주량, 중등도 이상 운동 유무를 보정하였을 때, 2013년 자료에서 메틸마뇨산과 감마글루타밀전이효소에서 유의한 양의 관련성이 관찰되었고, 2016년 자료에서 유의한 음의 관련성이 관찰되었으며, 그 외의 자료에서는 유의한 관련성이 관찰되지 않았다.



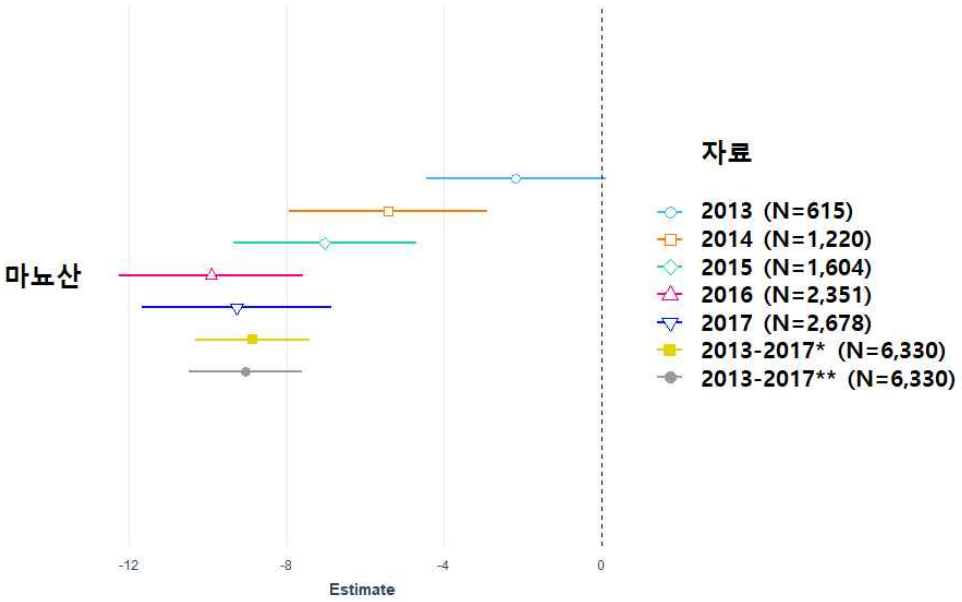
[그림 3-14] 다중 선형회귀분석을 이용한 요 중 메틸마뇨산(methyl hippuric acid)과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

그림 3-15는 N,N-디메틸포름아미드 단일 노출과 아스파테이트아미노전이효소 간의 관련성을 다중 선형회귀분석을 이용하여 분석한 결과이다. 2013년부터 2017년까지 각 해의 자료를 분석하였고, 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 가장 과거의 자료를 남기는 경우와 같은 사람의 자료 중 과거 자료를 중복 제거하여 가장 최근의 자료를 남기는 경우로 자료를 가공하여 추가로 분석하였다. 나이, 체질량지수, 흡연 상태, 음주량, 중등도 이상 운동 유무를 보정하였을 때, 2013년과 2014년 자료에서 N,N-디메틸포름아미드와 아스파테이트아미노전이효소는 양의 연관성이 유의하게 관찰되었으나, 그 외의 자료에서는 유의하지 않았다.



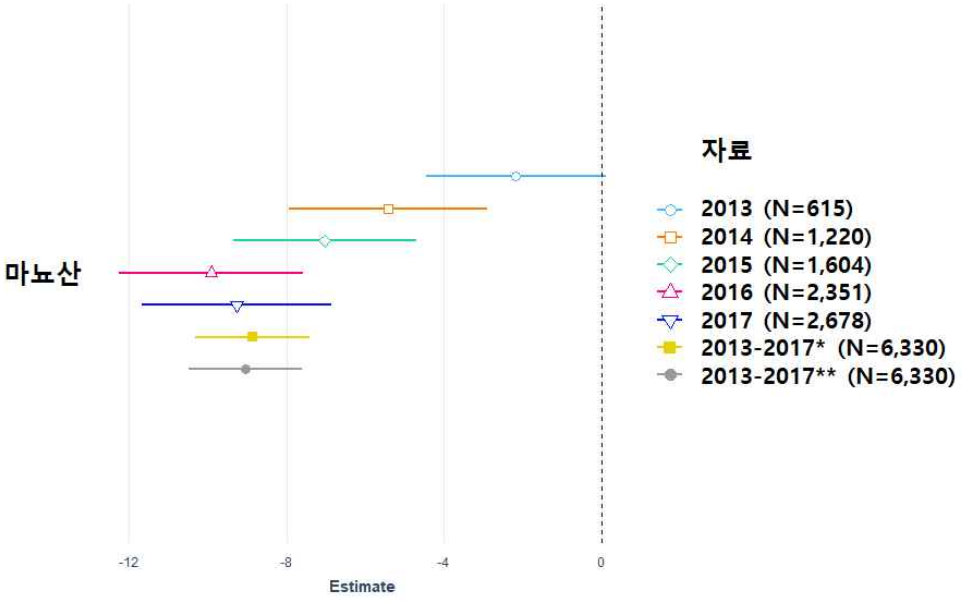
[그림 3-15] 다중 선형회귀분석을 이용한 요 중 N,N-디메틸포름아미드(N,N-dimethylformamide)와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

그림 3-16은 N,N-디메틸포름아미드 단일 노출과 알라닌아미노전이효소 간의 관련성을 다중 선형회귀분석을 이용하여 분석한 결과이다. 2013년부터 2017년까지 각 해의 자료를 분석하였고, 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 가장 과거의 자료를 남기는 경우와 같은 사람의 자료 중 과거 자료를 중복 제거하여 가장 최근의 자료를 남기는 경우로 자료를 가공하여 추가로 분석하였다. 나이, 체질량지수, 흡연 상태, 음주량, 중등도 이상 운동 유무를 보정하였을 때, 모든 자료에서 N,N-디메틸포름아미드와 알라닌아미노전이효소의 관련성은 유의하지 않았다.



[그림 3-16] 다중 선형회귀분석을 이용한 요 중 N,N-디메틸포름아미드(N,N-dimethylformamide)와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

그림 3-15는 N,N-디메틸포름아미드 단일 노출과 감마글루타밀전이효소 간의 관련성을 다중 선형회귀분석을 이용하여 분석한 결과이다. 2013년부터 2017년까지 각 해의 자료를 분석하였고, 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 가장 과거의 자료를 남기는 경우와 같은 사람의 자료 중 과거 자료를 중복 제거하여 가장 최근의 자료를 남기는 경우로 자료를 가공하여 추가로 분석하였다. 나이, 체질량지수, 흡연 상태, 음주량, 중등도 이상 운동 유무를 보정하였을 때, 2013년, 2014년, 2015년의 자료에서 N,N-디메틸포름아미드와 감마글루타밀전이효소의 관련성은 유의하지 않았으나, 그 외의 자료에서는 음의 연관성을 유의하게 보였다.



[그림 3-15] 다중 선형회귀분석을 이용한 혈 중 N,N-디메틸포름아미드(N,N-dimethylformamide)와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

R을 이용한 다중 선형회귀분석 및 결과 요약을 위한 그림을 그리는 R 코드는 다음과 같다.

```
#데이터 읽기#
data1<-read.csv("D:/연구/복합노출 연구/data/three2013y1.csv")
data2<-read.csv("D:/연구/복합노출 연구/data/three2014y1.csv")
data3<-read.csv("D:/연구/복합노출 연구/data/three2015y1.csv")
data4<-read.csv("D:/연구/복합노출 연구/data/three2016y1.csv")
data5<-read.csv("D:/연구/복합노출 연구/data/three2017y1.csv")
data6<-read.csv("D:/연구/복합노출 연구/data/three1317ally1.csv")
data7<-read.csv("D:/연구/복합노출 연구/data/three1317ally2.csv")

#다중선형회귀분석#
mod1<-(lm(y1~logx1+age+sex+bmi+smoking1+smoking2+drinking+exercise,data=data1))
mod2<-(lm(y1~logx1+age+sex+bmi+smoking1+smoking2+drinking+exercise,data=data2))
mod3<-(lm(y1~logx1+age+sex+bmi+smoking1+smoking2+drinking+exercise,data=data3))
mod4<-(lm(y1~logx1+age+sex+bmi+smoking1+smoking2+drinking+exercise,data=data4))
mod5<-(lm(y1~logx1+age+sex+bmi+smoking1+smoking2+drinking+exercise,data=data5))
mod6<-(lm(y1~logx1+age+sex+bmi+smoking1+smoking2+drinking+exercise,data=data6))
mod7<-(lm(y1~logx1+age+sex+bmi+smoking1+smoking2+drinking+exercise,data=data7))

#그림 그리기#
#install.packages("jtools")
library("jtools")

plot_summs(mod1, mod2, mod3, mod4, mod5, mod6, mod7,
           model.names=c("2013 (N=615)", "2014 (N=1,220)", "2015 (N=1,604)",
                         "2016 (N=2,351)", "2017 (N=2,678)",
                         "2013-2017_ascending (N=6,330)",
                         "2013-2017_descending (N=6,330)"),
           legend.title = "Data", coefs=c("Hippuric acid"="logx1"), scale=TRUE)

plot_summs(mod1, mod2, mod3, mod4, mod5, mod6, mod7,
            model.names=c("2013 (N=615)", "2014 (N=1,220)", "2015 (N=1,604)",
                          "2016 (N=2,351)","2017 (N=2,678)",
                          "2013-2017_ascending (N=6,330)",
                          "2013-2017_descending (N=6,330)"), legend.title = "Data",
            coefs=c("Hippuric acid"="logx1", "Age"="age", "Sex"="sex","BMI"="bmi",
                    "Past smoker"="smoking1", "Current smoker"="smoking2",
                    "Alcohol drinking habits"="drinking", "Physical activity"="exercise"),
            scale=TRUE)
```

3) 주성분분석(Principal component analysis, PCA)을 활용한 복합노출과 간기능검사 결과 간의 관련성 분석

주성분분석은 차원축소를 위한 대표적 기법으로, 상관관계가 있는 변수들 사이의 변동을 상관관계가 없는 주성분이라 부르는 새로운 변수들로 설명하고자 하는 차원축소 기법이다. 주성분은 원래 변수들의 선형결합이 되며 공분산행렬의 고유벡터가 선형결합을 위한 계수 벡터가 된다(김지현, 2015).

주성분분석은 변수의 차원을 축약하여 개체에 대한 정보를 얻기 쉽게 하며, 주성분변수가 상관계수가 0(독립)이라는 사실을 이용하여 회귀분석의 다중공선성 문제해결을 위한 방법으로 사용한다.

주성분분석에서 변수들이 측정단위가 다르거나 분산의 불균형이 심할 경우 변수를 표준화한 다음 분석할 것이 권장되어 본 연구에서는 왜도가 심한 노출변수의 특성을 고려하여 세 가지 노출변수 모두 로그변환을 한 뒤 평균 0, 표준편차 1로 표준화하여 분석에 사용하였다.

주성분의 개수를 정하는 기준은 일반적으로 70~80% 이상을 많이 쓰므로(김기영 외, 1997), 본 연구에서는 원변수 누적 기여율 70%이상을 기준으로 선택하였다.

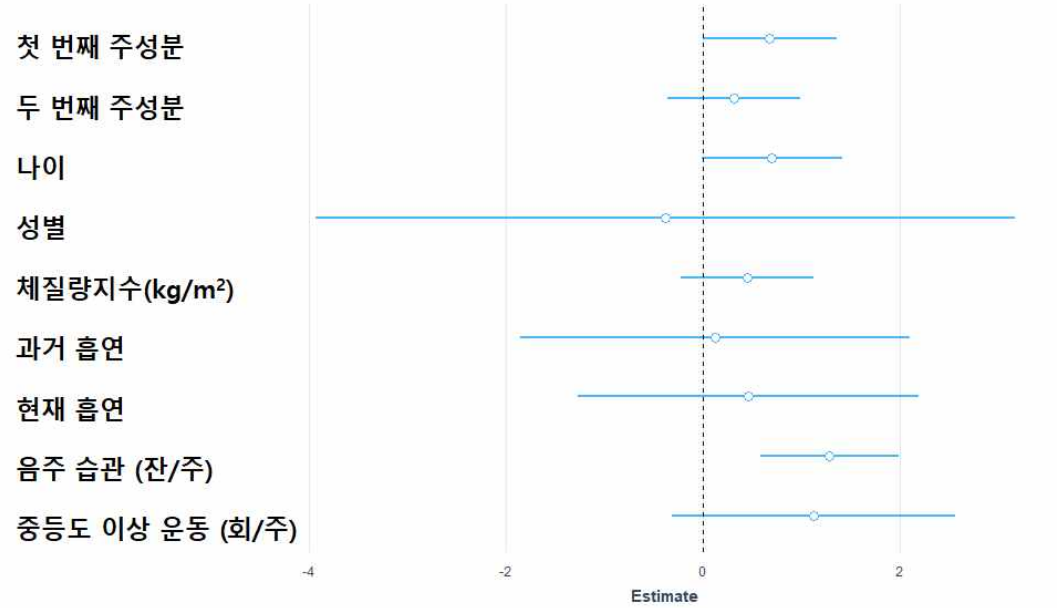
주성분분석을 위해 psych R package를 사용하였다.

표 3-9는 2013년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분으로 제 1 주성분과 제 2 주성분으로 나누었다. 첫 번째 주성분에는 마뇨산과 메틸마뇨산, N,N-디메틸포름아미드가 포함되었으며 두 번째 주성분에는 메틸마뇨산과 N,N-디메틸포름아미드가 포함되었다.

<표 3-9> 2013년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터

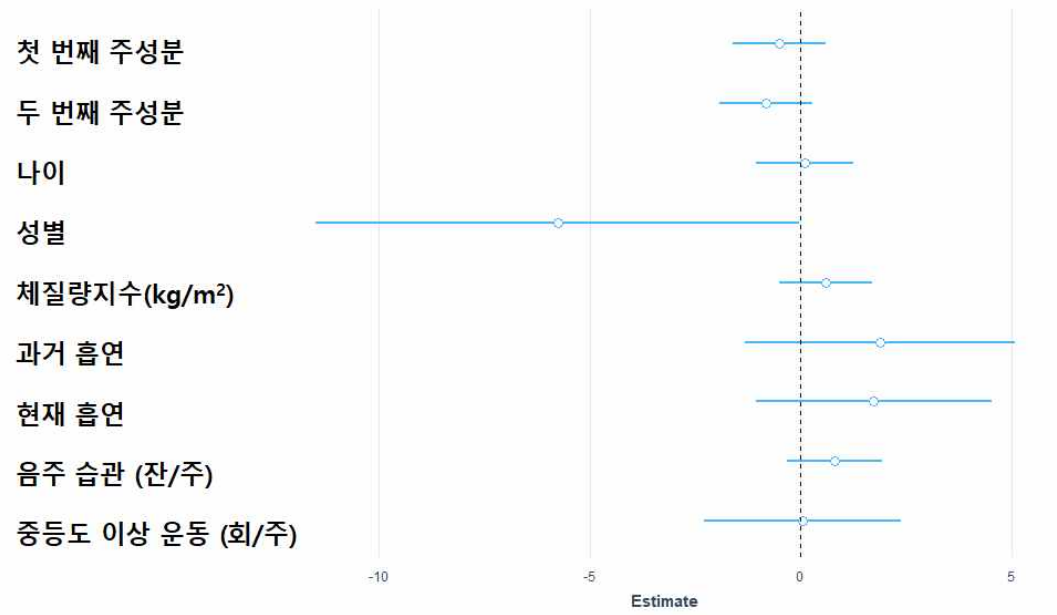
	제 1 주성분	제 2 주성분
마뇨산	0.852	
메틸마뇨산	0.543	-0.723
N,N-디메틸포름아미드	0.511	0.751

2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 첫 번째 주성분은 통계적으로 유의한 양의 관련성을 보였지만 두 번째 주성분에서는 유의한 연관성이 관찰되지 않았다(그림 3-16).



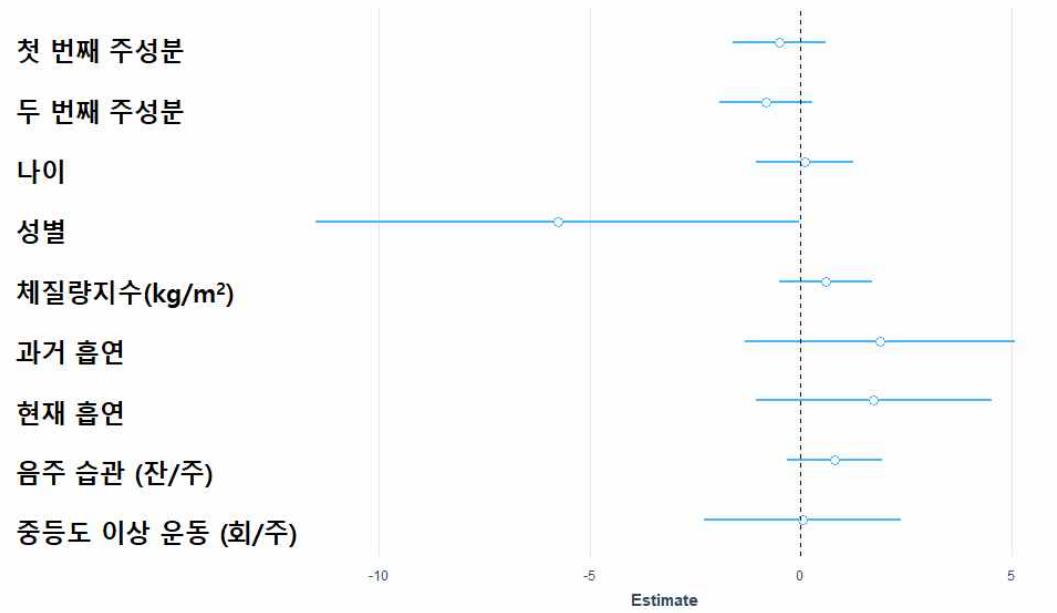
[그림 3-16] 2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 첫 번째 주성분과 두 번째 주성분 모두 통계적으로 유의한 연관성이 관찰되지 않았다(그림 3-17).



[그림 3-17] 2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 통계적으로 유의한 연관성을 보이지 않았지만 두 번째 주성분에서는 통계적으로 유의한 음의 연관성을 보였다(그림 3-18).



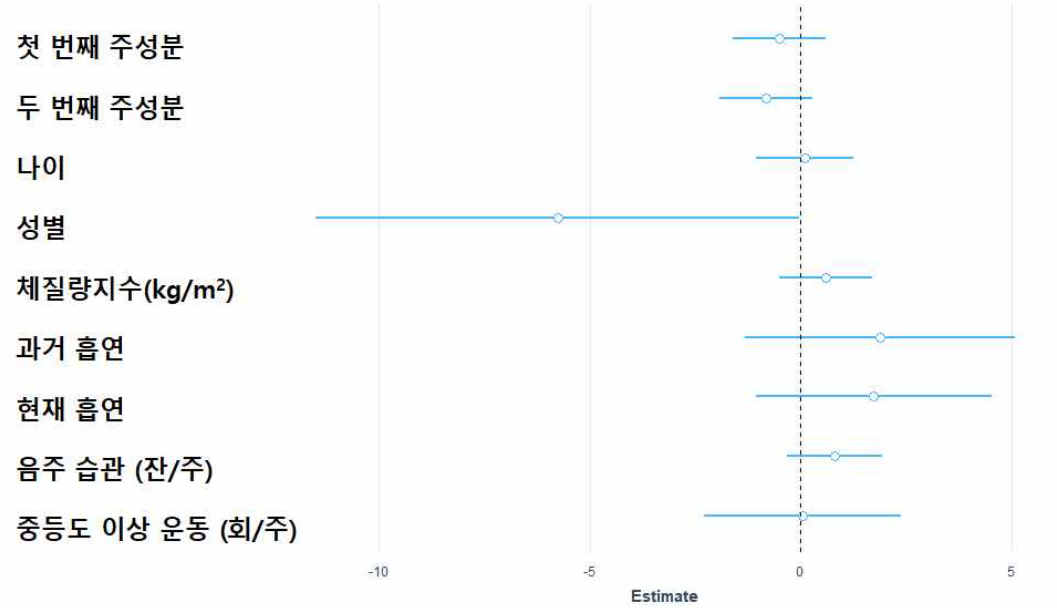
[그림 3-18] 2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

표 3-10은 2014년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분으로 제 1 주성분과 제 2 주성분으로 나누었다. 첫 번째 주성분에는 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었으며, 두 번째 주성분에는 마뇨산과 N,N-디메틸포름아미드가 포함되었다.

<표 3-10> 2014년자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터

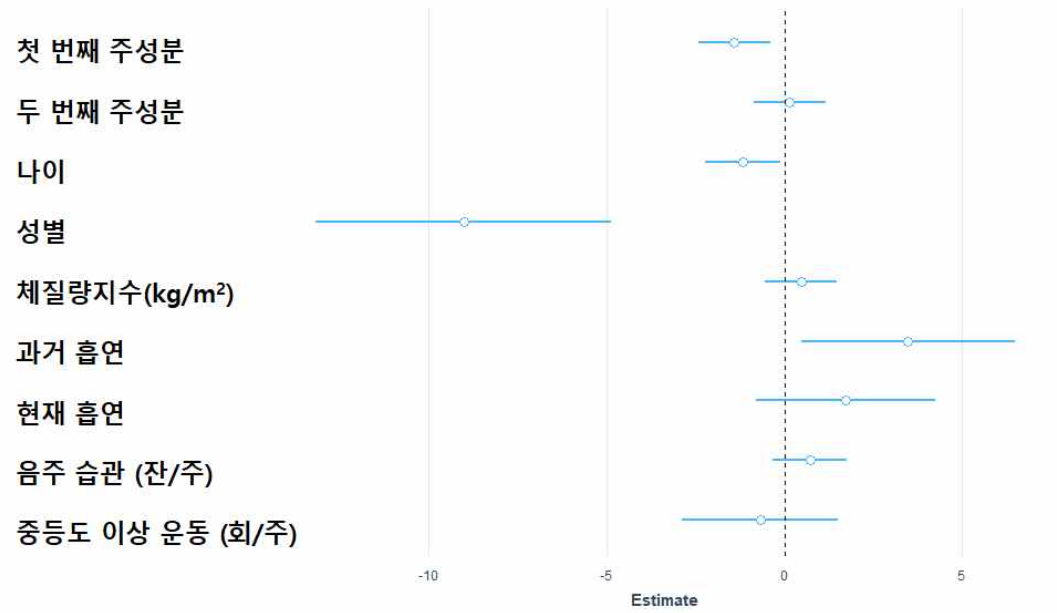
	제 1 주성분	제 2 주성분
마뇨산	0.800	0.168
메틸마뇨산	0.812	
N,N-디메틸포름아미드	-0.121	0.987

2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 통계적으로 유의한 음의 연관성을 보였지만 두 번째 주성분에서는 통계적으로 유의한 양의 연관성을 보였다(그림 3-19).



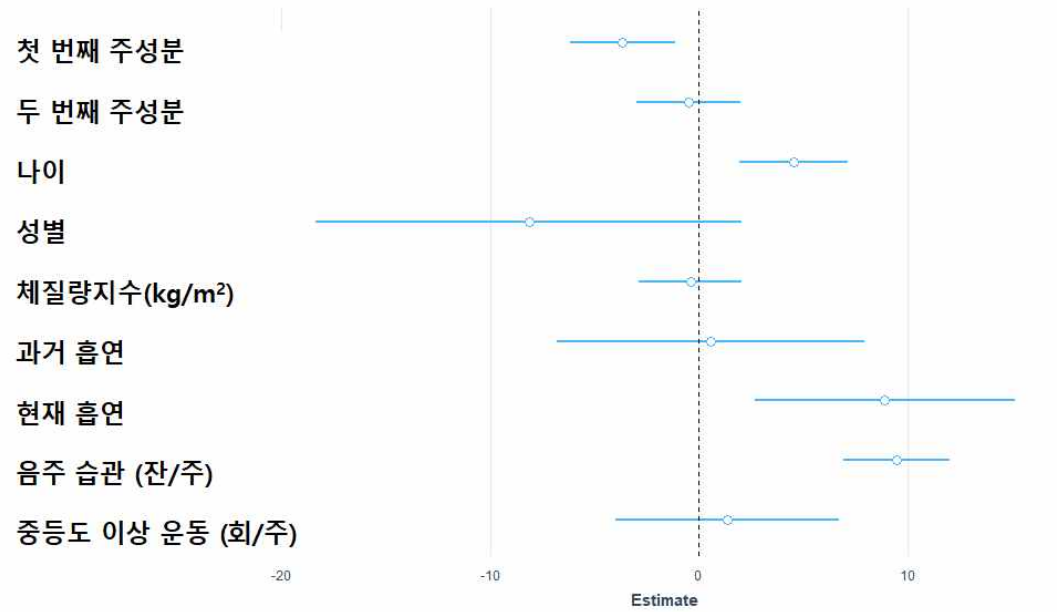
[그림 3-19] 2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 통계적으로 유의한 음의 연관성을 보였지만 두 번째 주성분에서는 유의한 연관성을 보이지 않았다 (그림 3-20).



[그림 3-20] 2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 통계적으로 유의한 음의 연관성을 보였지만 두 번째 주성분에서는 유의한 연관성을 보이지 않았다(그림 3-21).



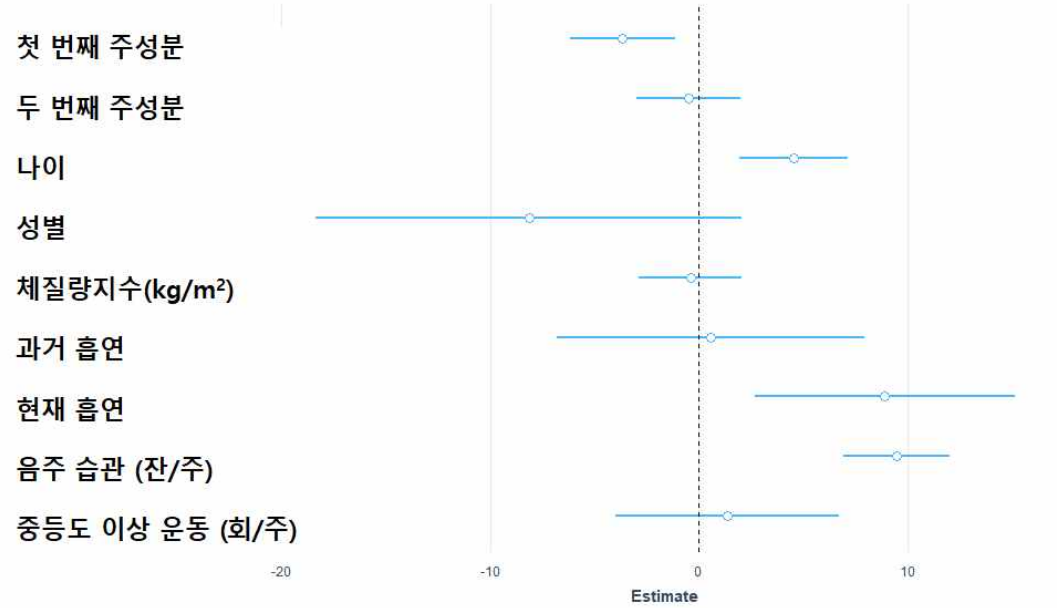
[그림 3-21] 2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

표 3-11은 2015년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분으로 제 1 주성분과 제 2 주성분으로 나누었다. 첫 번째 주성분에는 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었으며, 두 번째 주성분에도 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-11> 2015년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터

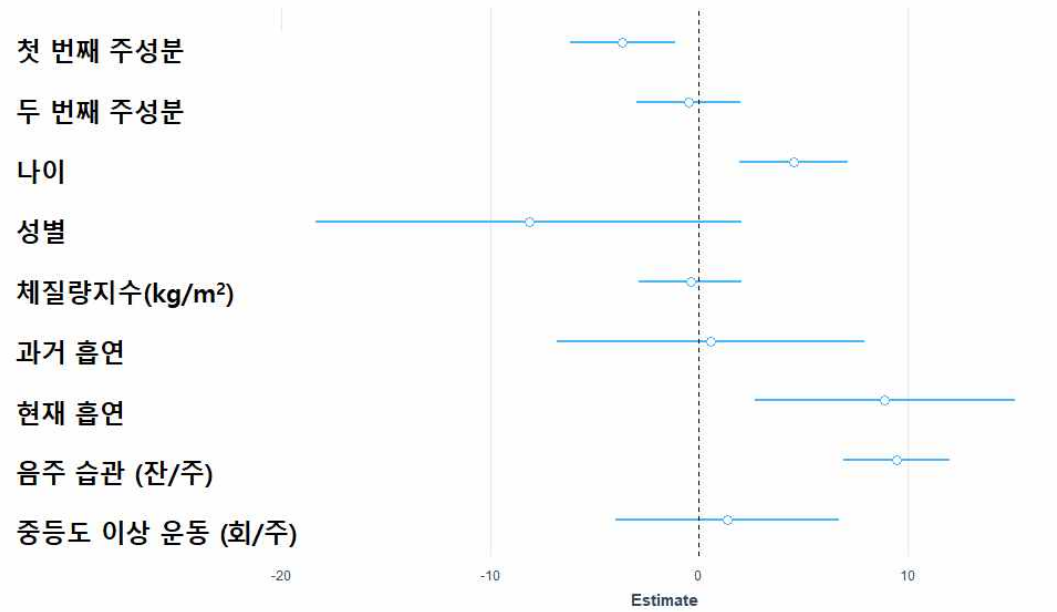
	제 1 주성분	제 2 주성분
마뇨산	0.750	-0.138
메틸마뇨산	0.732	-0.261
N,N-디메틸포름아미드	0.312	0.947

2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 통계적으로 유의한 음의 연관성을 보였지만 두 번째 주성분에서는 유의한 연관성을 보이지 않았다(그림 3-22).



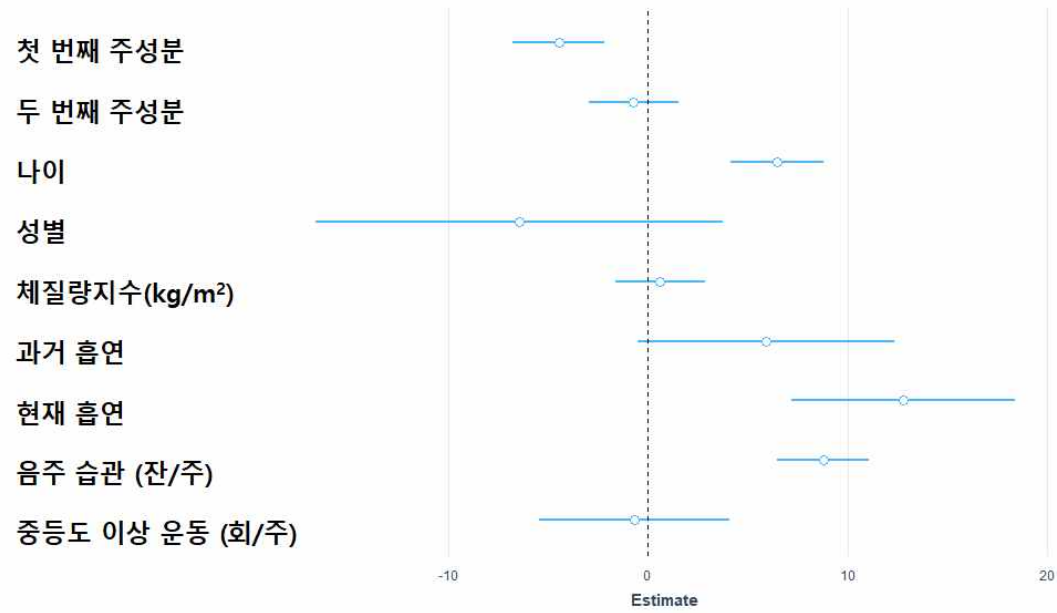
[그림 3-22] 2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 통계적으로 유의한 음의 연관성을 보였지만 두 번째 주성분에서는 유의한 연관성을 보이지 않았다 (그림 3-23).



[그림 3-23] 2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 통계적으로 유의한 음의 연관성을 보였지만 두 번째 주성분에서는 유의한 연관성을 보이지 않았다(그림 3-24).



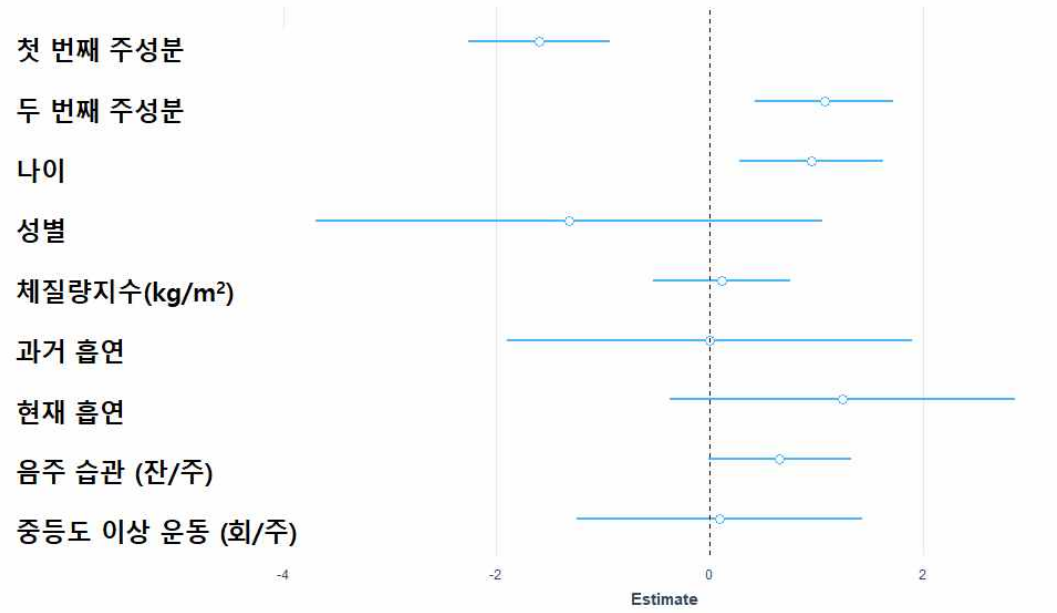
[그림 3-24] 2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

표 3-12는 2016년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분으로 제 1 주성분과 제 2 주성분으로 나누었다. 첫 번째 주성분에는 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었으며, 두 번째 주성분에는 메틸마뇨산과 N,N-디메틸포름아미드가 포함되었다.

<표 3-12> 2016년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터

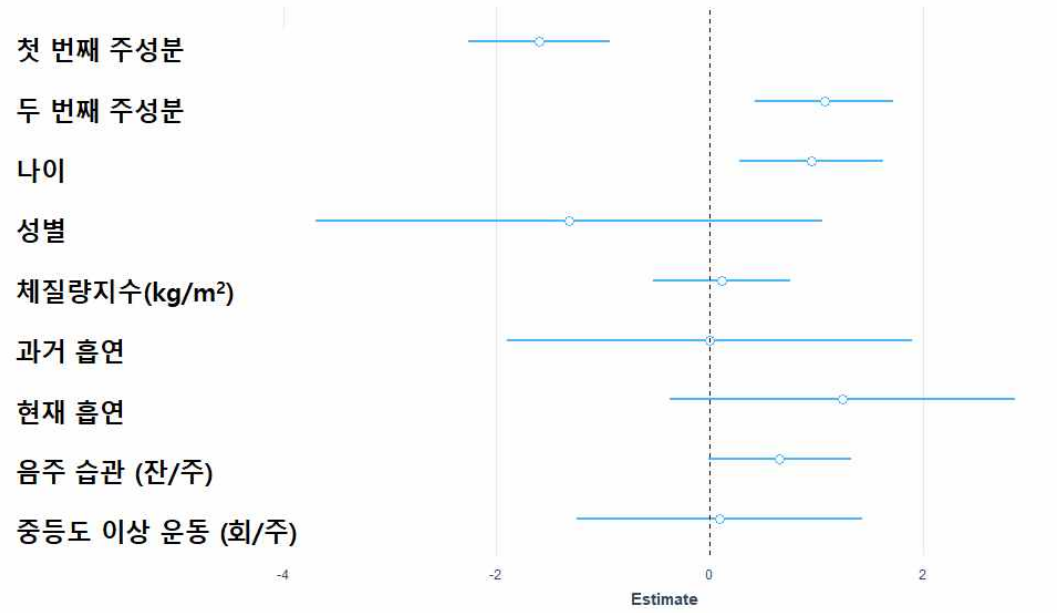
	제 1 주성분	제 2 주성분
마뇨산	0.852	
메틸마뇨산	0.543	-0.723
N,N-디메틸포름아미드	0.511	0.751

2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 통계적으로 유의한 음의 연관성을 보였지만 두 번째 주성분에서는 통계적으로 유의한 양의 연관성을 보였다(그림 3-25).



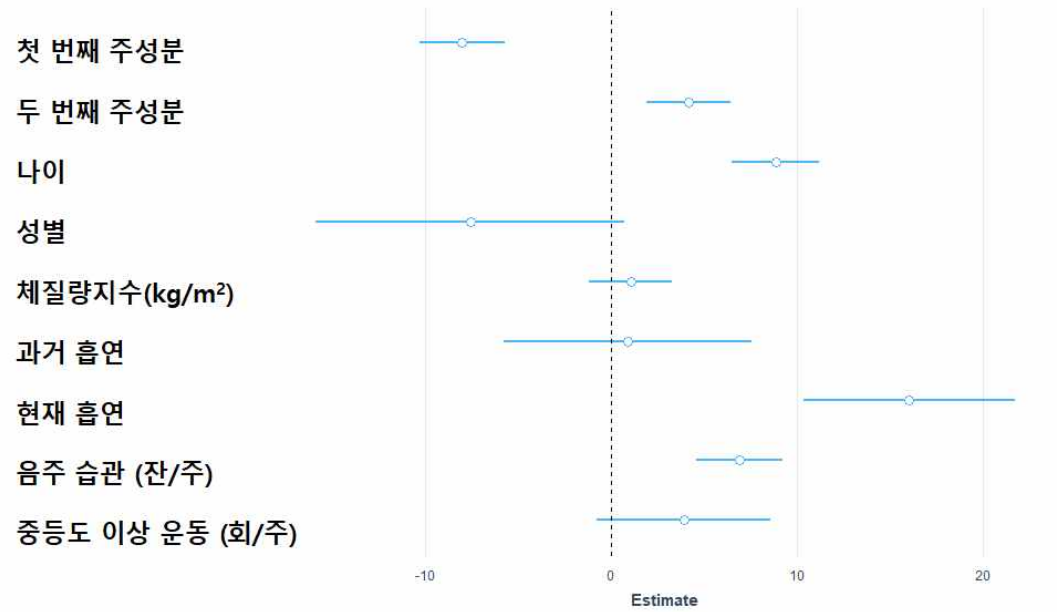
[그림 3-25] 2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 첫 번째 주성분과 두 번째 주성분 모두 통계적으로 유의한 연관성을 보이지 않았다(그림 3-27).



[그림 3-27] 2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 통계적으로 유의한 음의 연관성을 보였지만 두 번째 주성분에서는 통계적으로 유의한 양의 연관성을 보였다(그림 3-28).



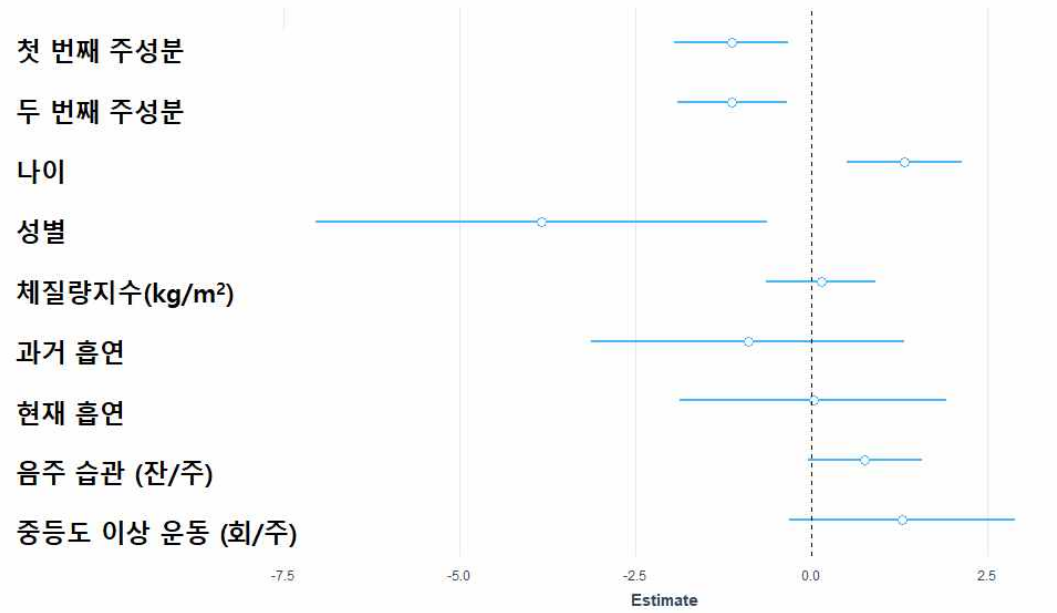
[그림 3-28] 2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

표 3-13은 2017년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분으로 제 1 주성분과 제 2 주성분으로 나누었다. 첫 번째 주성분에는 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었으며, 두 번째 주성분에는 마뇨산과 N,N-디메틸포름아미드가 포함되었다.

<표 3-13> 2017년 자료에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터

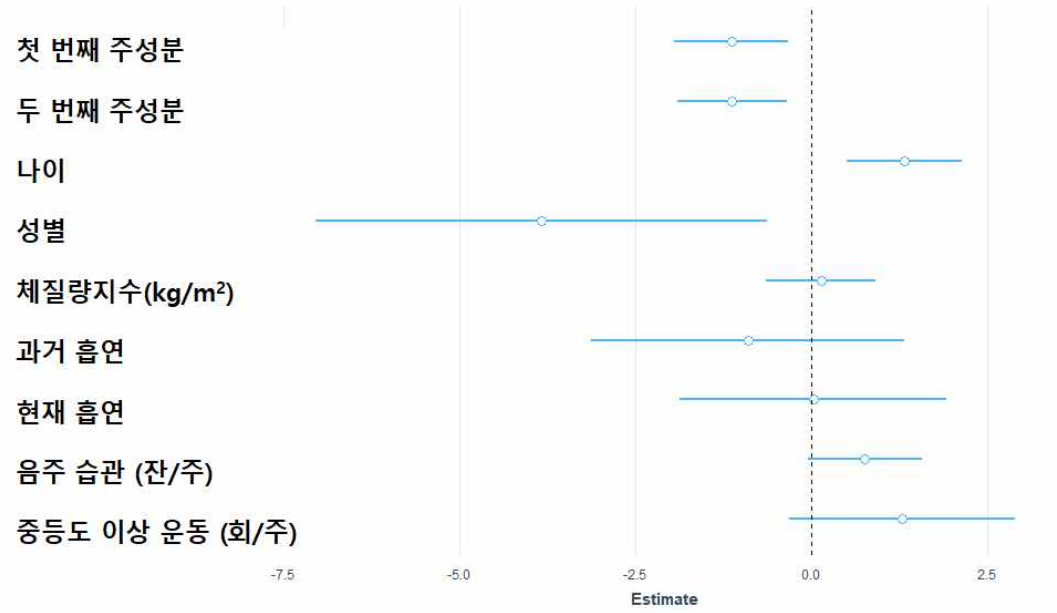
	제 1 주성분	제 2 주성분
마뇨산	0.492	0.746
메틸마뇨산	0.807	
N,N-디메틸포름아미드	-0.543	0.696

2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 첫 번째 주성분과 두 번째 주성분 모두에서 통계적으로 유의한 음의 연관성을 보였다(그림 3-29).



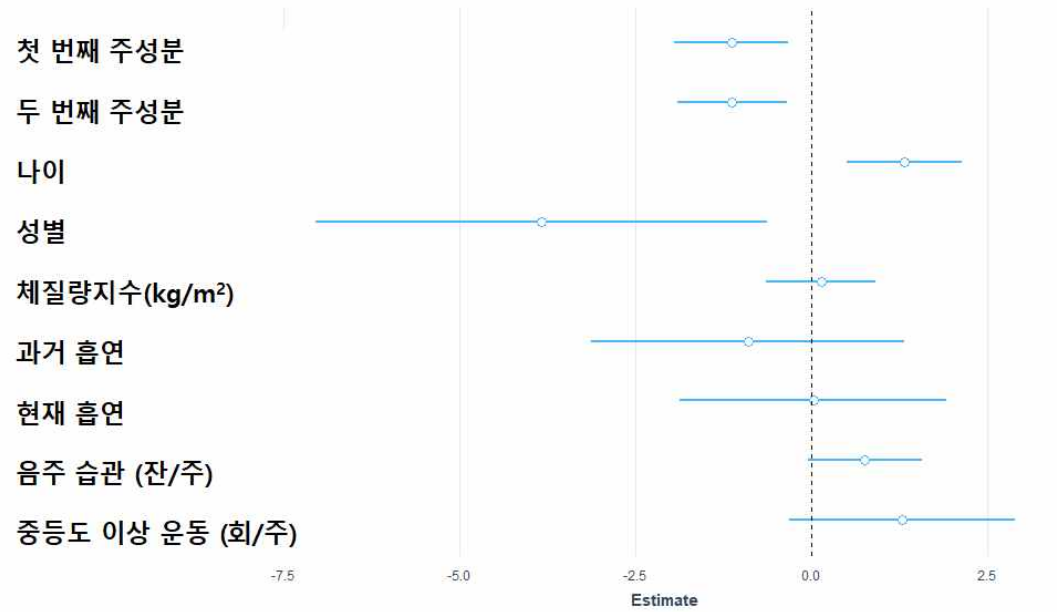
[그림 3-29] 2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 첫 번째 주성분과 두 번째 주성분 모두에서 통계적으로 유의한 음의 연관성을 보였다(그림 3-30).



[그림 3-30] 2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 유의한 연관성이 관찰되지 않았으나 두 번째 주성분에서는 통계적으로 유의한 음의 연관성을 보였다(그림 3-31).



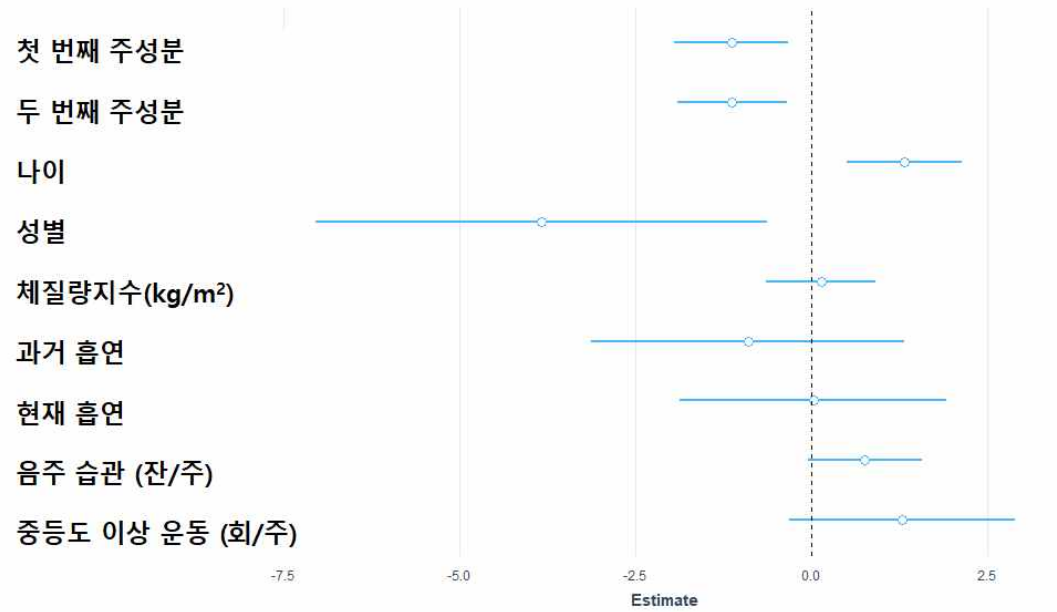
[그림 3-31] 2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

표 3-14은 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 과거의 자료를 남기는 방법으로 자료를 가공하였고, 세 가지 노출변수의 주성분분석 후 산출된 구성성분으로 제 1 주성분과 제 2 주성분으로 나누었다. 첫 번째 주성분에는 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었으며, 두 번째 주성분에는 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-14> 2013-2017년 자료(최근 자료 중복제거)에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터

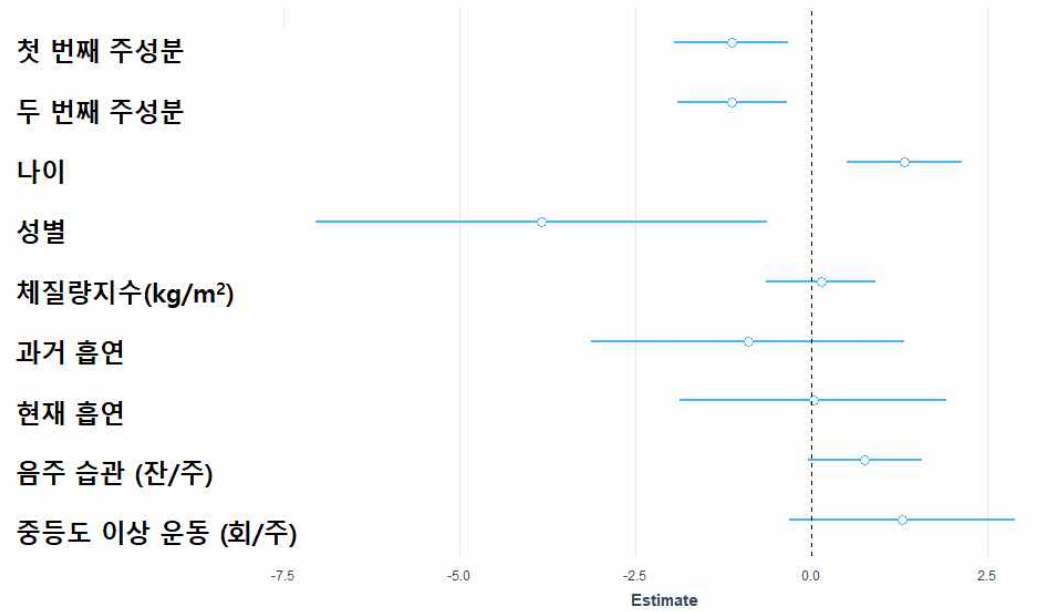
	제 1 주성분	제 2 주성분
마뇨산	0.750	-0.137
메틸마뇨산	0.706	-0.385
N,N-디메틸포름아미드	0.419	0.894

2013-2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 유의한 음의 연관성이 관찰되었으나 두 번째 주성분에서는 통계적으로 유의한 연관성을 보이지 않았다(그림 3-32).



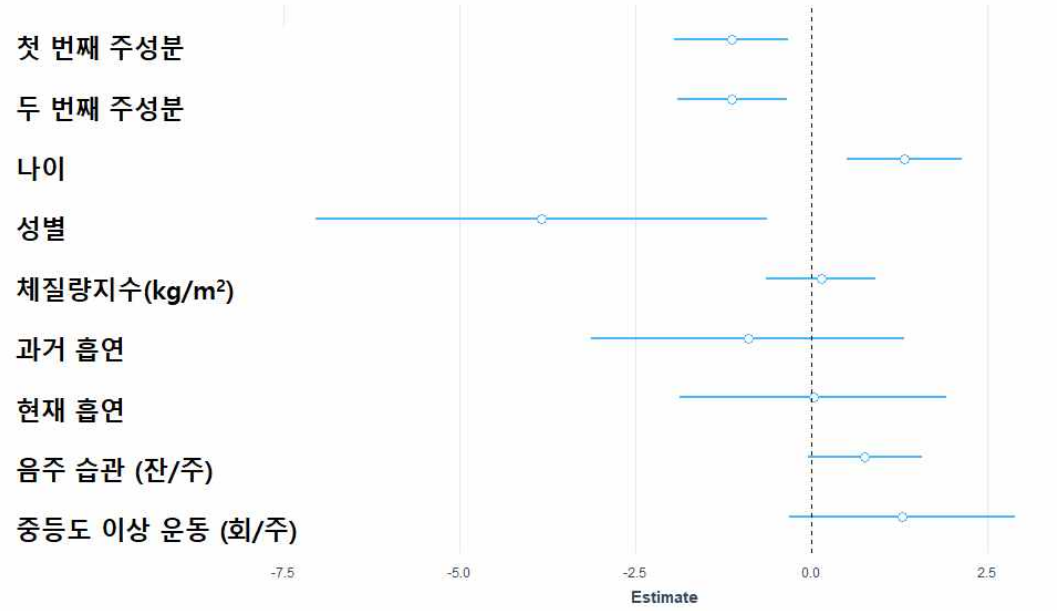
[그림 3-32] 2013-2017년 자료(최근 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

2013-2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 유의한 음의 연관성이 관찰되었으나 두 번째 주성분에서는 통계적으로 유의한 연관성을 보이지 않았다(그림 3-34).



[그림 3-34] 2013-2017년 자료(최근 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

2013-2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 첫 번째 주성분과 두 번째 주성분 모두 유의한 음의 연관성이 관찰되었다(그림 3-35).



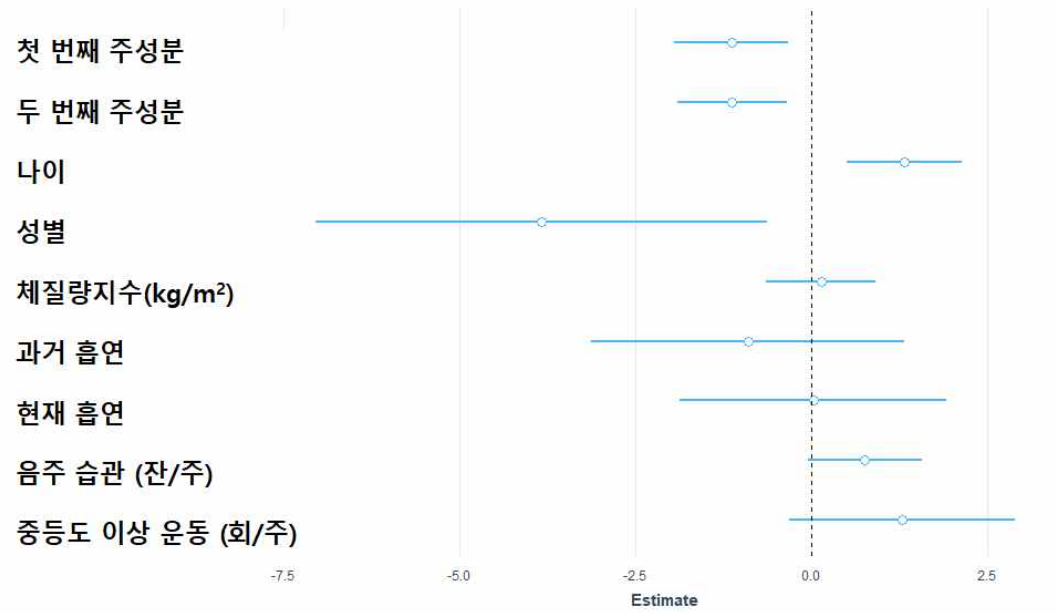
[그림 3-35] 2013-2017년 자료(최근 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

표 3-15는 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 과거 자료를 중복 제거하는 방법으로 자료를 가공하였고, 세 가지 노출변수의 주성분분석 후 산출된 구성성분으로 제 1 주성분과 제 2 주성분으로 나누었다. 첫 번째 주성분에는 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었으며, 두 번째 주성분에는 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-15> 2013-2017년 자료(과거 자료 중복제거)에서 세 가지 노출변수의 주성분분석 후 산출된 구성성분의 로딩벡터

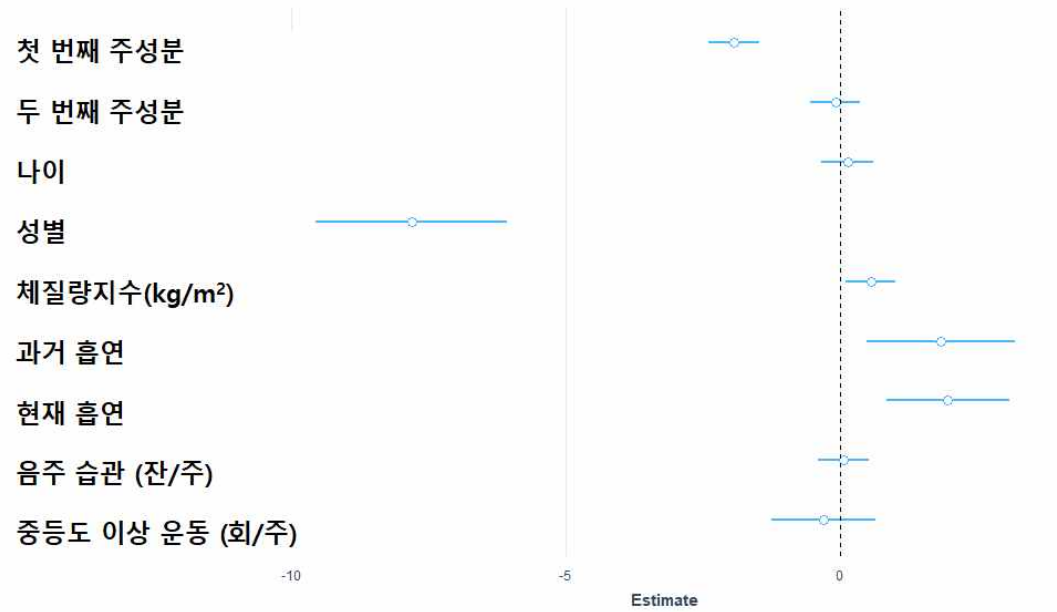
	제 1 주성분	제 2 주성분
마뇨산	0.769	
메틸마뇨산	0.706	-0.402
N,N-디메틸포름아미드	0.330	0.913

2013-2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 유의한 음의 연관성이 관찰되었지만 두 번째 주성분에선 유의한 연관성이 관찰되지 않았다(그림 3-36).



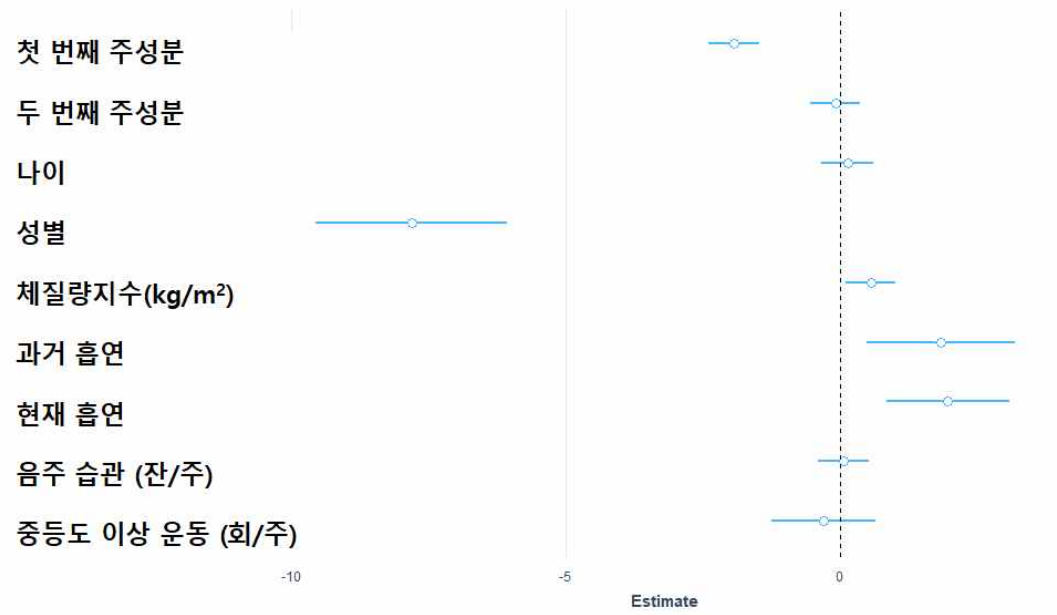
[그림 3-36] 2013-2017년 자료(과거 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

2013-2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 첫 번째 주성분에서는 유의한 음의 연관성이 관찰되었지만 두 번째 주성분에선 유의한 연관성이 관찰되지 않았다 (그림 3-37).



[그림 3-37] 2013-2017년 자료(과거자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

2013-2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 주성분 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 첫 번째 주성분과 두 번째 주성분 모두에서 유의한 음의 연관성이 관찰되었다(그림 3-38).



[그림 3-38] 2013-2017년 자료(과거자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

R을 이용한 주성분분석 및 결과 요약을 위한 그림을 그리는 R 코드는 다음과 같다.

```
#데이터 읽기#
data<-read.csv("D:/연구/복합노출 연구/data/three2013y1.csv")

data<-data[,-1] #id column 제거
head(data)

#주성분분석#
#install.packages("psych")
library(psych)
pca_mod<-principal(data[,1:3], 2, rotate = "none", scale=TRUE) #exposures data만 이용
pca_mod$loadings

head(pca_mod$scores) #score
summary(pca_mod$scores)
data2<-data.frame(data,pca_mod$scores)

#주성분을 노출로 다중선형회귀분석#
pcr_mod<-lm(y1~PC1+PC2+age+sex+bmi+smoking1+smoking2+drinking+exercise,data=data2)
summary(pcr_mod)

#그림 그리기#
#install.packages("jtools")
library("jtools")

plot_summs(pcr_mod, model.names=c("2013 (N=615)"),
           legend.title = "Data",
           coefs=c("1st principal component"="PC1", "2nd principal component"="PC2"),
           scale=TRUE)

plot_summs(pcr_mod, model.names=c("2013 (N=615)"),legend.title = "Data",
           coefs=c("1st principal component"="PC1", "2nd principal component"="PC2",
          "Age"="age", "Sex"="sex","BMI"="bmi", "Past smoker"="smoking1",
          "Current smoker"="smoking2", "Alcohol drinking habits"="drinking",
          "Physical activity"="exercise"), scale=TRUE)
```

4) 부분최소제곱 회귀(Partial least square regression)를 활용한 복합노출과 간기능검사 결과 간의 관련성 분석

부분최소제곱법은 노출과 건강 결과 간의 공분산을 최대화하는 성분들을 형성하여 잠재 변수를 구성하는 감독된 차원 축소 방법으로, 성분을 구성할 때 건강 결과와의 관련성을 반영하지 못하는 주성분분석의 단점을 보완한다.

부분최소제곱법은 차원 축소법과 다중 선형 회귀분석을 결합한 방법으로 고차원의 설명 변수를 갖는 예측 모델을 구성하는 유용한 방법이다(Geladi P 외, 1986; Wold S 외, 2001). 독성학적 측면에서는 부분최소제곱법과 주성분분석 등 차원 축소를 이용한 방법들이 환경유해물질 누적효과(cumulative effect)를 평가할 수 있는 것으로 알려져 있다(Braun JM 외, 2016; Maresca MM 외, 2016; Czarnota J 외, 2015). 부분최소제곱법은 설명 변수들 간의 상관관계 뿐만 아니라 설명 변수와 결과 변수와의 연관성까지 고려할 수 있기 때문에 주성분분석에 비하여 더 적은 수의 성분이 사용된다는 장점이 있다. 부분최소제곱법은 최적의 구성요소 개수를 교차검증을 통해 오분류율을 최소로 하거나 이미 정해진 임계 값을 넘어서지 않는 구성요소의 개수를 선택할 수 있으나, 본 연구에서는 가장 큰 설명력을 갖는 첫 번째 성분만을 선택하고, 첫 번째 성분에 의한 산출된 점수를 주 노출로 하여 다중 선형회귀분석을 수행하였다(이슬비, 2018).

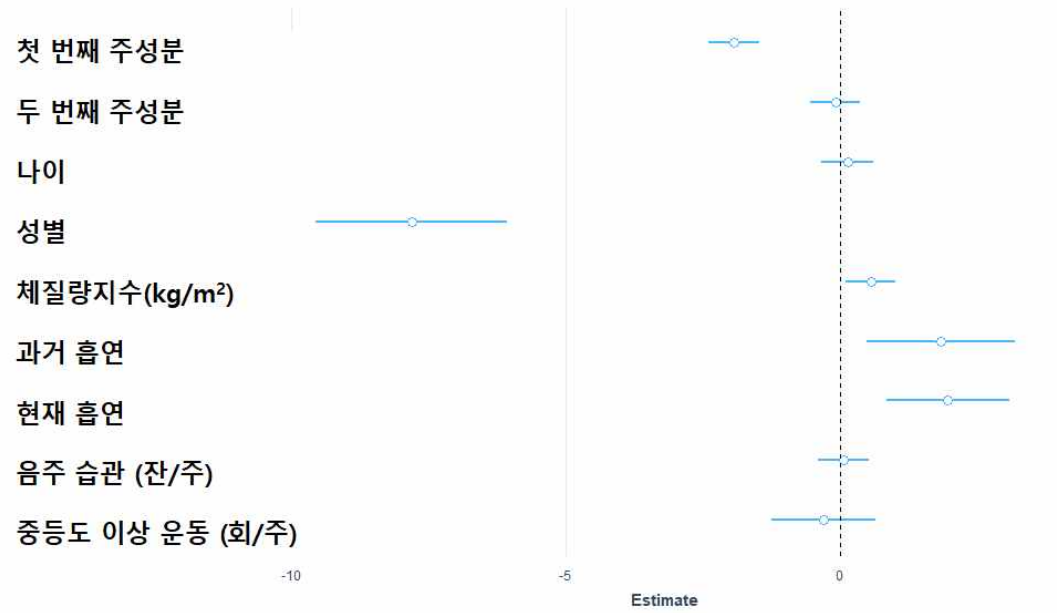
부분최소제곱법 회귀 분석을 위해 plsr R package를 사용하였다.

표 3-16은 2013년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-16> 2013년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	0.565
메틸마뇨산	0.410
N,N-디메틸포름아미드	0.750

2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-39).



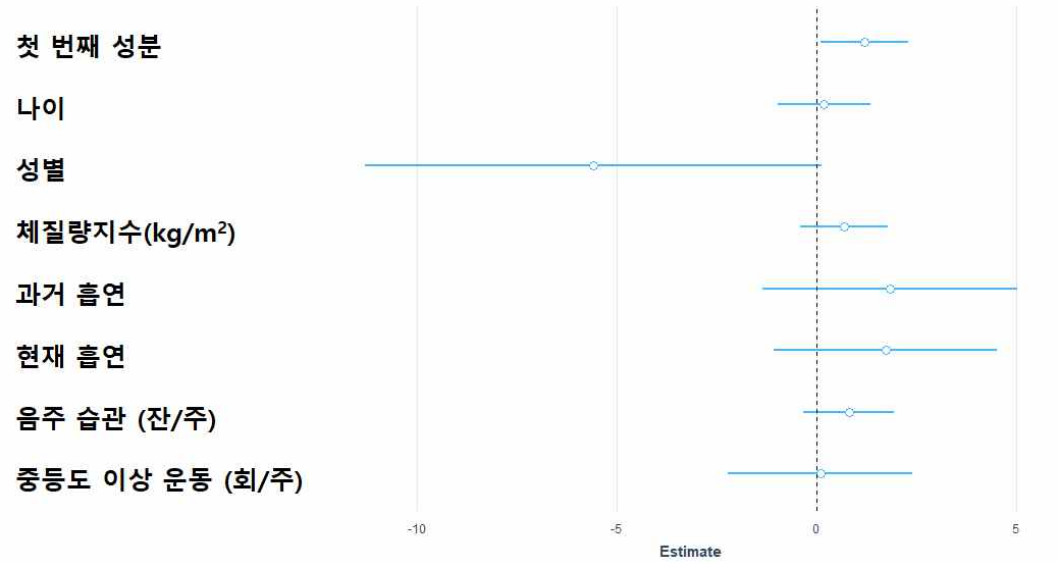
[그림 3-39] 2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

표 3-17은 2013년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-17> 2013년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.627
메틸마뇨산	0.407
N,N-디메틸포름아미드	-0.696

2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-40).



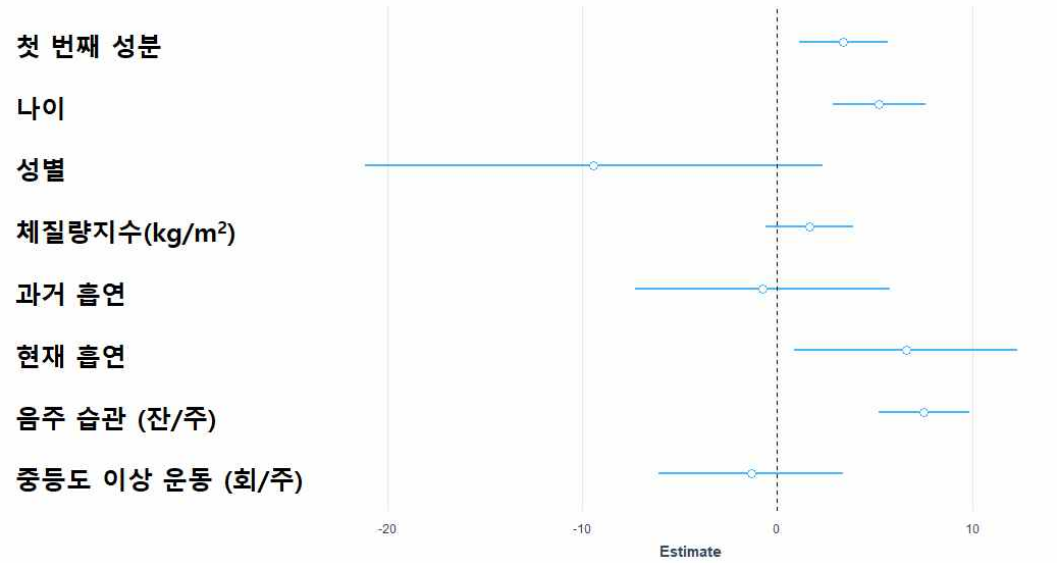
[그림 3-40] 2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

표 3-18은 2013년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-18> 2013년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.384
메틸마뇨산	0.797
N,N-디메틸포름아미드	-0.524

2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-41).



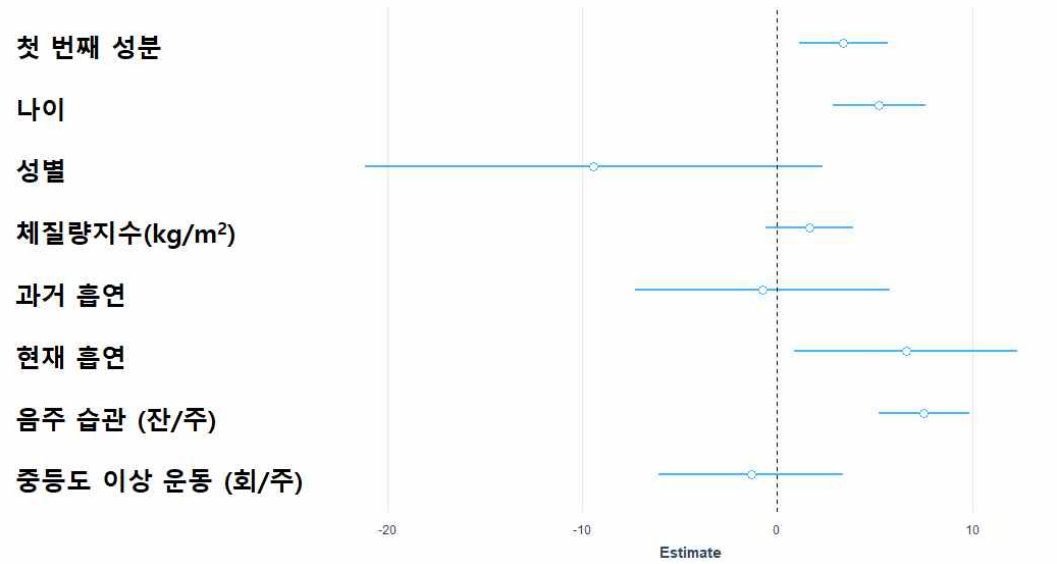
[그림 3-41] 2013년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

표 3-19는 2014년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-19> 2014년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.631
메틸마뇨산	-0.483
N,N-디메틸포름아미드	0.633

2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-42).



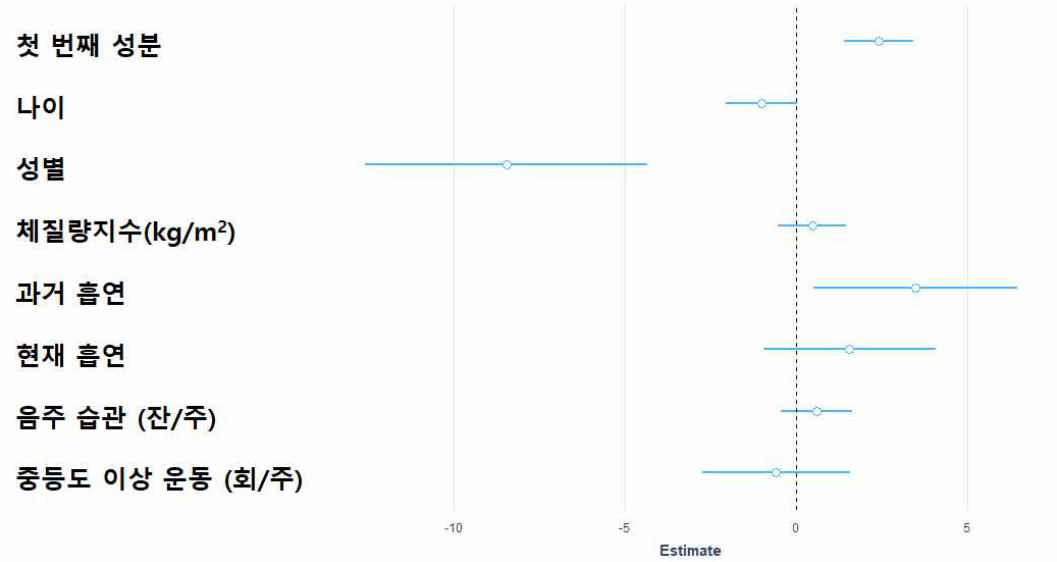
[그림 3-42] 2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

표 3-20은 2014년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-20> 2014년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.965
메틸마뇨산	-0.350
N,N-디메틸포름아미드	0.201

2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-43).



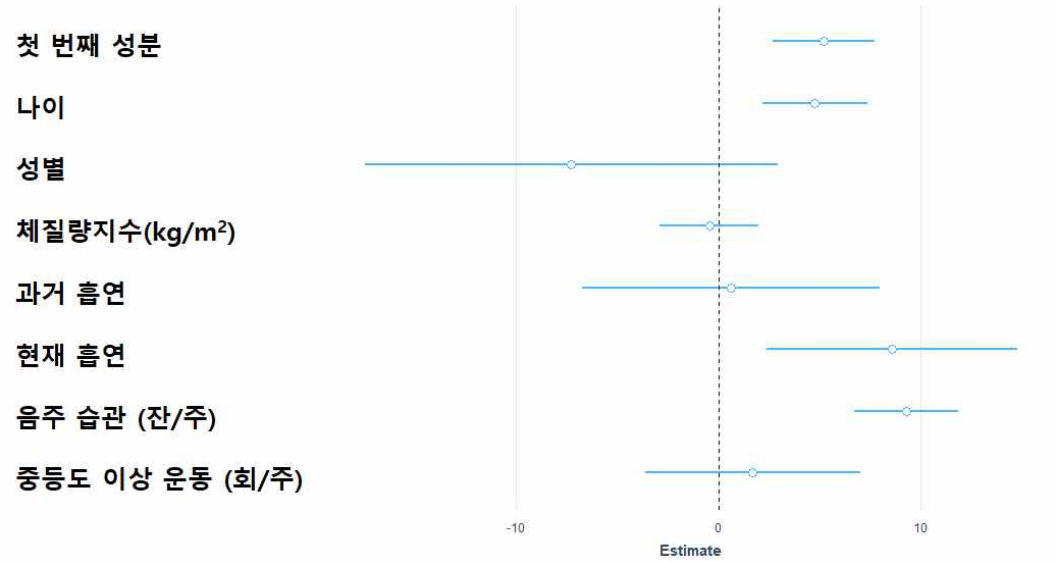
[그림 3-43] 2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

표 3-21은 2014년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-21> 2014년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.929
메틸마뇨산	-0.433
N,N-디메틸포름아미드	0.152

2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-44).



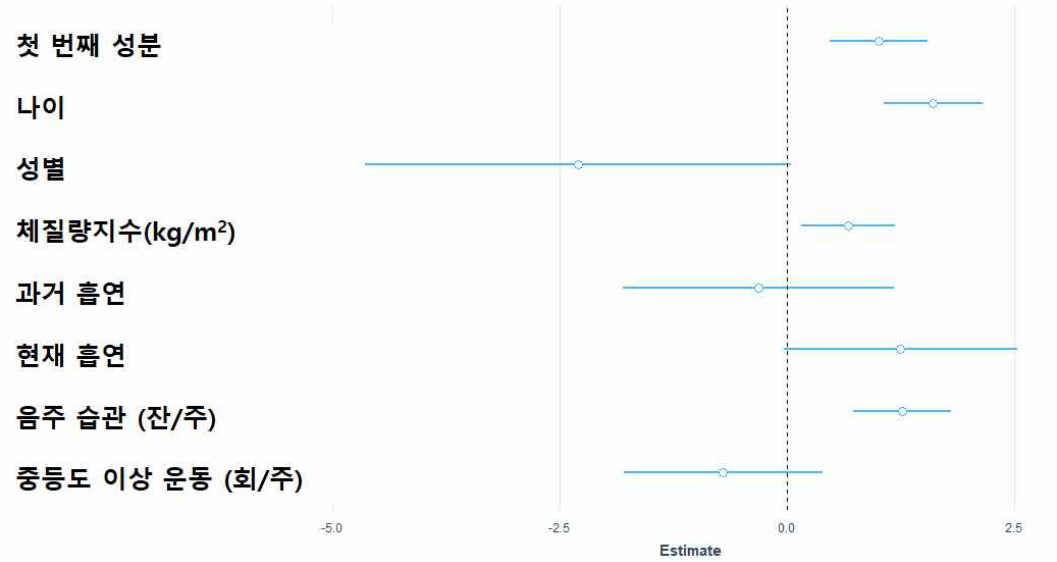
[그림 3-44] 2014년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

표 3-22는 2015년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산과 N,N-디메틸포름아미드가 포함되었다.

<표 3-22> 2015년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.936
메틸마뇨산	
N,N-디메틸포름아미드	-0.395

2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-45).



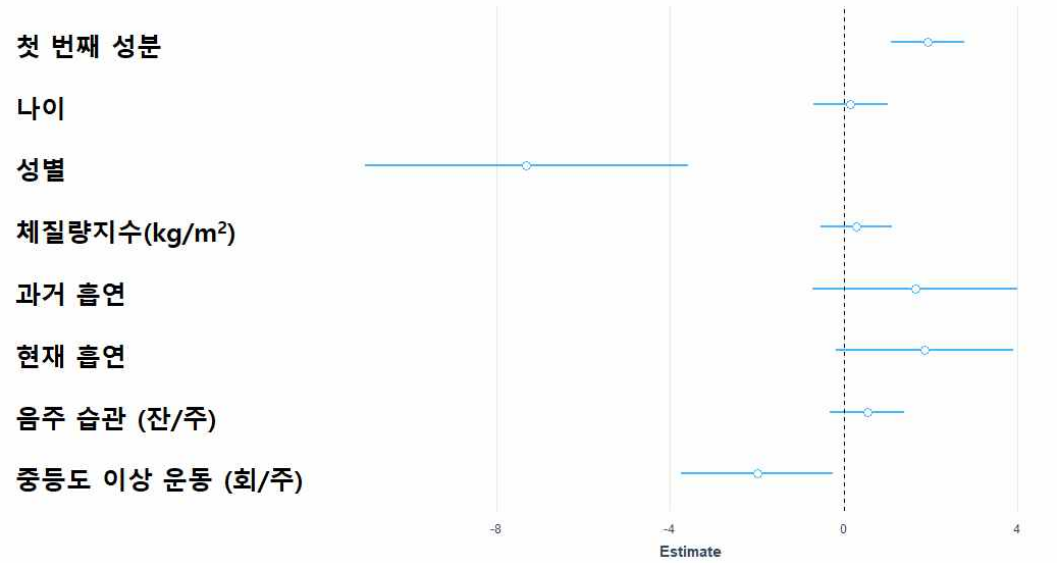
[그림 3-45] 2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

표 3-23은 2015년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산과 N,N-디메틸포름아미드가 포함되었다.

<표 3-23> 2015년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.974
메틸마뇨산	
N,N-디메틸포름아미드	-0.289

2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-46).



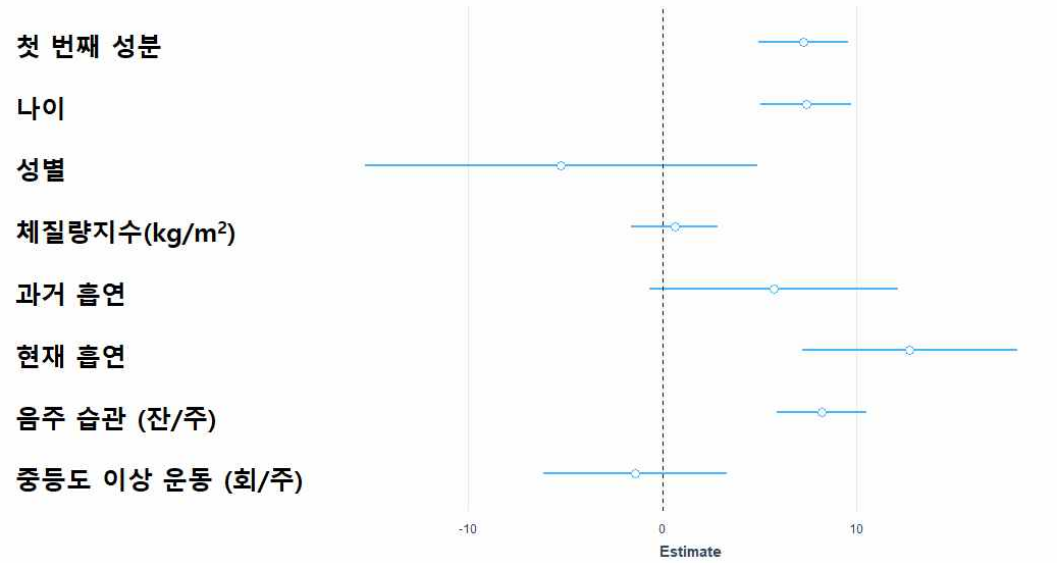
[그림 3-46] 2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

표 3-24는 2015년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산과 N,N-디메틸포름아미드가 포함되었다.

<표 3-24> 2015년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.960
메틸마뇨산	
N,N-디메틸포름아미드	-0.325

2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-47).



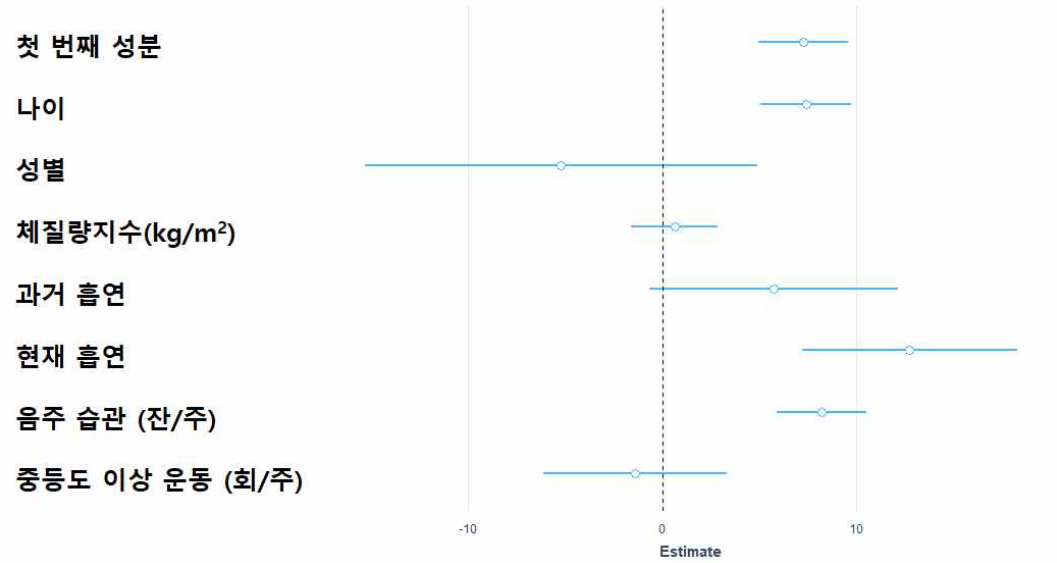
[그림 3-47] 2015년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

표 3-25는 2016년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-25> 2016년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.791
메틸마뇨산	-0.552
N,N-디메틸포름아미드	-0.321

2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-48).



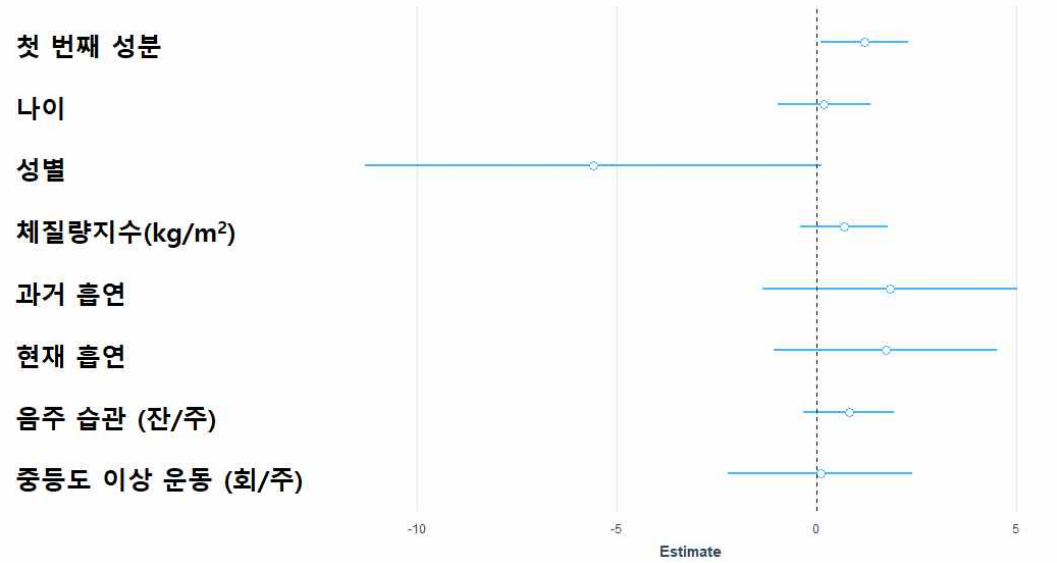
[그림 3-48] 2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

표 3-26은 2016년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-26> 2016년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.627
메틸마뇨산	0.407
N,N-디메틸포름아미드	-0.696

2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-49).



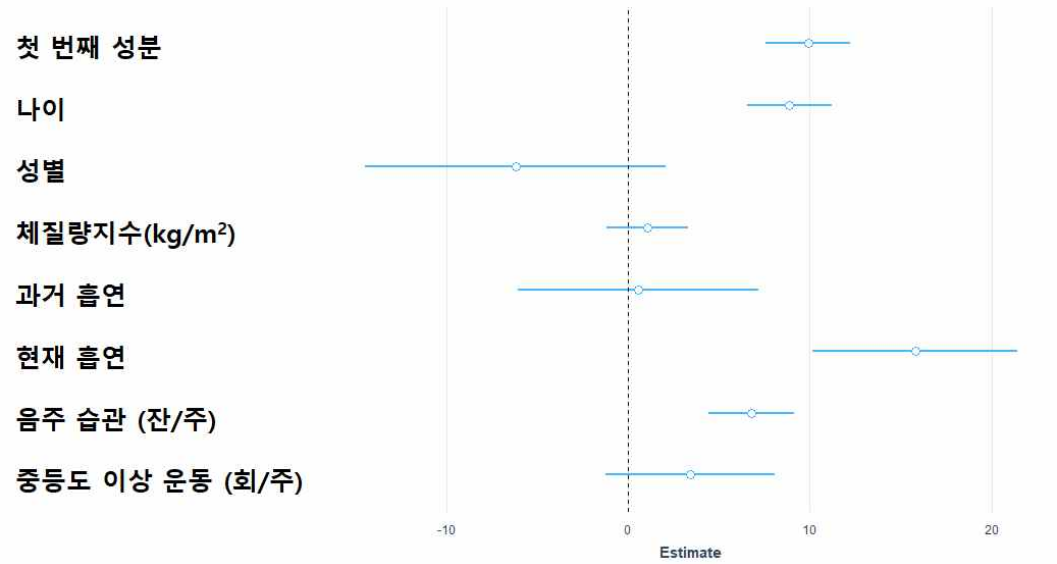
[그림 3-49] 2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

표 3-27은 2016년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-27> 2016년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.841
메틸마뇨산	-0.387
N,N-디메틸포름아미드	-0.432

2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-50).



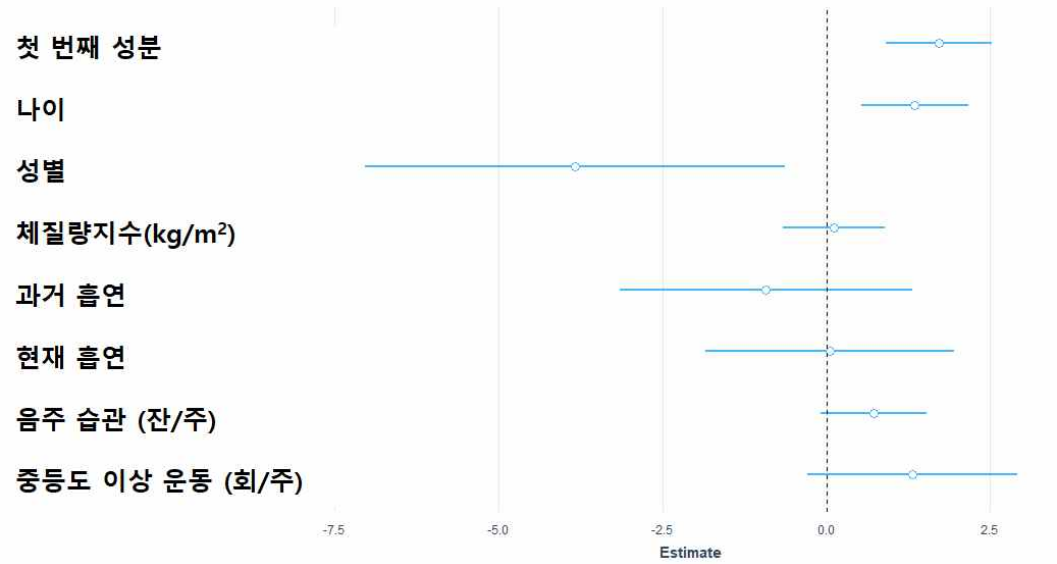
[그림 3-50] 2016년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

표 3-28은 2017년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산과 메틸마뇨산이 포함되었다.

<표 3-28> 2017년 자료에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.934
메틸마뇨산	-0.374
N,N-디메틸포름아미드	

2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-51).



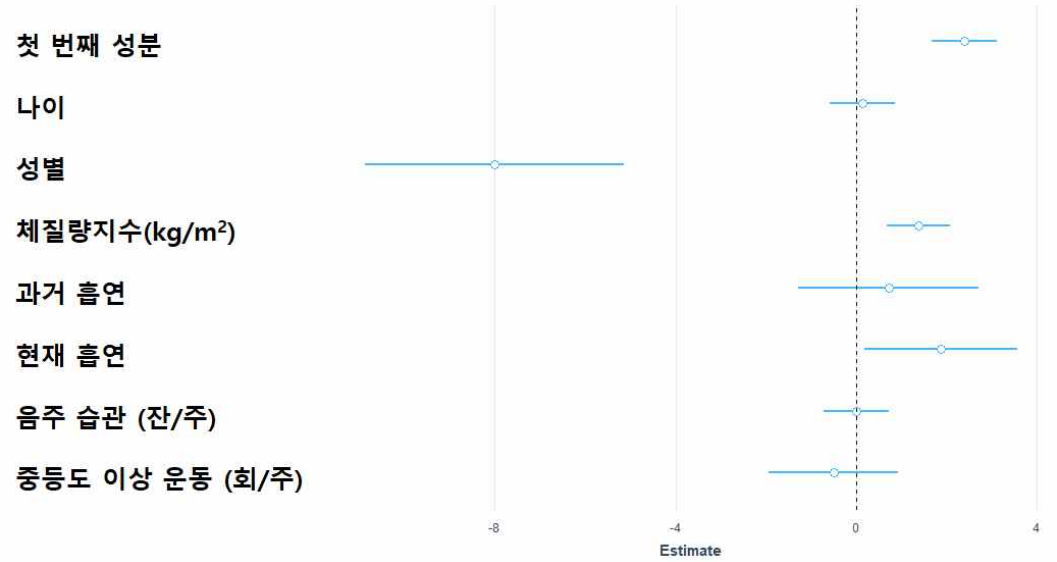
[그림 3-51] 2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

표 3-29는 2017년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-29> 2017년 자료에서 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.906
메틸마뇨산	-0.418
N,N-디메틸포름아미드	-0.106

2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-52).



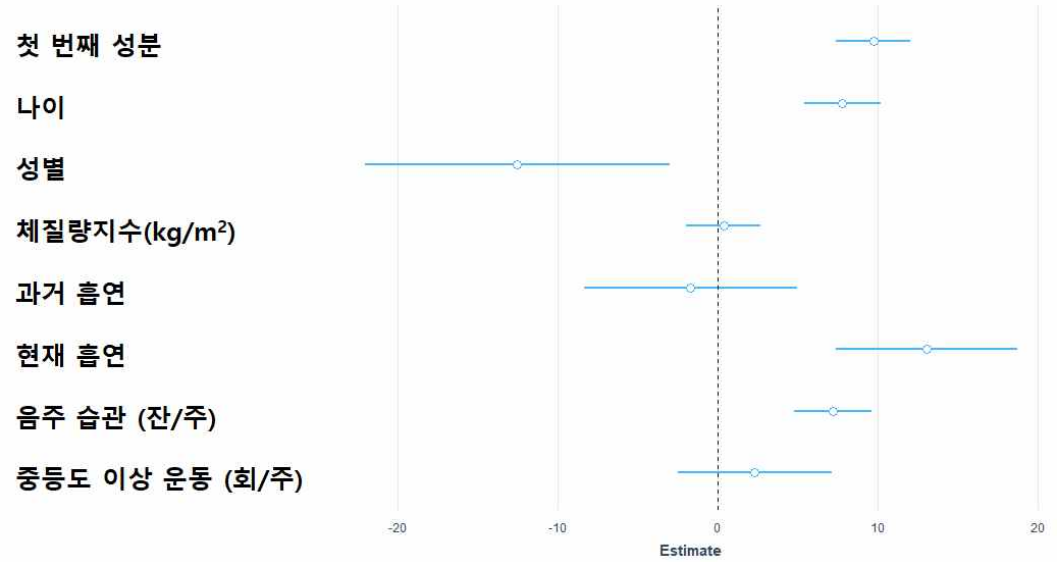
[그림 3-52] 2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

표 3-30은 2017년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산과 N,N-디메틸포름아미드가 포함되었다.

<표 3-30> 2017년 자료에서 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.884
메틸마뇨산	
N,N-디메틸포름아미드	-0.470

2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-53).



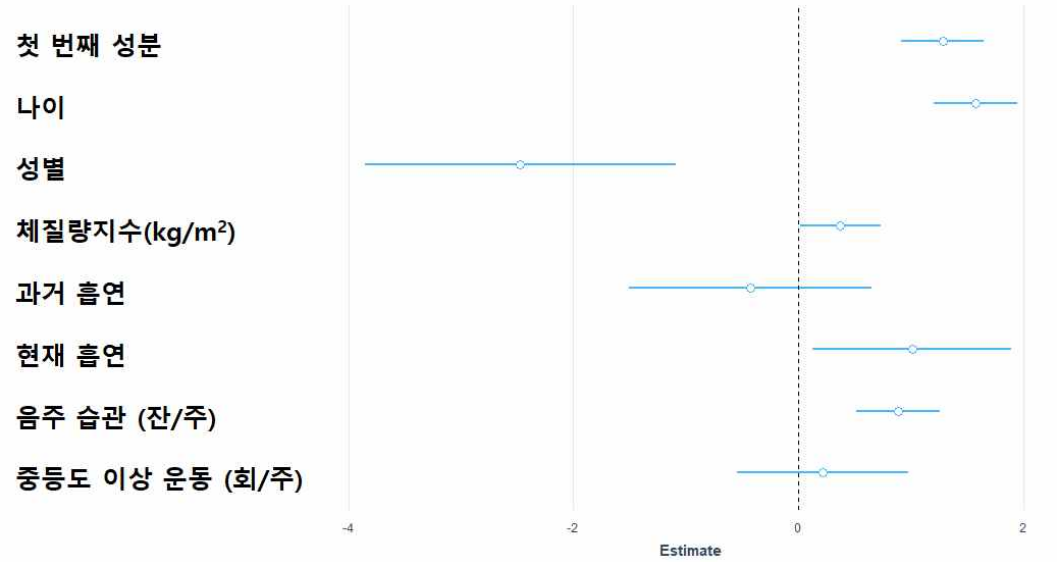
[그림 3-53] 2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

표 3-31은 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 과거의 자료를 남기는 방법으로 자료를 가공하였고, 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-31> 2013-2017년 자료(최근 자료 중복제거)에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.948
메틸마뇨산	-0.334
N,N-디메틸포름아미드	-0.165

2013-2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-54).



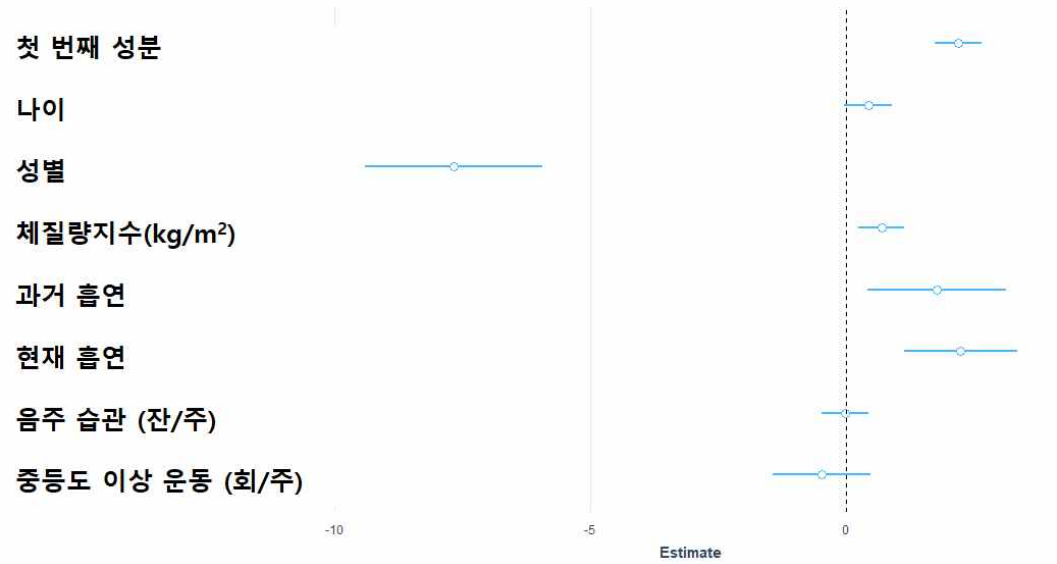
[그림 3-54] 2013-2017년 자료(최근 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

표 3-32는 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 과거의 자료를 남기는 방법으로 자료를 가공하였고, 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-32> 2013-2017년 자료(최근 자료 중복제거)에서 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.932
메틸마뇨산	-0.309
N,N-디메틸포름아미드	-0.269

2013-2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-55).



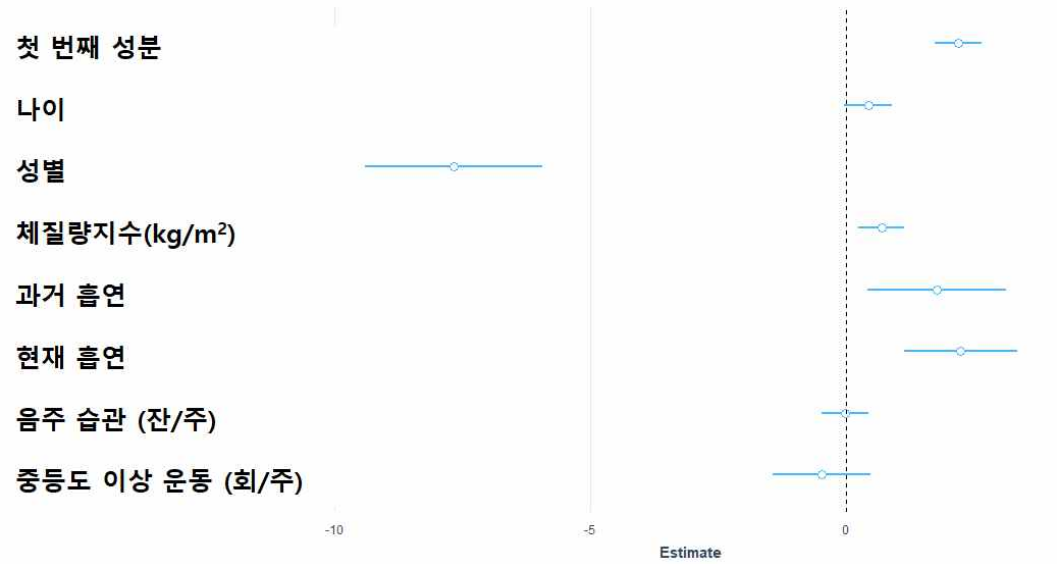
[그림 3-55] 2013-2017년 자료(최근 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 및 두 번째 주성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

표 3-33은 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 최근 자료를 중복 제거하여 과거의 자료를 남기는 방법으로 자료를 가공하였고, 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-33> 2013-2017년 자료(최근 자료 중복제거)에서 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.909
메틸마뇨산	-0.227
N,N-디메틸포름아미드	-0.400

2013-2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-56).



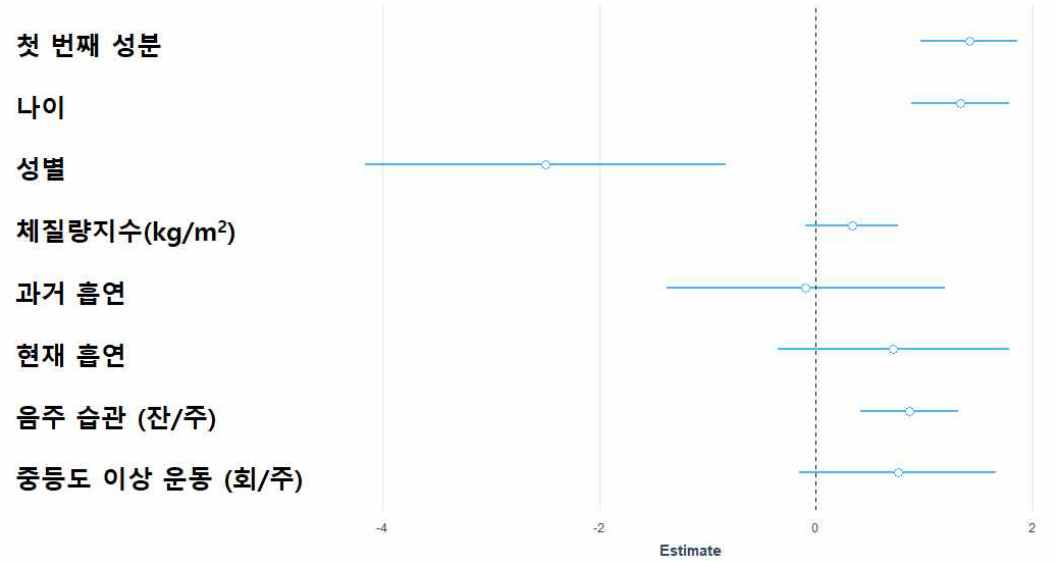
[그림 3-56] 2013-2017년 자료(최근 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

표 3-34는 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 과거 자료를 중복 제거하는 방법으로 자료를 가공하였고, 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-34> 2013-2017년 자료(과거 자료 중복제거)에서 세 가지 노출변수와 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.866
메틸마뇨산	-0.497
N,N-디메틸포름아미드	-0.130

2013-2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-57).



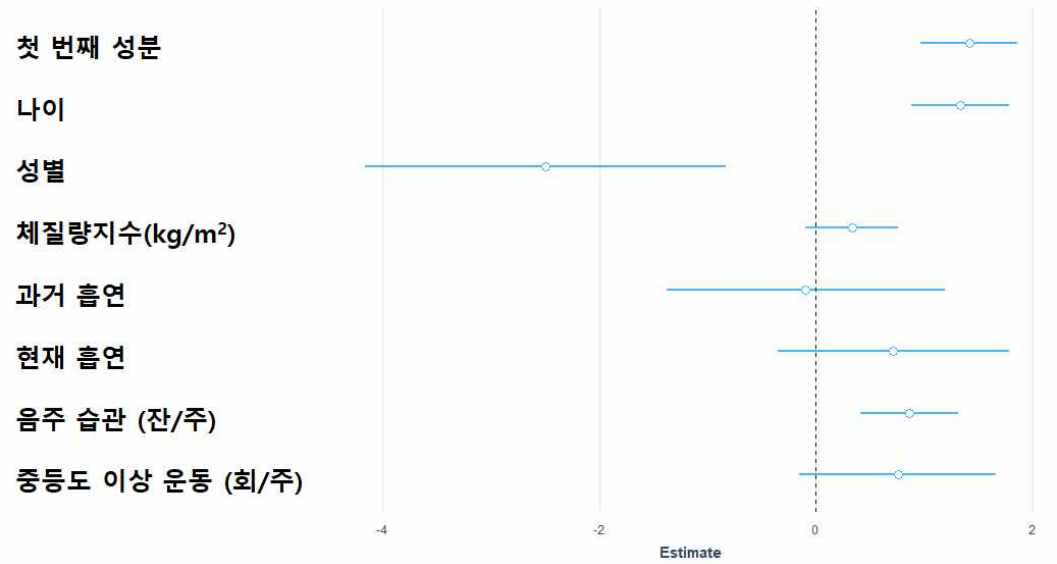
[그림 3-57] 2013-2017년 자료(과거 자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 관련성

표 3-35는 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 과거 자료를 중복 제거하는 방법으로 자료를 가공하였고, 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-35> 2013-2017년 자료(과거 자료 중복제거)에서 세 가지 노출변수와 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.915
메틸마뇨산	-0.368
N,N-디메틸포름아미드	-0.225

2013-2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-58).



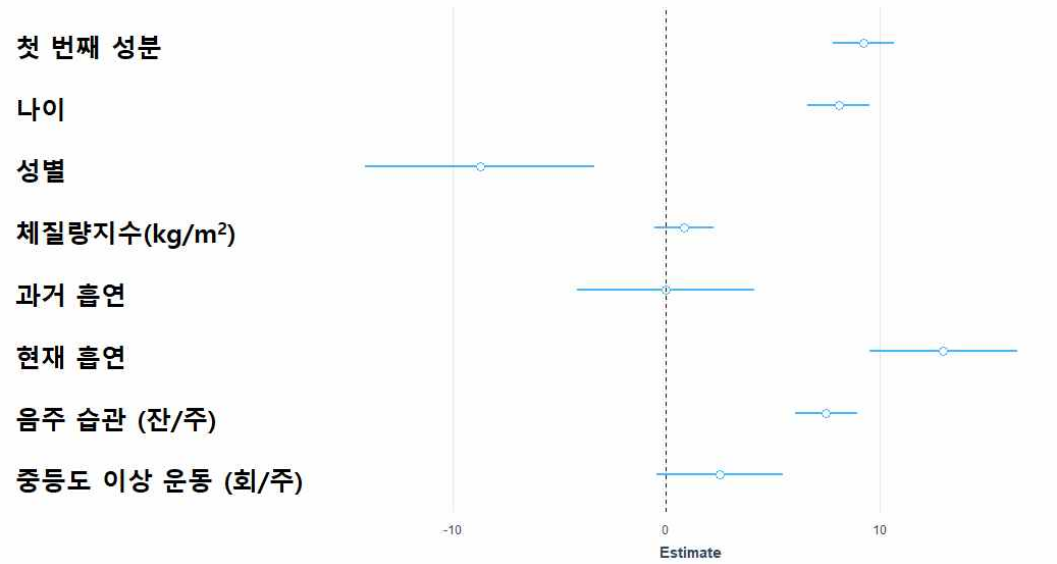
[그림 3-58] 2013-2017년 자료(과거자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 관련성

표 3-36은 2013년부터 2017년까지 자료를 합한 후 같은 사람의 자료 중 과거 자료를 중복 제거하는 방법으로 자료를 가공하였고, 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 주성분으로 마뇨산, 메틸마뇨산, N,N-디메틸포름아미드가 포함되었다.

<표 3-36> 2013-2017년 자료(과거 자료 중복제거)에서 세 가지 노출변수와 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 부분최소제곱법 회귀 후 산출된 첫 번째 성분의 로딩벡터

	첫 번째 성분
마뇨산	-0.919
메틸마뇨산	-0.164
N,N-디메틸포름아미드	-0.398

2013-2017년 자료에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성을 분석한 결과 유의한 양의 연관성이 관찰되었다(그림 3-59).



[그림 3-59] 2013-2017년 자료(과거자료 중복제거)에서 다중 선형회귀분석을 이용한 첫 번째 성분과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 관련성

R을 이용한 부분최소제곱법 회귀 및 결과 요약을 위한 그림을 그리는 R 코드는 다음과 같다.

```
#데이터 읽기#
data<-read.csv("D:/연구/복합노출 연구/data/three2013y1.csv")

data<-data[,-1] #id column 제거
head(data)

#부분최소제곱법 회귀 분석#
#install.packages("pls")
library(pls)
pls_mod<-plsr(y1~.,data=data[,1:4], ncomp=1, scale=TRUE) #exposures와 outcome data만 이용
round(pls_mod$loadings,3) #loading values

head(pls_mod$scores) #score
summary(pls_mod$scores)
data$pls_score<-pls_mod$scores

#첫 번째 PLS 성분으로 다중선형회귀분석#
plsr_mod<-lm(y1~pls_score+age+sex+bmi+smoking1+smoking2+drinking+exercise,data=data)
summary(plsr_mod)

#그림 그리기#
#install.packages("jtools")
library("jtools")

plot_summs(plsr_mod, legend.title = "Data",
           coefs=c("1st PLS component"="pls_score", "Age"="age", "Sex"="sex","BMI"="bmi",
                   "Past smoker"="smoking1", "Current smoker"="smoking2",
                   "Alcohol drinking habits"="drinking", "Physical activity"="exercise"),
           scale=TRUE)
```

5) 베이지안커널머신 회귀(Bayesian kernal machine regression, BKMR)를 활용한 복합노출과 간기능검사 결과 간의 관련성 분석

베이지안커널머신 회귀는 커널 함수의 적용을 통해 노출과 건강 결과 사이의 비선형 관계와 노출 간 상호 작용을 고려하여 유해물질들의 결합 효과(joint effect)를 평가할 수 있는 비모수 방법론이다. 본 연구에서는 세 가지 노출 물질과 간기능검사지표 사이의 비선형 관계와 노출 물질들 사이의 교호 작용을 반영하기 위하여 베이지안커널머신 회귀 모형을 적합시켰다. 각 연구대상자 $i=1,...,n$에 대하여 BKMR 모형은 아래와 같다(Bobb JF 외, 2018; Bobb JF 외, 2014).

$$Y_i = h(z_i) + x_i^T\beta + \epsilon_i$$

여기서 Y_i는 개인 i의 간기능검사 결과 값이며, $z_i = (z_{i1},...,z_{iM})^T$는 개인 i의 M번째 노출물질, h는 앞으로 추정될 미지의 노출-반응 함수이며, β는 공변량의 영향(x_i는 벡터임), 잔차 $\epsilon_i \sim N(0,\sigma^2)$는 서로 독립이고 같은 분산을 가지는 동일한 정규분포로 가정된다.

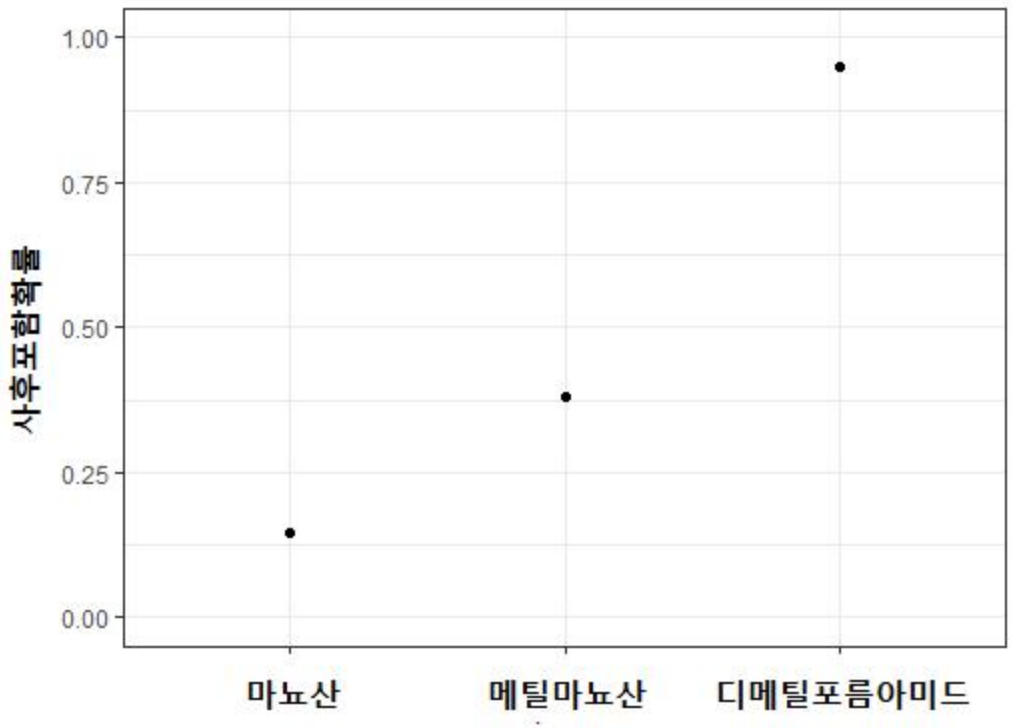
이때, $h()$함수는 노출물질 $(z_1,...,z_p)$ 사이의 교호작용 또는 노출물질과 건강결과 사이의 비선형관계를 반영할 수 있는 노출-반응함수이다. 이때 여러 가지 특정 커널 함수를 선택할 수 있는데, 본 연구에서는 실제 노출-반응 시나리오에 기초한 시뮬레이션 연구에서 잘 작동하는 것으로 나타난 가우시안 커널(Gaussian kernel) 함수를 사용하였으며, 이 커널은 $h()$에 대한 다양한 기본함수형식을 유연하게 잡아낼 수 있다.

본 연구에서는 베이지안커널머신 회귀의 적용을 통해 건강 결과에 대한 유해물질의 중요도를 대표할 수 있는 사후포함확률(posterior inclusion probability, PIP)을 확인하였으며, 각 노출 물질과 간기능검사지표의 관련성을 나타내는 노출-반응 함수를 통해 비선형 관계를 확인하였다. 또한, 세 가지 노출 물질들 사이의 교호작용 효과를 확인하기 위해 한 물질의 노출과 간기능검사지표의 연관성을 다른 물질 농도를 25 분위수, 50 분위수, 75 분위수로 고정한 뒤 살펴보았다. 마지막으로, 노출-반응 함수 $h()$를 이용하

여 세 가지 노출 물질의 복합노출 영향과 개별 물질 영향을 확인하였다.

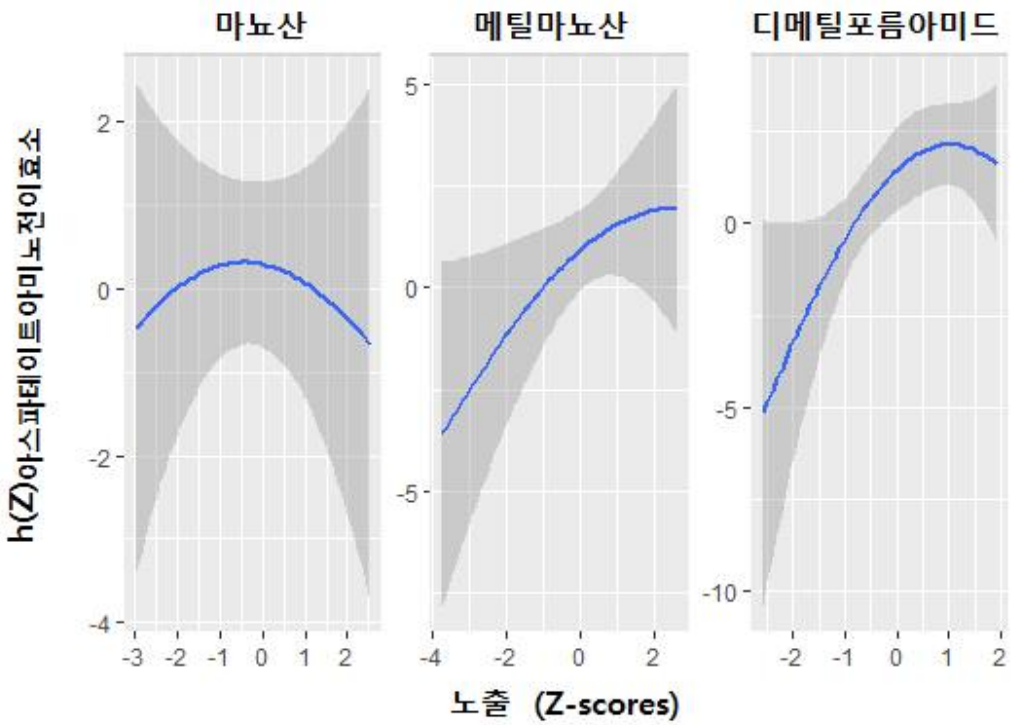
베이지안커널머신 회귀 분석을 위해 bkmr R package를 사용하였다.

2013년 자료에서 세 가지 노출 물질과 혈중 아스파테이트아미노전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 디메틸포름아미드, 메틸마뇨산, 마뇨산 순으로 높았다.



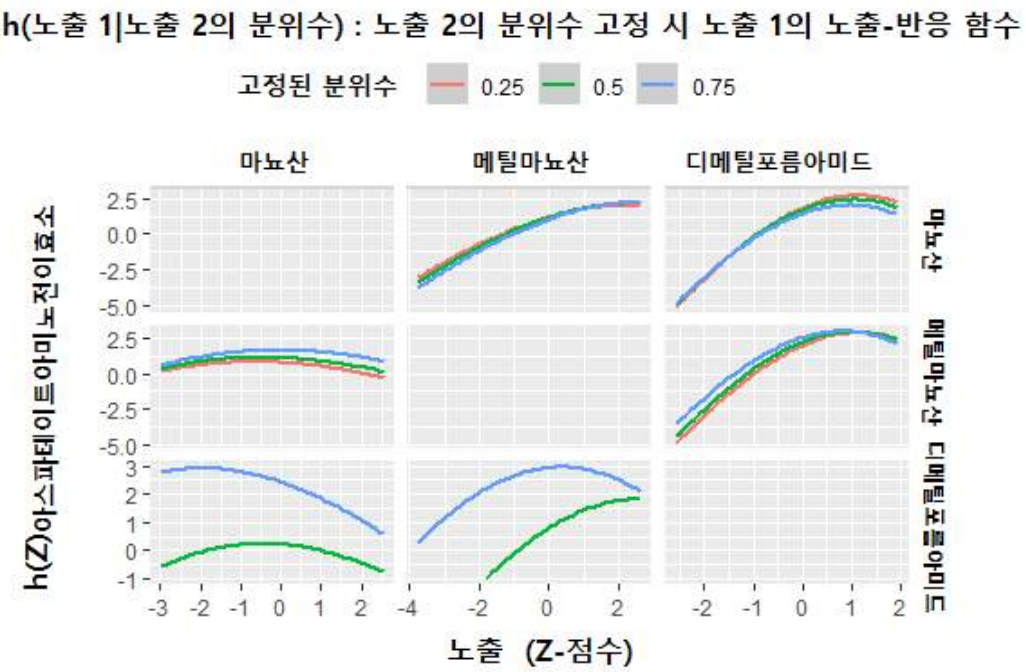
[그림 3-60] 2013년 자료에서 세 가지 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2013년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 메틸마뇨산과 디메틸포름아미드가 양의 선형에 가까운 연관성을 보였다. 마뇨산은 역 U 패턴을 보였다.



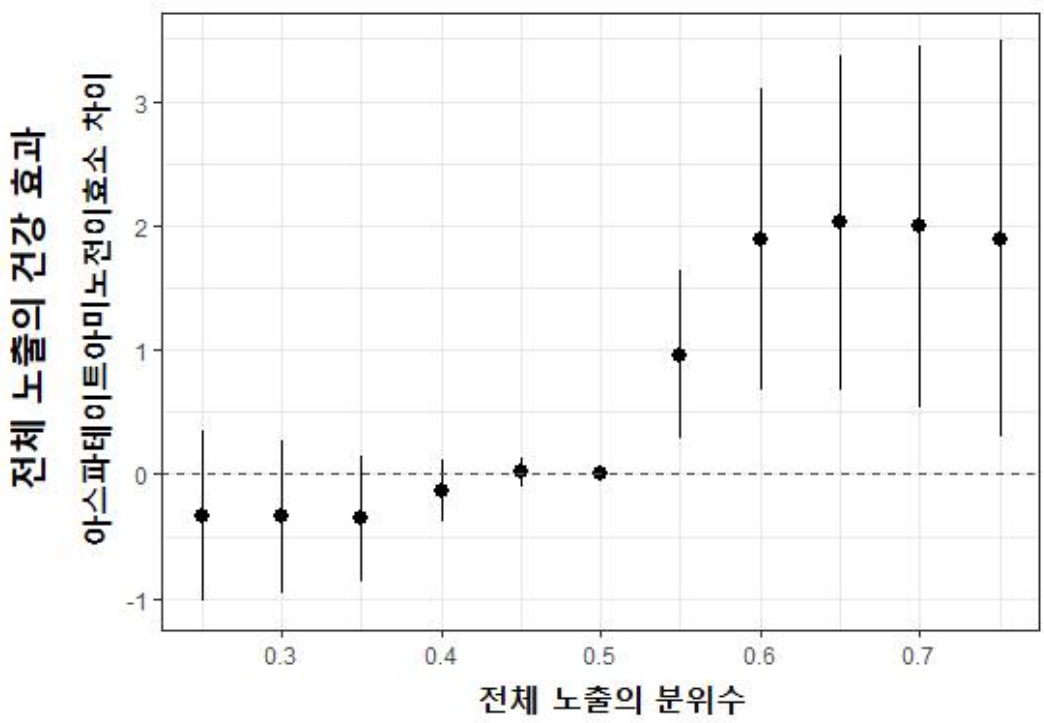
[그림 3-61] 2013년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 단변량 노출-반응 함수

2013년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 각 물질들 간에 교호 작용이 존재하는 것으로 나타났다.



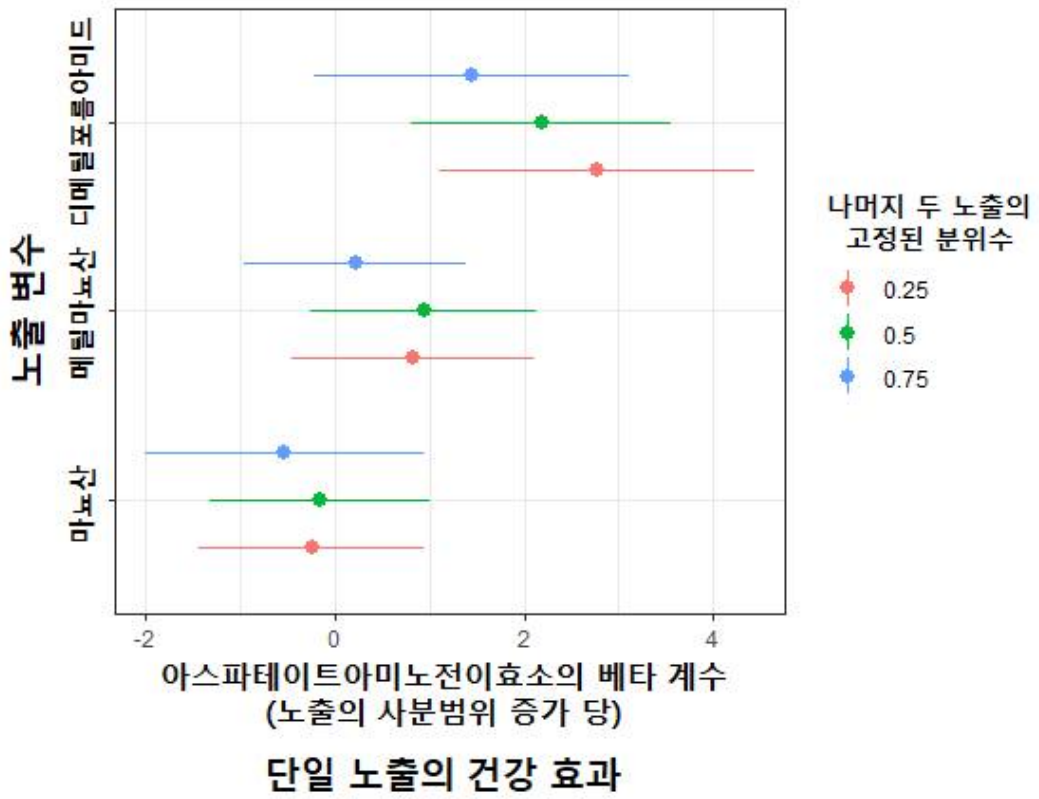
[그림 3-62] 2013년 자료에서 두 가지 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 이변량 노출-반응 함수

2013년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 중앙값을 기준으로 낮은 노출 수준에서 높은 노출 수준으로 올라감에 따라 혈중 아스파테이트아미노전이효소가 증가하는 경향이 유의하게 나타났다.



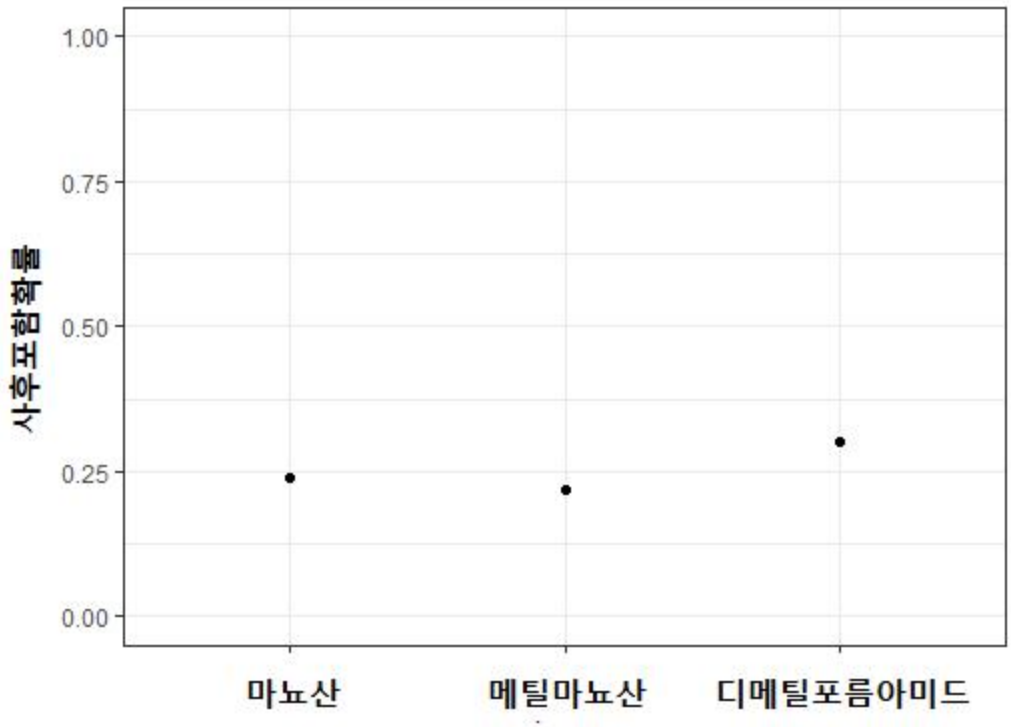
[그림 3-63] 2013년 자료에서 세 가지 노출 물질이 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)에 미치는 복합노출 영향

2013년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 디메틸포름아미드는 다른 노출이 50 백분위수 또는 75 백분위수 일 때 아스파테이트아미노전이효소의 높은 수준과 유의한 관련이 있었으며, 디메틸포름아미드의 단일 노출 추정 값은 75 백분위수로 고정되었을 때에 비해 다른 노출이 50 백분위수나 25 백분위수로 고정되었을 때 크기가 더 커서 디메틸포름아미드와 나머지 다른 노출 중 하나 이상이 교호 작용이 있음을 알 수 있었다.



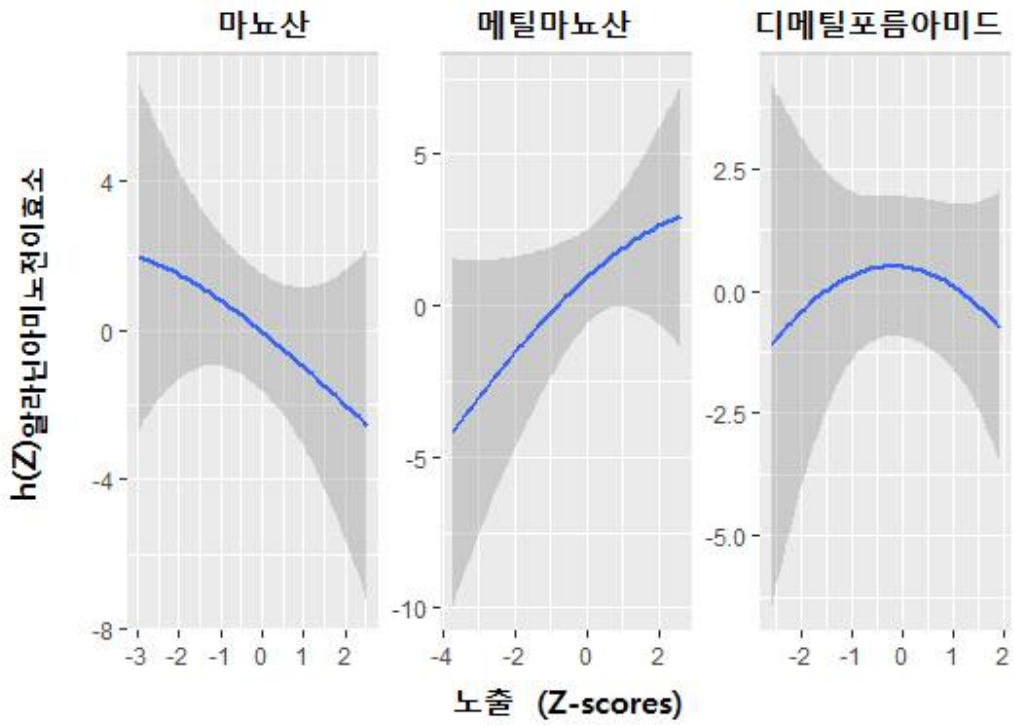
[그림 3-64] 2013년 자료에서 각 노출 물질이 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)에 개별적으로 미치는 영향

2013년 자료에서 세 가지 노출 물질과 혈중 알라닌아미노전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 디메틸포름아미드, 마뇨산, 메틸마뇨산 순으로 높았다.



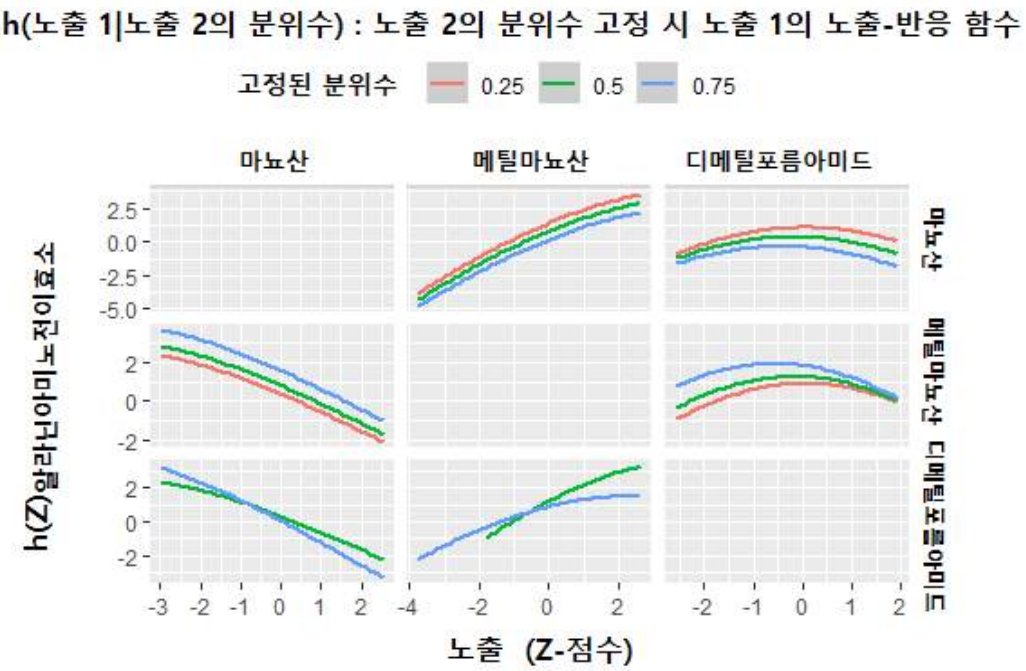
[그림 3-65] 2013년 자료에서 세 가지 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2013년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산이 음의 선형에 가까운 연관성을 보였고, 메틸마뇨산은 양의 선형에 가까운 연관성을 보였으며, 디메틸포름아미드는 역 U 패턴을 보였다.



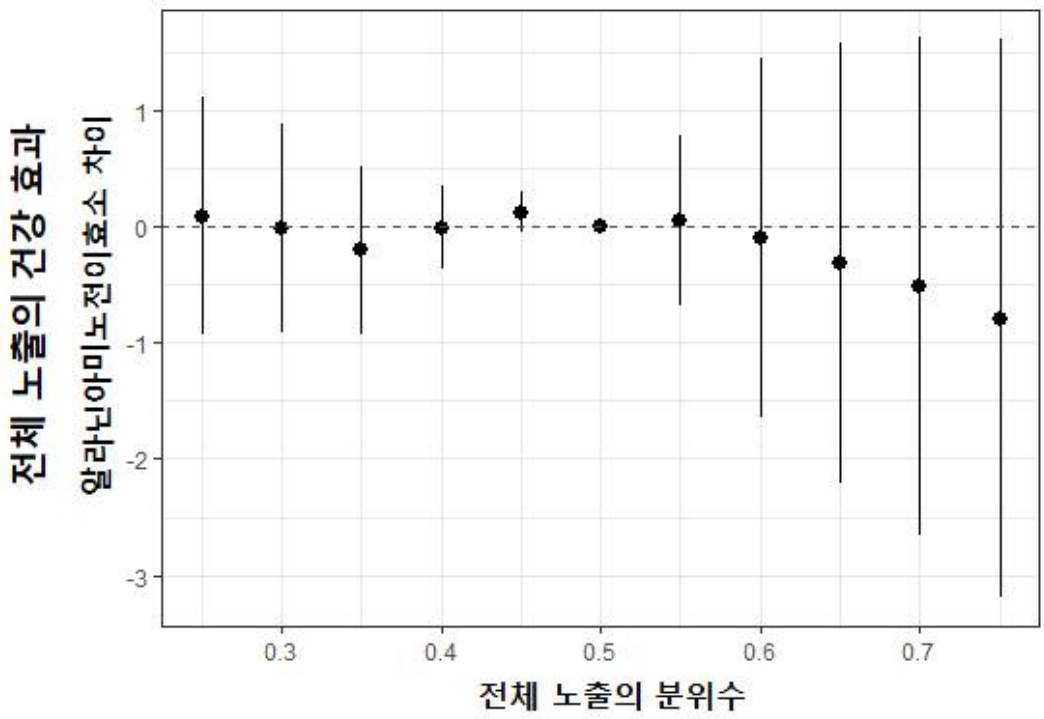
[그림 3-66] 2013년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)간의 단변량 노출-반응 함수

2013년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 마뇨산과 디메틸포름아미드, 메틸마뇨산과 디메틸포름아미드 간에 교호 작용이 존재하는 것으로 나타났다. 마뇨산과 메틸마뇨산 간에는 교호작용이 존재하지 않았다.



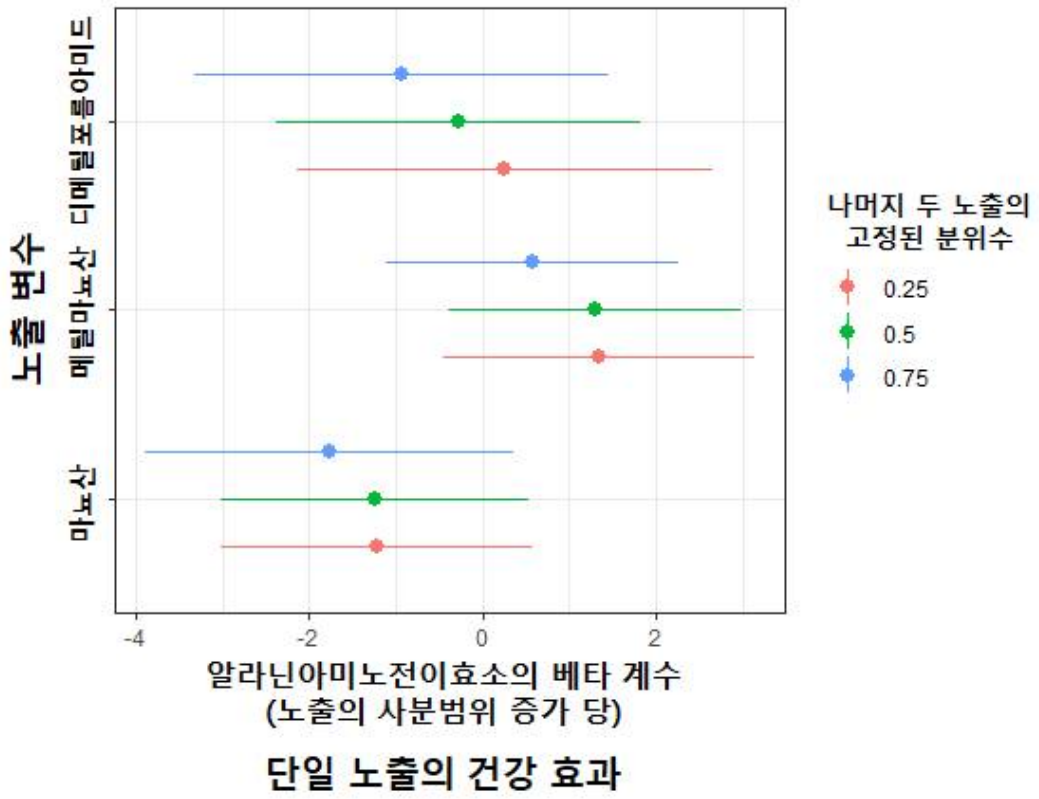
[그림 3-67] 2013년 자료에서 두 가지 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 이변량 노출-반응 함수

2013년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 혈중 알라닌 아미노전이효소에 미치는 영향은 유의하지 않았다.



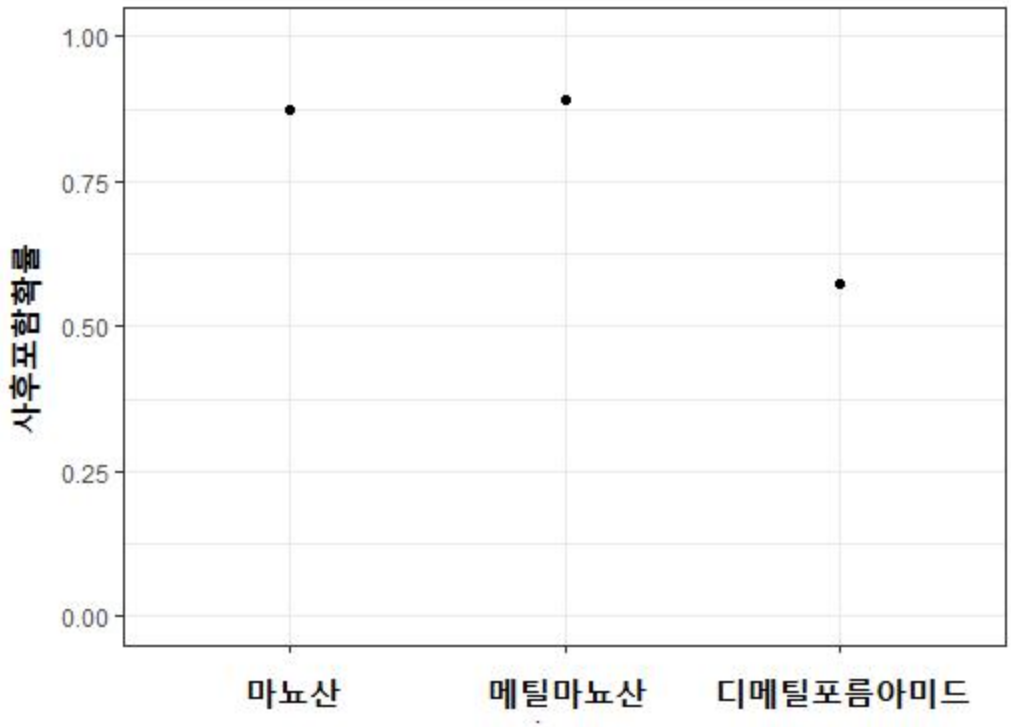
[그림 3-68] 2013년 자료에서 세 가지 노출 물질이 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)에 미치는 복합노출 영향

2013년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 단일 노출 중 알라닌아미노전이효소와 유의한 연관성을 나타내는 노출은 없었다.



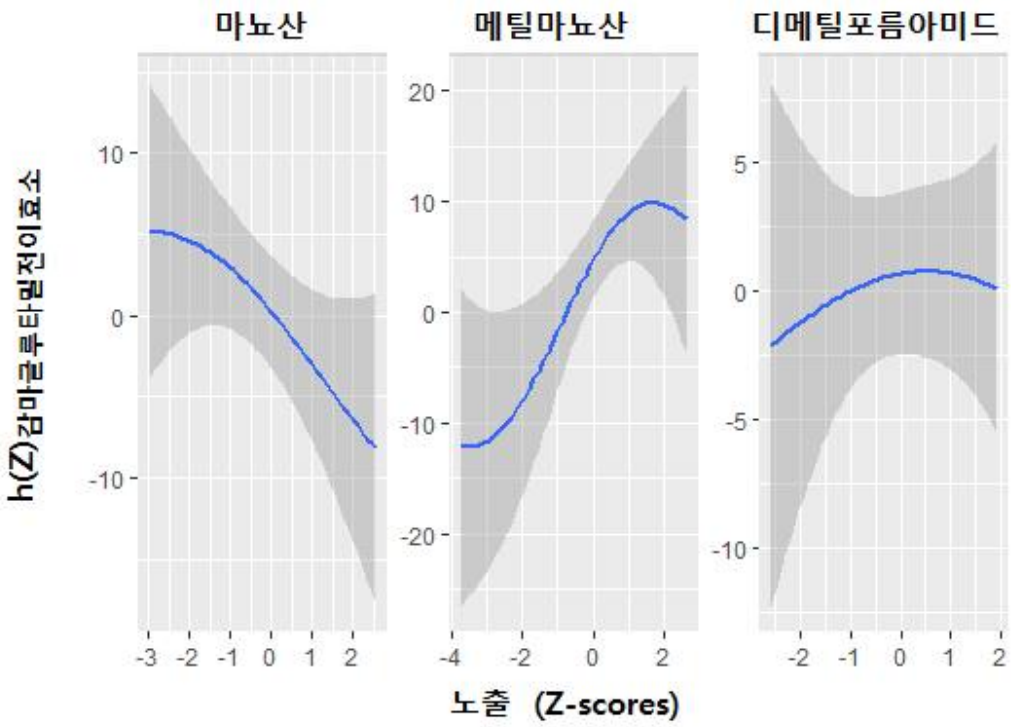
[그림 3-69] 2013년 자료에서 각 노출 물질이 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)에 개별적으로 미치는 영향

2013년 자료에서 세 가지 노출 물질과 혈중 감마글루타밀전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 메틸마뇨산, 마뇨산, 디메틸포름아미드 순으로 높았다.



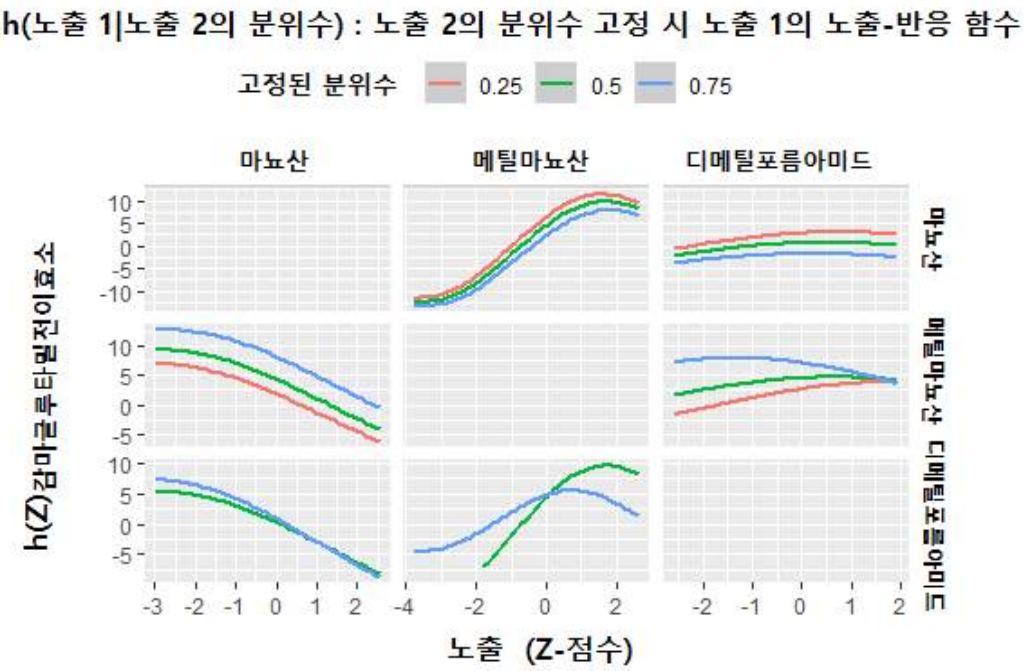
[그림 3-70] 2013년 자료에서 세 가지 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2013년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산이 음의 선형에 가까운 연관성을 보였고, 메틸마뇨산은 양의 선형에 가까운 연관성을 보였으며, 디메틸포름아미드는 역 U 패턴에 가까운 연관성을 보였다.



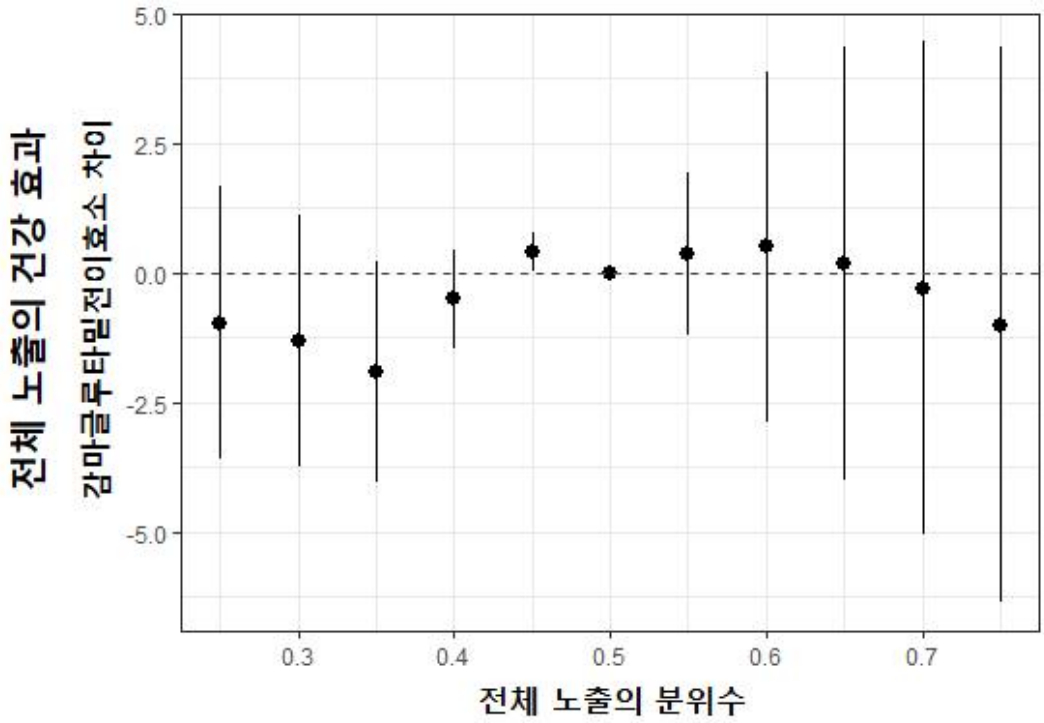
[그림 3-71] 2013년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 단변량 노출-반응 함수

2013년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 마뇨산과 디메틸포름아미드, 메틸마뇨산과 디메틸포름아미드 간에 교호 작용이 존재하는 것으로 나타났다. 마뇨산과 메틸마뇨산은 간에는 교호작용이 존재하지 않았다.



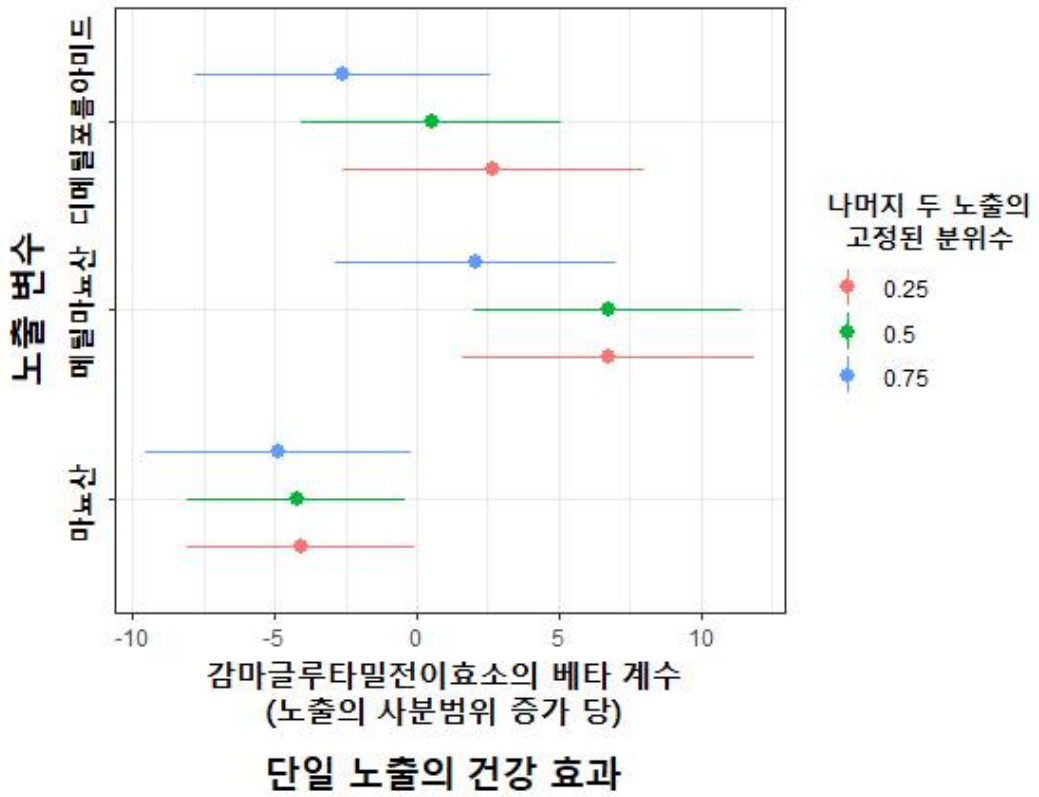
[그림 3-72] 2013년 자료에서 두 가지 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 이변량 노출-반응 함수

2013년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 혈중 감마글루타밀아미노전이효소에 미치는 영향은 유의하지 않았다.



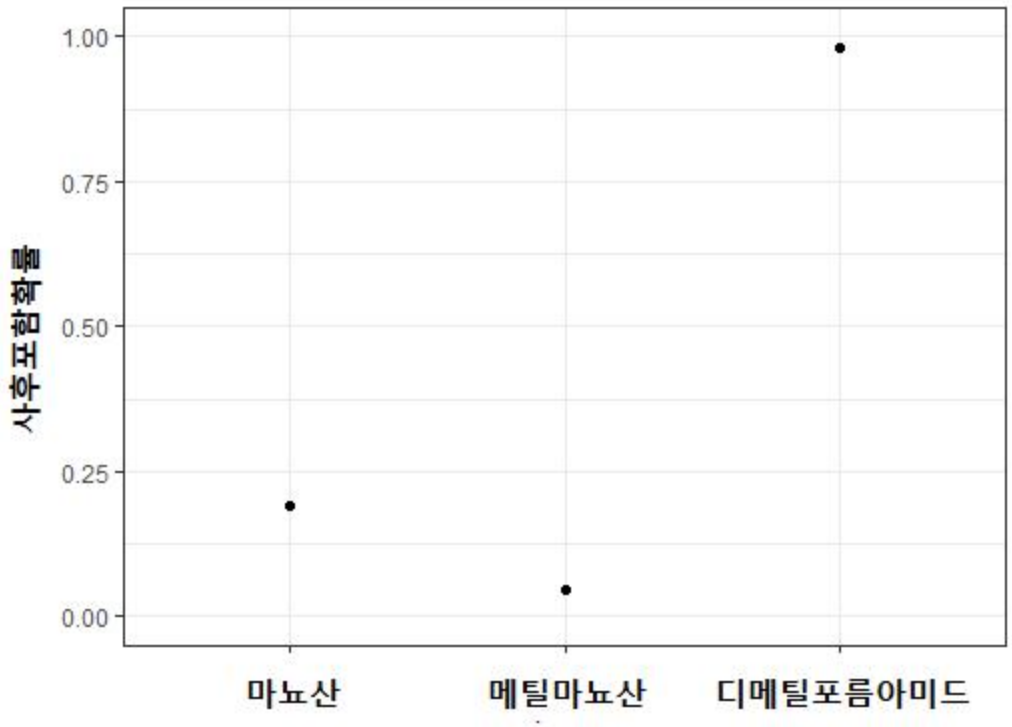
[그림 3-73] 2013년 자료에서 세 가지 노출 물질이 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)에 미치는 복합노출 영향

2013년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 메틸마뇨산은 다른 노출이 50 백분위수 또는 75 백분위수 일 때 감마글루타밀전이효소의 높은 수준과 유의한 관련이 있었으며, 메틸마뇨산의 단일 노출 추정 값은 75 백분위수로 고정되었을 때에 비해 다른 노출이 50 백분위수나 25 백분위수로 고정되었을 때 크기가 더 커서 디메틸포름아미드와 나머지 다른 노출 중 하나 이상이 교호 작용이 있음을 알 수 있었다.



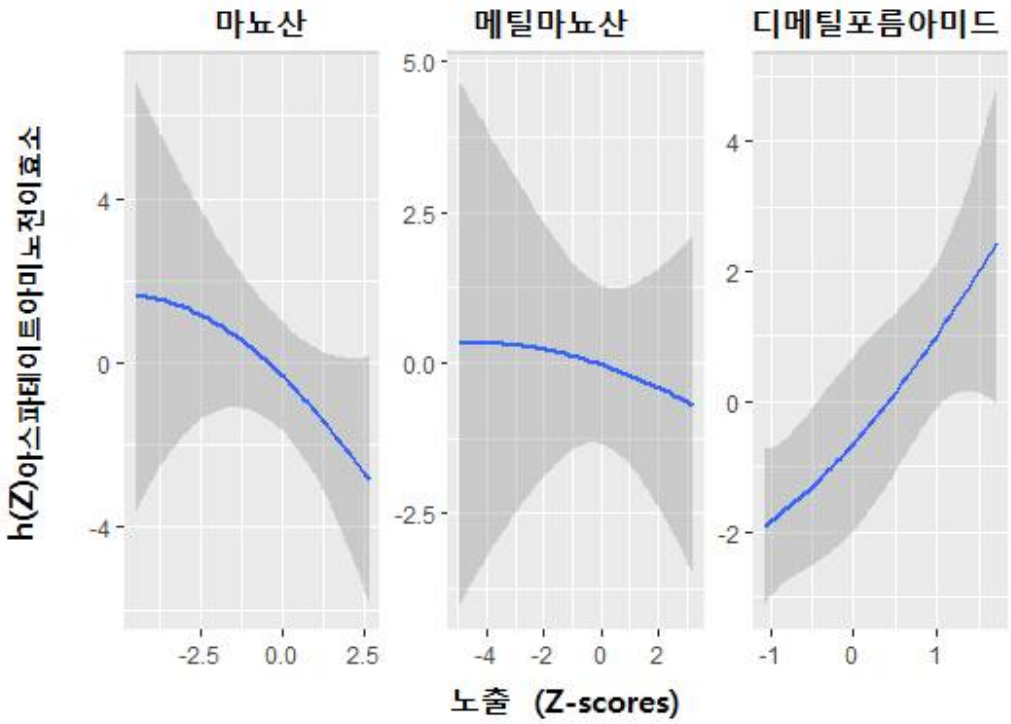
[그림 3-74] 2013년 자료에서 각 노출 물질이 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 에 개별적으로 미치는 영향

2014년 자료에서 세 가지 노출 물질과 혈중 아스파테이트아미노전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 디메틸포름아미드, 마뇨산, 메틸마뇨산 순으로 높았다.



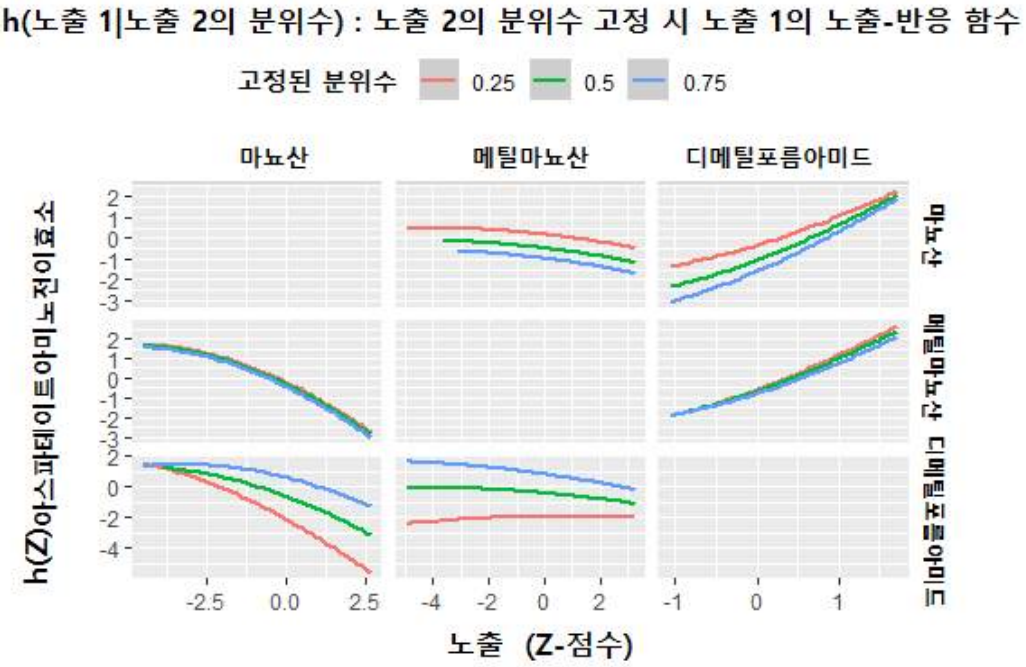
[그림 3-75] 2014년 자료에서 세 가지 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2014년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산과 메틸마뇨산이 음의 선형에 가까운 연관성을 보였고, 디메틸포름아미드는 양의 선형에 가까운 연관성을 보였다.



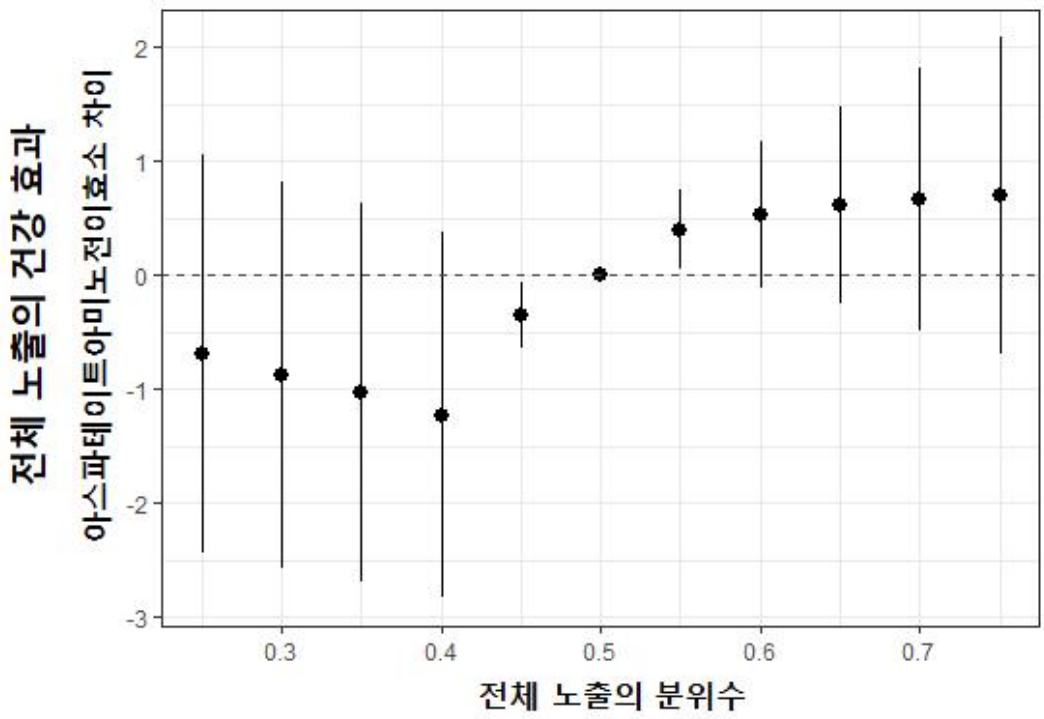
[그림 3-76] 2014년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 단변량 노출-반응 함수

2014년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 각 물질들 간에 교호 작용이 존재하는 것으로 나타났다.



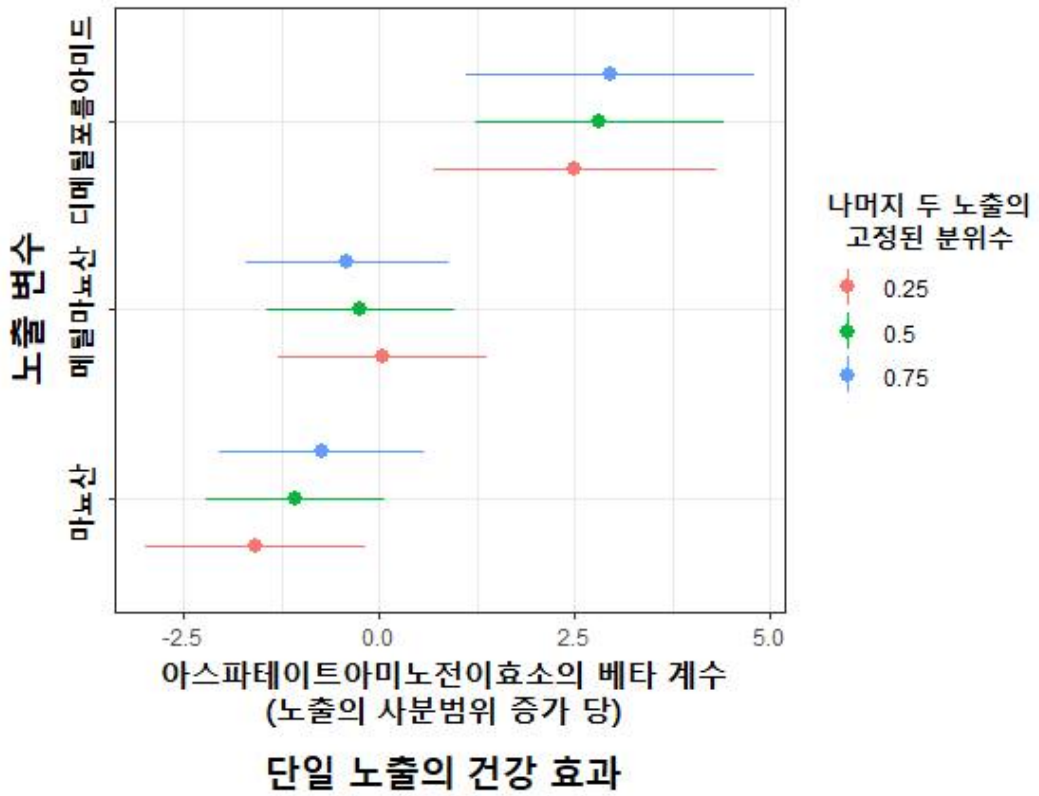
[그림 3-77] 2014년 자료에서 세 가지 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 베이지안커널머신 회귀 분석 후 노출 물질 사이의 교호작용

2014년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 혈중 아스파테이트아미노전이효소에 미치는 영향은 유의하지 않았다.



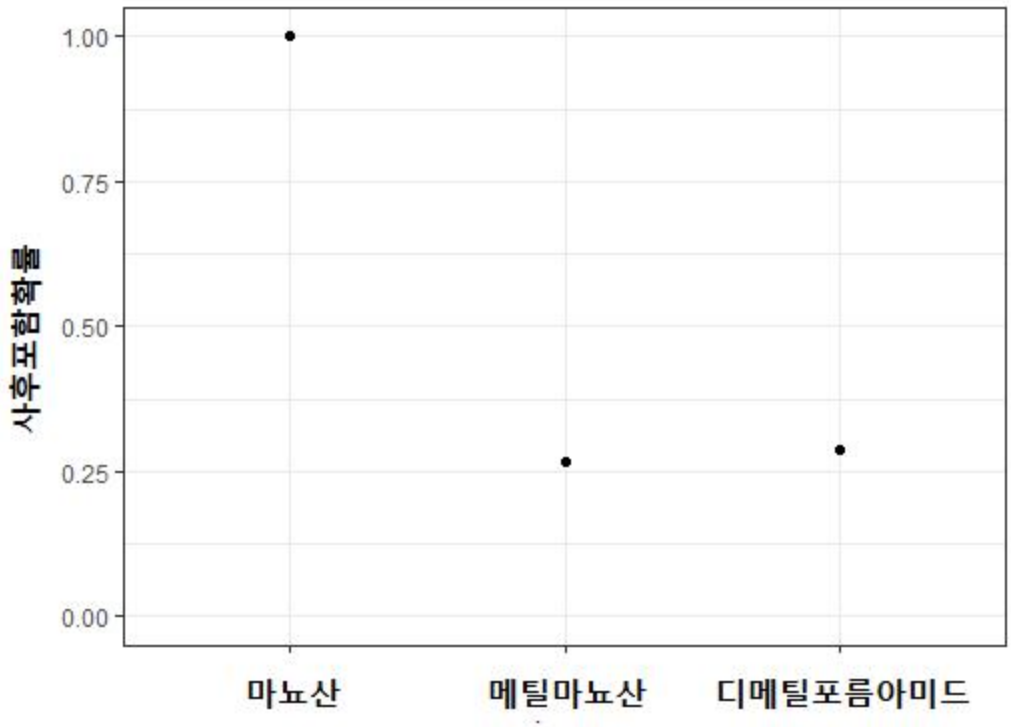
[그림 3-78] 2014년 자료에서 세 가지 노출 물질이 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)에 미치는 복합노출 영향

2014년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 디메틸포름아미드의 노출 증가는 아스파테이트아미노전이효소의 증가와 유의한 관련이 있었으며, 디메틸포름아미드의 단일 노출 추정 값은 25 백분위수로 고정되었을 때에 비해 다른 노출이 50 백분위수나 25 백분위수로 고정되었을 때 크기가 더 커서 디메틸포름아미드와 나머지 다른 노출 중 하나 이상이 교호 작용이 있음을 알 수 있었다.



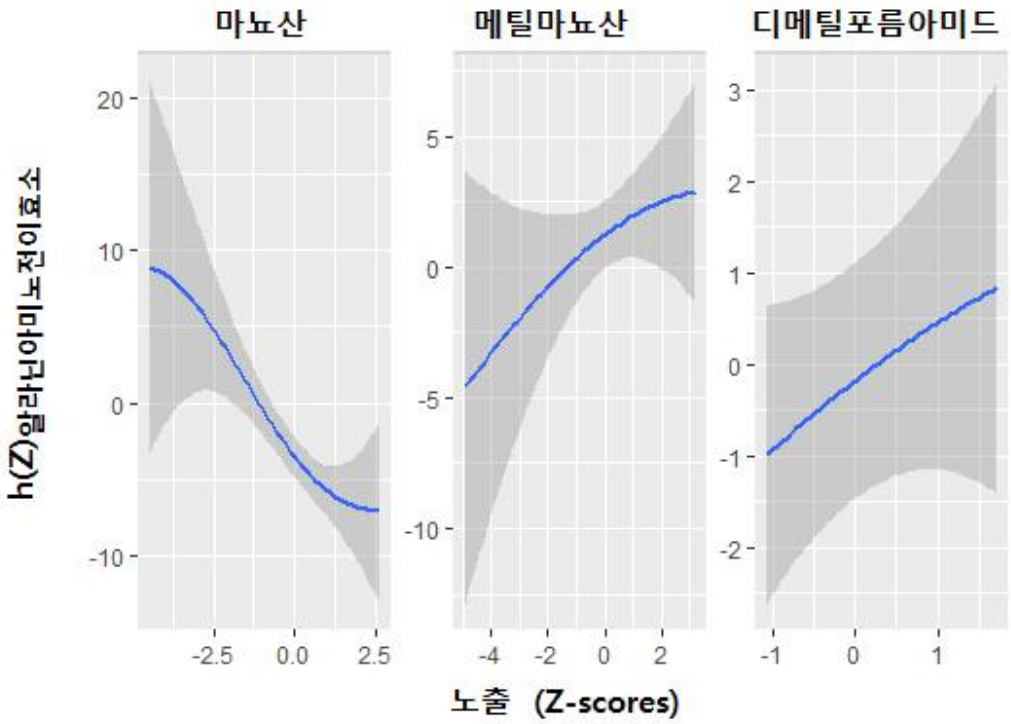
[그림 3-79] 2014년 자료에서 각 노출 물질이 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)에 개별적으로 미치는 영향

2014년 자료에서 세 가지 노출 물질과 혈중 알라닌아미노전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 마뇨산, 디메틸포름아미드, 메틸마뇨산 순으로 높았다.



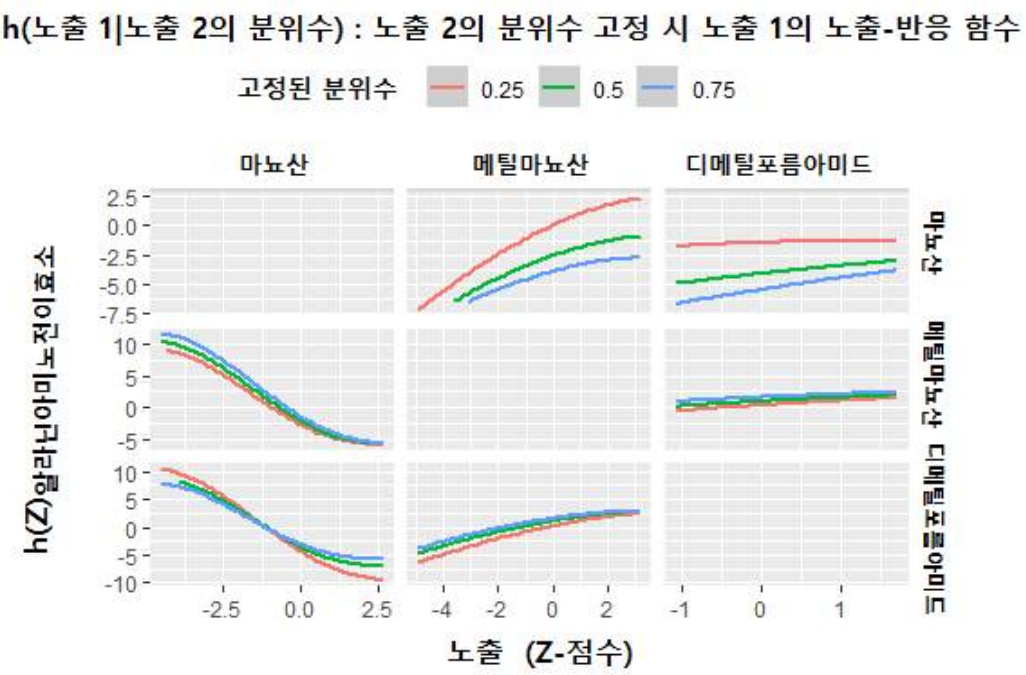
[그림 3-80] 2014년 자료에서 세 가지 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2014년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산이 음의 선형에 가까운 연관성을 보였고, 메틸마뇨산과 디메틸포름아미드는 양의 선형에 가까운 연관성을 보였다.



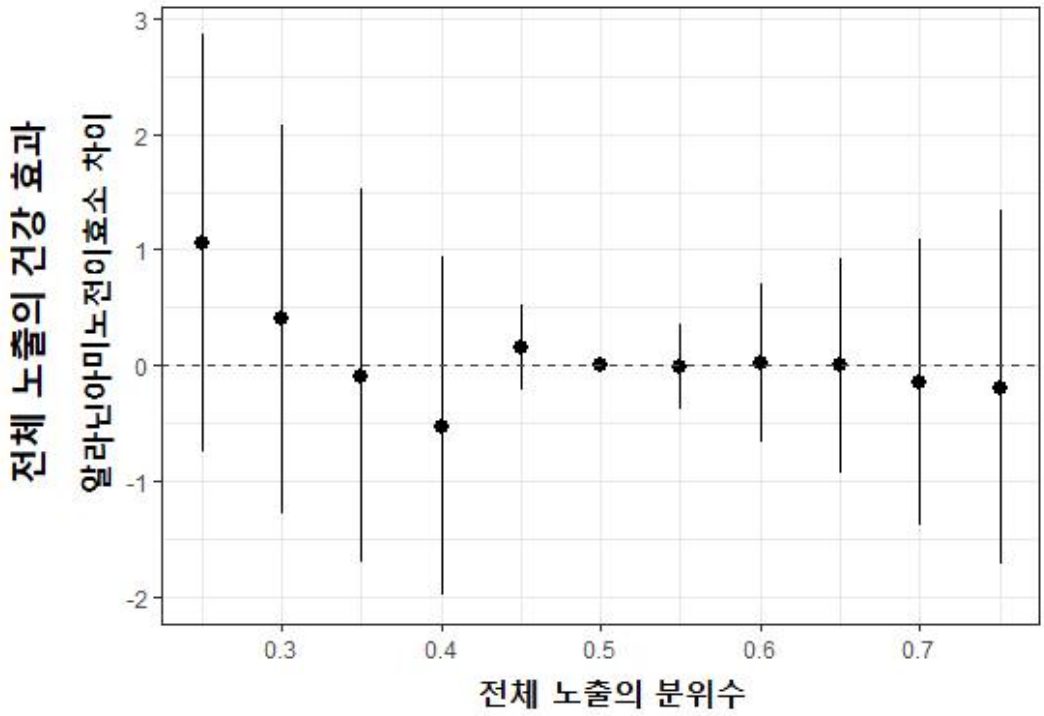
[그림 3-81] 2014년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)간의 단변량 노출-반응 함수

2014년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 각 물질들 간에 교호 작용이 존재하는 것으로 나타났다.



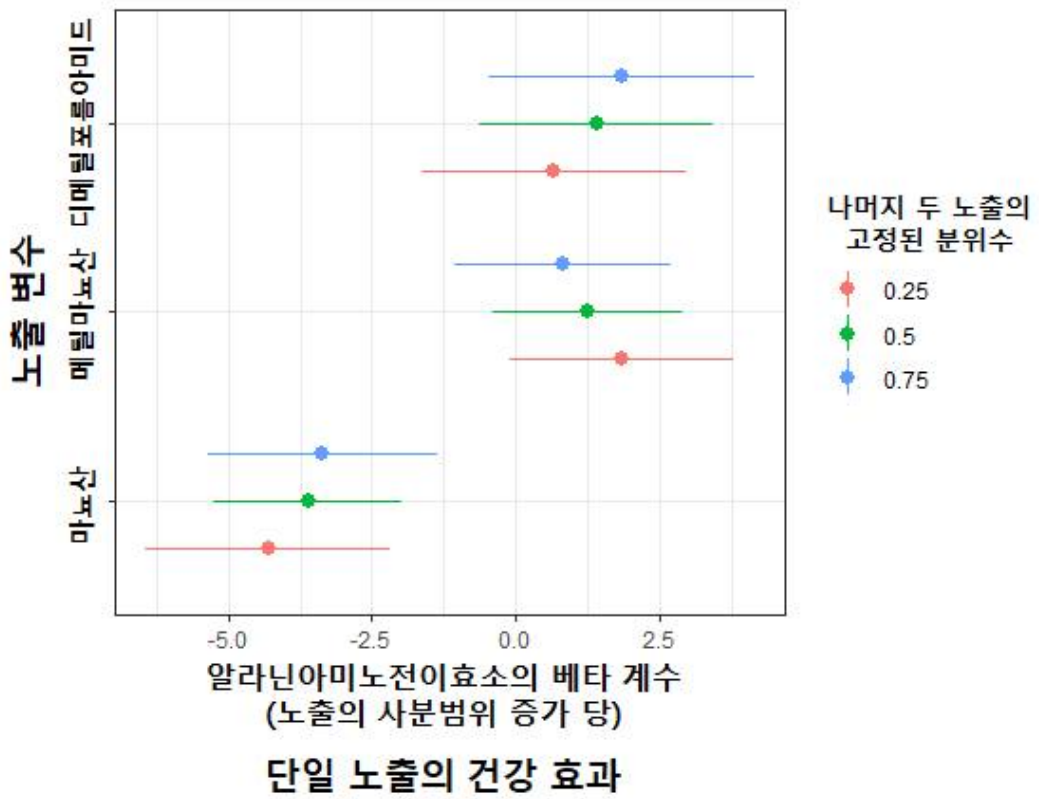
[그림 3-82] 2014년 자료에서 세 가지 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 이변량 노출-반응 함수

2014년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 혈중 알라닌아미노전이효소에 미치는 영향은 유의하지 않았다.



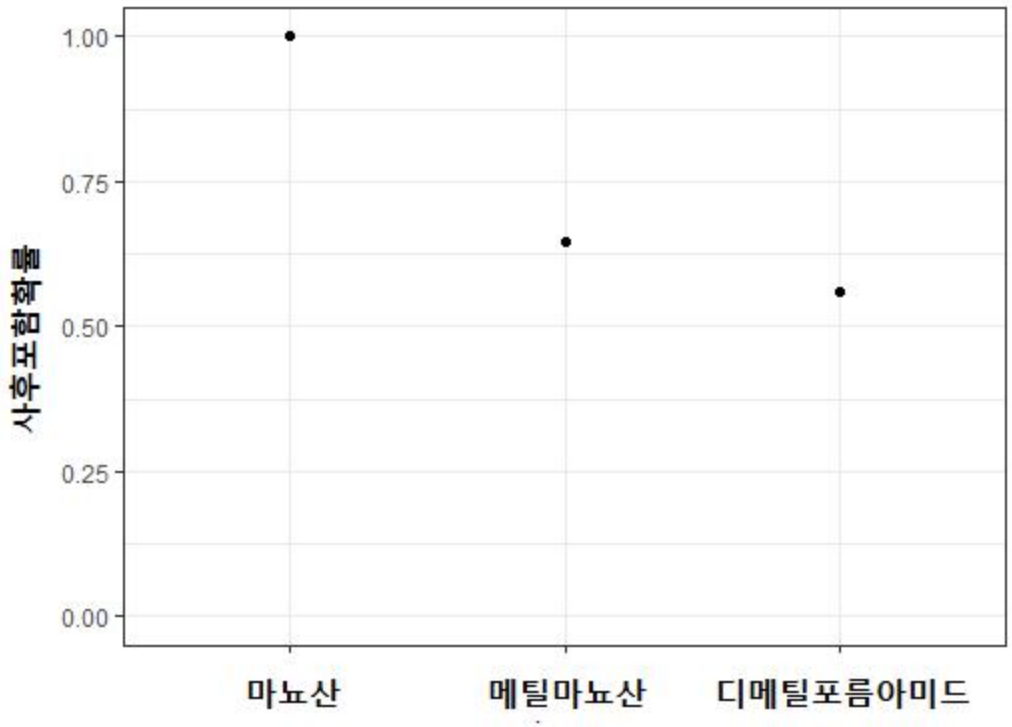
[그림 3-83] 2014년 자료에서 세 가지 노출 물질이 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)에 미치는 복합노출 영향

2014년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 단일 노출 중 알라닌아미노전이효소의 증가와 관련된 노출은 없었다.



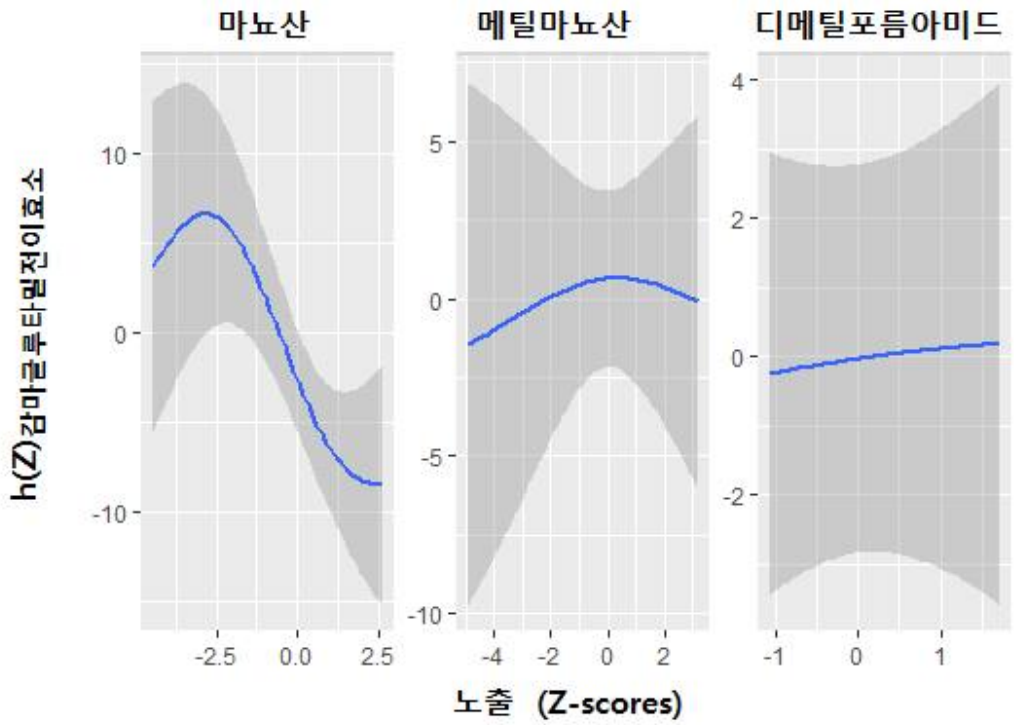
[그림 3-84] 2014년 자료에서 각 노출 물질이 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)에 개별적으로 미치는 영향

2014년 자료에서 세 가지 노출 물질과 혈중 감마글루타밀전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 마뇨산, 메틸마뇨산, 디메틸포름아미드 순으로 높았다.



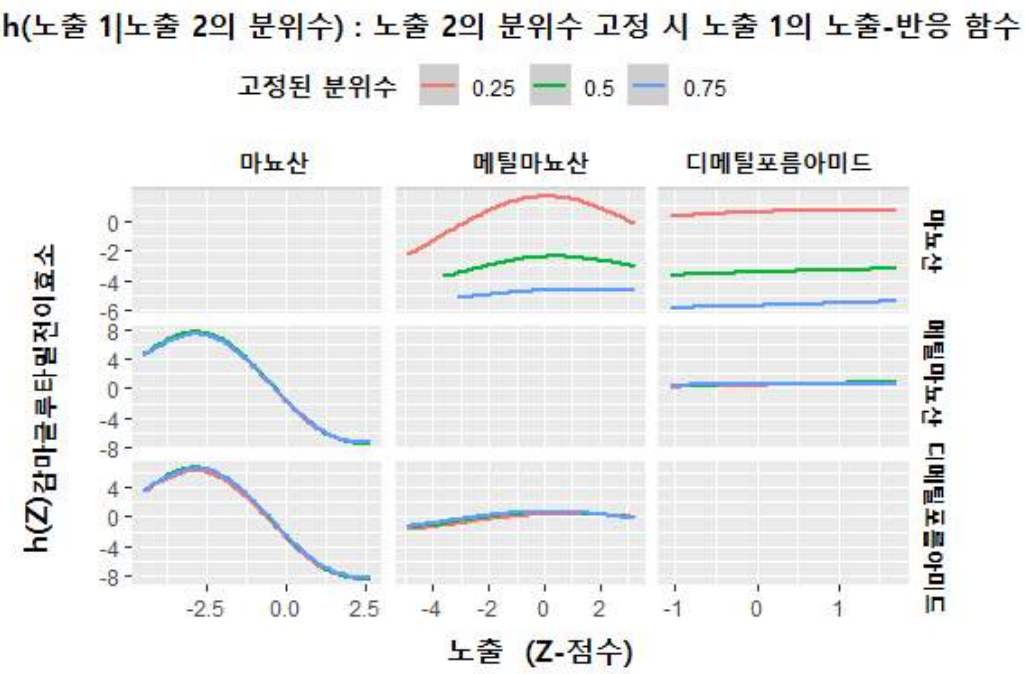
[그림 3-85] 2014년 자료에서 세 가지 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2014년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산이 음의 선형에 가까운 연관성을 보였고, 메틸마뇨산은 역 U 패턴을 보였으며, 디메틸포름아미드는 약하게 양의 선형에 가까운 연관성을 보였다.



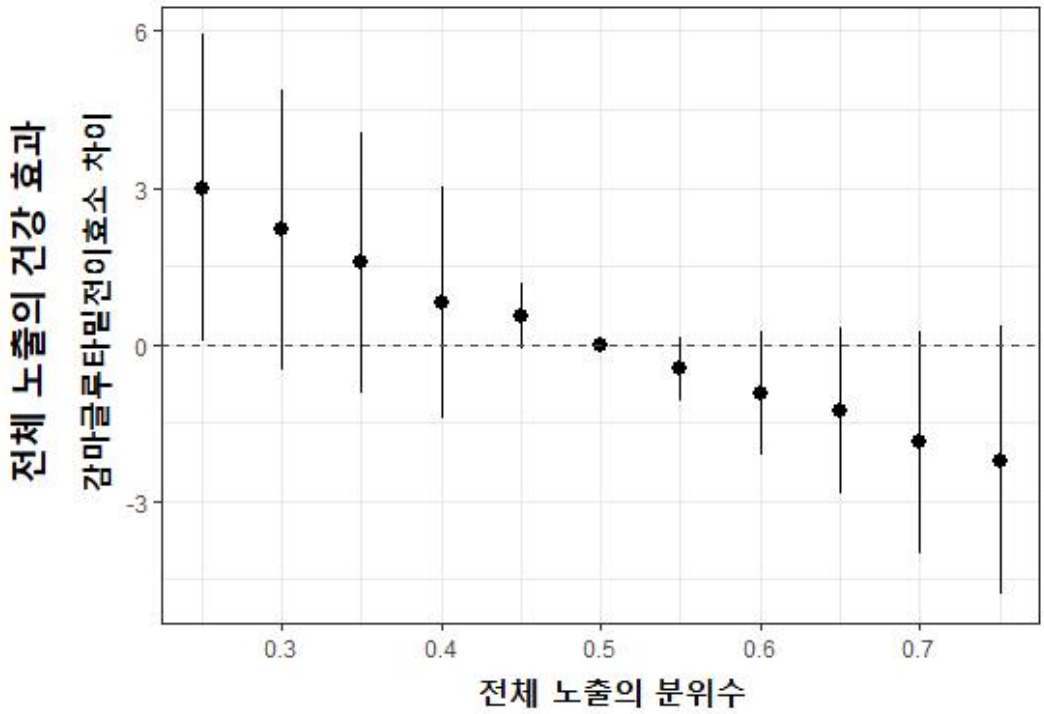
[그림 3-86] 2014년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 단변량 노출-반응 함수

2014년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 각 물질들 간에 교호 작용이 존재하는 것으로 나타났다.



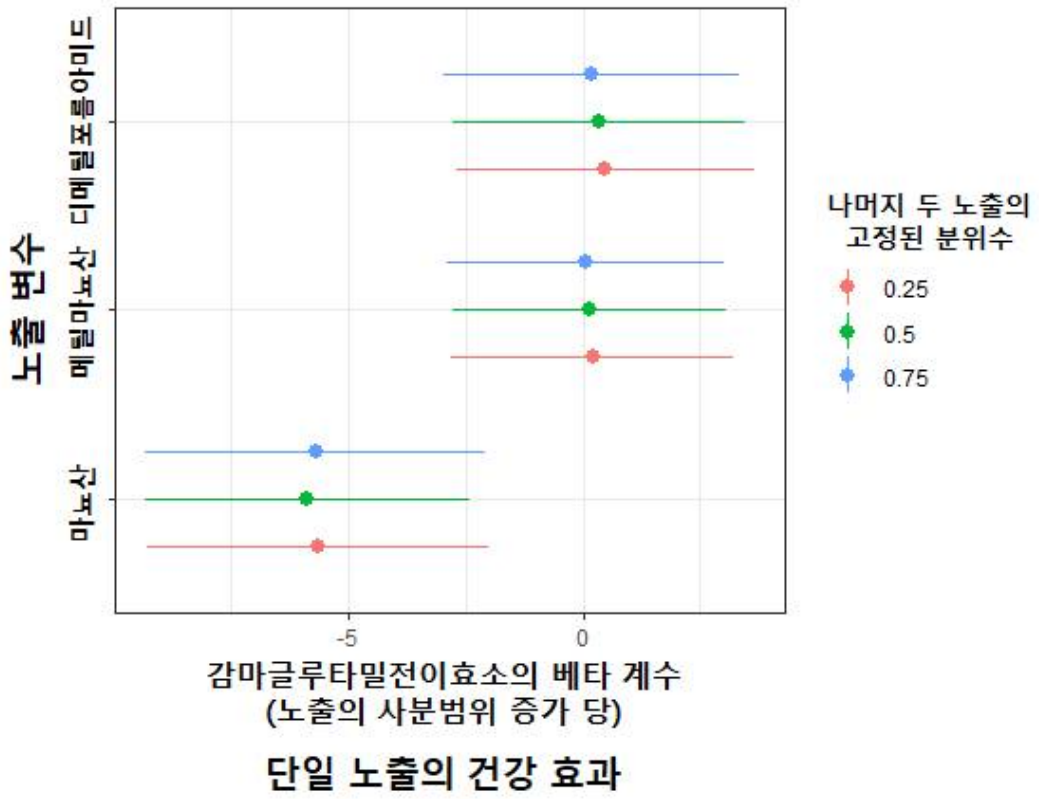
[그림 3-87] 2014년 자료에서 두 가지 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 이변량 노출-반응 함수

2014년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 증가 할수록 혈중 감마글루타밀전이효소가 낮아지는 경향이 있었으나, 유의하지는 않았다.



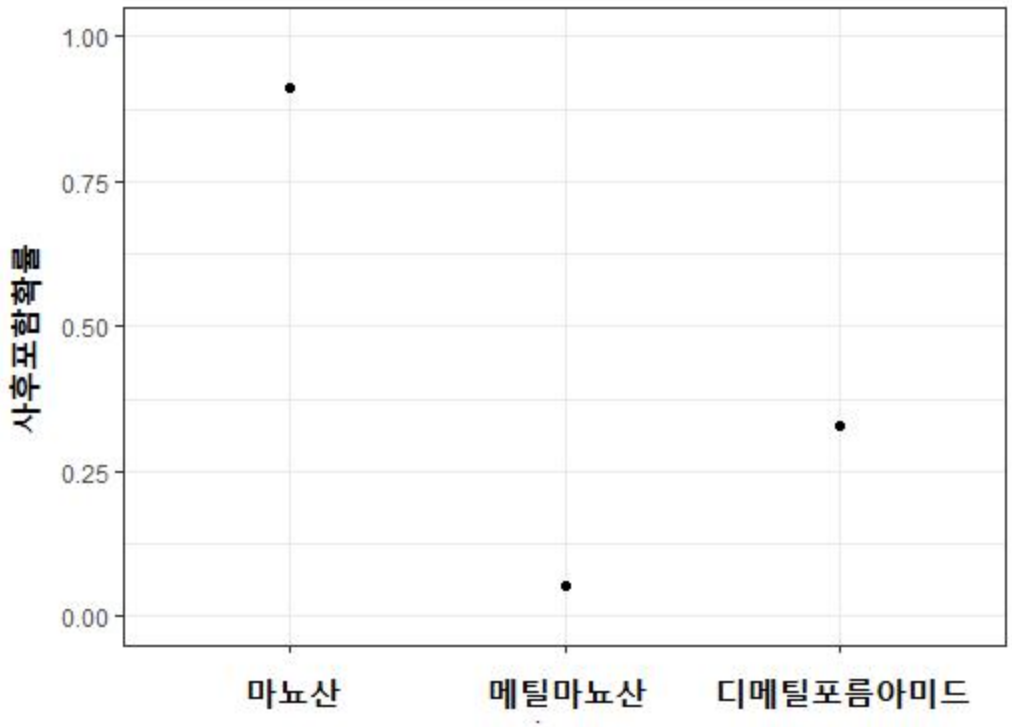
[그림 3-88] 2014년 자료에서 세 가지 노출 물질이 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)에 미치는 복합노출 영향

2014년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 단일 노출 중 감마글루타밀전이효소의 증가와 관련된 노출은 없었다.



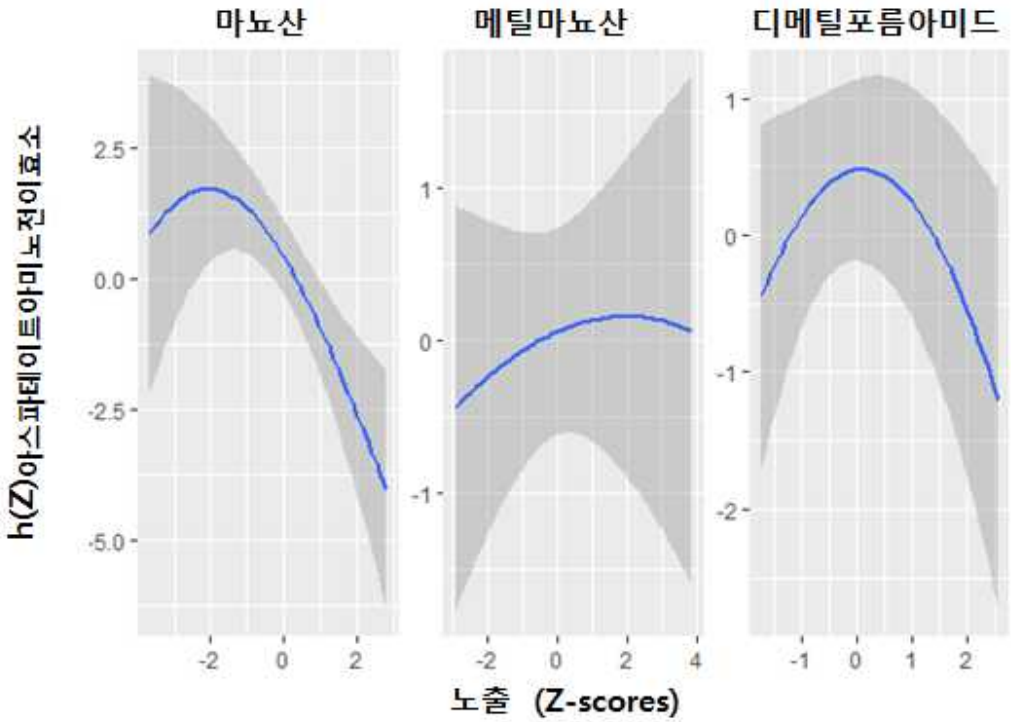
[그림 3-89] 2014년 자료에서 각 노출 물질이 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)에 개별적으로 미치는 영향

2015년 자료에서 세 가지 노출 물질과 혈중 아스파테이트아미노전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 마뇨산, 디메틸포름아미드, 메틸마뇨산 순으로 높았다.



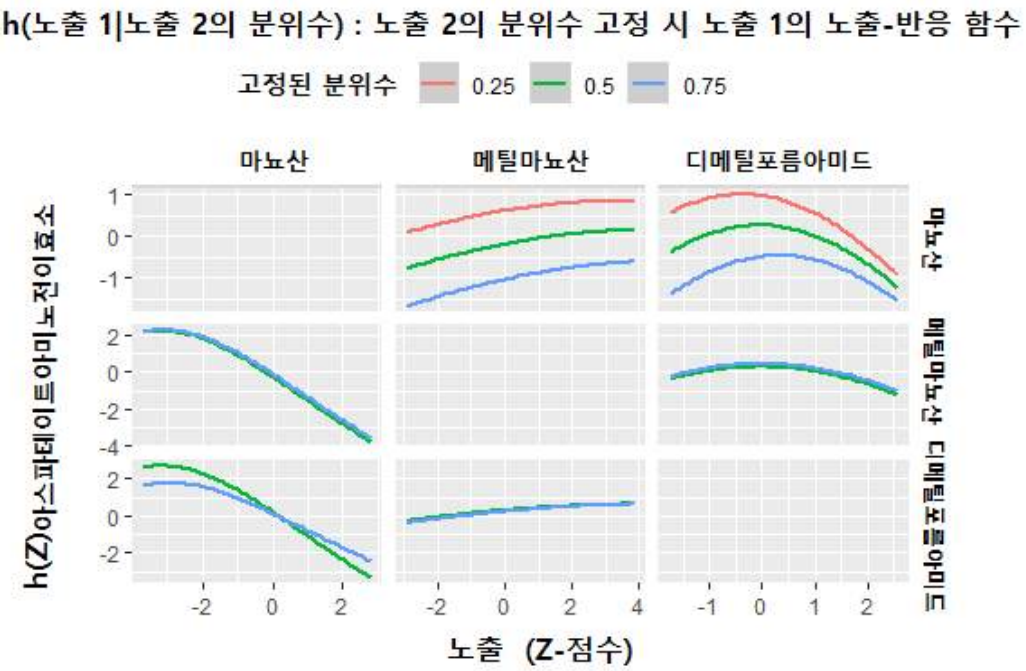
[그림 3-90] 2015년 자료에서 세 가지 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2015년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산, 메틸마뇨산, 디메틸포름아미드가 역 U 패턴에 가까운 연관성을 보였다.



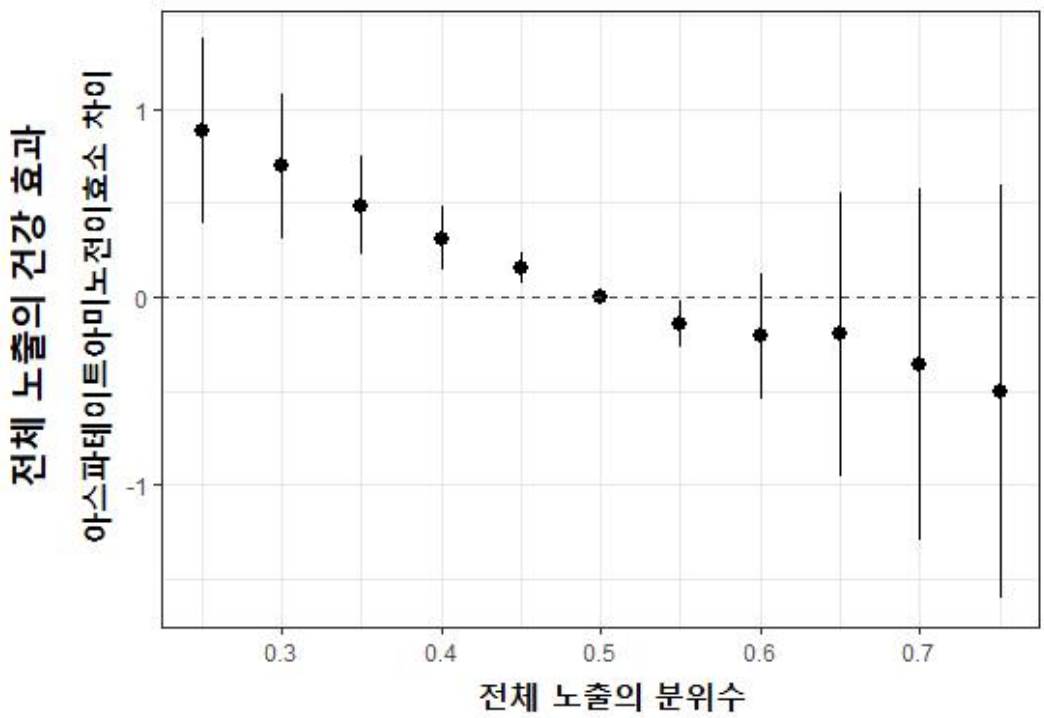
[그림 3-91] 2015년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 단변량 노출-반응 함수

2015년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 각 물질들 간에 교호 작용이 존재하는 것으로 나타났다.



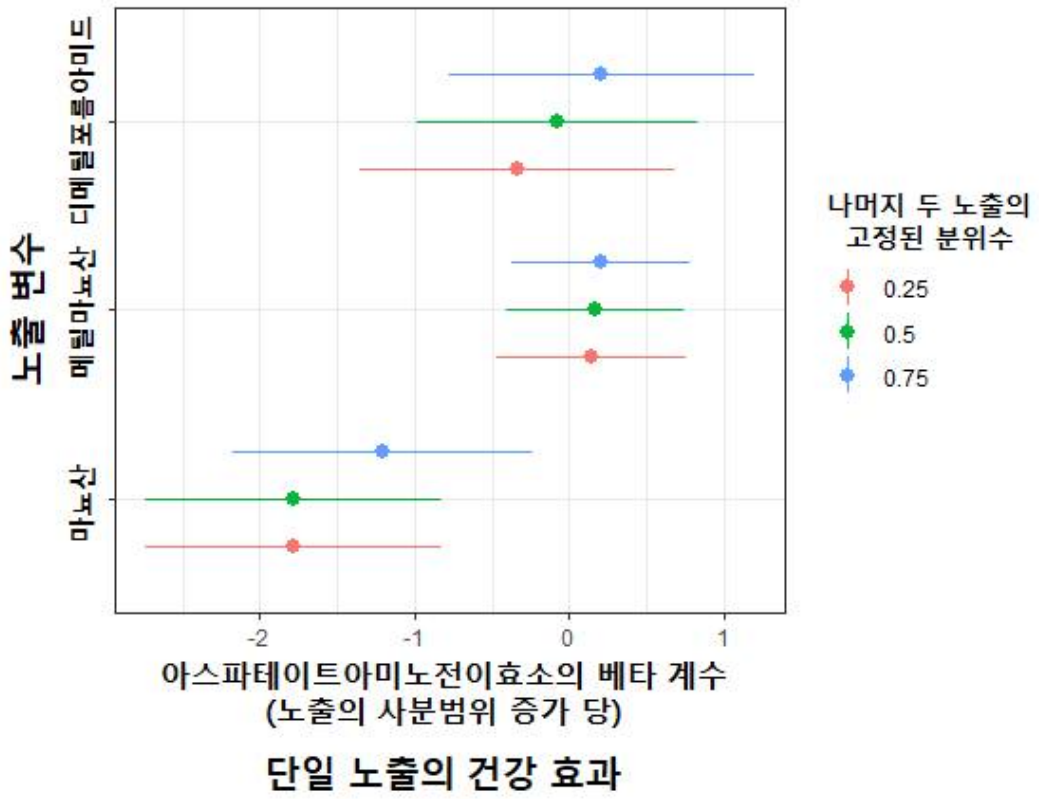
[그림 3-92] 2015년 자료에서 두 가지 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 이변량 노출-반응 함수

2015년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 중위값을 기준으로 감소할수록 혈중 감마글루타밀전이효소가 높아지는 경향이 유의하였다.



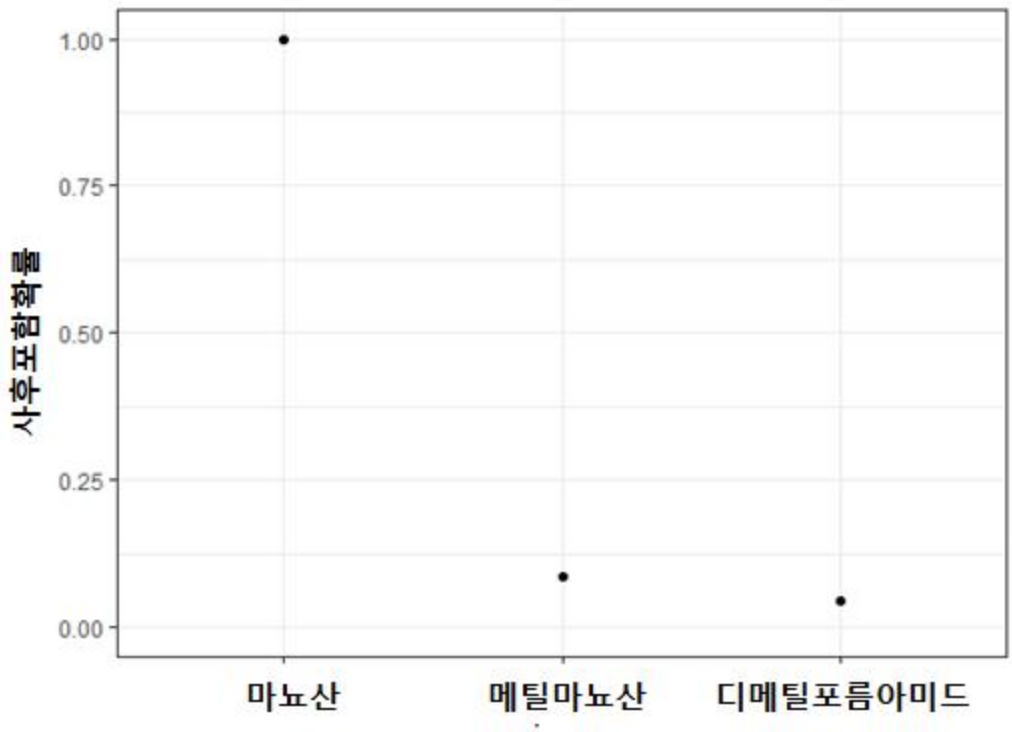
[그림 3-93] 2015년 자료에서 세 가지 노출 물질이 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)에 미치는 복합노출 영향

2015년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 단일 노출 중 아스파테이트아미노전이효소의 증가와 관련된 노출은 없었다.



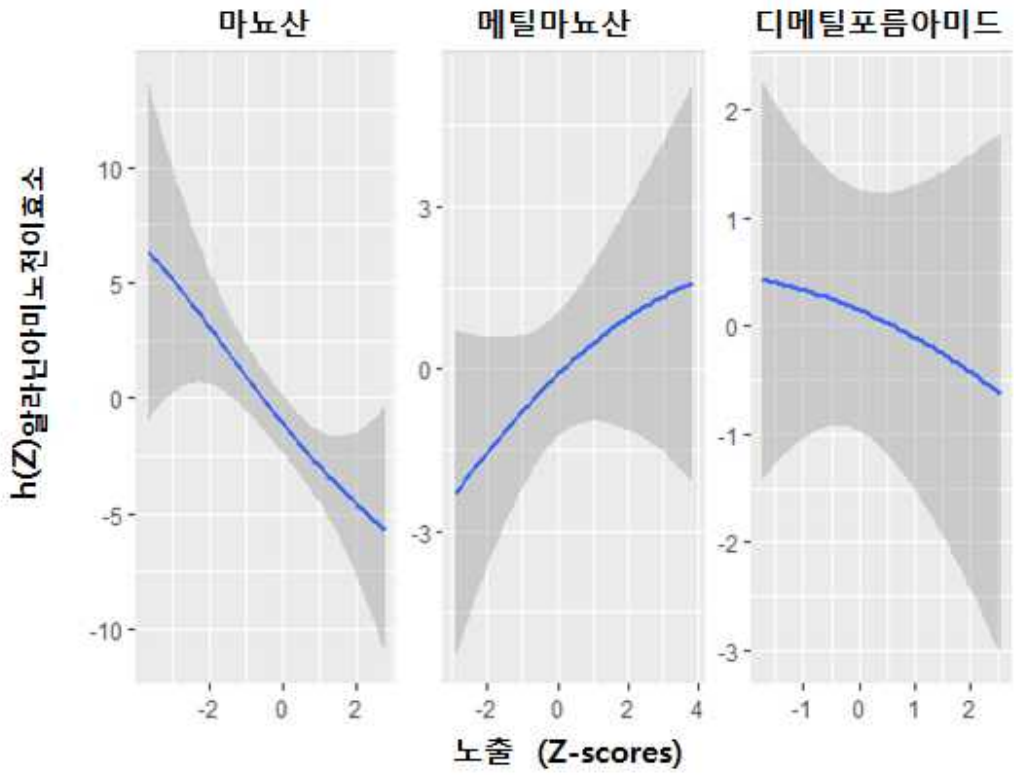
[그림 3-94] 2015년 자료에서 각 노출 물질이 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)에 개별적으로 미치는 영향

2015년 자료에서 세 가지 노출 물질과 혈중 알라닌아미노전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 마뇨산, 메틸마뇨산, 디메틸포름아미드 순으로 높았다.



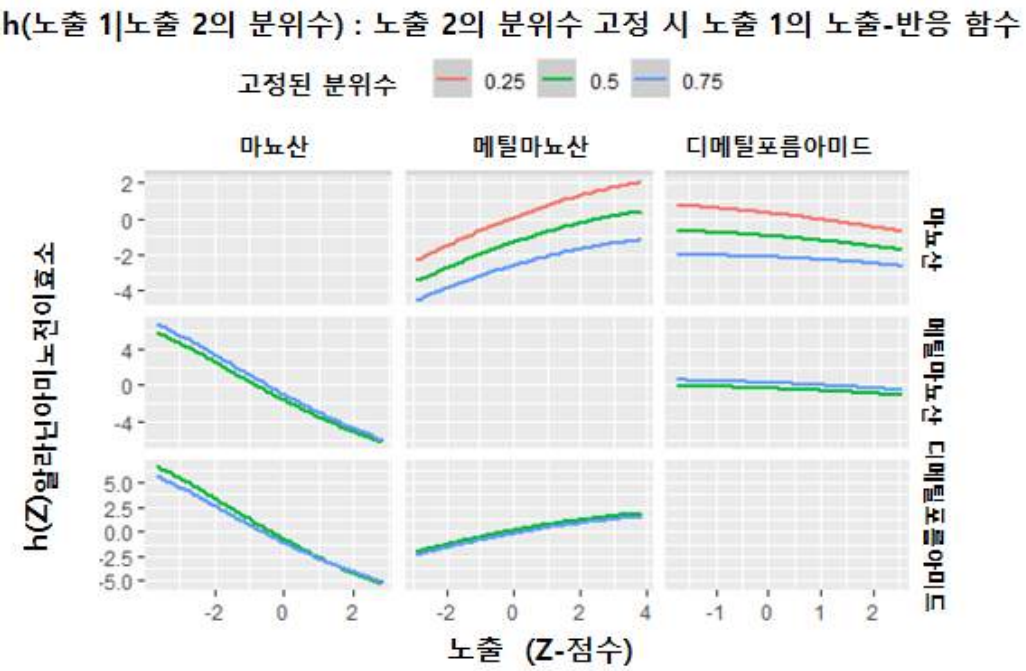
[그림 3-95] 2015년 자료에서 세 가지 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2015년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산과 디메틸포름아미드가 음의 선형에 가까운 연관성을 보였고, 메틸마뇨산은 양의 선형에 가까운 연관성을 보였다.



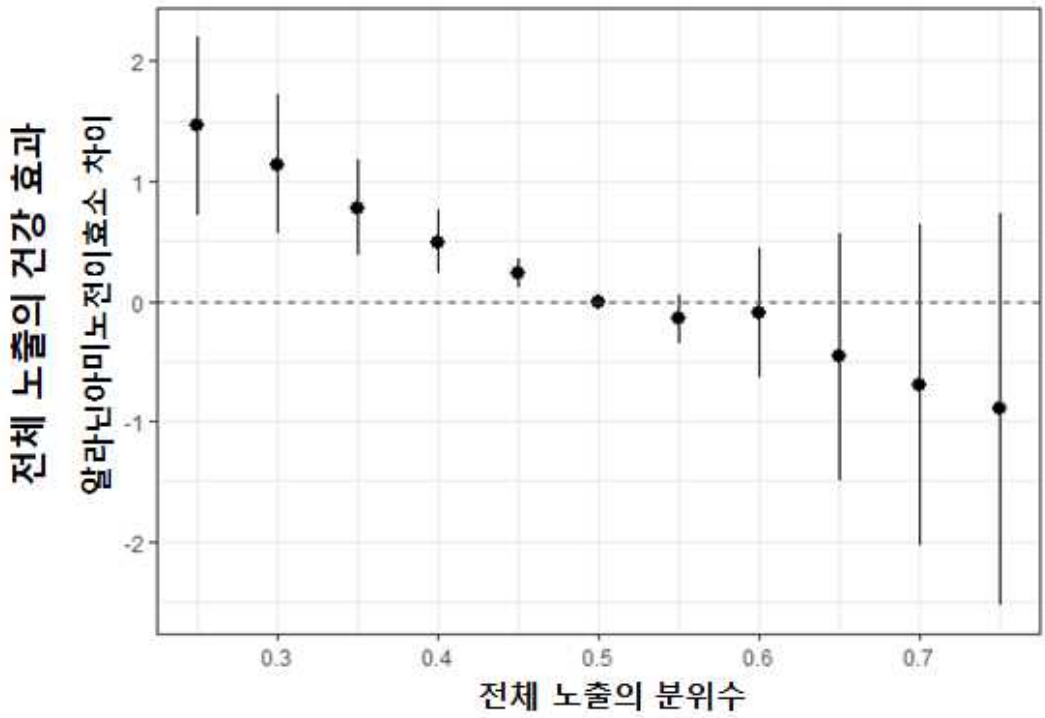
[그림 3-96] 2015년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)간의 단변량 노출-반응 함수

2015년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 마뇨산과 디메틸포름아미드 간에 교호 작용이 존재하는 것으로 나타났다. 메틸마뇨산과 디메틸포름아미드, 마뇨산과 메틸마뇨산 간에는 교호작용이 존재하지 않았다.



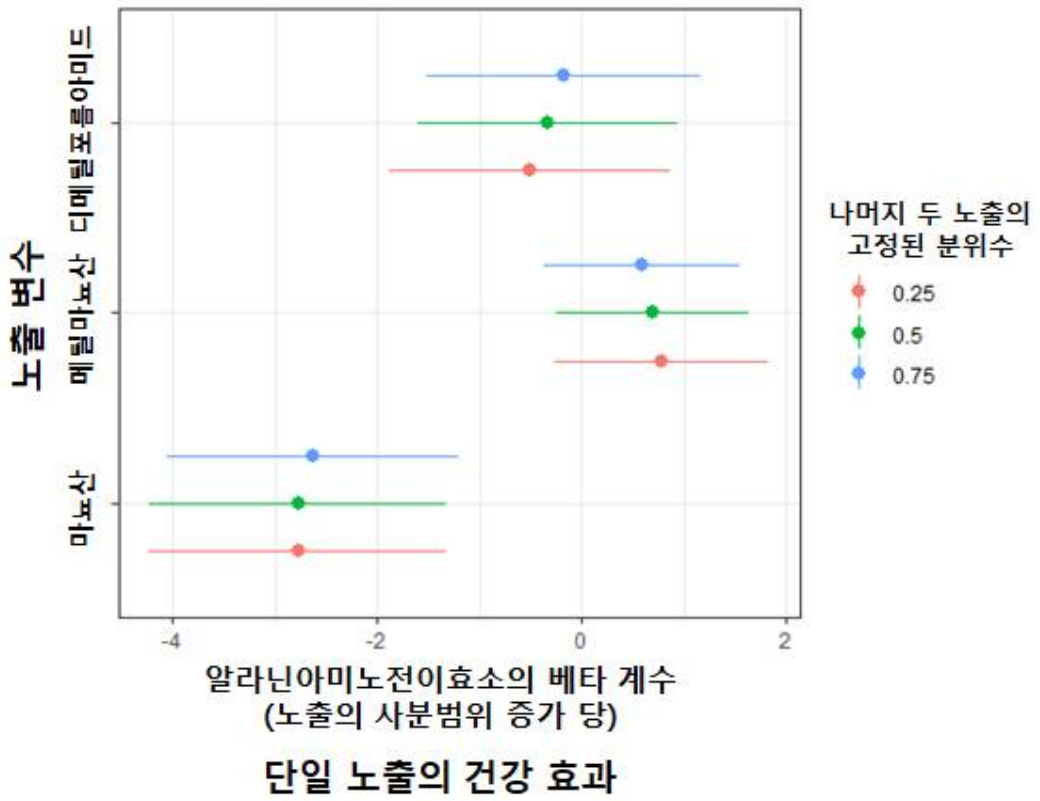
[그림 3-97] 2015년 자료에서 두 가지 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 이변량 노출-반응 함수

2015년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 중위값을 기준으로 감소할수록 혈중 알라닌아미노전이효소가 높아지는 경향이 유의하였다.



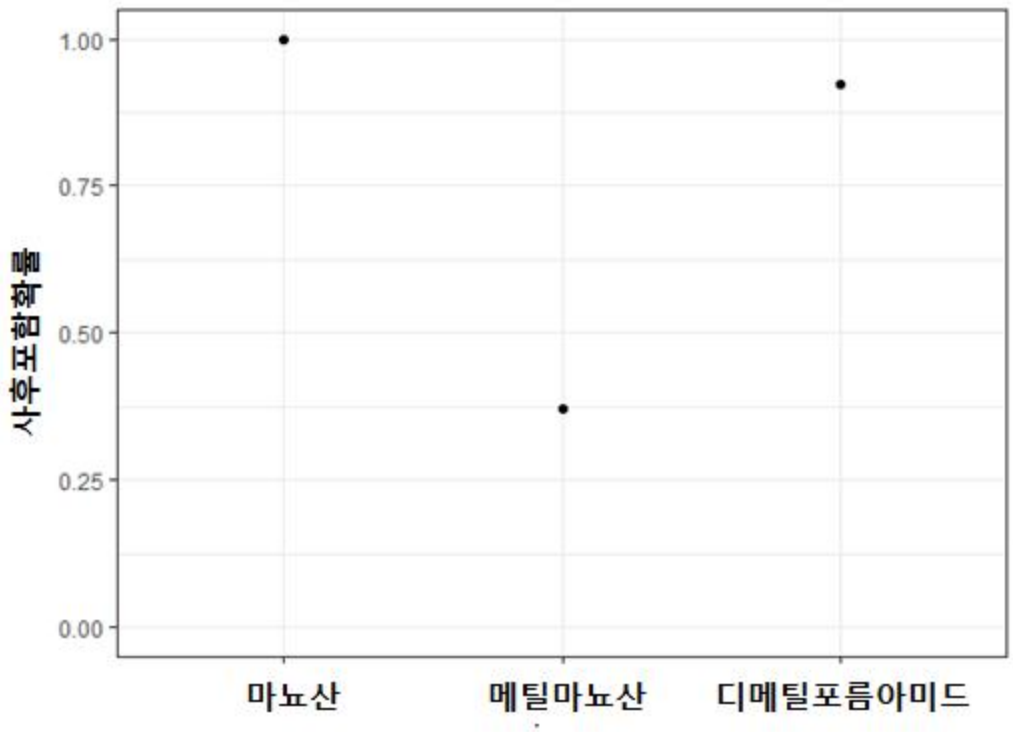
[그림 3-98] 2015년 자료에서 세 가지 노출 물질이 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)에 미치는 복합노출 영향

2015년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 단일 노출 중 알라닌아미노전이효소의 증가와 관련된 노출은 없었다.



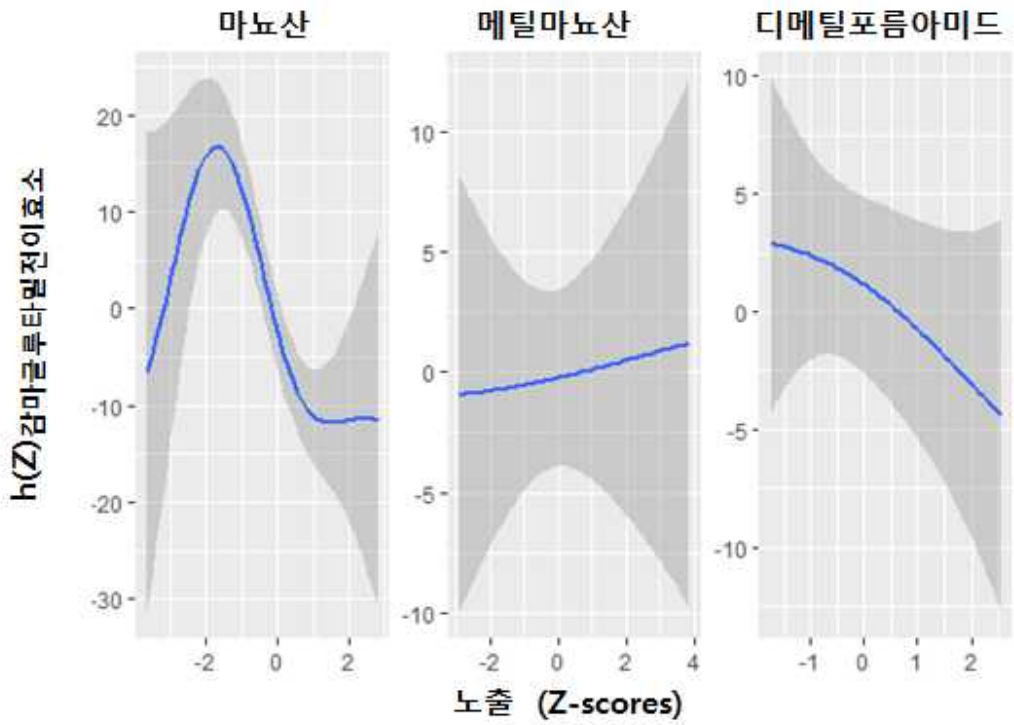
[그림 3-99] 2015년 자료에서 각 노출 물질이 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)에 개별적으로 미치는 영향

2015년 자료에서 세 가지 노출 물질과 혈중 감마글루타밀전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 마뇨산, 디메틸포름아미드, 메틸마뇨산 순으로 높았다.



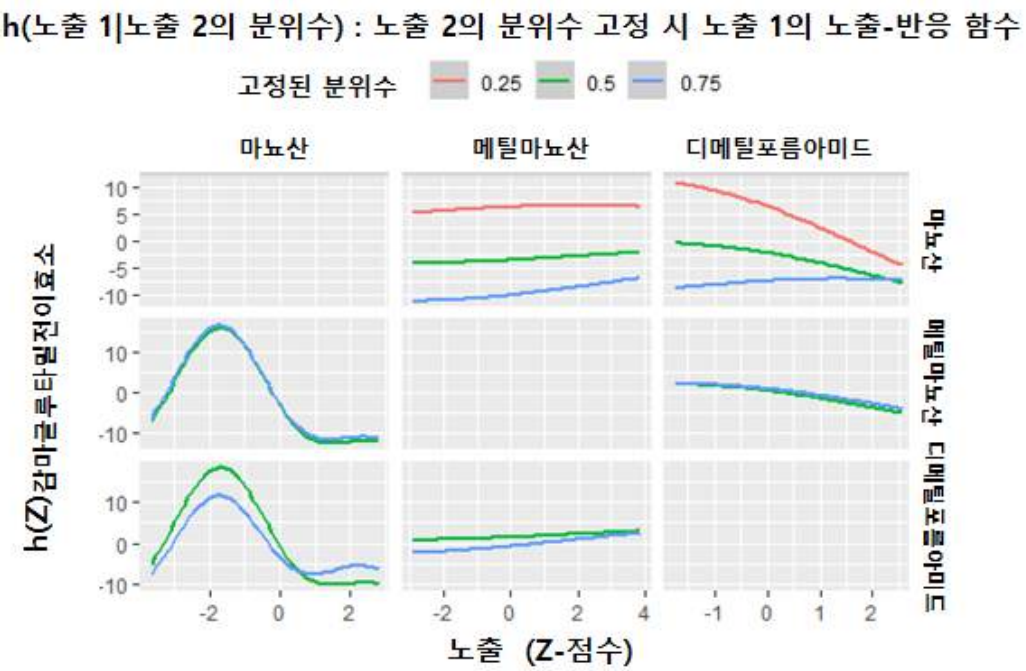
[그림 3-100] 2015년 자료에서 세 가지 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2015년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산이 역 U 패턴을 보였고, 메틸마뇨산은 약한 양의 선형 연관성을 보였으며, 디메틸포름아미드는 음의 선형에 가까운 연관성을 보였다.



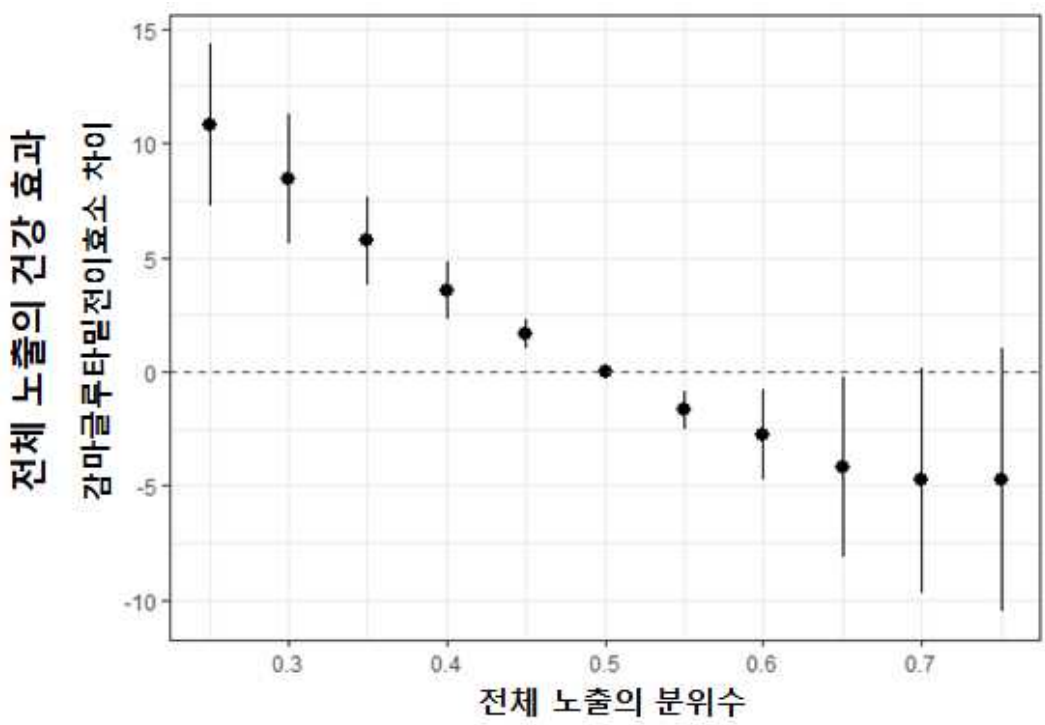
[그림 3-101] 2015년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 단변량 노출-반응 함수

2015년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 마뇨산과 디메틸포름아미드, 메틸마뇨산과 디메틸포름아미드 간에 교호 작용이 존재하는 것으로 나타났다. 마뇨산과 메틸마뇨산은 간에는 교호작용이 존재하지 않았다.



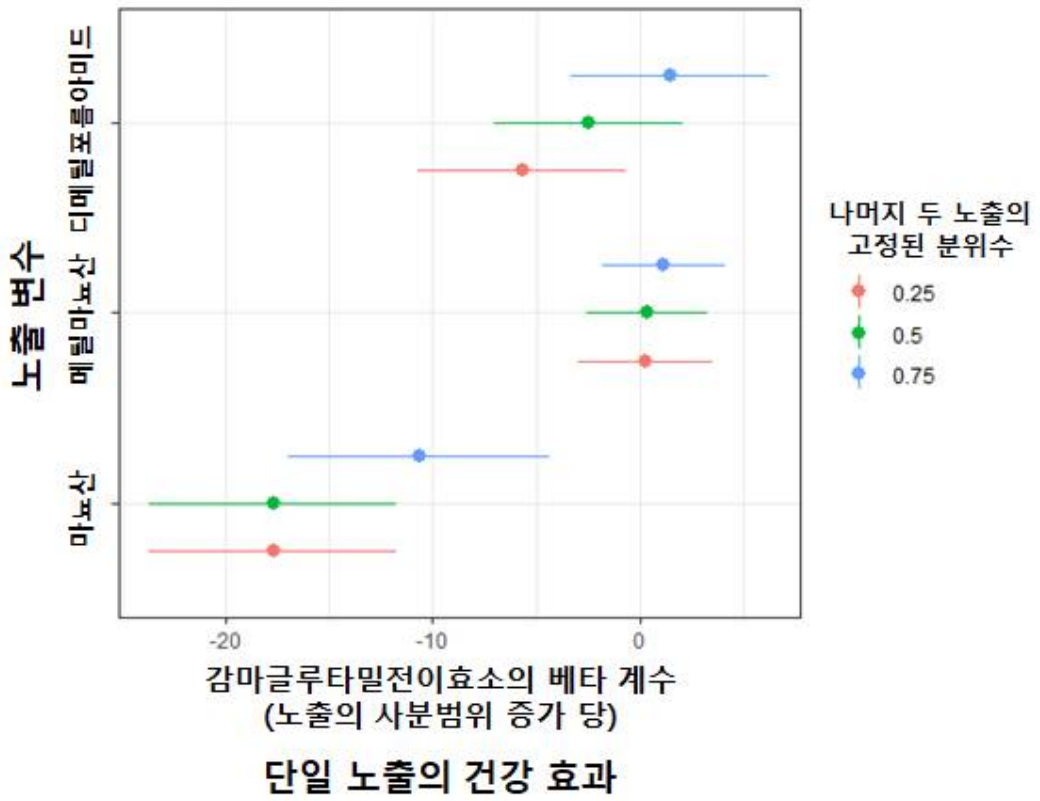
[그림 3-102] 2015년 자료에서 두 가지 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 이변량 노출-반응 함수

2015년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 중위값을 기준으로 감소할수록 혈중 감마글루타밀전이효소가 높아지는 경향이 유의하였다.



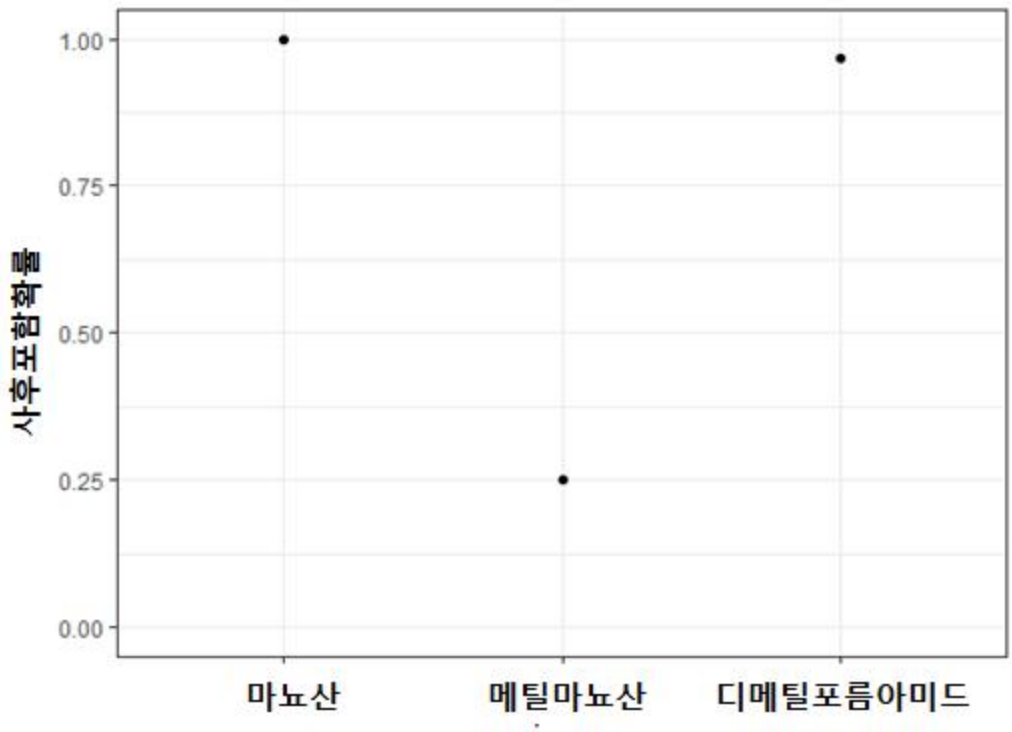
[그림 3-103] 2015년 자료에서 세 가지 노출 물질이 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)에 미치는 복합노출 영향

2015년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 단일 노출 중 감마글루타밀전이효소의 증가와 관련된 노출은 없었다.



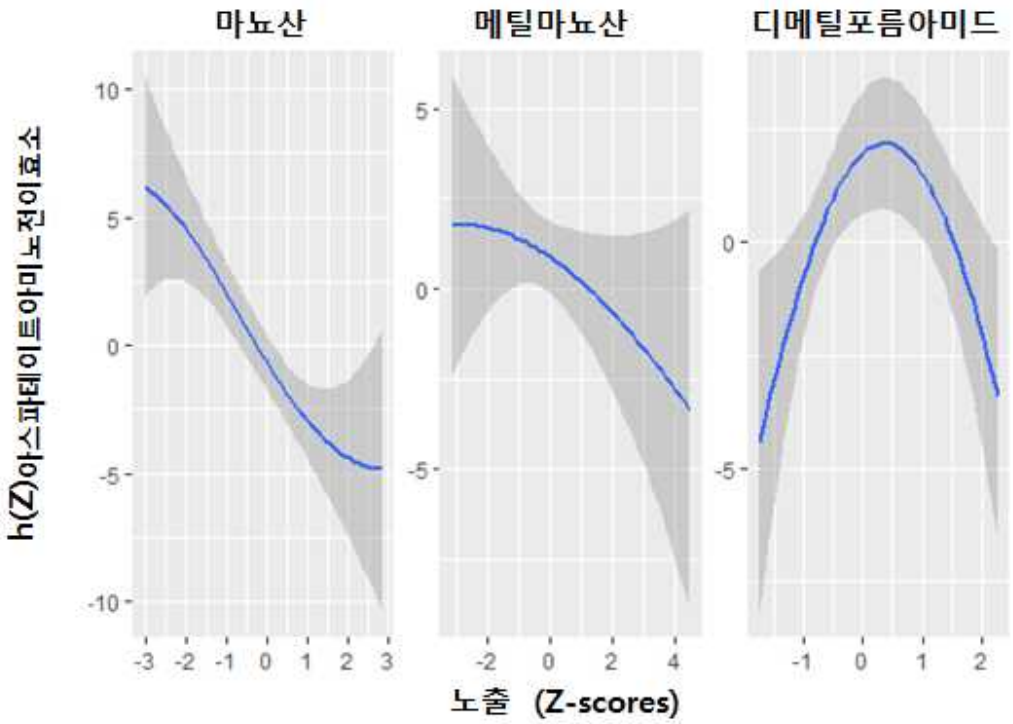
[그림 3-104] 2015년 자료에서 각 노출 물질이 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)에 개별적으로 미치는 영향

2016년 자료에서 세 가지 노출 물질과 혈중 아스파테이트아미노전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 마뇨산, 디메틸포름아미드, 메틸마뇨산 순으로 높았다.



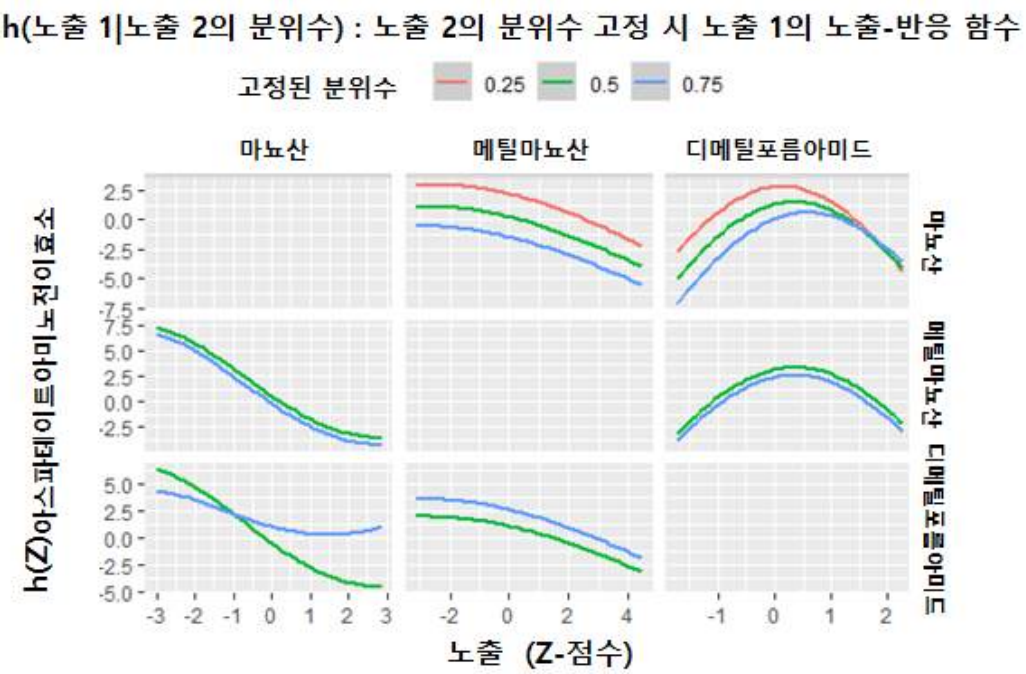
[그림 3-105] 2016년 자료에서 세 가지 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2016년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산과 메틸마뇨산이 음의 선형에 가까운 연관성을 보였고, 디메틸포름아미드는 역 U 패턴에 가까운 연관성을 보였다.



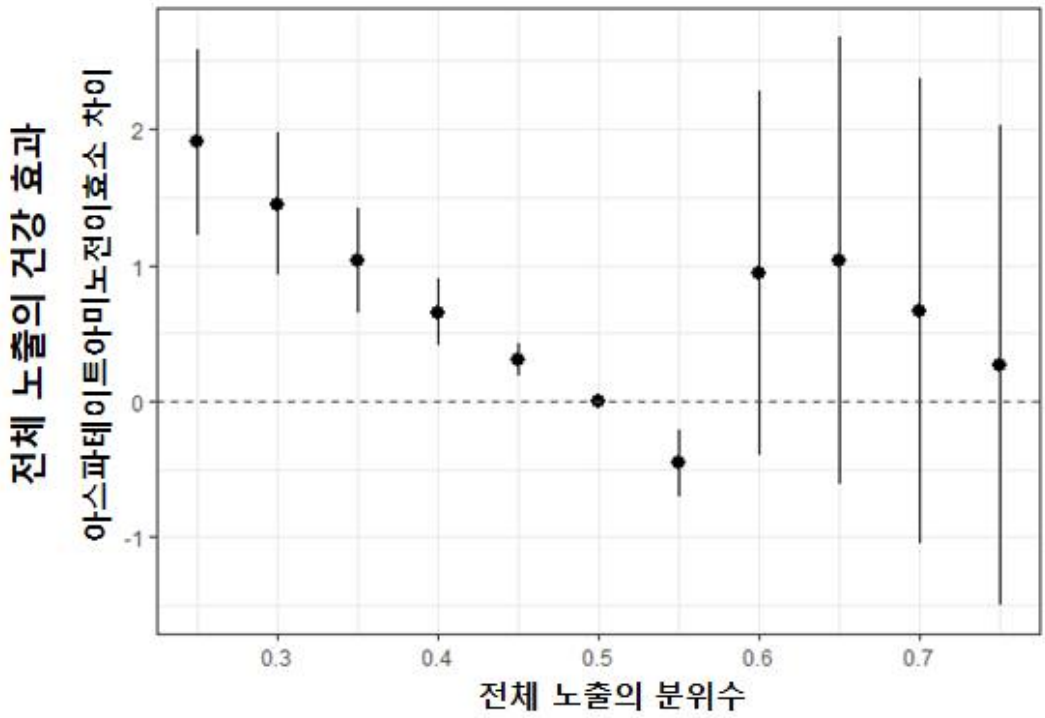
[그림 3-106] 2016년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 단변량 노출-반응 함수

2016년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 마뇨산과 디메틸포름아미드 간에 교호 작용이 존재하는 것으로 나타났다. 메틸마뇨산과 디메틸마뇨산, 마뇨산과 메틸마뇨산 간에는 교호작용이 존재하지 않았다.



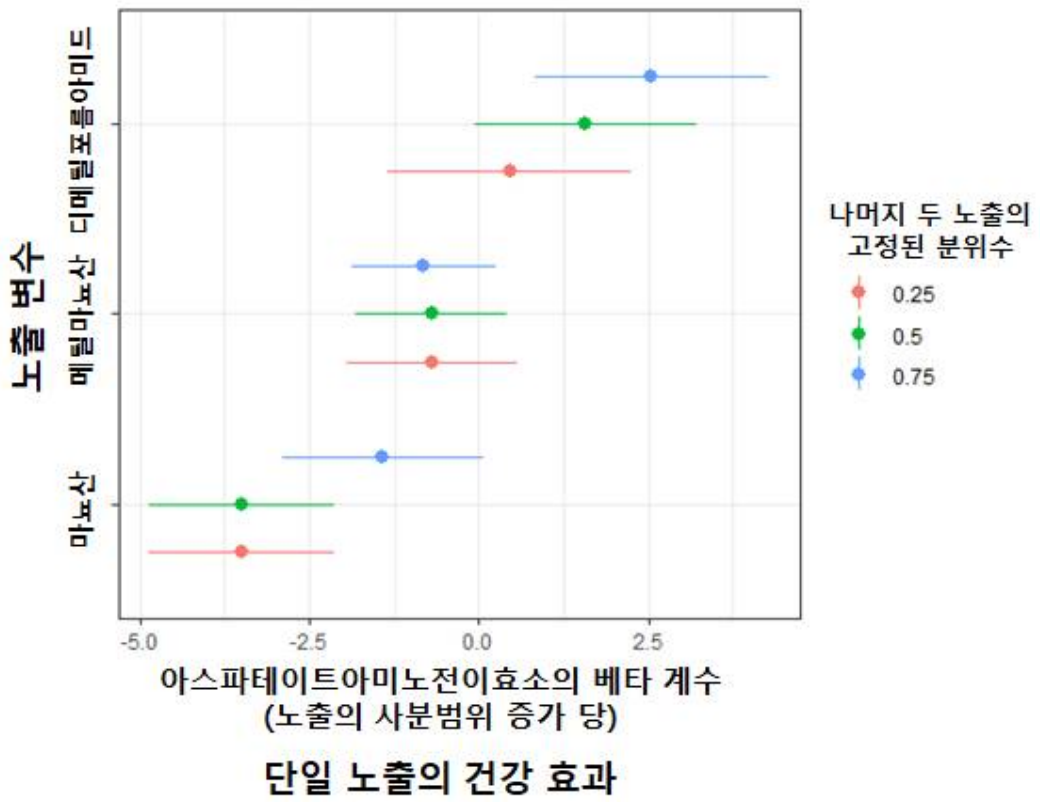
[그림 3-107] 2016년 자료에서 두 가지 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 이변량 노출-반응 함수

2016년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 중위값을 기준으로 감소할수록 혈중 아스파테이트아미노전이효소가 높아지는 경향이 유의하였다.



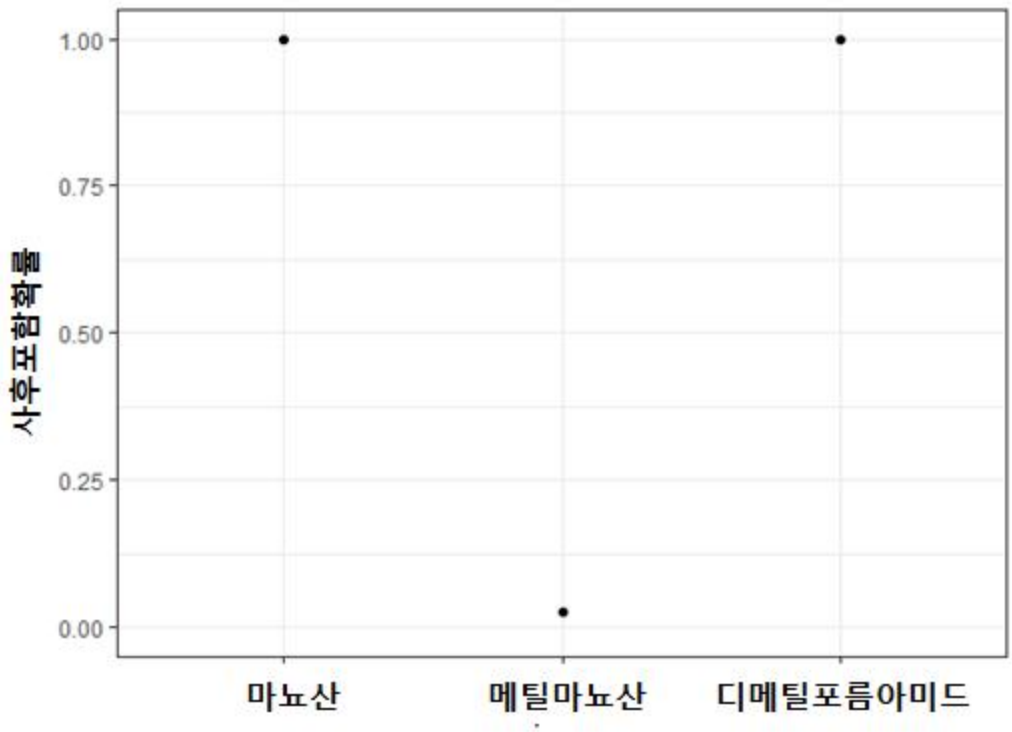
[그림 3-108] 2016년 자료에서 세 가지 노출 물질이 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)에 미치는 복합노출 영향

2016년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 디메틸포름아미드는 다른 노출이 50 백분위수 또는 75 백분위수 일 때 아스파테이트아미노전이효소의 높은 수준과 유의한 관련이 있었으며, 디메틸포름아미드의 단일 노출 추정 값은 25 백분위수로 고정되었을 때에 비해 다른 노출이 50 백분위수나 75 백분위수로 고정되었을 때 크기가 더 커서 디메틸포름아미드와 나머지 다른 노출 중 하나 이상이 교호 작용이 있음을 알 수 있었다.



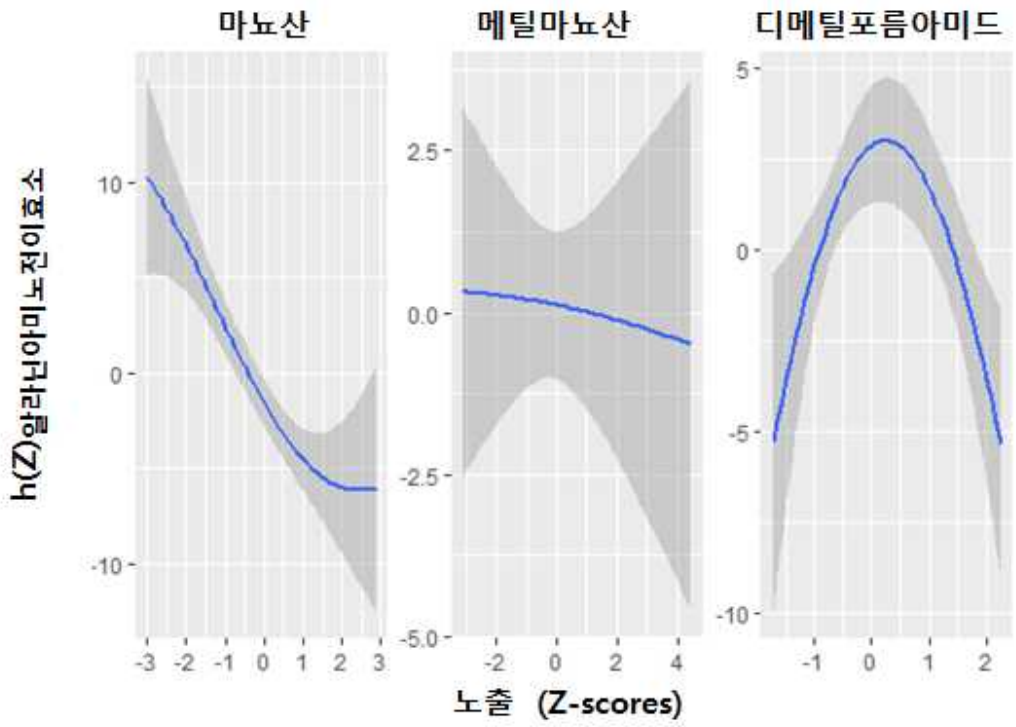
[그림 3-109] 2016년 자료에서 각 노출 물질이 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)에 개별적으로 미치는 영향

2016년 자료에서 세 가지 노출 물질과 혈중 알라닌아미노전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 마뇨산, 디메틸포름아미드, 메틸마뇨산 순으로 높았다.



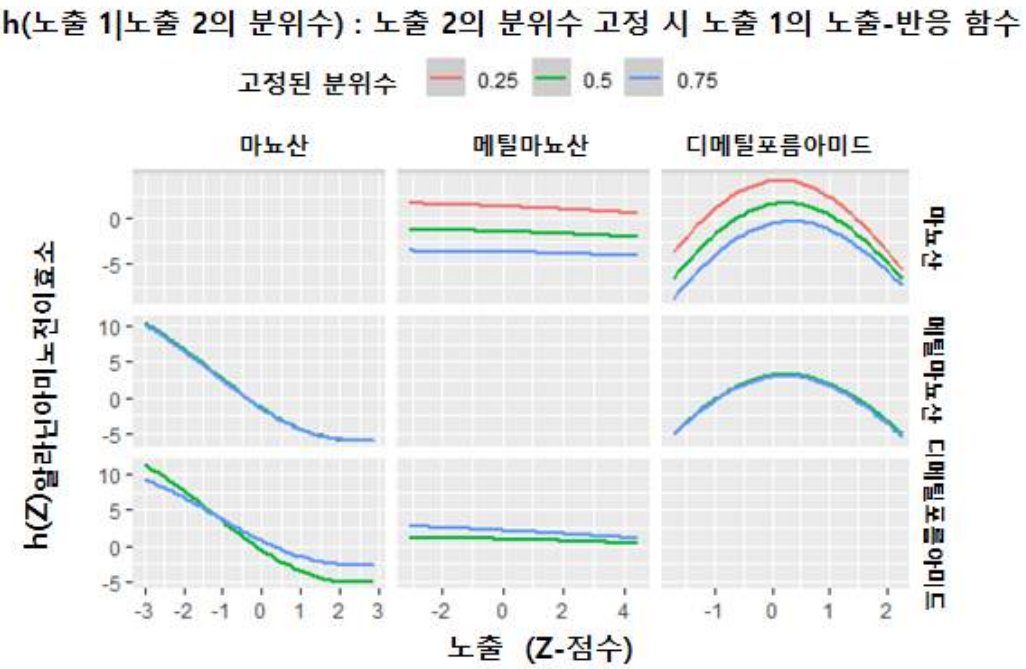
[그림 3-110] 2016년 자료에서 세 가지 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2016년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산과 메틸마뇨산이 음의 선형에 가까운 연관성을 보였고, 디메틸포름아미드는 역 U 패턴을 보였다.



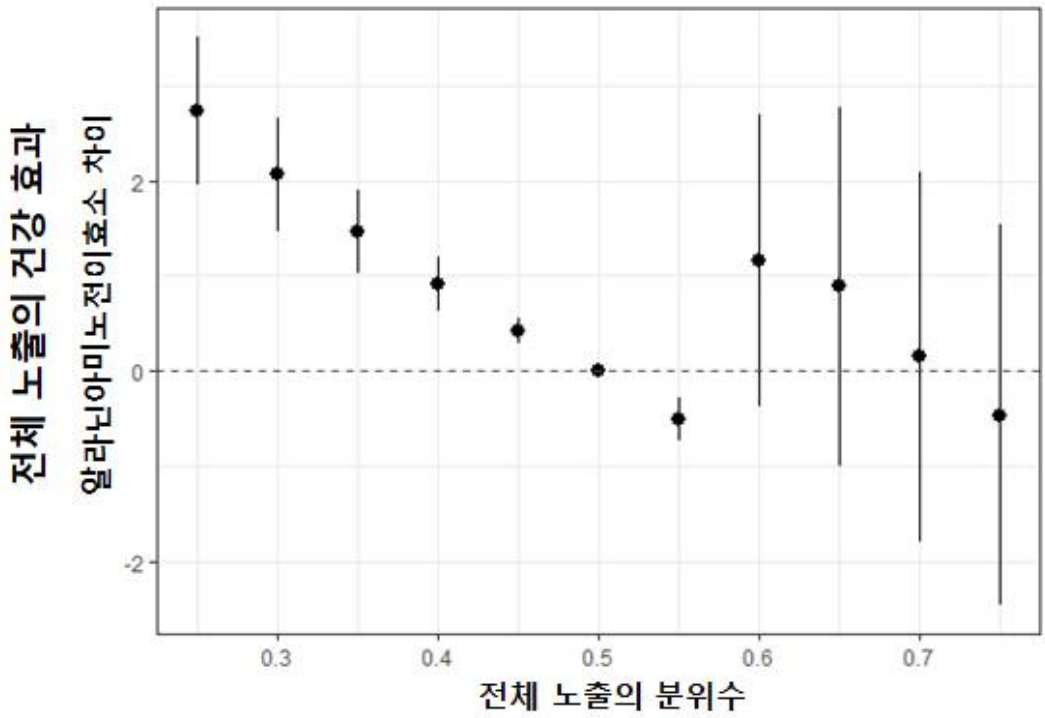
[그림 3-111] 2016년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)간의 단변량 노출-반응 함수

2016년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 마뇨산과 디메틸포름아미드 간에 교호 작용이 존재하는 것으로 나타났다. 메틸마뇨산과 디메틸포름아미드, 마뇨산과 메틸마뇨산 간에는 교호작용이 존재하지 않았다.



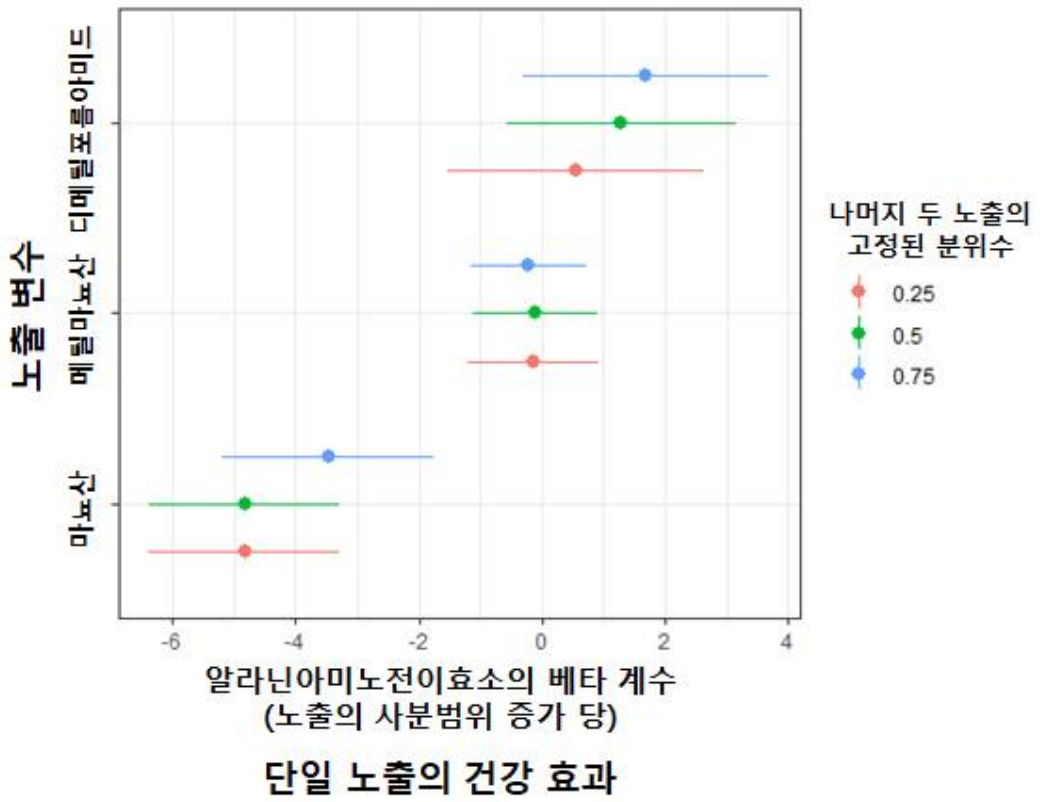
[그림 3-112] 2016년 자료에서 두 가지 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 이변량 노출-반응 함수

2016년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 중위값을 기준으로 감소할수록 혈중 알라닌아미노전이효소가 높아지는 경향이 유의하였다.



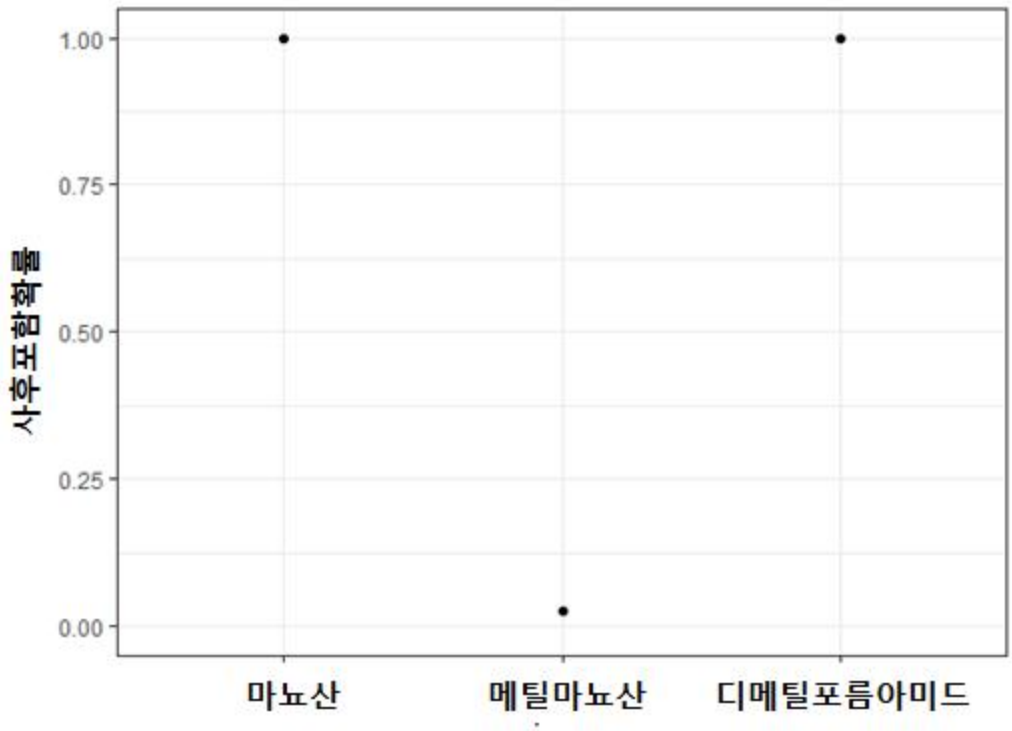
[그림 3-113] 2016년 자료에서 세 가지 노출 물질이 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)에 미치는 복합노출 영향

2016년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 단일 노출 중 알라닌아미노전이효소의 증가와 관련된 노출은 없었다.



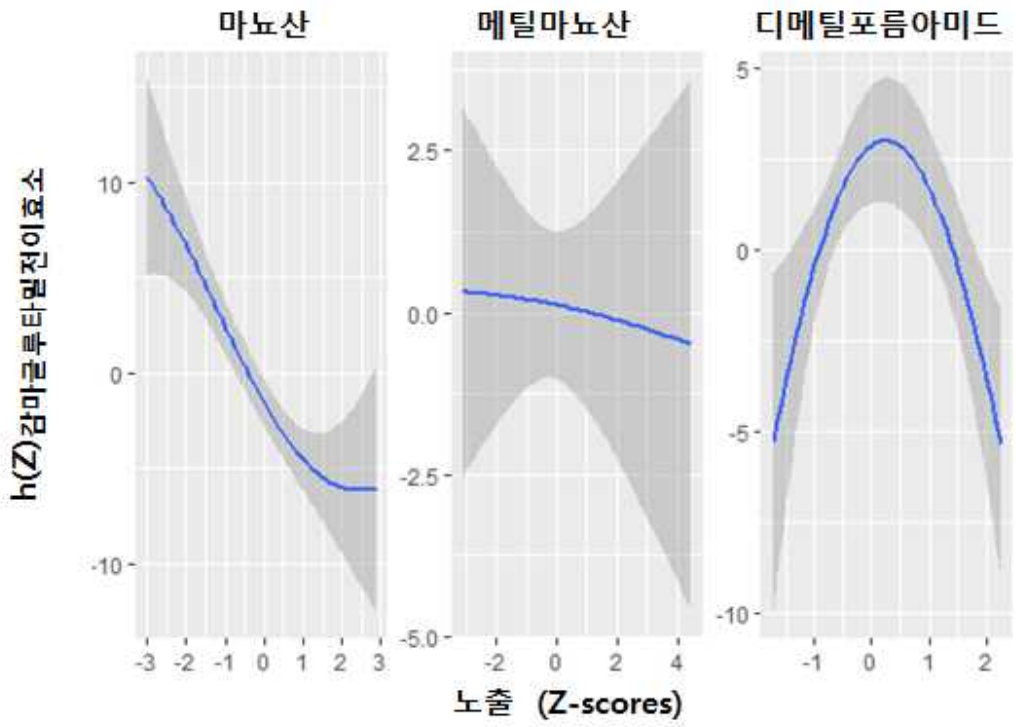
[그림 3-114] 2016년 자료에서 각 노출 물질이 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)에 개별적으로 미치는 영향

2016년 자료에서 세 가지 노출 물질과 혈중 감마글루타밀전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 마뇨산, 디메틸포름아미드, 메틸마뇨산 순으로 높았다.



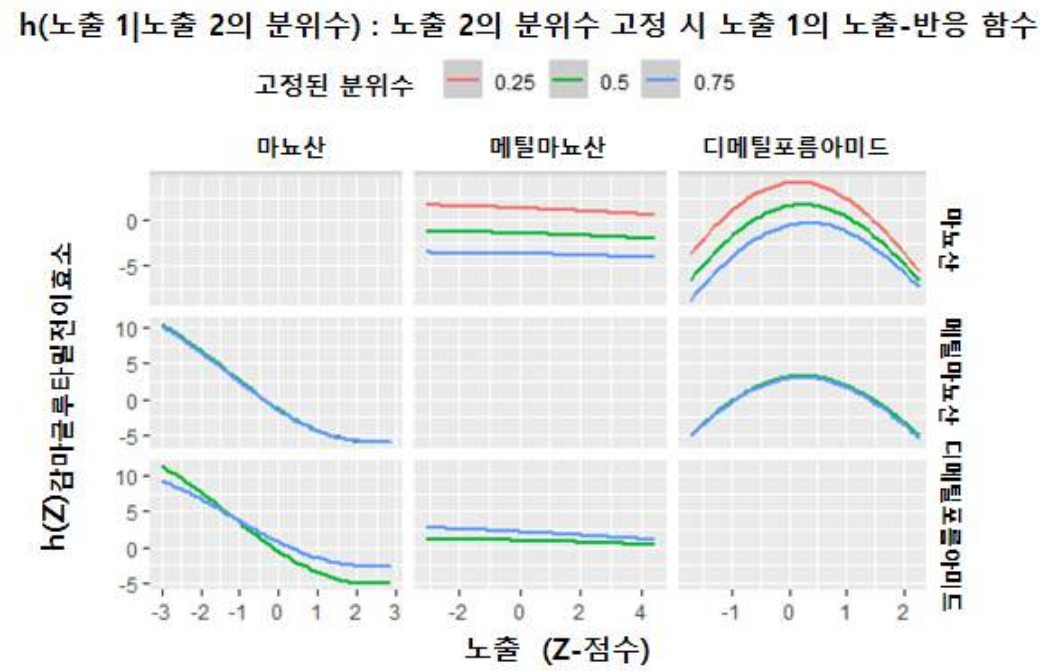
[그림 3-115] 2016년 자료에서 세 가지 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2016년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산과 메틸마뇨산이 음의 선형에 가까운 연관성을 보였고, 디메틸포름아미드는 역 U 패턴을 보였다.



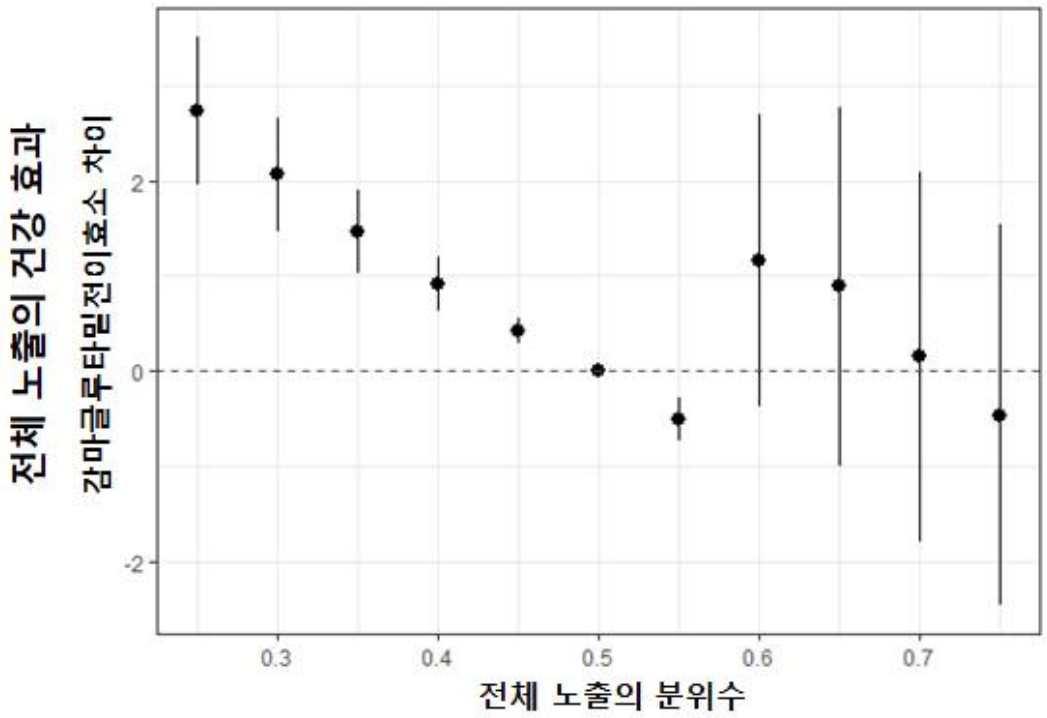
[그림 3-116] 2016년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 단변량 노출-반응 함수

2016년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 마뇨산과 디메틸포름아미드 간에 교호 작용이 존재하는 것으로 나타났다. 메틸마뇨산과 디메틸포름아미드, 마뇨산과 메틸마뇨산 간에는 교호작용이 존재하지 않았다.



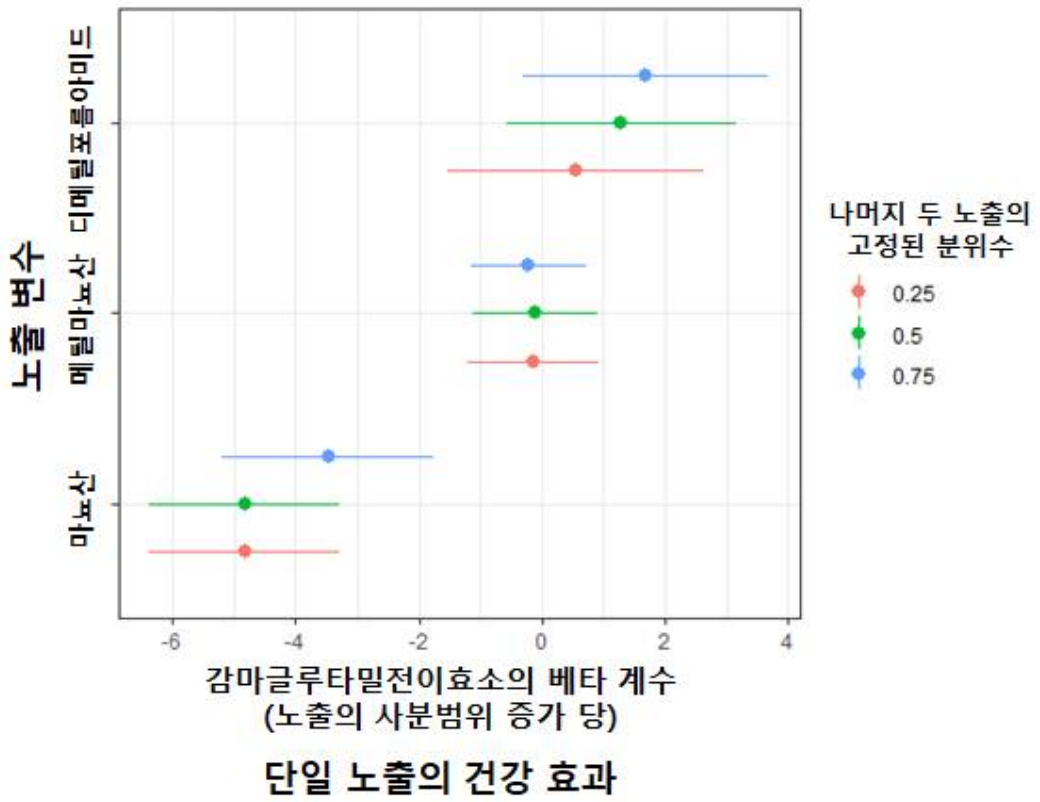
[그림 3-117] 2016년 자료에서 두 가지 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 이변량 노출-반응 함수

2016년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 중위값을 기준으로 감소할수록 혈중 감마글루타밀전이효소가 높아지는 경향이 유의하였다.



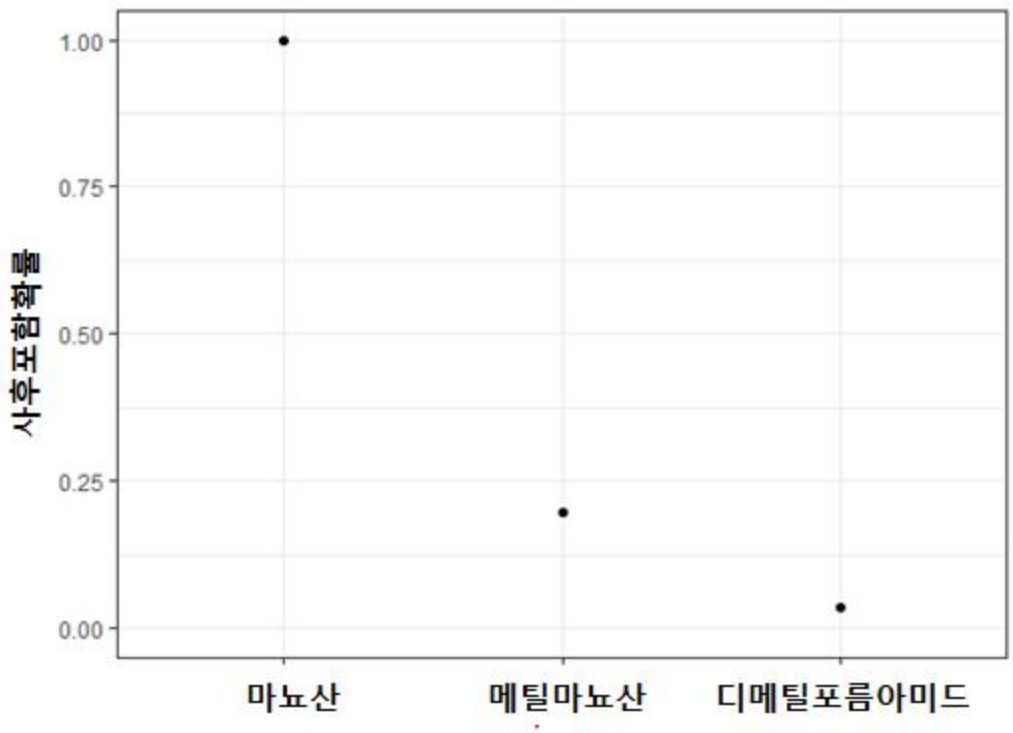
[그림 3-118] 2016년 자료에서 세 가지 노출 물질이 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)에 미치는 복합노출 영향

2016년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 단일 노출 중 감마글루타밀전이효소의 증가와 관련된 노출은 없었다.



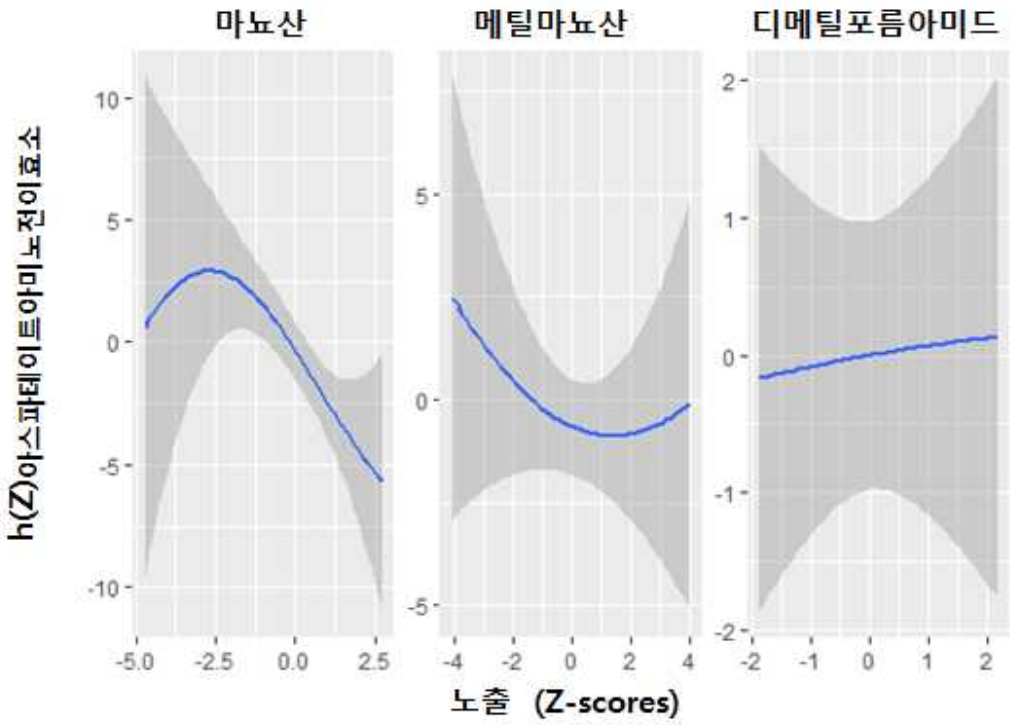
[그림 3-119] 2016년 자료에서 각 노출 물질이 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)에 개별적으로 미치는 영향

2017년 자료에서 세 가지 노출 물질과 혈중 아스파테이트아미노전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 마뇨산, 메틸마뇨산, 디메틸포름아미드 순으로 높았다.



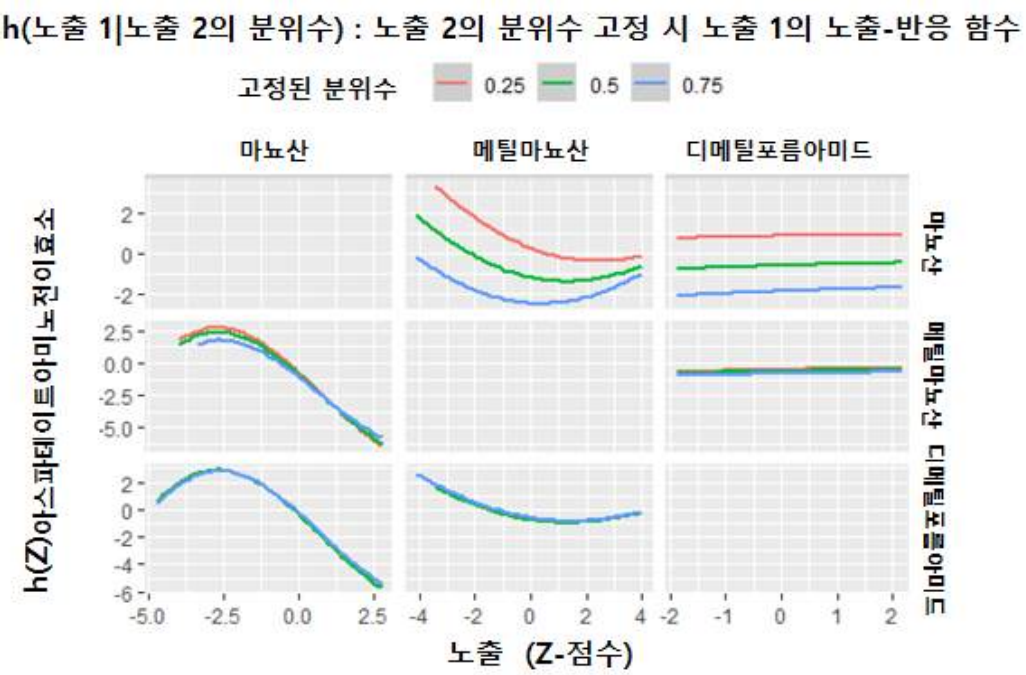
[그림 3-120] 2017년 자료에서 세 가지 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2017년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산아 역 U 패턴을 보였고, 메틸마뇨산이 U 패턴을 보였으며, 디메틸포름아미드는 양의 선형 연관성을 보였다.



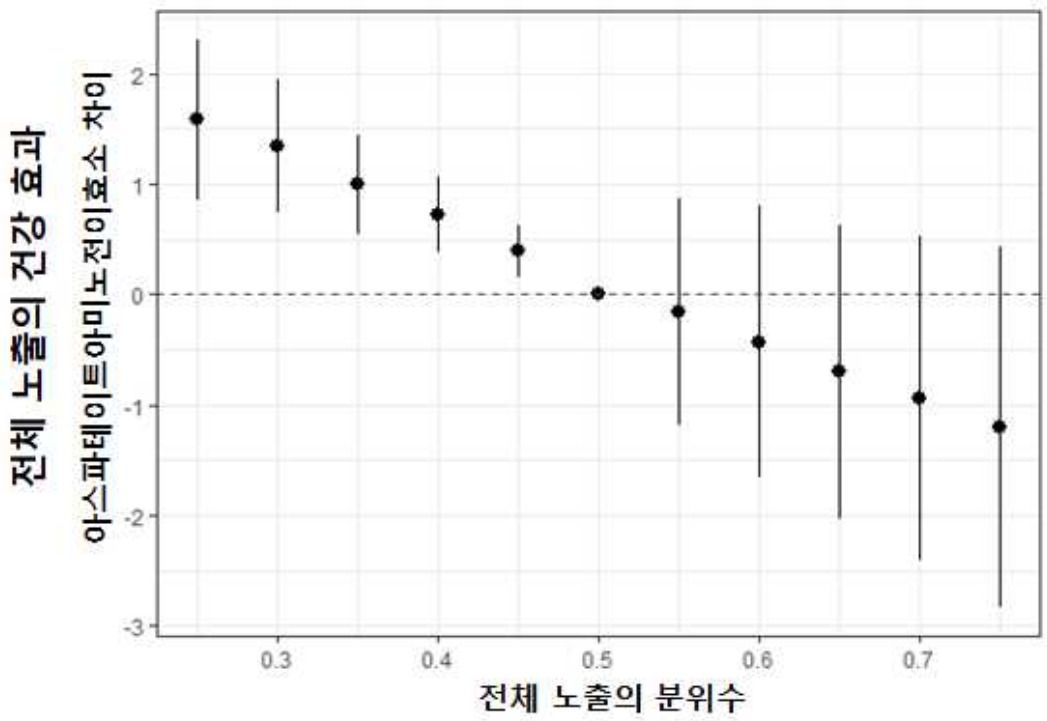
[그림 3-121] 2017년 자료에서 각 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 단변량 노출-반응 함수

2017년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 마뇨산과 메틸마뇨산, 마뇨산과 디메틸포름아미드 간에 교호 작용이 존재하는 것으로 나타났다. 메틸마뇨산과 디메틸포름아미드 간에는 교호작용이 존재하지 않았다.



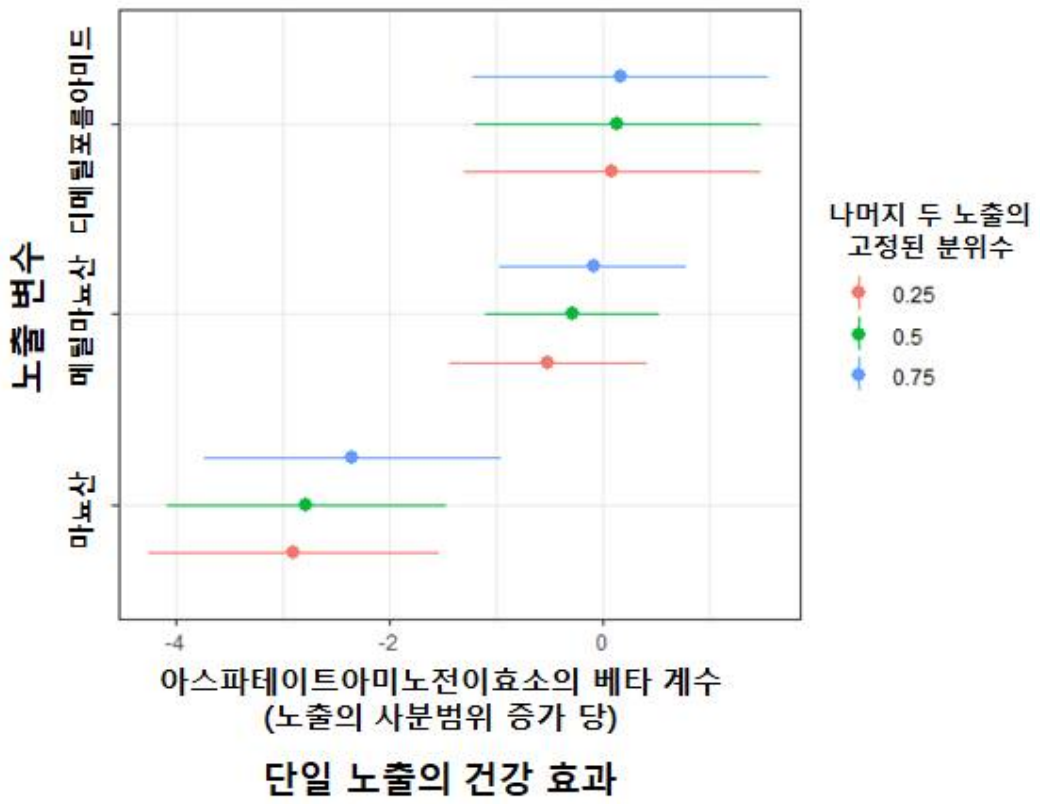
[그림 3-122] 2017년 자료에서 두 가지 노출 물질과 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST) 간의 이변량 노출-반응 함수

2017년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 중위값을 기준으로 감소할수록 혈중 아스파테이트아미노전이효소가 높아지는 경향이 유의하였다.



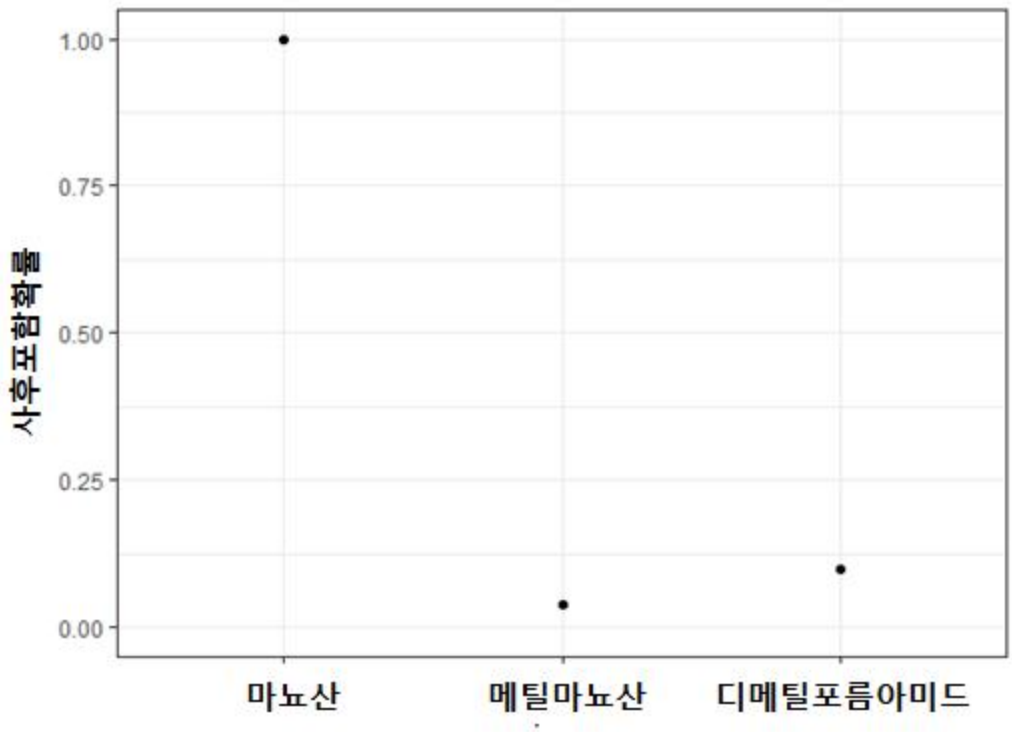
[그림 3-123] 2017년 자료에서 세 가지 노출 물질이 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)에 미치는 복합노출 영향

2017년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 단일 노출 중 아스파테이트아미노전이효소의 증가와 관련된 노출은 없었다.



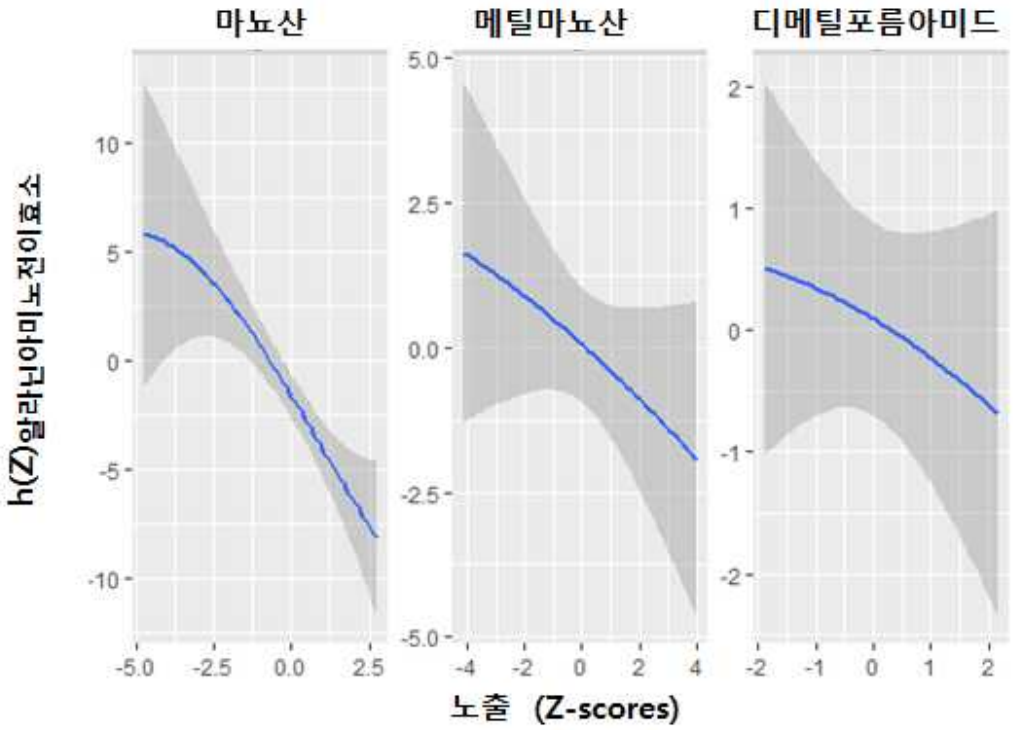
[그림 3-124] 2017년 자료에서 각 노출 물질이 혈중 아스파테이트아미노전이효소(aspartate transaminase, AST)에 개별적으로 미치는 영향

2017년 자료에서 세 가지 노출 물질과 혈중 알라닌아미노전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 마뇨산, 디메틸포름아미드, 메틸마뇨산 순으로 높았다.



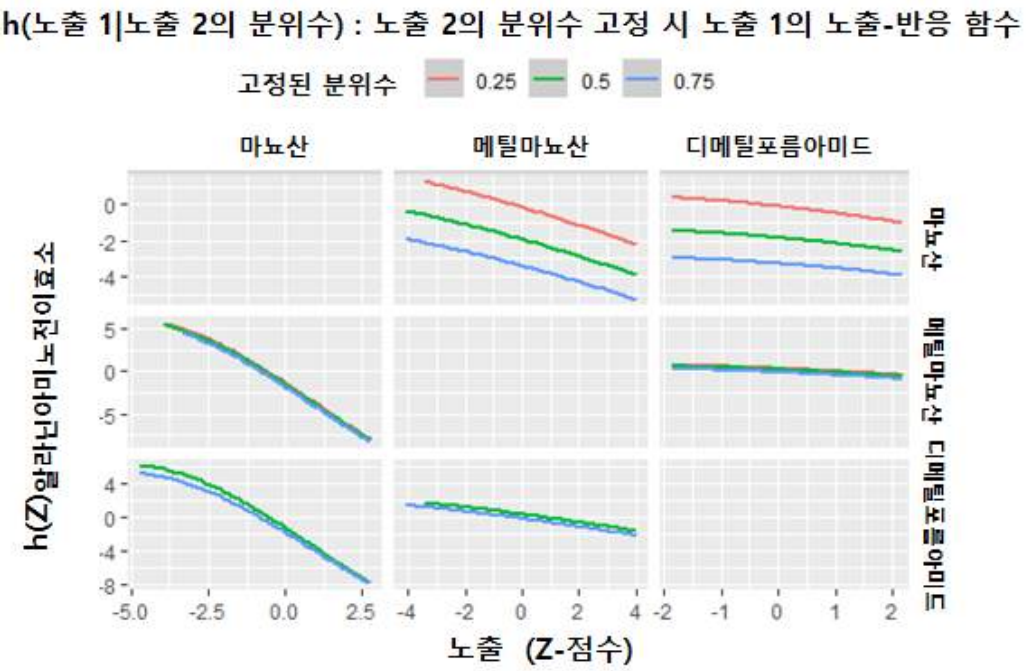
[그림 3-125] 2017년 자료에서 세 가지 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2017년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산, 메틸마뇨산, 디메틸포름아미드는 음의 선형에 가까운 연관성을 보였다.



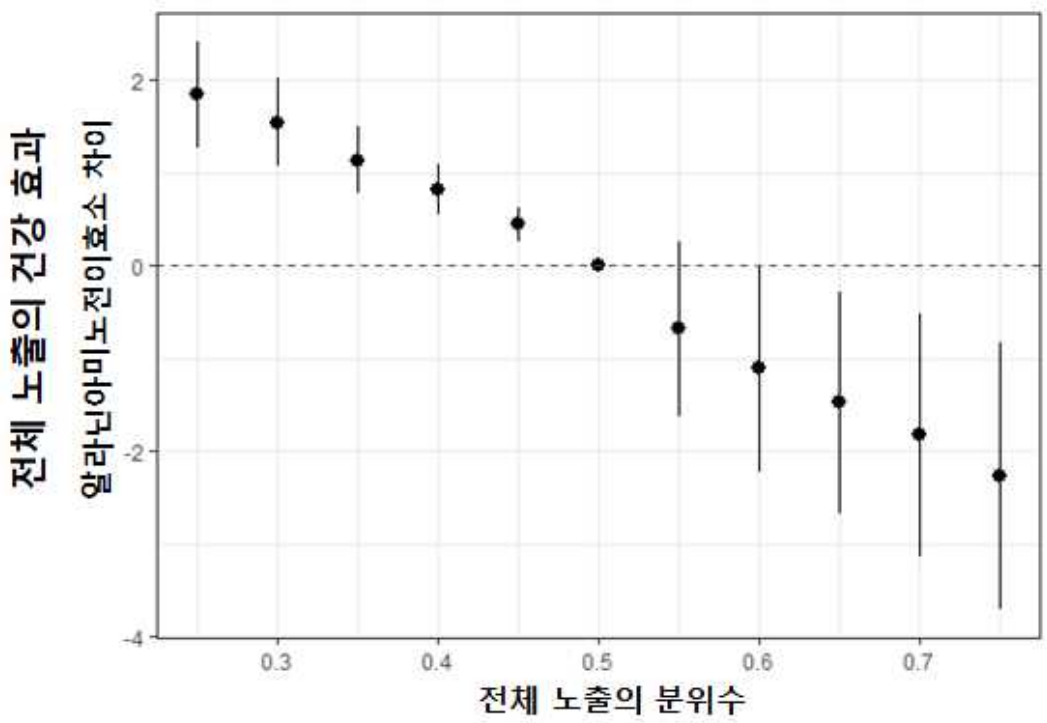
[그림 3-126] 2017년 자료에서 각 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)간의 단변량 노출-반응 함수

2017년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 마뇨산과 메틸마뇨산, 마뇨산과 디메틸포름아미드 간에 교호 작용이 존재하지 않았다.



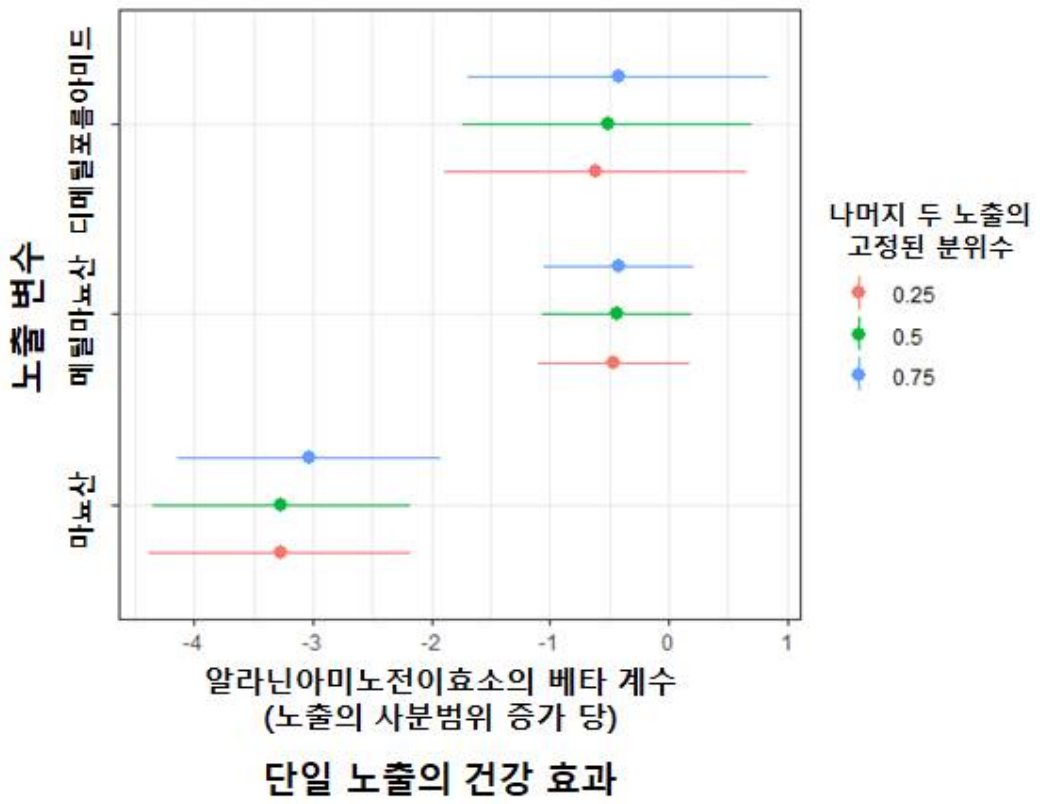
[그림 3-127] 2017년 자료에서 두 가지 노출 물질과 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT) 간의 이변량 노출-반응 함수

2017년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 증가할수록 혈중 알라닌아미노전이효소가 감소하는 경향이 유의하였다.



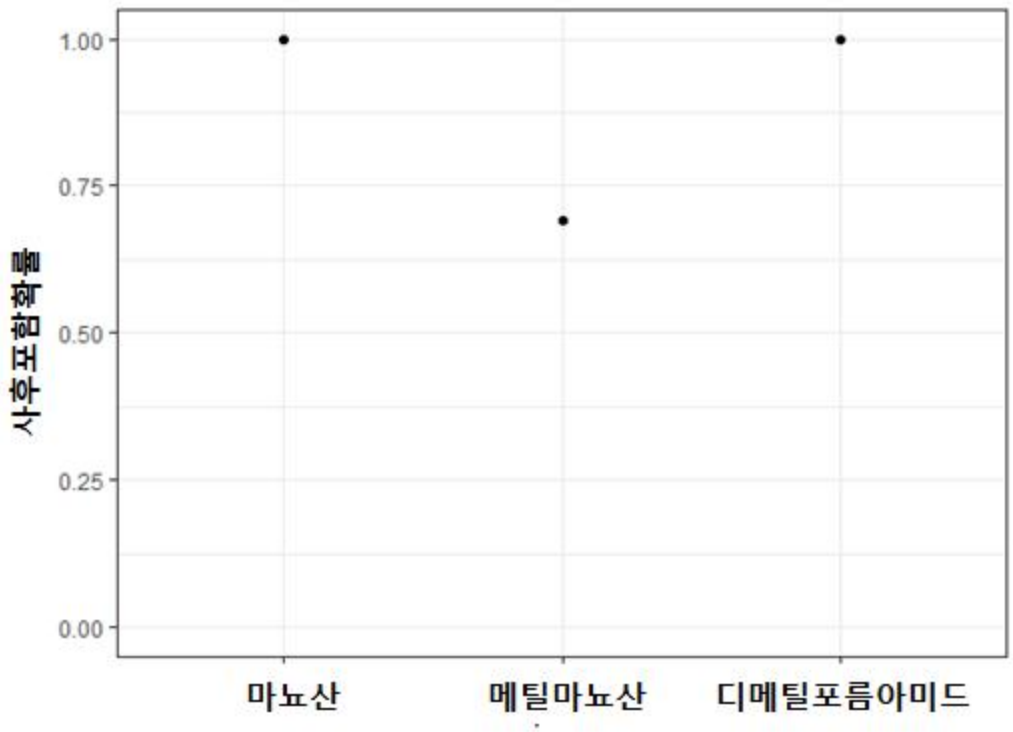
[그림 3-128] 2017년 자료에서 세 가지 노출 물질이 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)에 미치는 복합노출 영향

2017년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 단일 노출 중 알라닌아미노전이효소의 증가와 관련된 노출은 없었다.



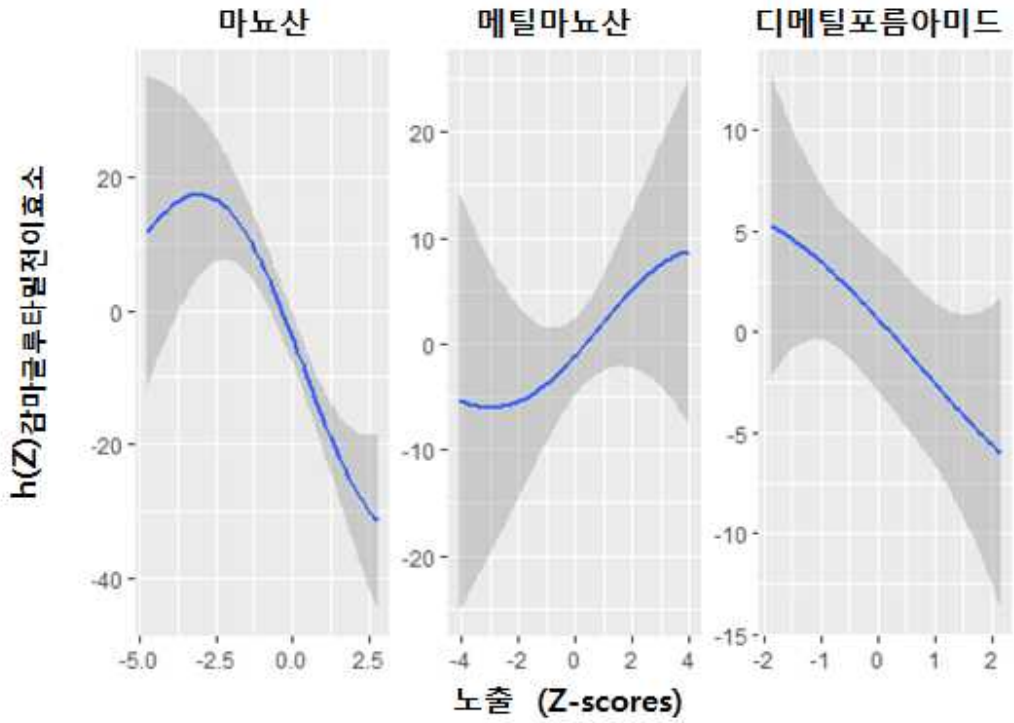
[그림 3-129] 2017년 자료에서 각 노출 물질이 혈중 알라닌아미노전이효소(alanine aminotransferase, ALT)에 개별적으로 미치는 영향

2017년 자료에서 세 가지 노출 물질과 혈중 감마글루타밀전이효소 간의 베이지안커널머신 회귀분석 후 각 노출물질의 사후포함확률을 산출하였으며, 이는 모델 내에서 각 물질의 중요도로 해석할 수 있다. 그 결과 사후포함확률은 마뇨산, 디메틸포름아미드, 메틸마뇨산 순으로 높았다.



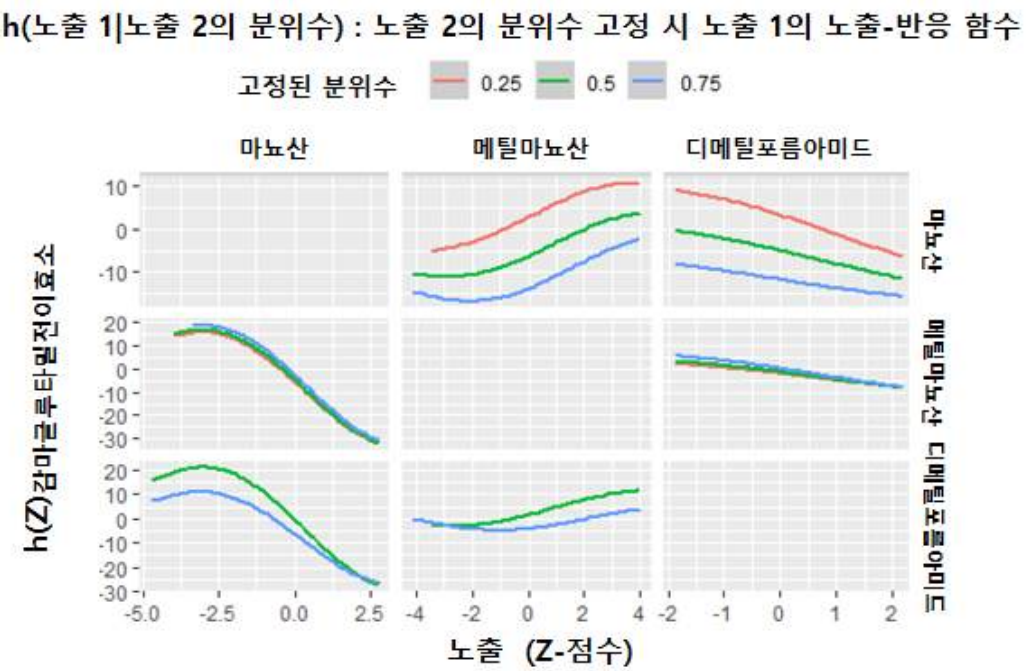
[그림 3-130] 2017년 자료에서 세 가지 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)의 베이지안커널머신 회귀 분석 후 각 노출 물질의 사후포함확률(Posterior inclusion probability, PIP)

2017년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소 간의 비선형 관계를 확인하기 위해 각 노출 물질의 노출-반응 함수를 살펴보면(다른 두 노출은 50분위수로 고정), 마뇨산은 역 U 패턴에 가까웠고, 메틸마뇨산은 양의 선형에 가까운 연관성, 디메틸포름아미드는 음의 선형에 가까운 연관성을 보였다.



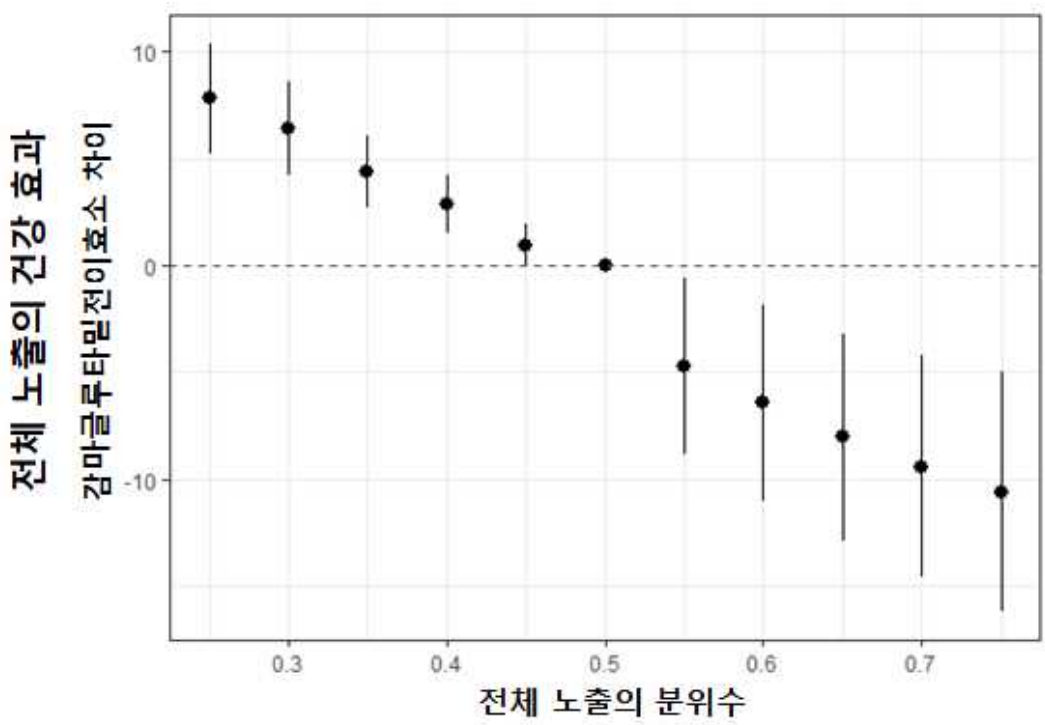
[그림 3-131] 2017년 자료에서 각 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 단변량 노출-반응 함수

2017년 자료에서 위와 같이 노출 물질 사이의 교호작용 효과를 확인해 보았을 때, 마뇨산과 디메틸포름아미드, 메틸마뇨산과 디메틸포름아미드 간에 교호 작용이 존재하였고, 마뇨산과 메틸마뇨산 간에는 교호 작용이 존재하지 않았다.



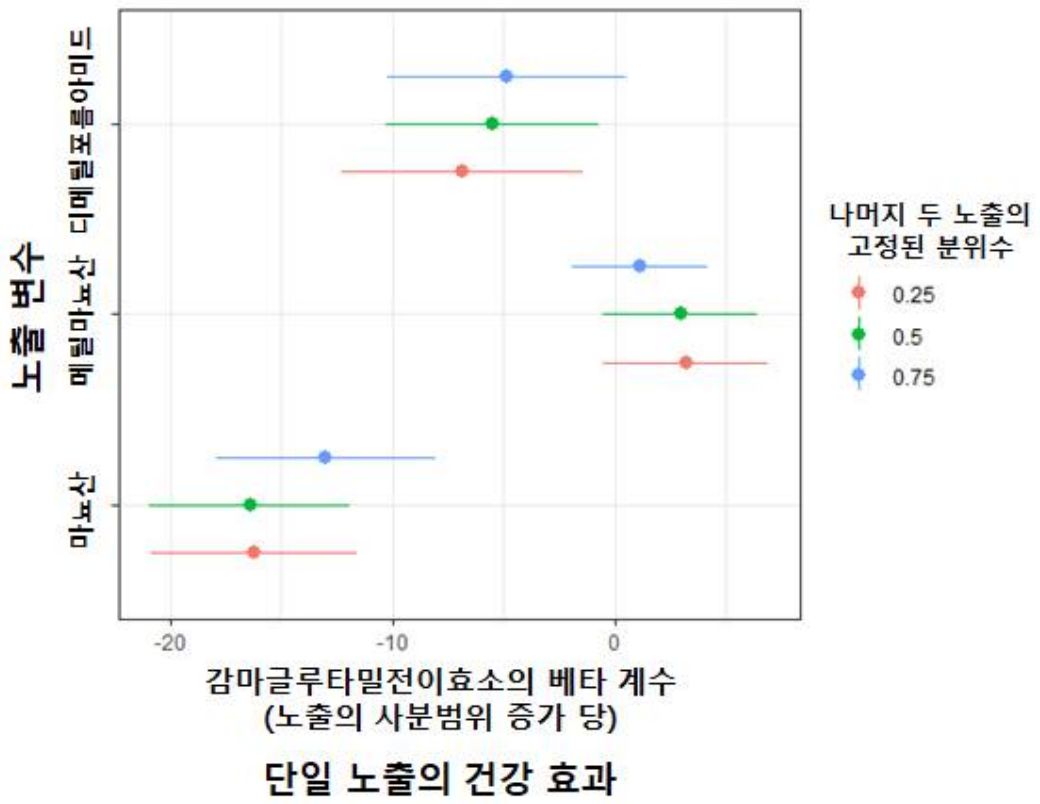
[그림 3-132] 2017년 자료에서 두 가지 노출 물질과 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT) 간의 이변량 노출-반응 함수

2017년 자료에서 노출 물질과 결과 변수 간의 비선형 관계와 물질 간 교호 작용을 고려하여 전체적인 복합 노출의 영향을 평가하였을 때, 세 가지 노출물질의 복합노출이 증가할수록 혈중 감마글루타밀전이효소가 감소하는 경향이 유의하였다.



[그림 3-133] 2017년 자료에서 세 가지 노출 물질이 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)에 미치는 복합노출 영향

2017년 자료에서 전체 노출에 대한 개별 노출의 기여도를 추정하기 위해 단일 노출 효과를 추정하였다. 단일 노출 중 감마글루타밀전이효소의 증가와 관련된 노출은 없었다.



[그림 3-134] 2017년 자료에서 각 노출 물질이 혈중 감마글루타밀전이효소(gamma-glutamyl transferase, GGT)에 개별적으로 미치는 영향

베이지안커널머신 회귀 분석은 데이터 기반 분석법으로 알고리즘이 수렴하여 결과의 재현성을 얻기 위해서는 50,000번의 반복 횟수(iteration)를 제안하고 있다(Bobb JF, 2018). 이전에 베이지안커널머신 회귀 분석을 사용한 선행연구에서도(Zhang Y 외, 2019) 반복횟수를 50,000번으로 수행하였다.

하지만, 본 연구에서는 반복 횟수가 커지거나 자료의 크기가 커질 경우 분석 시간이 오래 걸리는 베이지안커널머신 회귀 분석의 단점으로 인해 반복횟수를 <표 3-37>과 같이 1,000번 또는 200번으로 줄여서 분석을 수행하였다. 본 연구에서 베이지안커널머신 회귀 분석은 중앙처리장치(CPU)가 Intel(R) Core(TM) i7-6600U CPU @ 2.60 GHz 2.81 GHz이고 램(RAM)이 16.0 GB인 노트북을 사용하였고, 분석 자료 및 반복횟수 별 분석시간은 <표 3-37>과 같다.

<표 3-37> 베이지안커널머신 회귀 분석 속도

분석 자료	노출변수	결과변수	대상자 수	반복 횟수	분석 시간
2013년	마뇨산	아스파테이트아미노전이효소	615	1,000	20분
2013년	메틸마뇨산	알라닌아미노전이효소	615	1,000	22분
2013년	디메틸포름아미드	감마글루타밀전이효소	615	1,000	25분
2014년	마뇨산	아스파테이트아미노전이효소	1,220	1,000	3시간
2014년	메틸마뇨산	알라닌아미노전이효소	1,220	1,000	3시간
2014년	디메틸포름아미드	감마글루타밀전이효소	1,220	1,000	3시간
2015년	마뇨산	아스파테이트아미노전이효소	1,604	1,000	13시간
2015년	메틸마뇨산	알라닌아미노전이효소	1,604	1,000	7시간
2015년	디메틸포름아미드	감마글루타밀전이효소	1,604	1,000	7시간
2016년	마뇨산	아스파테이트아미노전이효소	2,351	1,000	1일 8시간
2016년	메틸마뇨산	알라닌아미노전이효소	2,351	1,000	22시간
2016년	디메틸포름아미드	감마글루타밀전이효소	2,351	1,000	1일 2시간
2017년	마뇨산	아스파테이트아미노전이효소	2,678	1,000	1일 20시간
2017년	메틸마뇨산	알라닌아미노전이효소	2,678	200	1시간
2017년	디메틸포름아미드	감마글루타밀전이효소	2,678	200	6시간

R을 이용한 베이지안커널머신 회귀 분석 및 결과 요약을 위한 그림을 그리는 R 코드는 다음과 같다.

```
#데이터 읽기#
data<-read.csv("D:/연구/복합노출 연구/data/three2013y1.csv")

data<-data[,-1] #id column 제거
head(data)

#BKMR 분석#
#install.packages("bkmr")
library(bkmr)
y.train<-data$y1
x.train<-data[,c(1:3)]
covariate<-data[,c(5:11)]

bkmr_mod<-kmbayes(y = y.train, Z = x.train, X=covariate, iter = 1000, verbose = FALSE,
varsel = TRUE,,est.h=TRUE)

summary(bkmr_mod)

#install.packages("ggplot2")
library(ggplot2)
pips <- ExtractPIPs(bkmr_mod)
pips$exposure <- as.factor(1:nrow(pips))
ggplot(pips, aes(exposure, PIP)) +
  geom_point() +
  ylab("PIP") +
  ylim(0, 1)+theme_bw()

pred.resp.univar <- PredictorResponseUnivar(fit = bkmr_mod)

ggplot(pred.resp.univar, aes(z, est, ymin = est - 1.96*se, ymax = est + 1.96*se)) +
  geom_smooth(stat = "identity") +
  facet_wrap(~ variable,scales="free") +
  ylab("h(z)")

pred.resp.bivar <- PredictorResponseBivar(fit = bkmr_mod, min.plot.dist = 1)
pred.resp.bivar.levels <- PredictorResponseBivarLevels(pred.resp.df = pred.resp.bivar, Z =
x.train, qs = c(0.25, 0.5, 0.75))
pred.resp.bivar.levels$variable1_1<-pred.resp.bivar.levels$variable1
```

```
ggplot(pred.resp.bivar.levels, aes(z1, est)) +
  geom_smooth(aes(col = quantile), stat = "identity") +
  facet_grid(variable2 ~ variable1,scales="free") +
  ggtitle("h(expos1 ¦ quantiles of expos2)") +
  xlab("expos1")+theme(legend.position="top")

risks.overall <- OverallRiskSummaries(fit = bkmr_mod,
                                      qs = seq(0.25, 0.75, by = 0.05),
                                      q.fixed = 0.5, method = "approx")
risks.overall

ggplot(risks.overall, aes(quantile, est, ymin = est - 1.96*sd, ymax = est + 1.96*sd)) +
  geom_pointrange()+theme_bw()+geom_hline(yintercept = 0,linetype=2,col="grey30")


risks.singvar <- SingVarRiskSummaries(fit = bkmr_mod,
                                      qs.diff = c(0.25, 0.75),
                                      q.fixed = c(0.25, 0.50, 0.75),
                                      method = "approx")
risks.singvar

ggplot(risks.singvar, aes(variable, est, ymin = est - 1.96*sd, ymax = est + 1.96*sd, col =
q.fixed)) +
  geom_pointrange(position = position_dodge(width = 0.75)) + theme_bw()+
  coord_flip()
```

Ⅳ. 고찰 및 결론

본 연구에서는 먼저 유해물질 복합노출에 의한 건강영향을 평가하는 통계기법들을 고찰하고, 각 방법을 사용한 역학연구의 방법론을 검토하여, 특수건강진단 자료에 적합한 방법을 선정하여 예비연구를 수행하였다.

유해물질 복합노출을 평가하는 통계적 방법을 선택하기 위해 먼저 국내 노동자의 건강을 평가할 수 있는 이차자료의 특성을 고려하여 분석의 목표를 다음과 같이 설정하였다. 첫째, 여러 노출물질의 복합적인 건강영향을 평가할 수 있어야 하고, 둘째, 종적자료를 분석할 수 있다. 셋째, 대용량 자료를 분석할 수 있어야 한다. 넷째, 복합적인 영향 중 각 노출물질이 건강영향에 기여하는 정도를 확인할 수 있어야 한다. 다섯째, 분석 결과를 생물학적으로 해석이 가능해야 한다. 하지만 이 외에도, 각 통계방법 마다 여러 가지 추가적인 장단점이 있어 내부 연구진 회의를 통해 추가적으로 통계방법의 장단점을 평가하였고, 그 결과 가중분위수합 회귀, 주성분분석, 부분최소제곱 회귀, 구조방정식 모델링, 베이지안커널머신 회귀가 연구 목적에 맞게 활용할 가치가 있다고 평가되었다.

이 중 본 연구에서는 주성분분석(principle component anlaysis), 부분최소제곱 회귀(partial least square regression, PLS), 베이지안커널머신 회귀(Bayesian kernal machine regression, BKMR)을 실제로 특수건강진단 자료에 활용해보고, 각 통계방법의 특성을 살펴보았다. 실제 적용 결과, 노출물질의 기여도 순위, 노출-결과의 비선형성, 노출-결과의 교호 작용, 복합노출의 건강영향 평가, 단일노출의 건강영향 평가를 동시에 할 수 있는 베이지안커널머신 회귀가 가장 좋은 통계방법으로 생각되었으나, 분석시간이 오래 걸리는 단점이 있어 자료의 용량이 커질수록 적용하기 어려울 것으로 생각된다. 반면, 주성분분석과 부분최소제곱 회귀의 경우 분석시간이 빠르고, 전통적인 차원축소 방법이라는 점에서 장점이 있으나, 로딩벡터가 마이

너스 값이 나올 경우 노출과 결과 간의 연관성을 설명하기 어렵다는 단점이 있었고, 산출된 성분의 단위를 측정하기 어려워 결과 해석이 어려운 단점이 있었다. 또한, 특수건강검진 자료의 경우 노출 물질들 간의 연관성이 높지 않아 차원축소 방법인 주성분분석과 부분최소제곱 회귀의 적용이 적절하지 못한 점이 있었으나 노동자 대상 역학연구에서 노출물질들 간의 연관성이 높게 관찰될 경우 다중공선성을 피하기 위해 고려할 수 있는 방법이다.

결론적으로 위 방법들은 각각 단점이 있으나, 다중공선성에 강하고 복합노출의 건강영향 추정치, 비선형성, 교호작용 등의 정보를 제공해줄 수 있다는 점에서 전통적인 회귀분석과는 다른 장점이 있으므로, 노동자 대상 역학연구에 적극적으로 활용할 가치가 있다.

V. 참고문헌

Agay-Shay K, Martinez D, Valvi D et al. Exposure to endocrine-disrupting chemicals during pregnancy and weight at 7 years of age: A multi-pollutant approach. Environ Health Perspect. 2015; 123(10): 1030-1037

Arlinghaus A, Lombardi DA, Willetts JL et al. A structural equation modeling approach to fatigue-related risk factors for occupational injury. Am J Epidemiol. 2012 Oct 1;176(7):597-607.

Bobb JF, Claus Henn B, Valeri L et al. Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. Environ Health. 2018 Aug 20;17(1):67.

Bobb JF, Valeri L, Claus Henn B et al. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics. 2015 Jul;16(3):493-508.

Bobb JF. Example using the bkmr R package for probit regression with simulated data. Available from: URL: https://jenfb.github.io/bkmr/ProbitEx.html

Braun JM, Gennings C, Hauser R et al. What Can Epidemiological Studies Tell Us about the Impact of Chemical Mixtures on

Human Health? Environ Health Perspect. 2016 Jan;124(1):A6-9.

Buckley JP, Engel SM, Braun JM et al. Prenatal Phthalate Exposures and Body Mass Index Among 4- to 7-Year-old Children: A Pooled Analysis. Epidemiology. 2016 May;27(3):449-58.

Casas M, Valvi D, Ballesteros-Gomez A et al. Exposure to Bisphenol A and Phthalates during Pregnancy and Ultrasound Measures of Fetal Growth in the INMA-Sabadell Cohort. Environ Health Perspect. 2016 Apr;124(4):521-8.

Cavieres MF, Jaeger J, Porter W. Developmental toxicity of a commercial herbicide mixture in mice: I. Effects on embryo implantation and litter size. Environ Health Perspect. 2002; 110(11): 1081-1085

Chiu YH, Bellavia A, James-Todd T et al. Evaluating effects of prenatal exposure to phtlalate mixtures on birth weight: a comparison of three statistical approaches. Environ Int. 2018; 113: 231-239.

Czarnota J, Gennings C, Colt JS et al. Analysis of Environmental Chemical Mixtures and Non-Hodgkin Lymphoma Risk in the NCI-SEER NHL Study. Environ Health Perspect. 2015 Oct;123(10):965-70.

Forns J, Mandal S, Iszatt N et al. Novel application of statistical methods for analysis of multiple toxicants identifies DDT as a

risk factor for early child behavioral problems. Environ Res. 2016 Nov;151:91-100.

Gascon M, Casas M, Morales E et al. Prenatal exposure to bisphenol A and phthalates and childhood respiratory tract infections and allergy. J Allergy Clin Immunol. 2015 Feb;135(2):370-8.

Geladi P, B.R. Kowalski, Partial least-squares regression: a tutorial. Analytica chimica acta, 1986. 185: p. 1-17.

Govarts E, Remy S, Bruckers L et al. Combined effects of prenatal exposures to environmental chemicals on birth weight. Int J Environ Res Public Health. 2016; 13(5): pii: E495.

Kim TW, Koh DH, Park CY. Decision tree of occupational lung cancer using classification and regression analysis. Saf Health Work. 2010 Dec;1(2):140-8.

Lazarevic N, Barnett AG, Aly PD et al. Statistical methodology in studies of prenatal exposure to mixtures of endocrine-disrupting chemicals: A review of existing approaches and new alternatives. Environ Health Perspect. 2019; 127(2): 26001.

Lenters V, Portengen L, Rignell-Hydbom A et al. Prenatal Phthalate, Perfluoroalkyl Acid, and Organochlorine Exposures and Term Birth Weight in Three Birth Cohorts: Multi-Pollutant Models Based on Elastic Net Regression. Environ Health

Perspect. 2016 Mar;124(3):365-72.

Lenters V, Portengen L, Smit LA et al. Phthalates, perfluoroalkyl acids, metals and organochlorines and reproductive function: a multipollutant assessment in Greenlandic, Polish and Ukrainian men. Occup Environ Med. 2015 Jun;72(6):385-93.

Luo L, Hudson LG, Lewis J et al. Two-step approach for assessing the health effects of environmental chemical mixtures: application to simulated datasets and real data from the Navajo Birth Cohort Study. Environ Health. 2019 May 9;18(1):46.

Maresca MM, Hoepner LA, Hassoun A et al. Prenatal Exposure to Phthalates and Childhood Body Size in an Urban Cohort. Environ Health Perspect. 2016 Apr;124(4):514-20.

Menon R, Bhat G, Saade GR, Spratt H. Multivariate adaptive regression splines analysis to predict biomarkers of spontaneous preterm birth. Acta Obstet Gynecol Scand. 2014 Apr;93(4):382-91.

Millar J, Psychas P, Abuaku B et al. Detecting local risk factors for residual malaria in northern Ghana using Bayesian model averaging. Malar J. 2018 Sep 29;17(1):343.

Nguyen TH, Bertin M, Bodin J et al. Multiple exposures and coexposures to occupational hazards among agricultural workers: a systematic review of observational studies. Saf Health Work. 2018 Sep;9(3):239-248.

Porru S, Pavanello S, Carta A et al. Complex relationships between occupation, environment, DNA adducts, genetic polymorphisms and bladder cancer in a case-control study using a structural equation modeling. PLoS One. 2014 Apr 10;9(4):e94566.

Rajapakse N, Silva E, Kortenkamp A. Combining xenoestrogens at levels below individual no observed effect concentrations dramatically enhances steroid hormone action. Environ Health Perspect. 2002; 110(9): 917-921

Romano ME, Eliot MN, Zoeller RT et al. Maternal urinary phthalate metabolites during pregnancy and thyroid hormone concentrations in maternal and cord sera: The HOME Study. Int J Hyg Environ Health. 2018 May;221(4):623-631.

Sun Z, Tao Y, Li S et al. Statistical strategies for constructing health risk models with multiple pollutants and their interactions: Possible choices and comparisons. Environ Health. 2013; 12(1): 85

Swartz MD, Cai Y, Chan W et al. Air toxics and birth defects: a Bayesian hierarchical approach to evaluate multiple pollutants and spina bifida. Environ Health. 2015 Feb 9;14:16.

Talbott EO, Marshall LP, Rager JR et al. Air toxics and the risk of autism spectrum disorder: the results of a population based

case-control study in southwestern Pennsylvania. Environ Health. 2015 Oct 6;14:80.

Tayler KW, Joubert BR, Braun JM et al. Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology: Lessons from an innovative workshop. Environ Health Perspect. 2016; 124(12): A227-A229

Valeri L, Mazumdar MM, Bobb JF et al. The Joint Effect of Prenatal Exposure to Metal Mixtures on Neurodevelopmental Outcomes at 20-40 Months of Age: Evidence from Rural Bangladesh. Environ Health Perspect. 2017 Jun 26;125(6):067015.

Welshons WV, Thayer KA, Judy BM et al. Large effects from small exposures. I. Mechanisms for endocrine-disrupting chemicals with estrogen activity. Environ Health Perspect. 2003; 111(8): 994-1006

Wold S, M. Sjöström, L. Eriksson. PLS-regression: a basic tool of chemometrics. Chemometrics and intelligent laboratory systems, 2001. 58(2): p. 109-130.

Woodruff TJ, Zeise L, Axelrad DA et al. Meeting report: moving upstream-evaluating adverse upstream end points for improved risk assessment and decision-making. Environ Health Perspect. 2008; 116(11): 1568-1575

Xin Fang, Runkui Li, Haidong Kan et al. Bayesian model

averaging method for evaluating associations between air pollution and respiratory mortality: a time-series study. BMJ Open. 2016; 6(8): e011487.

Zhang Y, Dong T, Hu W et al. Association between exposure to a mixture of phenols, pesticides, and phthalates and obesity: Comparison of three statistical models. Environ Int. 2019 Feb;123:325-336.

김기영, 전명식.「SAS 주성분분석」, 고려대학교 통계연구소 통계분석강의 총서7. 자유아카데미. 1997

김지현. 변동계수행렬을 이용한 주성분분석. 응용통계연구. 2015 28권 3호. 385-392.

이미주, 유럽 위원회에서 발간한 “화학물질 혼합물의 독성과 평가에서 언급된 평가법 소개와 향후과제. 산업보건. 2016. 343: 28-32.

이슬비. 임신 중 복합 환경유해물질 노출이 6개월 영유아 아토피 피부염 발생에 미치는 영향. 이화여자대학교 대학원 박사학위 청구논문. 2018

VI. Abstract

Exploratory analysis for estimating the health effects from multiple hazardous materials exposure in occupational health field

Objectives : Recently exposome, which means the sum of all exposures from conception to the present, and birth cohort studies that repeatedly measure environmental exposure from the womb to the present, are actively conducted in the environmental epidemiology fields. Many research is being conducted to evaluate the health effects of environmental multiple exposures. On the other hand, there is a lack of research on the health effects of multiple exposures in occupational fields where the exposure to hazardous substances is decreasing in the domestic working environment and frequent multiple exposures occur. Therefore, it is necessary to review the appropriate statistical techniques to analyze the health effects of the multiple hazardous materials exposure in the occupational health fields.

Methods and Results : Through review of literatures, the statistical techniques, characteristics, advantages and disadvantages, and the possibility of applying occupational disease epidemiological studies were examined. Workers' Specific Health Examination data were analyzed in a pilot study on the selected statistical method. First,

toluene, xylene, and dimethylformamide exposures were individually evaluated. Principle component analysis (PCA), partial least square (PLS) method, Bayesian kernel machine regression (BKMR) were used to assess the exposure of multiple hazardous materials and health effects. Liver enzyme levels were used to assess health effects.

Conclusion : It needs to consider analytical methods that take into account the interactions and nonlinearities of hazardous materials in occupational epidemiologic fields. This study can be used as a guideline in selecting statistical methods for each sitution.

Keywords : Multiple exposure, Toluene, Xylene, Dimethylformamide, Workers' Specific Health Examination

VI. 부록

부록 1. 각 통계적 방법을 고려한 역학 논문

1. Millar J, Psychas P, Abuaku B et al. Detecting local risk factors for residual malaria in northern Ghana using Bayesian model averaging. Malar J. 2018 Sep 29;17(1):343.

1) 배경

- 사하라 사막 이남 아프리카(sub-Saharan Africa)의 잔여 말라리아에 기여하는 지역적인 기저 요인을 종합적으로 검토할 필요가 있음.

2) 목적

- 잔여 말라리아와 관련된 잠재적 위험 및 보호 요인을 확인하고 비교하기 위하여 베이지안 모형 평균(Bayesian model averaging, BMA)를 이용하였음.

3) 방법 : 시계열 연구

(1) 연구자료 및 연구대상자

- 계절성 말라리아가 자주 발생하는 가나 북동쪽의 기니 사바나에 있는 Bunkpurugu-Yunyoo 지구에서 자료를 수집함. 이 지구는 Bunkpurugu(인구: 7436 명)와 Nakpanduri(인구: 5783 명)의 2 개의 도심과 농사와 목축업을 하는 농촌으로 구성됨.

- 주요 민족-언어 그룹은 Bimoba(약 60%)와 Konkomba(약 30%)이고, Mamprusi, Kusasis, Dagombas, Fulanis 등의 작은 집단이 있음. Bimoba는 2개의 도심을 포함한 북부, 동부의 고지대에 많은 경향이 있고, Konkomba는 목축업을 하는 남부, 서부의 저지대에 우세함. Konkomba는 지리적, 경제적으로 소외된 경향이 있고 일반적으로 교육 수준이 낮음.

- 개인 수준의 종적 데이터셋은 가나 대학(University of Ghana)이 수행한 연구 과정에서 수집함.

- 2010년-2013년 사이 6-59개월 사이의 어린이를 대상으로 우기 3회 및 건기 3회에 걸쳐 조사를 실시함.

- 새로운 대표 샘플은 다단계무작위 군집 샘플링(multi-stage randomized cluster sampling technique)을 이용하여 선택함. 가나 보건 서비스(Ghana Health Service)의 지역 사회 명단을 기반으로 대표 집단(representative communities)을 무작위로 선정하였음. 이렇게 선정한 샘플은 2010년 인구조사 자료를 기반으로 할 때, 각각 조사 당 5세 이하 인구의 약 20%를 차지함.

(2) 노출 평가

- 설문지를 통해 인구통계학적, 사회경제적, 말라리아 개입변수에 대한 정보를 수집함.

- 6개월 미만의 경우 모성 면역의 영향 때문에 분석에서 제외함.

(3) 건강영향평가

- 혈액도말표본을 현미경(blood-film microscopy)으로 관찰하여 말라리아 상태를 확인하였음.

(4) 보정변수

- 잠재적 위험 인자들의 상관 관계를 계산하여, 상관 관계가 높은 경우($R^2 > 0.49$), 단일 대표 공변량(single representive covariate)이 선택 되었고, 모든 모형에서 공변량은 농업인, 지난 7개월간의 실내 잔류 분무(IRS, indoor residual spraying), 평균 주간 육지 표면 온도,

정규 식생 지수(NDVI, normalized difference vegetation index), 누적 강우량 및 역사적 강우량 추세였음.

- 베이지안 모형 평균(BMA)은 각 개인에게 모든 공변량에 대한 정보가 필요하므로, 공변량이 누락된 개인은 분석에서 제외 되어 총 10,029 명(전체 데이터의 84%)을 대상으로 분석함.

(5) 통계분석 방법

(5-1) 기본 모형

- 모든 말라리아 위험 모형은 일반 베이지안 틀(general Bayesian framework)을 이용해 구성함. y_{ijt} 를 이진법적 현미경 결과로 설정하고(1=양성, 0=음성, i=개인(individual), j=지역사회(community), t=시간(time)), 베이지안 프로빗 회귀 모형(Bayesian probit regression model)을 이용하여 $z_{ijt}>0$ 인 경우, $y_{ijt}=1$ 로 가정함. 즉, $z_{ijt}>0$ 인 경우에만 말라리아에 양성임.

$$z_{ijt} \sim N\left(x_{ijt}^{T}\beta, 1\right)$$

- x^{T}_{ijt} 는 절편과 잠재적 위험 인자의 벡터이며, β 는 해당 회귀 매개변수가 있는 벡터임. 사전확률(priors)는 다음과 같음. $\beta \sim N\left(0, \sigma^2 \Sigma\right)$ $\sigma \sim Unif\,(0, 100)$

- 사전확률 β 에 포함된 행렬 Σ 는 $diag(\Sigma)=[100\ 1\ .\ .\ .\ 1]$의 대각 행렬임.

(5-2) 복잡한 모형: 계절적 차이와 비선형 연관성

- 설계벡터 x^{T}_{ijt} 에 추가요소를 포함하여 건기와 우기의 위험 인자의 영향이 다른지 평가하기 위한 모형을 구축하여 각 공변량에 대한 매개변수(parameter, 혹은 모수) 추정치가 계절에 따라 다를 수 있도록 함.

- 이 프레임워크는 가장 가까운 도시 중심까지의 거리와 가장 가까운 보건 시설과의 거리라는 2 개의 관련 연속 변수의 잠재 비선형

패턴을 설명하기 위하여 선형 스플라인을 통해 사용됨. 선형 스플라인의 생성은 공변량 x 의 영역 내의 m 값의 집합으로 구성되며 $k_1,...,k_m$ 같은 매듭(knot)으로 불려짐. 각각의 매듭에 대하여 새롭게 파생된 공변량 x_d 는 다음과 같음.

$$x_d = \begin{cases} 0, & x < k_d \\ x - k_d, & x \geq k_d \end{cases}$$

- 각 스플라인 공변량에 대하여 m 개의 추가 유도 변수가 생성됨. 매듭값은 관찰된 변수의 20, 40, 60, 80 %분위수에서 선택됨. 이러한 스플라인을 포함하면 변수의 효과가 매듭값에서 이동하여 특정 위험 인자의 비선형 연관성을 나타낼 수 있음.

(5-3) 베이즈 모형 평균(Bayesian model averaging)

- 지금까지 언급한 각각의 모형은 마르코프 연쇄 몬테카를로 방법(Markov chain Monte Carlo, MCMC) 알고리즘을 이용할 수 있음. 변수 선택은 MCMC 알고리즘에 통합되었음.

- 가능한 공변량의 부분집합을 포함하는 모형으로 MCMC 를 초기화 함. MCMC 의 각 반복에서 무작위로 선택된 이동을 통해 새로운 후보 모형을 제안함. 후보 모형은 한계 로그 우도(marginal log-likelihood)에 따라 허용되거나 거부됨. 유익한 공변량은 한계 우도를 증가 시켜 선택 과정에서 유지되는 경향이 있음. 공변량 q 의 부분집합에 의해 정의된 모형 M_q 와 관련된 한계 확률은 연관된 회귀 매개변수 β_q 를 적분한 후 닫힌 형태로 계산될 수 있음.

$$p\left(M_q|z,\sigma^2\right) \propto \int N\left(z|X_q\beta_q, I\right)\mathrm{N}\left(\beta_q|0,\sigma^2\Sigma\right)d\beta_q$$

$$\propto \left(\sigma^2\right)^{-\frac{pq+1}{2}} exp\left(-\frac{1}{2}\left[-\mu_q^T \mathrm{T}_q^{-1}\mu_q\right]\right)|T_q|^{\frac{1}{2}}$$

$$\mathrm{T}_q^{-1} = \left\{X_q^T X_q + \tfrac{1}{\sigma^2}\Sigma_q^{-1}\right\} \text{and}\, \mu_q = \mathrm{T}_q X_q^T z.$$

- 알고리즘은 새로운 공변량의 출현이나 기존 공변량의 죽음 또는 교환을 무작위로 제안하여 모형 공간을 탐구함. 이런 제안된 이동은 standard Metropolis-Hastings acceptance ratio에 따라 허용되거나 거부됨.

$$\min\left\{1, \frac{p(M_{q^*}|z,\sigma^2)p(M_{q^*})}{p(M_q|z,\sigma^2)p(M_q)}\right\} = \min\left\{1, \frac{e(\sigma^2)^{-\frac{p_{q^*}+1}{2}} exp\left(-\frac{1}{2}\left[-\mu_{q^*}^T \mathrm{T}_{q^*}^{-1}\mu_{q^*}\right]\right)|T_{q^*}|^{\frac{1}{2}}}{(\sigma^2)^{-\frac{pq+1}{2}} exp\left(-\frac{1}{2}\left[-\mu_q^T \mathrm{T}_q^{-1}\mu_q\right]\right)|T_q|^{\frac{1}{2}}} \times R\right\}$$

- R은 일반적으로 1과 같고, M_{q*}와 M_q는 각각 제안된 모형과 현재 모형임. 각 모형은 깁스 표본기법(Gibbs sampler)을 이용하여 첫 1,000번의 불태움 기간(burn-in period)을 포함하여 10,000번 반복 실행되었으며 매개변수 추정치의 수렴은 추적 플롯을 이용하여 확인하였음.

- 전통적인 베이지안 회귀(traditional Bayesian regression)와 유사하게 각 공변량에 대한 회귀 계수(β)는 사후추정치(posterior draws)를 기준으로 추정되며 95% 신뢰구간에 0이 포함되지 않으면 통계적으로 유의하다고 판단함. 개별 사후 표본이 다른 매개변수 공간을 기반으로 하기 때문에 베이지안 모형 평균(BMA)의 "모형 평균화" 구성요소이며 이를 통해 변수 선택과 관련된 불확실성을 매개변수 추정에 통합 가능함.

(5-4) 표본외 예측(Out-of-sample predictions)

- 모형의 더 큰 유연성을 위한 허용치는 몇 가지 파생된 공변량과 관련 매개변수를 추가하여 얻을 수 있음. 기본 모형은 29개의 공변량을 포함하고, 상호작용 조건을 추가하여 공변량을 56개로 늘렸으며 선형 스플라인을 추가하여 73개의 공변량을 포함함. 매개변수의 수를 늘리면 초과맞춤(overfitting)이 발생할 수 있으므로 변수 선택이 점점 더 중요한 작업이 됨.

- 베이지안 모형 평균(BMA)의 표본외 예측 성능을 평가하기 위해 특정 연도의 데이터에 대한 모형과 미래의 말라리아 상태에 대한

예측을 수행하여 표준 로지스틱 회귀분석(standard logistic regression) 및 최소 절대 수축과 선택 연산자 회귀(least absolute shrinkage and selection operator (Lasso) regression)와 비교함.

- 공변량의 수에 비례하여 수행된 방법을 보여주기 위해 기본 모형과 확장 모형에 대한 계절별 예측을 수행하였고 로그 우도(log likelihood)의 합계에 따라 표본외 예측 기술을 평가함. 로그 우도 합계가 높은 모형이 좋은 예측 능력을 가졌다고 간주됨.

4) 결과

- 말라리아 유병률에 대한 인구통계학적, 사회경제적, 환경적, 말라리아 개입변수의 상대적 영향은 베이지안 모형 평균(BMA)을 사용하여 변수 선택함. 말라리아 위험과 가장 가까운 도심 및 보건시설과의 거리 사이에는 명백한 비선형적 연관성을 보임.

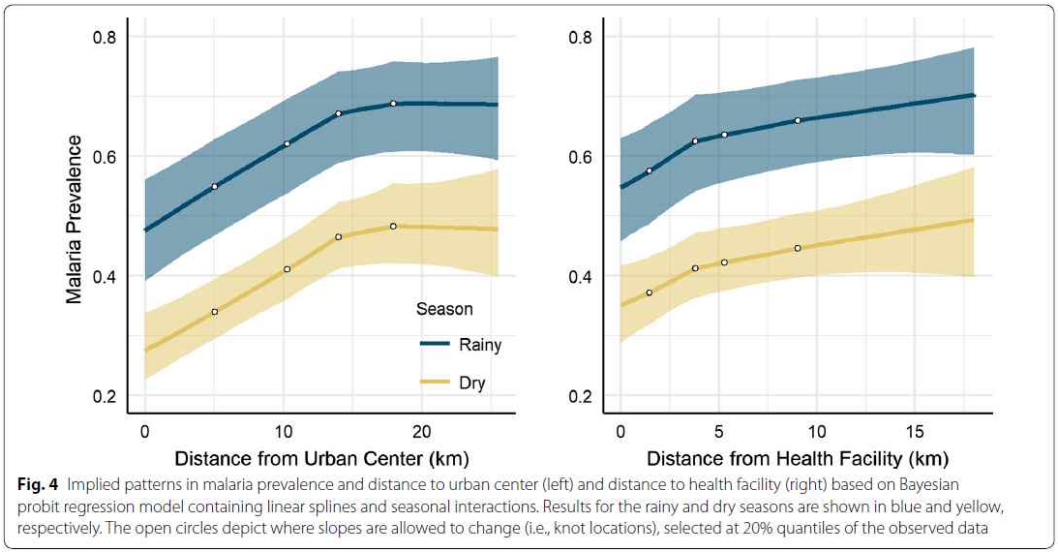


Fig. 4 Implied patterns in malaria prevalence and distance to urban center (left) and distance to health facility (right) based on Bayesian probit regression model containing linear splines and seasonal interactions. Results for the rainy and dry seasons are shown in blue and yellow, respectively. The open circles depict where slopes are allowed to change (i.e., knot locations), selected at 20% quantiles of the observed data

[그림 7-1] 말라리아 유병률과 도시 중심까지 거리(왼쪽)와 보건시설까지 거리(오른쪽)

- 말라리아 위험과 나이 및 일부 종족 간의 연관성은 우기와 건기에 유의미한 차이가 있었음.

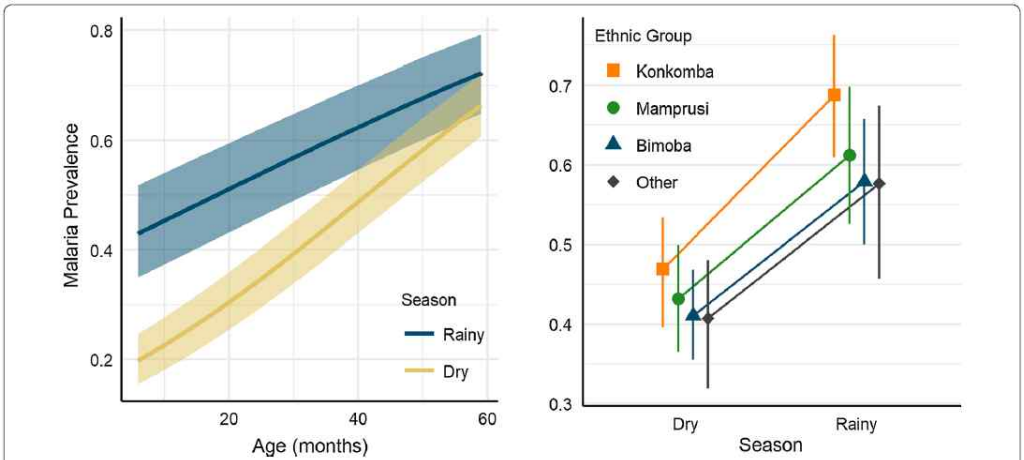


Fig. 3 Modelled patterns in malaria risk factors based on Bayesian probit regression containing seasonal interaction terms. The left panel depicts mean slope estimate (lines) and 95% credible intervals (polygons) for the predicted malaria prevalence based on age in the rainy and dry seasons. The right panel depicts the mean (points) and 95% credible intervals (vertical bars) for the predicted malaria prevalence based on ethnic group in the rainy and dry seasons

[그림 7-2] 계절별 상호작용을 포함하는 베이지안 프로빗 회귀를 기반으로 한 말라리아 위험요소의 모형화된 패턴

- BMA는 계절에 따른 유효성 검사 접근법을 사용한 표본외 예측 능력에서 다른 회귀 접근법보다 뛰어났음.

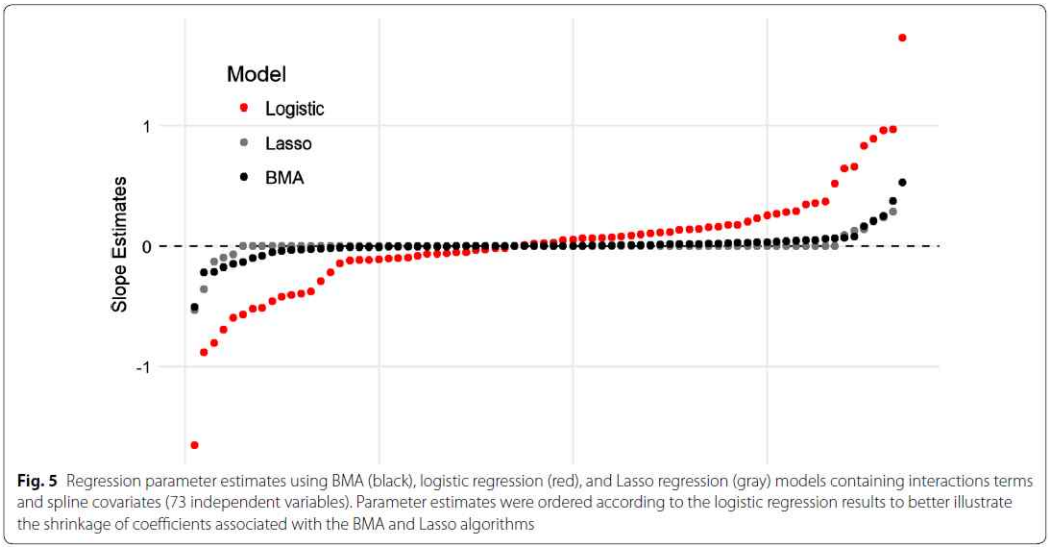


Fig. 5 Regression parameter estimates using BMA (black), logistic regression (red), and Lasso regression (gray) models containing interactions terms and spline covariates (73 independent variables). Parameter estimates were ordered according to the logistic regression results to better illustrate the shrinkage of coefficients associated with the BMA and Lasso algorithms

[그림 7-3] 상호작용 항과 스플라인 공변량을 포함하는 베이지안 모형 평균(검정), 로지스틱 회귀(빨강), 올가미 회귀(회색)을 사용한 회귀 모수 추정

5) 고찰 및 결론

- 베이지안 모형 평균은 말라리아의 위험 인자를 비롯하여 기타 매개체 감염 및 환경 매개 질환의 위험인자를 이해하기 위한 유익한 모형을 구축하는데 유용한 도구임.

2. Menon R, Bhat G, Saade GR, Spratt H. Multivariate adaptive regression splines analysis to predict biomarkers of spontaneous preterm birth. Acta Obstet Gynecol Scand. 2014 Apr;93(4):382-91.

1) 목적

- 아프리카 계 미국인과 백인에서 발생한 조산과 관련된 인구통계학적/ 임상적 요인 및 생물학적 지표를 포함한 자료의 분류 모델을 개발하고자 함.

2) 방법

- 이 분석에 사용 된 생물학적 지표 자료는 조산과 관련된 유전자 지표와 지표의 인종적 차이를 연구하기 위해 설립 된 바이오 뱅크 인 내쉬빌 출생 코호트 (Nashville Birth Cohort)의 샘플을 사용하여 발표 된 이전 연구에서 얻은 것임.
- 18 세에서 40 세 사이의 임산부가 참여할 수 있었고, 여성은 조산에 대한 유전 적 위험 인자를 평가하기 위한 부모 사례-대조 연구로부터 모집됨. 모든 여성은 2003 년 9 월부터 12 월 사이에 테네시 주 내슈빌 센 테니얼 여성 병원에 모집 됨.
- 재태연령(gestational age)은 마지막 생리 기간에 의해 결정되었고 초음파에 의해 확증됨.
- 조산을 한 환자(24+0 주에서 36+0 주 사이)가 사례(case)군으로 포함되었음. 대조군으로는 임신과 관련된 합병증이 없이 37+0 주에서 42+0 주 사이에 질 분만을 한 여성이 포함됨.
- 태아의 인종은 태아의 어머니와 아버지, 부모 및 조부모의 인종에 의해 결정되었고, 설문지를 통해 확인함.
- 혼합 인종의 여성은 모계 및 부계 인종과 관련하여 여성 간의 동질성을 유지하기 위해 연구에서 제외됨.

- 히스패닉 이외의 출신으로 확인 된 아프리카계 미국인 또는 백인 참가자 만 포함됨.

- 다태 임신, 자간전증, 전치태반, 태아 기형, 임신부의 의학적 합병증(예 : 임신성 당뇨병) 또는 외과적 합병증이 있는 여성은 제외되었음.

- 임신 중 수술 절차를 받았거나 조산을 위해 치료 받거나 양수 내 감염이 의심되는 여성은 대조군에서 배제되었음.

- 분만 또는 조산을 위해 입원한 191 명의 여성에서 모체 혈장, 제대 혈장, 양수를 수집함.

- 이 중 양수는 막 파열 전 경질 양수천자(transvaginal amniocentesis)를 통해 수집됨.

- 환자 인터뷰에서 인구통계학적 자료를 수집하였고, 환자의 의료 기록에서 임상 자료를 수집하였음. 수집된 자료에는 연령, 사회경제적상태(교육, 연간 소득, 보험 상태 및 결혼 상태), 행동 요인(흡연 상태) 및 의학적 및 산과 병력이 포함됨.

- 191 명의 여성의 36 개의 생물학적 지표에 대한 자료를 다변량 적응 회귀 스플라인(MARS, Multivariate Adaptive Regression Spline)을 활용하여 축소한 다음 아프리카계 미국인과 백인의 조산을 위한 예측 모델을 개발하기 위한 자료 분석을 수행함.

- 생물학적 지표 자료의 분석을 위해 감독 된 머신 러닝 알고리즘(supervised machine learning algorithm) 방법 인 다변량 적응 회귀 스플라인(MARS, Multivariate Adaptive Regression Spline)를 사용함.

- 다변량 적응 회귀 스플라인(MARS)은 축소 및 모델 구축에 강한 모델링 방법임. 다변량 적응 회귀 스플라인(MARS)은 예측 변수의 스플라인 함수의 시리즈(series of spline functions)에 의해 복잡한 비선형 관계를 추정 할 수 있는 비모수(nonparametric), 다변량 회귀 분석 방법임.

- 회귀 스플라인은 변수 간 관계의 임계값(thresholds)과 중단지점을 찾으려고 하며 개인의 동작 변화 또는 시간이 지남에 따라 변화하는 과정을 알아내는 데 매우 적합함.

- 다변량 적응 회귀 스플라인(MARS)은 연속형 변수 또는 범주형 변수 등 많은 형태의 예측 변수를 모델링 할 수 있고 많은 수의 입력 예측 변수를 허용 할 수 있다는 장점이 있음.

- 비모수적 접근법으로서 다변량 적응 회귀 스플라인(MARS)은 관심있는 예측 변수의 분포에 대한 기본 가정을 하지 않음. 이 특성은 많은 사이토카인 값이 정규 분포를 따르지 않기 때문에 모델링에서 매우 중요한 점임.

- 스플라인 모델의 기본 개념은 서로 다른 간격으로 분석물의 불연속 선형(discrete linear) 또는 비선형(nonlinear) 함수를 사용하여 모델링하는 것임.

- 스플라인이라고 하는 결과 조각 별 곡선(resulting piecewise curve)은 모델 내 기본 함수로 나타남. 여기서 말하는 기본 함수란 독립변수의 조합을 뜻함.

- 다변량 적응 회귀 스플라인(MARS)은 최적 분류 모델을 구성하기 위해 2 단계 과정을 거침. 먼저 단일 변수 변환 또는 다변량 교호 작용 항을 나타내는 기본 함수를 추가하여 지나치게 큰 모형(overly large model)을 만듦. 두 번째 단계에서 다변량 적응 회귀 스플라인(MARS)은 최적의 모델에 도달 할 때까지 모델에 가장 적게 기여하는 기본 함수를 삭제함.

- 모델의 과적합을 피하기 위해 교차 검증 기술을 사용함. 하나 이상의 상호 작용 항을 포함하는 모델로 모델을 제한하기로 선택함.

- 모체 혈장, 제대 혈장, 양수의 각 구획에서 각 인종 범주(전체, 아프리카계 미국인 및 백인)에 대해 100 개의 무작위로 정렬 된 데이터 세트가 만들어 짐.

- 데이터 세트는 다변량 적응 회귀 스플라인(MARS) 애플리케이션에 별도로 업 로드되었으며, 데이터 세트의 30 %가 테스트(testing)로, 70 %가 교육(training)으로 분류되었음.
- 최대 기본 함수는 입력 예측 변수 수의 3 배로 제한되었고, 최대 상호 작용 수는 1 개로 제한되었으며 가장 정확한 속도가 선택되었음.
- 변수 중요도, 전체 모델에 대한 각 변수의 기여를 평가하는 상대 지표 (0 부터 100 %)를 100 회 실행마다 기록하였음.
- 최대 100 회 이상 실행 된 후 평균화된 변수를 가지고 데이터 세트을 만듦.
- 그런 다음 교차 검증 옵션을 사용하여 10 배 교차 검증, 10 개의 입력 예측 변수, 하나의 최대 상호 작용 항, 30 개의 최대 기본 함수, 가장 정확한 속도를 사용하여 단일 모델을 만듦.
- 다변량 적응 회귀 스플라인(MARS) 모델은 훈련 데이터(traning data)로부터 학습 된 알려진 특성에 기초 한 예측(testing)을 위해 수신자판단특성곡선(receiver operating curve, ROC) 아래의 면적(AUC)으로 특징지어짐.

3) 결과

- 다변량 적응 회귀 스플라인은 전체 및 인종 계층화 생물학적 지표 자료에 대한 모델을 생성함.
- 임상 및 인구통계학적 자료는 모델에 기여하지 않음.
- 인종으로 층화한 데이터는 세 구획 모두에서 별개의 모델을 생성함.
- 아프리카계 미국인의 모체 혈장에서는 IL 1RA, TNF α, angiopoietin 2, TNFRI, IL 5, MIP1α, IL 1β 및 TGF α 이 조산예측을 위해 모델화되었고 AUC train: 0.98, AUC test: 0.86 의 결과가 확인됨. 백인에서는 TNFR1, ICAM 1 및 IL 1RA 이 모델에 기여하였음(AUC train: 0.84, AUC test: 0.68). 아프리카계 미국인의

제대혈장에서는 IL 12P70 과 IL 8 이 기여하였음(AUC train: 0.82, AUC test: 0.66). 백인의 제대혈장에서는 IGFII, PDGFBB, TGF β1, IL 12P70, 및 TIMP1 이 모델에 사용되었음(AUC train: 0.99, AUC test: 0.82). 아프리카계 미국인의 양수에서는 FasL, TNFRII, RANTES, KGF, IGFI 이 모델에 기여함(AUC train: 0.95, AUC test: 0.89). 백인에서는 TNF α, MCP3, TGF β3, TNFR1 및 angiopoietin 2 이 모델에 기여함(AUC train: 0.94 AUC test: 0.79).

<표 7-1> 다변량 적응 회귀 스플라인(MARS) 기본 함수: 인종을 합한 모성 혈장

B_m	Definition	a_m	Variable descriptor
BF1	$(TNFRI-76.99)_+$	–0.0002	TNFR-I
BF3	$(251-IL8)_+$	0.0016	IL-8
BF4	$(TIMP1-13465.3)_+$	0.0000	TIMP1
BF7	$(75.72-TNFALPHA)_+$	0.0088	TNF-α
BF9	$(238.22-EOTAXIN)_+$	–0.0019	Eotaxin
BF10	$(VEGF-4293)_+$	0.0000	VEGF
BF12	$(FGFBASIC-4041)_+$	–0.0008	FGFBasic
BF15	$(75.68-IL1RA)_+$	0.0040	IL-1RA
BF16	$(FGFBASIC-7079)_+$	0.0003	FGFBasic
BF18	$(FGFBASIC-2765)_+$	0.0005	FGFBasic
BF22	$(IL8-1054.88)_+$	0.0002	IL-8
BF24	$(IL8-873.25)_+$	–0.0031	IL-8

각 구간에서 조산에 대한 다변량 적응 회귀 스플라인(MARS)의 기본 함수(Basis functions, BF). Bm-각 개인의 기본 함수, am-기본 함수의 계수; $(y)_+ - \max(0,y)$.

<표 7-2> 다변량 적응 회귀 스플라인(MARS) 기본 함수: 흑인 모성 혈장

B_m	Definition	a_m	Variable descriptor
BF4	$(230.52\text{-IL1BETA})_+$	0.0013	IL-1β
BF7	$(100\text{-TNFALPHA})_+$	0.0043	TNF-α
BF10	$(39951\text{-ANGPT2})_+$	0.0000	Angiopoietin 2
BF12	$(293\text{-IL5})_+$	−0.0012	IL-5
BF13	$(\text{TNRI-}733)_+$	0.0005	TNFR-I
BF14	$(733\text{-TNFRI})_+$	0.0011	TNFR-I
BF15	$(\text{TNRI-}1720.07)_+$	−0.0007	TNFR-I
BF17	$(\text{MIP1ALPHA-}271)_+$	−0.0002	MIP1-α
BF18	$(\text{TGFALPHA-}38.06)_+$	−0.0017	TGF-α
BF20	$(\text{TGFALPHA-}316)_+$	0.0025	TGF-α
BF22	$(\text{IL1RA-}474.38)_+$	−0.0018	IL-1RA
BF24	$(\text{IL1RA-}169.82)_+$	0.0019	IL-1RA

각 구간에서 조산에 대한 다변량 적응 회귀 스플라인(MARS)의 기본 함수(Basis functions, BF). Bm-각 개인의 기본 함수, am-기본 함수의 계수; (y)+ - max(0,y).

<표 7-3> 다변량 적응 회귀 스플라인(MARS) 기본 함수: 백인 모성 혈장

B_m	Definition	a_m	Variable descriptor
BF2	(72.82-IL1RA)	0.0067	IL-1RA
BF3	$(\text{TNFRI-}745)_+$	−0.0003	TNFR-I
BF7	$(\text{ICAM1-}4.99)_+$	0.0000	ICAM-1

각 구간에서 조산에 대한 다변량 적응 회귀 스플라인(MARS)의 기본 함수(Basis functions, BF). Bm-각 개인의 기본 함수, am-기본 함수의 계수; (y)+ - max(0,y).

<표 7-4> 다변량 적응 회귀 스플라인(MARS) 기본 함수: 인종을 합한 태아 혈장

B_m	Definition	a_m	Variable descriptor
BF17	$(199.14\text{-FASLIGAND})_+$	−0.0032	FASL
BF19	$(554.6\text{-TFBETA1})_+$	0.0008	TGF-β_1
BF21	$(169.2\text{-IL1RA})_+$	0.0011	IL-1RA
BF22	$(\text{ICAM1-}33171.0)_+$	0.0001	ICAM-1
BF25	$(21101\text{-ICAM1})_+$	−0.0001	ICAM-1
BF26	$(\text{IGFII-}60377.7)_+$	0.0000	IGF-II

각 구간에서 조산에 대한 다변량 적응 회귀 스플라인(MARS)의 기본 함수(Basis functions, BF). Bm-각 개인의 기본 함수, am-기본 함수의 계수; (y)+ - max(0,y).

<표 7-5> 다변량 적응 회귀 스플라인(MARS) 기본 함수: 흑인 태아 혈장

B_m	Definition	a_m	Variable descriptor
BF4	$(150.39\text{-IL8})_+$	0.0041	IL-8
BF12	$(222.12\text{-IL12P70})_+$	−0.0012	IL-12P70

각 구간에서 조산에 대한 다변량 적응 회귀 스플라인(MARS)의 기본 함수(Basis functions, BF). Bm-각 개인의 기본 함수, am-기본 함수의 계수; (y)+ - max(0,y).

<표 7-6> 다변량 적응 회귀 스플라인(MARS) 기본 함수: 백인 태아 혈장

B_m	Definition	a_m	Variable descriptor
BF17	$(\text{TGFBETA1-}555.13)_+$	0.0126	TGF-β_1
BF18	$(555.13\text{-TGFBETA1})_+$	−0.0019	TGF-β_1
BF19	$(\text{TGFBETA1-}339.95)_+$	−0.0069	TGF-β_1
BF21	$(\text{TGFBETA1-}644.55)_+$	−0.0058	TGF-β_1
BF26	$(\text{PDGFBB-}1433.48)_+$	−0.0026	PDGFBB
BF28	$(\text{PDGFBB-}452.67)_+$	−0.0016	PDGFBB

각 구간에서 조산에 대한 다변량 적응 회귀 스플라인(MARS)의 기본 함수(Basis functions, BF). Bm-각 개인의 기본 함수, am-기본 함수의 계수; (y)+ - max(0,y).

<표 7-7> 다변량 적응 회귀 스플라인(MARS) 기본 함수: 인종을 합한 양수

B_m	Definition	a_m	Variable descriptor
BF1	$(MCP3-2000)_+$	0.0001	MCP-3
BF2	$(2000-MCP3)_+$	0.0001	MCP-3
BF4	$(4476.13-ANGPT2)_+$	0.0001	Angiopoietin 2
BF5	$(ICAM1-30000)_+$	0.0001	ICAM-1
BF8	$(ICAM1-17660)_+$	0.0000	ICAM-1
BF10	$(ANGPT2-36618.3)_+$	0.0000	ANGPT2
BF13	$(MCP1-375)_+$	−0.0003	MCP-1

각 구간에서 조산에 대한 다변량 적응 회귀 스플라인(MARS)의 기본 함수(Basis functions, BF). Bm-각 개인의 기본 함수, am-기본 함수의 계수; (y)+ - max(0,y).

<표 7-8> 다변량 적응 회귀 스플라인(MARS) 기본 함수: 흑인 양수

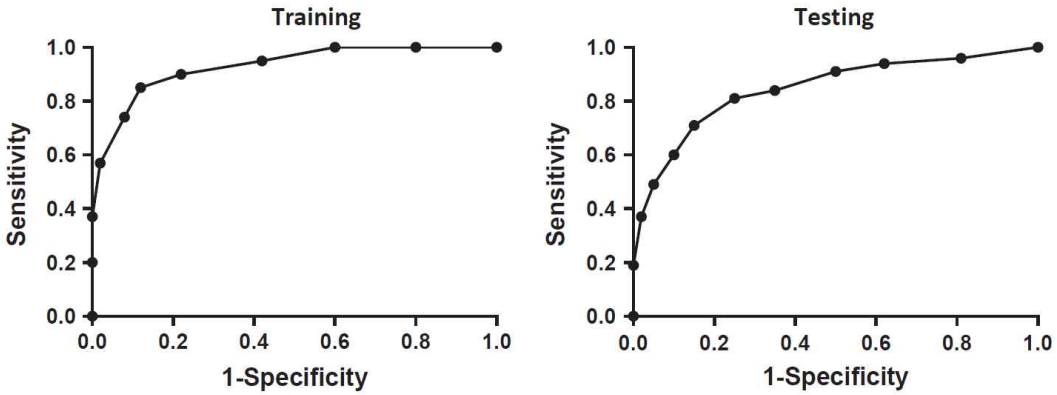
B_m	Definition	a_m	Variable descriptor
BF2	$(179-FASLIGAND)_+$	0.0032	FASLigand
BF6	$(1193-RANTES)_+$	0.0000	RANTES
BF9	$(IGFI-431)_+$	0.0008	IGF-I
BF11	$(KGF-5.00)_+$	−0.0001	KGF
BF23	$(TNFRII-1299)_+$	−0.0003	TNFRII
BF25	$(TNFRII-2300)_+$	0.0004	TNFRII

각 구간에서 조산에 대한 다변량 적응 회귀 스플라인(MARS)의 기본 함수(Basis functions, BF). Bm-각 개인의 기본 함수, am-기본 함수의 계수; (y)+ - max(0,y).

<표 7-9> 다변량 적응 회귀 스플라인(MARS) 기본 함수: 백인 양수

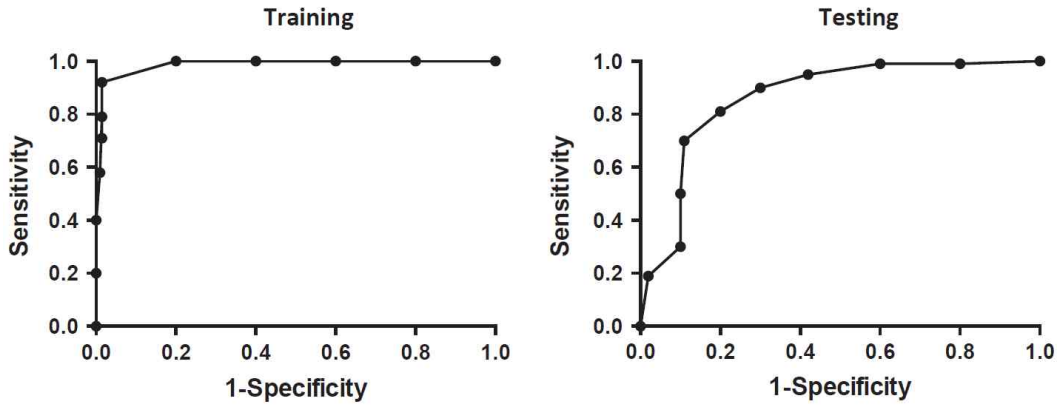
B_m	Definition	a_m	Variable descriptor
BF3	$(TNRI-247)_+$	0.0139	TNFRI
BF7	$(TGFBETA3-122)_+$	0.0001	TGF-β_3
BF9	$(23-TNFALPHA)_+$	0.0201	TNF-α
BF12	$(ANGPT2-19.99)_+$	−0.0001	Angiopoietin 2
BF13	$(ANGPT2-3257.93)_+$	0.0001	Angiopoietin 2
BF15	$(MCP3-4.99)_+$	0.0001	MCP-3
BF20	$(TNFI-208)_+$	−0.0138	TNFRI

각 구간에서 조산에 대한 다변량 적응 회귀 스플라인(MARS)의 기본 함수(Basis functions, BF). Bm-각 개인의 기본 함수, am-기본 함수의 계수; (y)+ - max(0,y).



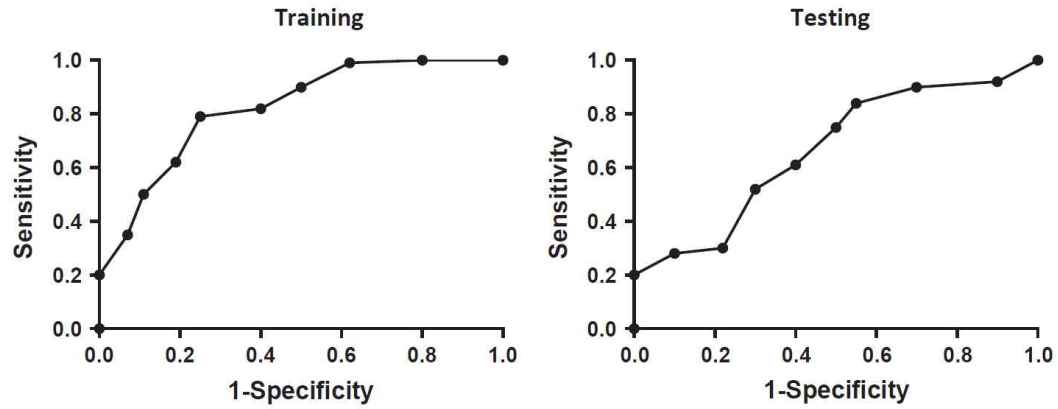
[그림 7-4] 조산 예측 모델에 대한 수신자 작동 특성(reciever operating characteristic, ROC) 곡선: 인종을 합한 모성 혈장

Y 축 = 민감도; X axis = 1 특이도. 훈련 수신자 작동 특성 곡선은 왼쪽이고 시험 곡선은 오른쪽임.



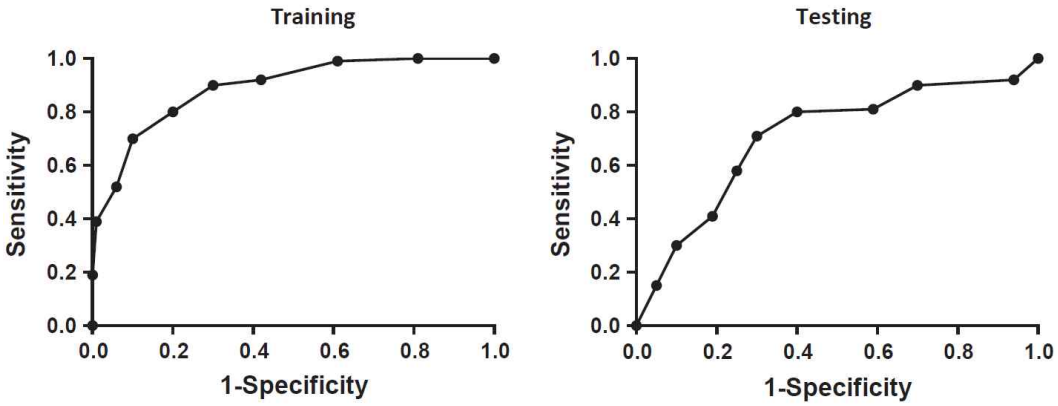
[그림 7-5] 조산 예측 모델에 대한 수신자 작동 특성(reciever operating characteristic, ROC) 곡선: 흑인 모성 혈장

Y 축 = 민감도; X axis = 1 특이도. 훈련 수신자 작동 특성 곡선은 왼쪽이고 시험 곡선은 오른쪽임.



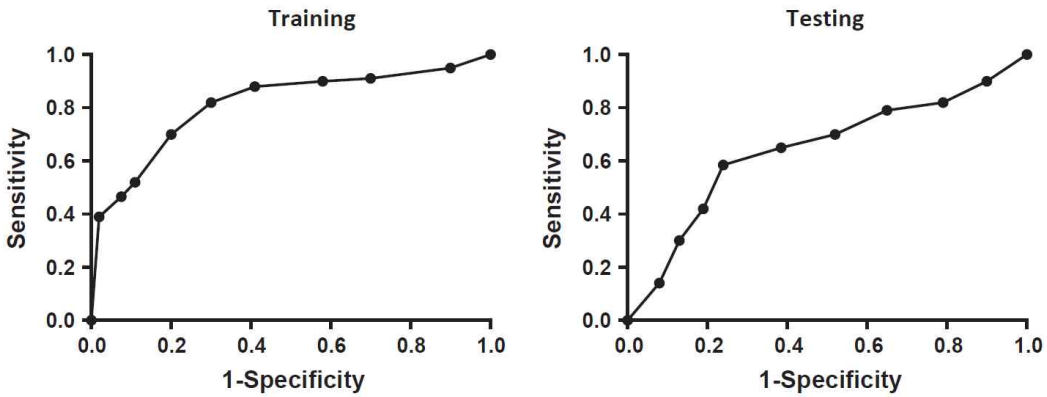
[그림 7-6] 조산 예측 모델에 대한 수신자 작동 특성(reciever operating characteristic, ROC) 곡선: 백인 모성 혈장

Y 축 = 민감도; X axis = 1 특이도. 훈련 수신자 작동 특성 곡선은 왼쪽이고 시험 곡선은 오른쪽임.



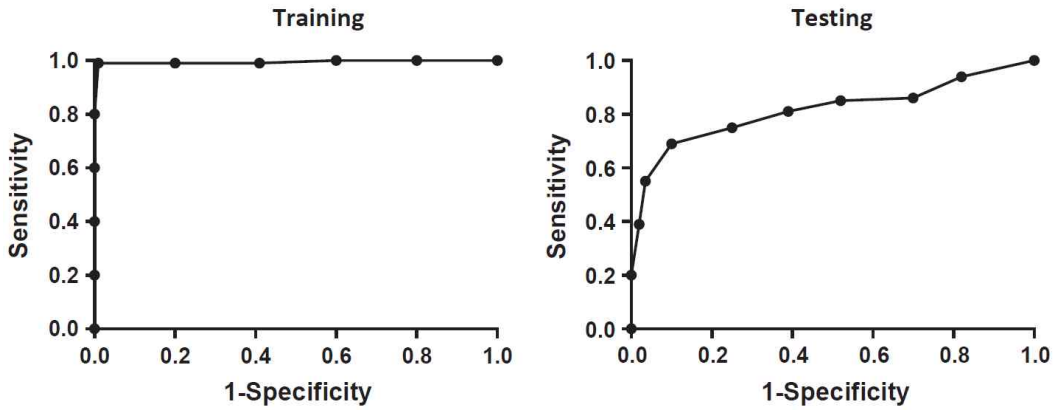
[그림 7-7] 조산 예측 모델에 대한 수신자 작동 특성(reciever operating characteristic, ROC) 곡선: 인종을 합한 태아 혈장

Y 축 = 민감도; X axis = 1 특이도. 훈련 수신자 작동 특성 곡선은 왼쪽이고 시험 곡선은 오른쪽임.



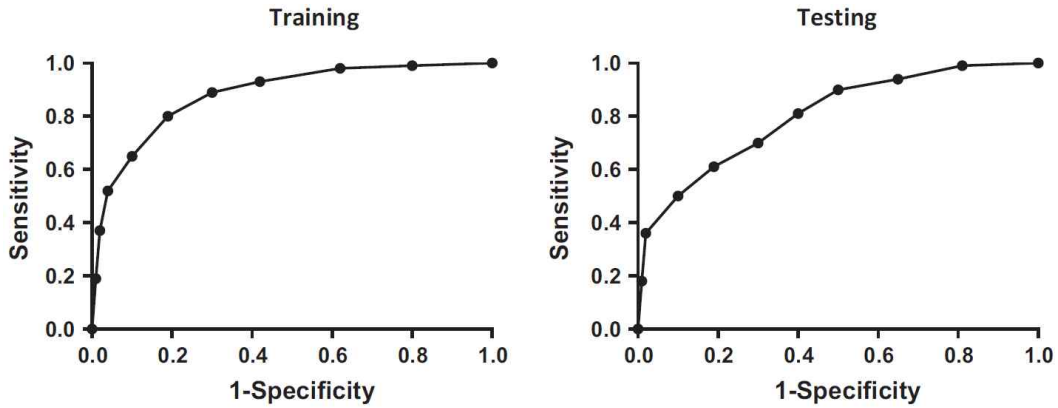
[그림 7-8] 조산 예측 모델에 대한 수신자 작동 특성(reciever operating characteristic, ROC) 곡선: 흑인 태아 혈장

Y 축 = 민감도; X axis = 1 특이도. 훈련 수신자 작동 특성 곡선은 왼쪽이고 시험 곡선은 오른쪽임.



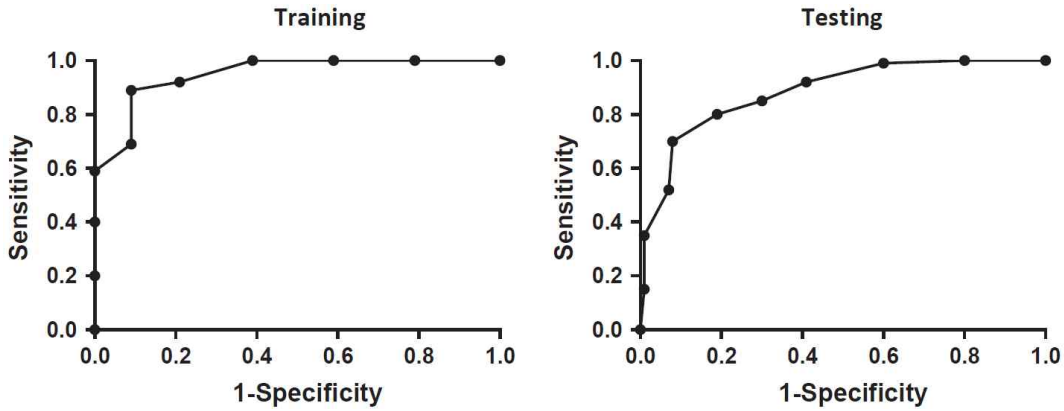
[그림 7-9] 조산 예측 모델에 대한 수신자 작동 특성(reciever operating characteristic, ROC) 곡선: 백인 태아 혈장

Y 축 = 민감도; X axis = 1 특이도. 훈련 수신자 작동 특성 곡선은 왼쪽이고 시험 곡선은 오른쪽임.



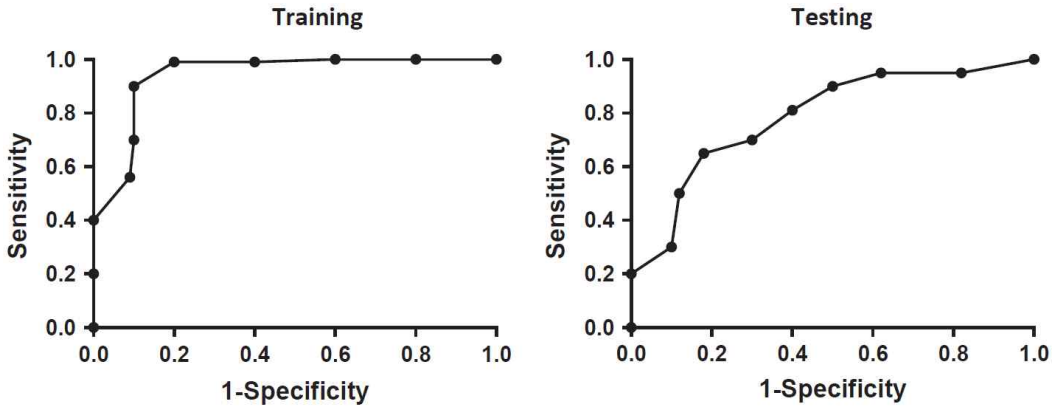
[그림 7-10] 조산 예측 모델에 대한 수신자 작동 특성(reciever operating characteristic, ROC) 곡선: 인종을 합한 양수

Y 축 = 민감도; X axis = 1 특이도. 훈련 수신자 작동 특성 곡선은 왼쪽이고 시험 곡선은 오른쪽임.



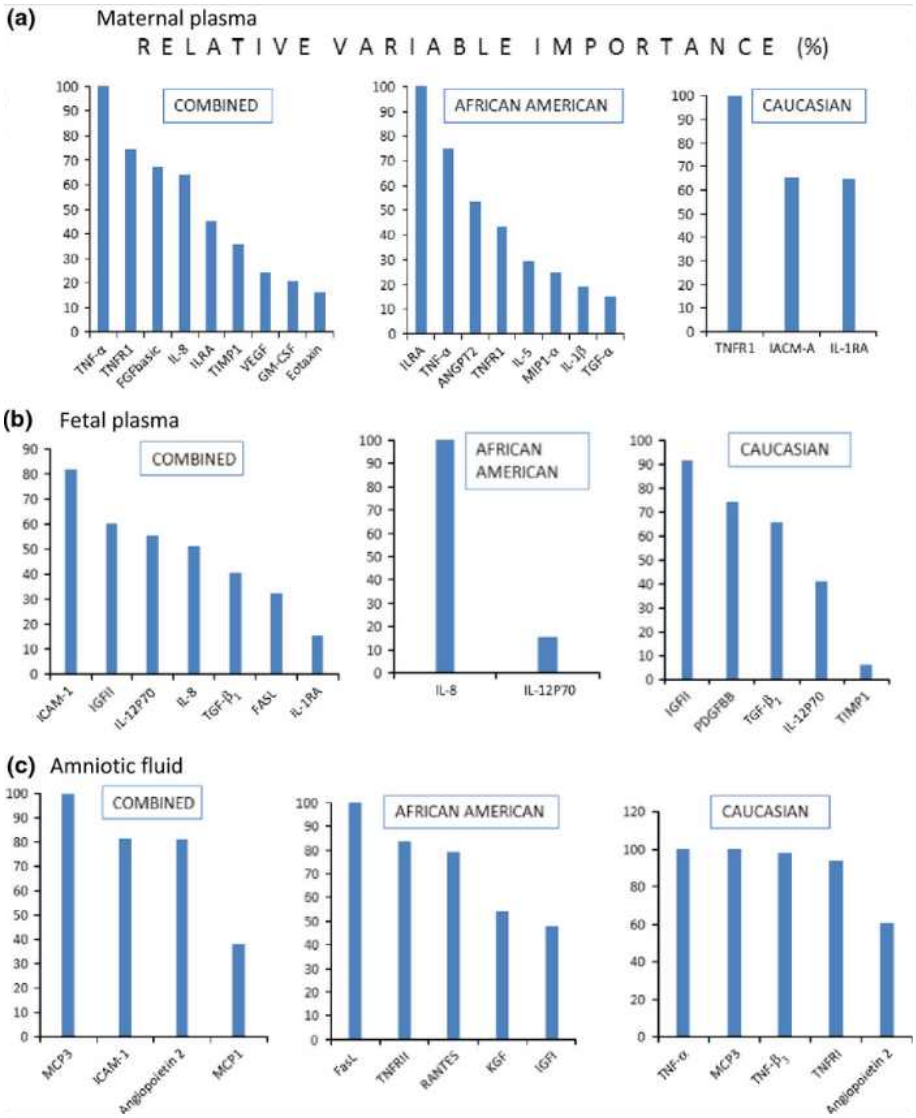
[그림 7-11] 조산 예측 모델에 대한 수신자 작동 특성(reciever operating characteristic, ROC) 곡선: 흑인 양수

Y 축 = 민감도; X axis = 1 특이도. 훈련 수신자 작동 특성 곡선은 왼쪽이고 시험 곡선은 오른쪽임.



[그림 7-12] 조산 예측 모델에 대한 수신자 작동 특성(reciever operating characteristic, ROC) 곡선: 백인 양수

Y 축 = 민감도; X axis = 1 특이도. 훈련 수신자 작동 특성 곡선은 왼쪽이고 시험 곡선은 오른쪽임.



[그림 7-13] 모든 구획/인종에 대한 상대 변수 중요성의 그래픽 표현

FGF basic, fibroblast growth factor basic; GMCSF, granulocyte macrophage colony stimulating factor; ICAM1, intercellular adhesion molecule; IGF, insulin like growth factor; IL; interleukin; IL1RA, interleukin 1 receptor antagonist; KGF, keratinocyte growth factor; MCP, monocyte chemotactic protein; MIP1α, macrophage inflammatory proteins alpha; PDGFBB, platelet derived growth factor; RANTES, regulated on activation, normal T cell expressed and secreted; TGF, transforming growth factor; TIMP1, tissue inhibitor of metalloproteinase 1; TNFR, tumor necrosis factor receptor; TNFα, tumor necrosis factor alpha; VEGF, vascular endothelial growth factor.

4) 결론

- 다변량 적응 회귀 스플라인은 조산과 관련된 여러 생물학적 지표를 모델링하고 인종적 차이를 보여줌.

3. Fang X, Li R, Kan H, Bottai M, Fang F, Cao Y. Bayesian model averaging method for evaluating associations between air pollution and respiratory mortality: a time-series study. BMJ Open. 2016 Aug 16;6(8):e011487.

1) 배경

- 여러 시계열 연구에서 PM_{10}을 포함한 흡입 가능한 거친 입자와 일일 호흡기계 사망자 수가 양의 상관관계를 나타냄..

2) 목적

- 시계열 연구에서 흡입 가능한 거친 입자(PM_{10})와 호흡기계 사망률 간의 연관성을 평가함.

3) 방법: 시계열 연구

- 2009년 1월 1일부터 2010년 12월 31일까지 중국 북부의 대도시 대기 모니터링 센터를 보유한 8개 지구의 일일 호흡기계 사망자 수, 대기 질(PM_{10}, NO_x, CO), 기상 조건(일일 평균 기온, 상대습도, 풍속, 기압) 데이터를 사용함.

- 호흡기 사망률 데이터는 지역별 사망원인 기록부(CDR, Caused of Death Registry)에서 얻음.

- 모든 CDR 사망자는 ICD-10 (International Classification of Diseases)에 따라 분류되었고, J00-J98인 경우 호흡기 질환으로 인한 사망으로 확인함.

- 1,038만명의 영주권자와 9,559명의 호흡기계 사망자가 포함되었음.

- 일일 호흡기계 사망자수는 과대산포(overdispersion)를 설명하기 위하여 분산(variance)이 평균과 같다는 분포가정을 완화하여 quasi-Poisson 분산을 따른다고 가정하였음. 일일 호흡기계 사망자

수와 역일(calendar day), 기온, 기압과의 비선형 관계를 고려할 때, 이러한 비모수 성분(non-parametric components)과 지구(district) 수준의 무작위 효과를 설명하기 위해 일반 부가 혼합 모형(GAMM, generalised additive mixed models)을 사용함.

- 자연 스플라인(natural splines)은 잠재적 교란요인인 기상 조건 및 요일(DOW, day of the week)을 조정하여 사망률의 비선형 추세를 맞추기 위하여 사용됨.

- 단일 오염물질에 대한 완전한(full) 일반 부가 혼합 모형(GAMM)은 미세먼지(PM_{10}), 상대 습도, 풍속, 요일(DOW)과 역일, 온도, 기압, 지구(district)의 무작위 효과에 대한 평활함수(smoothing functions)가 포함되어 있음.

$$\begin{aligned}\mathrm{Log}(E(y_{i,t})) = {} & \beta_0 + \beta_1 \times \mathrm{PM}_{10_{i,t}} + \beta_2 \\ & \times \text{relative humidity}_t + \beta_3 \\ & \times \text{wind speed}_t + \beta_4 \times \mathrm{DOW}_t \\ & + s(\text{day}_t, n_1/\text{year}) \\ & + s(\text{temperature}_t, n_2) \\ & + s(\text{Barometric pressure}_t, n_3) + \beta \\ & \times \text{district}_i + Z_i\mu + \log(\text{population}_i) \quad (1)\end{aligned}$$

- $E(y_{i,t})$는 t 일째 지구 i 에서 예상되는 사망자 수이고, DOW 는 요일에 대한 더미변수, i 는 8 개 지구에 대한 더미변수이고, Z_i 는 지구에 대한 무작위 절편(random intercept)임.

- s(.)는 자연 입방 스플라인(natural cubic splines)에 의해 실현되는 평활함수로, 장기적 시간 추세를 보정하기 위해 n_1 매듭을, 온도를 보정하기 위해 n_2 매듭을, 대기압을 보정하기 위해 n_3 매듭을 사용함.

- 자연 입방 스플라인은 이계도함수(second derivatives)가 0인 극한(limits)에서 유인성이 떨어지지만, 극한 주변에서 큰 분산(variance)을 나타냄.

- Poisson 회귀 모형에서 지구의 연간 평균 인구 크기를 오프셋(offset)으로 사용하였고 NO_x와 CO를 포함하는 다중 오염물질 모형을 사용하여 잠재적 교란요인의 가능성을 보정하였음.

- 역일, 온도, 상대 습도에 가장 적합한 매듭 수의 최적 GAMM은 Akaike information criterion (AIC)를 최소화하여 결정함.

- 식 1에 의해 주어진 모형 M_k를 고려하면 β가 관심계수(coefficient of interest)인 경우, 데이터 D의 사후분포는 다음과 같음.

$$\mathrm{pr}(\beta|D) = \sum_{k=1}^{K} \mathrm{pr}(\beta|M_k, D)\,\mathrm{pr}(M_k|D) \qquad (2)$$

- 이는 사후확률 모형에 의해 가중된 GAMM 모형에 의해 고려된 사후분포의 평균으로 M_1, M_k는 고려된 모형들임. 사후확률 모형 M_k는 다음과 같음.

$$\mathrm{pr}(M_k|D) = \frac{\mathrm{pr}(D|M_k)\,\mathrm{pr}(M_k)}{\sum_{l=1}^{K}\,\mathrm{pr}(D|M_l)\,\mathrm{pr}(M_l)} \qquad (3)$$

- M_k는 사전확률 $\mathrm{pr}(M_k)$가 참인 경우의 잠재적 기저 모형 중 하나임.

$$\mathrm{pr}(D|M_k) = \int \mathrm{pr}(D|\beta_k,\, M_k)\,\mathrm{pr}(\beta_k|M_k)\,d\beta_k \qquad (4)$$

- 식 4에서 는 M_k의 계수이며, $\mathrm{pr}(\beta_k|M_k)$은 모형 M_k에서의 β_k의 사전밀도, $\mathrm{pr}(D|M_k)$은 우도(likelihood)이며 모든 확률은 조건부임. β의 사후확률 평균과 분산은 다음과 같음.

$$E[\beta|D] = \sum_{k=1}^{K} \hat{\beta}_k \, pr(M_k|D) \qquad (5)$$

$$Var[\beta|D] = \sum_{k=1}^{K} (Var[\beta|D, M_k] + (\hat{\beta}_k - E[\beta|D])^2) \, pr(M_k|D), \qquad (6)$$

- β 의 95% 베이즈 신뢰구간(Bayesian credible interval)은 다음과 같음.

$$E[\beta|D] \pm 1.96\sqrt{Var[\beta|D]} \qquad (7)$$

- 각 모형에 대한 사후확률은 다음과 같음.

$$pr(M_k|D) = \frac{pr(M_k) \cdot e^{-0.5(BIC_{M_k} - \overline{BIC})}}{\sum_{l=1}^{K} pr(M_k) \cdot e^{-0.5(BIC_{M_l} - \overline{BIC})}} \qquad (8)$$

- $BICM_k$ 는 모형 k 의 베이즈 정보기준(BIC, Bayesian information criterion)이며, $\overline{BIC}$ 는 BIC_{Ml} 의 평균임.

- 식 5 의 $\hat{\beta}$ 및 식 6 의 Var(β|D)는 관심모수(parameter of interested)의 사후평균과 분산이며, 식 8 에서 BIC 를 사용하여 베이즈 요인(Bayes factors)를 근사화함.

- $pr(M_k)$는 균등분포로부터 가정됨.

$$pr(M_k) = \frac{1}{K} \qquad (9)$$

- 최고의 모형은 식 8 에 의해 크게 가중치가 부여된 식 5 이며 절약모형(parsimonious model)을 선호함.

- PM_{10}, NO_x, CO 의 상관관계를 고려하기 위해 주성분분석(PCA, principal component analysis)에서 유도된 주성분(principal components)를 다중 오염물질 모형에 도입하여 3 가지 오염물질 간의 동일선상의 영향을 배제함.

4) 결과

- 일반 부가 혼합 모형(GAMM)+베이지안 모형 평균(BMA)과 최적(optimal) 일반 부가 혼합 모형(GAMM)를 비교할 때, PM_{10}에 의한 일일 호흡기계 사망률은 1 사분위 범위(IQR, interquartile range) 당 1.38% vs 1.39%, 1.81% vs 1.83%, 0.87% vs 0.88% 상승함.

- 하지만 일반 부가 혼합 모형(GAMM)+베이지안 모형 평균(BMA)는 단일 오염물질 모형과 주성분 모형(PCs-based model)에 대하여 신뢰구간이 더 넓음. 다중 오염물질 모형에 대해서는 비슷함.

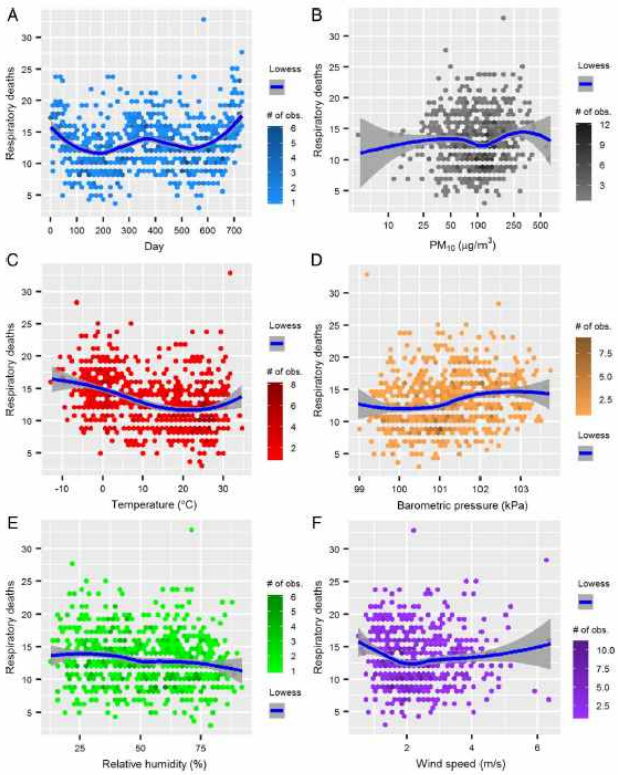


Figure 2 Relationship between number of daily respiratory deaths and (A) days; (B) PM_{10} concentrations and (C–F) meteorological conditions. Lowess; locally weighted scatterplot smoothing.

[그림 7-14] 일일 호흡기계 사망 수와 (A) 일자; (B) PM_{10} 농도와 (C-F)기상 조건

<표 7-10> 일반화 선형 혼합 모형, 최적 일반화 부가 모형과 일반부가혼합모형+베이지안 모형 평균으로부터 PM_{10}의 사분위수 증가와 연관된 일일 호흡기계 사망률의 백분율 증가

Table 5 Per cent increase in daily respiratory MR associated with an IQR increase in PM_{10} concentration from GLMM, optimal GAMM and GAMM+BMA

	Single-pollutant		Multipollutant		Multipollutant (PCA)	
Model	Per cent	95% CI	Per cent	95% CI	Per cent	95% CI
GLMM	3.07	(0.91 to 5.27)	1.94	(−0.80 to 4.75)	1.47	(−1.17 to 4.17)
Optimal GAMM*	1.39	(−1.08 to 3.93)	1.83	(−1.11 to 4.83)	0.88	(−2.03 to 3.88)
GAMM+BMA	1.38	(−1.09 to 4.28)	1.81	(−1.12 to 4.85)	0.87	(−2.23 to 4.07)

*Knots for days, temperature and barometric pressure are 14, 7 and 4, respectively.
BMA, Bayesian model averaging; GAMM, generalised additive mixed model; GLMM, generalised linear mixed model; MR, mortality rate; PCA, principal component analysis.

5) 고찰 및 결론

- 베이지안 모형 평균(BMA)은 시계열 연구에서 대기오염의 치명적 건강영향을 평가할 때 유용한 도구임.

4. Joan Forns, Siddhartha mandal et al. Novel application of statistical methods for analysis of multiple toxicants identifies DDT as a risk factor for early child behavioral problems. Environmental Research. 2016.July; 151(2016) 91-100.

1) 배경

- 지속적 유기 오염물질(persistent organic pollutants, POPs)은 잠재적 독성을 가지고 있는 것으로 잘 알려진 환경 화학물질로 자연에서 반감되지 않고 인간과 야생에서 축적되며 음식체인을 통해 농축됨. 지속적 유기 오염물질(POPs)은 PCB, 1,1,1-trichloro-2,2-bis 등 다양한 화학물질을 포함함. 신생아는 태반과 모유를 통해 지속적 유기 오염물질(POPs)에 노출됨. 많은 지속적 유기 오염물질(POPs)이 전세계적으로 금지되어 있지만 긴 반감기나 환경에서의 연속성 때문에 현재에도 혈액에서 감지가 됨.

- 주로 자궁에서 단일 오염원의 노출과 신경생물학적 발달사이의 관계 대한 연구에 기초하여 PCB, PBDEs, p,p-DDE, HCB는 발달적 신경독성물질로 간주됨. 그러나 아이들은 노출의 근원에 함께 존재하는 다중 오염원에 노출됨. 오늘날 연구에서 오염원과 건강사이의 관계에 대한 연구를 위하여 다중오염원 접근방법이 사용됨. 일반적으로 혼합물의 잠재적 신경독성에 대한 지식은 결여되어 있고, 특히 모유를 통한 노출로부터의 결과도 그러함.

2) 목적

- 상관관계가 있는 노출자료를 다룰 수 있는 통계적 방법을 사용하여 모유에서 측정된 24개의 지속적 유기 오염물질(POPs)에 출생 후 노출과 12개월과 24개월에 나타나는 이른 행동 문제의 관계를 평가함

3) 연구방법

(1) 연구자료 및 연구대상자

- 노르웨이의 HUMIS 코호트 자료 사용 : 2002-2009년 사이 모집된 모자 코호트.

- 병동에서 모집된 산모를 제외하고는, 노르웨이에서 엄마가 된 사람들에게 수행하는 일반적인 가정방문을 통해 분만 후 2주안에 공중보건간호사에 의해 모집되었음. 코호트의 참가자는 2606명이나 재정적인 문제로 612명의 모유를 분석하였고, 498명은 랜덤하게 추출되었으며, 54명은 조산한 사람에서, 60명은 신생아의 발달이 빠른 사람에서 과 추출(oversampling)함.

(2) 노출평가

- 참가자들은 25mL 모유를 8일 동안 매일 아침 샘플링함. WHO권고에 의하면 가급적 2주에서 2달 사이에 샘플링하도록 되어있으나 그렇지 않아도 수용함. 냉장보관하면서 250mL 박스에 모았으며, 채집 날짜와 시간을 작성하도록 함. 수유 펌프를 사용해도 됨. 250mL 박스가 다 차면 냉동이 유지되도록 하여 일반우편으로 NIPH에 보냄.

- 612명의 산모의 모유에서 13종의 PCB, 6종의 PBDE, 5개의 유기염소 화합물을 포함한 24개의 다른 POPs 물질을 측정함

(3) 건강영향평가

- 산모들이 수행한 Infant/toddler symptom checklist (ITSC) long version의 하위 항목을 사용하여 12개월과 24개월에 아동행동문제를 평가함. ITSC는 자기제어와 기질양상을 평가하고 야단법석떨거나 훌쩍거림에서 크게 소리지르며 우는 것으로 빠르게 넘어가는 행동을 일으키는 제어 문제와 7-30개월에 수면과 음식섭취가 어려운 것을 확인함. 본 연구에서는 자기제어, 집중, 수면, 섭취, 옷입힐 때와 목욕 시 접촉, 언어듣기의 하부척도에 대한 문항을 포함함.

각 항목에 대하여 아동이 설명에 '절대 하지않는다'거나 '가끔' 부합한다

면 0점, 과거에 설명에 부합했다면 1점, 거의 모든 시간 설명과 같다면 2점을 체크함. 각 항목의 점수는 더하여 총점으로 계산함. 더 높은 점수는 더 나쁜 행동 문제를 나타냄. 현 연구에서 12개월의 ITSC는 총 28개 항목을 포함(평균:2.73, SD:3.41 범위:0-24)하고, 24개월의 ITSC는 총 33개 항목을 포함(평균: 2.66, SD:2.67, 범위:0-22)함. 본 연구에서 12개월과 24개월의 ITSC의 내적 일관성은 충분함. 12개월과 24개월의 ITSC의 spearman 상관계수는 0.34임.

(4) 보정변수

- 신생아 성별, 출생 몸무게, 재태주 수, 임신 중 흡연, 출산 당시 산모 나이, 교육정도, 산과력, 출산 간격, 모유 기간, 임신 전 산모 체질량지수(body mass index, BMI), 임신 중 해산물 섭취량, 산모 스트레스 정도

(5) 통계분석 방법

- 공변량의 집합과 결과 사이의 이변량 관계에 대해 t검정(t-test)이나 분산분석(anova)으로 테스트하여 12개월과 24개월의 행동문제의 결정요인을 연구함.

- 24개의 지속적 유기 오염물질(POPs) 사이의 스피어만(spearman) 상관관계를 평가함

- 상관관계가 있는 변수를 적은 수의 비상관 주성분으로 변형하는 방법인 주성분 분석(PCA, principal component analysis)을 시행, 고유값(Eigenvalue)>1인 PCA 구성성분은 남김, 회귀분석에서 남겨진 주성분을 사용함. (주성분 분석은 각 변수의 가중치를 찾는 방법임. 이때 이 가중치는 데이터의 분산이 가장 크도록 찾아짐. 주성분 분석의 핵심은 찾아진 가중치에 의해 변환된 새로운 값이 최대의 분산을 갖도록 가중치를 찾는다는 점, 특징이 없는 변수의 가중치< 특징이 많은 변수의 가중치)

- 그 후 두가지 변수 선택 기술을 시행 : 탄성 그물 & 베이지안 모형 평균(Elastic net & Bayesian model averaging)

- 탄성 그물(E-net)은 예측치 사이의 극도의 강력한 상관관계가 있는 능선(Ridge)와 올가미(Lasso) 페널티의 결합형인 하이브리드 페널티 회귀분석으로, 예측치가 연관성이 높을 때 Lasso방법의 불안정성을 해결하여 고차원 데이터를 분석하는데 그 목적이 있음. 계수의 절대치의 합인 Lasso 페널티(람다)는 0을 향해 회귀계수를 축소시킴.

- Ridge 회귀분석은 회귀계수의 제곱을 페널티함. 0을 향해 비례하는 상관관계 변수의 계수를 축소하지만 절대 정확히 0이 되진 않음. Lasso는 상관관계가 있는 변수의 그룹으로부터 한 개의 변수를 선택하는 경향이 있는 반면에 E-net은 공선성있는 변수의 그룹을 선택할 수 있음.

- 두 번째 튜닝 매개변수(알파)는 Lasso(1)와 Ridge(0) 페널티 사이의 밸런스를 컨트롤함. 우리는 교차 검증을 사용하여 알파와 람다를 최대한 적합하게 만듦.

- 다음으로 적절한 변수를 선택하는 동안 모델의 불확실성을 다루기 위하여 BMA를 수행함. 모델 불확실성을 무시하는 전통적인 변수 선택방법과 비교하면 BMA는 가장 좋은 한가지 모델을 선택하는 것이 아니라 가능성이 있는 모델들의 범위의 평균에 의하여 추론함. BMA에 기초한 신뢰구간은 모델 내 샘플링한 편차와 모델간의 불확실성 둘 다를 고려하여 보정함. (classic한 베이지안의 평균 분석 방법은 모델 간 편차만 고려함)

- BMA에 의한 모델의 선택은 가장 높은 사후모형확률(PMP, posterior model probability)을 사용함. 또한 사후포함확률 PIP(posterior inclusion probability)의 결과를 분석하였음. 변수의 PIP는 데이터가 주어진 모든 모델에 대한 평균화된 모델에 포함되는 변수의 사후확률임.

- 변수선택 과정 후 선택된 혼란변수를 보정하고 Enet과 BMA에 의해 선택된 유해물질을 12개월과 24개월의 문제행동에 대한 가중치 다

중선형회귀분석에 적용함. 신생아의 빠른성장과 조산으로 인한 과표집(oversampling)을 설명하기 위해 가중치 회귀분석을 사용함

- 최종적으로 가중치 다변량 선형회귀분석을 한 이유는 다음과 같음 1) 공변량을 보정하면서 결과에서 선택된 변수들의 결합효과를 측정하기 위해서 2) 변수축소선택은 편향된 효과측정을 제공하고 신뢰구간 계산이 어려울 수 있음 3) 일반적인 최소제곱 매개변수는 설명이 쉬움

- 회귀분석은 선택된 유해물질을 사분위수 증가에 따른 연속변수로 포함하여 수행되어짐

- 문헌에 기록된 관계를 나타내는 비순환 그래프에 기초하여 혼란변수를 선택함 : 모유 샘플링 당시 아동의 나이(일), 임신중 해산물 섭취(횟수/개월), 산모나이, 산모 교육상태(12년이하, 또는 12년 초과), 집중적으로 모유수유한 기간(주), 부분적으로 모유수유한 기간(주), 산과력(초산, 경산), 임신전 PMI(연속변수), 아동성별, 임신전 흡연(전혀/가끔/계속), 조산여부

- 민감도 분석에서 출생 후 첫 30일 이내에 수집된 모유샘플로 제한하여 주 분석을 반복함.

- 모유 수집방법과 6개월째 산모의 스트레스를 보정하여 주 분석을 반복함.

- 통계분석과 수학적 모델링은 stata13 과 R을 사용함.

- 결과는 사분위 범위 당 증가임.

<표 7-11> 각 통계 방법의 목적과 사용 이유

	주목적	강점	약점	현 연구에 사용한 이유
주성분 분석 (PCA)	설명된 변이도 (variability)에 기초한 변수 선택	차원 축소에 유용하며 최소 추정 주성분이 독립적임	구성성분의 설명가능성이 결여	노출에서 설명 변수인 유해물질의 그룹을 확인하기 위해
탄성그물회귀 (ENET)	많은 예측치중 소수만이 인과관계가 있다는 가정하에 얻은 결과에 대해 희소성을 보장함	예측치(predictors)의 수 > 관측치(observations)의 수일 때 특히 가치있음. 변수의 그룹화를 지원함 여기서 강한 상관관계가 있는 예측자들은 모델 안에 함께 있거나 함께 빠지는 경향이 있음.	선형관계만 고려하고 상호작용은 고려하지 않음. 공선성을 처리하지 않으므로 상관관계가 있는 변수의 그룹을 선택할 수 있음 축소 예측으로 인한 편향된 계수 값	결과와 가장 강하게 관련된 유해물질을 선택하기 위해
베이지안 모형 평균(BMA)	사후확률에 기초한 모델 선택	모델의 불확실성을 설명함. 변수의 모든 조합으로	결과는 설명하기에 복잡할 수 있음.	결과와 가장 강하게 관련된 유해물질을 선택하

		구성된 전체 모델의 평균	선형 모델 또는 회귀모델에 제한적임	기 위해
가중된 다중회귀(weighted multiple regression)	변수선택의 효과를 끌어내기 위해서 확증적 분석	예측에 잘 알려진 모델, 설명하기 쉬운 추론 통계가 가능함	다중공선성에 민감함 선형관계에 있지 않을 때 덜 민감함 outlier에 민감함	앞의 통계방법에의해 선택된 유해물질간의 관계를 수치화하기 위해

4) 결과

- 남자 아이가 미미하게 여자아이보다 12개월에 더 문제 행동이 나타남. 그러나 24개월에서는 차이가 없음. 문제적 행동은 12개월에 조산아 사이에서 더 높게 나타남. 24개월에서는 차이 없음. 젊은 산모와 덜 교육된 산모의 아이가 12개월과 24개월에 더 문제적 행동을 보임.
- 일반적으로 모든 유해물질 사이에 통계적으로 유의한 상관관계를 발견함 : 가장 높은 상관관계를 유기 염소 살충제와 다른 PCBs 물질 사이에 있었고, 가장 낮은 상관관계는 유기염소 살충제와 PBDEs 사이에 있었음.

○ PCA

- PCA의 결과 4개 주 성분은 고유값(eigenvalue)>1임. 4개의 주성분이 전체 분산의 80%를 설명함. 주성분 1은 PCBs 물질, 주성분 2는 PBDE물질로 구성, 주성분 3은 유기염소물질 살충제로 구성, 주성분4는 HCB, PCB99, PCB 194로 구성.

- 주성분 회귀분석에서 PC3 와 12개월째 행동문제 사이에 양의 상관관계가 있음. 24개월째 문제행동은 어떤 주 성분과도 관계가 없었음.

<표 7-12> 주성분 분석 상관관계

Component	Variance	Difference	Proportion	Cumulative
Comp1	11.63	6.92	0.48	0.48
Comp2	4.71	3.08	0.20	0.68
Comp3	1.63	0.28	0.07	0.75
Comp4	1.36	0.42	0.06	0.81

Comp: 요소(Component)

<표 7-13> 0.25 이상 값을 가진 주성분(고유벡터)

Variable	Comp1	Comp2	Comp3	Comp4	Unexplained
PBDE28		0.26			0.62
PBDE47		0.43			0.09
PBDE99		0.44			0.05
PBDE100		0.44			0.04
PBDE153		0.37			0.28
PBDE154		0.44			0.07
HCB			0.28		0.57
B-HCH			0.29	0.61	0.19
oxyCD					0.26
p,p'-DDE			0.38		0.33
p,p'-DDT			0.53		0.35
PCB74					0.32
PCB99				-0.26	0.14
PCB105					0.17
PCB114	0.27				0.14
PCB118	0.26				0.13
PCB138	0.27				0.12
PCB153	0.28				0.04
PCB156	0.27				0.05
PCB157	0.25				0.15
PCB170			-0.25		0.21
PCB180	0.27				0.07
PCB189	0.25		-0.28		0.12
PCB194				0.44	0.16

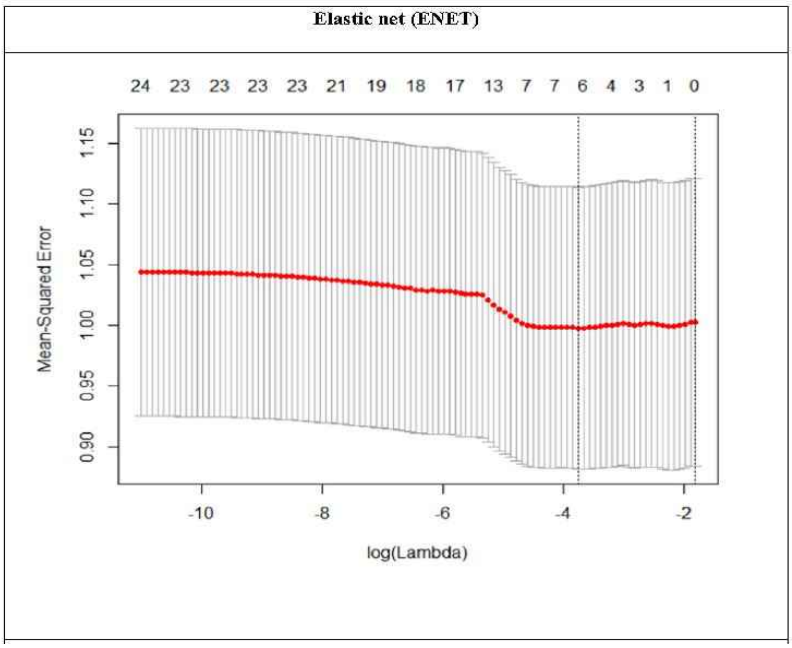
<표 7-14> 영유아 증상 체크리스트(ITSC)로 평가한 12주와 24주의 행동 문제에 대한 주성분 분석

Table 4
Principal component regression analysis with behavioral problems at 12 and 24 months assessed with the infant toddler symptom checklist (ITSC).

	PC1 (PCBs)			PC2 (PBDEs)			PC3 (Organochlorine Pesticides)			PC4 (HCB, PCB99 and PCB194)		
	Coef	95%CI		Coef	95%CI		Coef	95%CI		Coef	95%CI	
ITSC12	0.01	(−0.08,	0.09)	−0.03	(−0.15,	0.09)	0.50	(0.19,	0.81)	−0.07	(−0.26,	0.12)
ITSC24	−0.06	(−0.14,	0.01)	−0.02	(−0.14,	0.11)	0.08	(−0.09,	0.25)	−0.02	(−0.15,	0.11)

○ elastic net

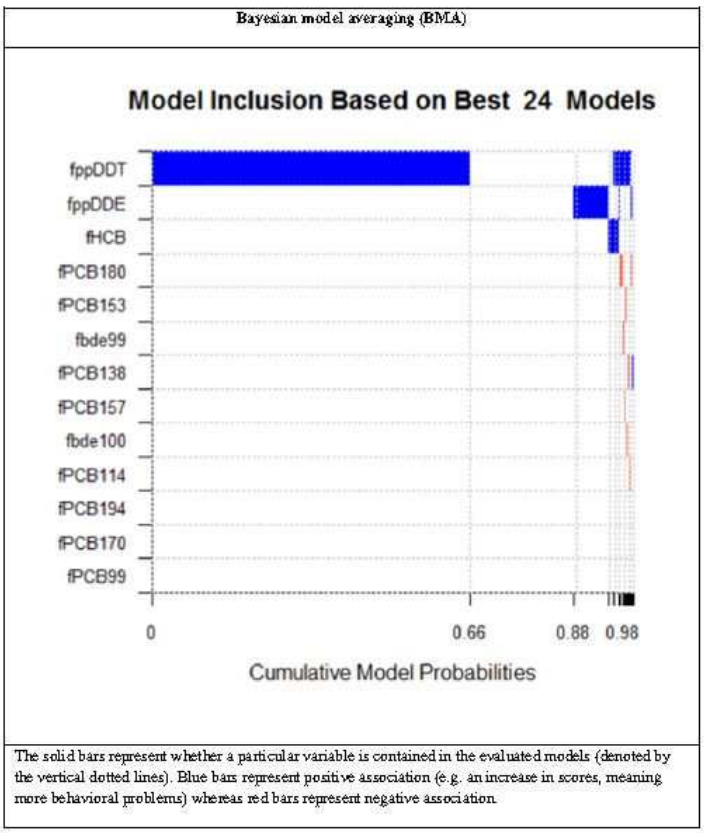
12개월 아동의 문제 행동과 관련있는 p,p'-DDT, p,p-DDE, HCB, PCB156, PBDE28, PBDE47이 최소 평균제곱 오차(MSE) 모델에서 선택됨. 그러나 오차 곡선은 결정적이지 않았음. MSE의 유효성을 최소화하는 고유 튜닝 매개변수가 없었기 때문.



[그림 7-15] 영유아 증상 체크리스트(ITSC)로 평가한 12개월의 행동 문제와 연관된 독성물질을 탄성그물 방법을 사용한 결과

○ BMA

- BMA에서는 최대 사후 확률로 p,p-DDT만 선택됨 (사후모델확률 PMP=0.66) 사후모델확률의 측면에서 다음 모델은 절편만 포함되었음. p,p-DDT가 12개월의 행동문제와 관련있는 단독 변수라는 것은 PIP의 결과들로 확인됨(PIP=0.749)



[그림 7-16] 영유아 증상 체크리스트(ITSC)로 평가한 24개월의 행동문제와 연관된 독성물질을 베이지안 모형 평균을 사용한 결과

<표 7-15> 탄성 그물과 베이지안 모형 평균으로 선택된 특정 주요 유기 오염물질을 포함하여 12개월의 행동 문제에 대한 가중 다중 선형 회귀

	PIP	Posterior Mean	Posterior SD
p-p'-DDT	0.7490	0.1637	0.1068
p-p'-DDE	0.0560	0.0005	0.0022
PCB153	0.0350	-0.0002	0.0020
BDE100	0.0327	-0.0038	0.0355
HCB	0.0320	0.0014	0.0087
PCB180	0.0243	-0.0004	0.0036
PCB138	0.0137	-0.0001	0.0018
PCB114	0.0090	-0.0021	0.0760
PCB194	0.0030	-0.0003	0.0090
BDE99	0.0027	-0.0002	0.0064
PCB99	0.0013	0.0003	0.0083
PCB170	0.0013	0.0000	0.0025
PCB157	0.0007	-0.0002	0.0113

24개월에 행동문제와 관련있는 유해물질을 선택하는데 사용된 다양한 변수선택방법은 일관성이 없었음. Enet은 두 개의 PCB물질(PCB170, 194)를 선택한 반면 BMA는 최고 사후 확률 모델에서 절편만 포함함. 게다가 BMA로부터 결과는 가장 높은 PIP값을 가지는 변수는 PCB118 이었는데 그 값이 매우 낮았음. 이러한 결과로 24개월의 결과와 어떤 유해물질이 관련이 있다는 증거는 부족한 것으로 생각되나 PCB 179와 194를 최종 다중선형회귀 분석에서 테스트해봄.

○ 행동문제에 대한 선택된 유해물질의 효과 평가

12개월의 ITSC의 중간값은 2임(IQR=4). 선형회귀분석 결과 p,p'-DDT 1사분위수 증가와 관련된 12개월의 ITSC의 전체 점수는 0.62 증가 됨. 삼분위에서 p,p'-DDT를 테스트했으나 양-반응 관계는 관찰되지 않음. p,p'-DDT와 12개월의 문제행동사이의 관찰된 관계는 산모 교

육수준에 의해 보정됨. 교육수준이 낮은 산모 그룹(ß=0.59; 95% CI : 0.38, 0.81)이 교육수준이 높은 그룹(ß=0.14; 95%CI : -0.05, 0.34)보다 더 연관성 높게 나타남. 12개월 아동의 문제 행동과 p,p'-DDT사이의 관련성에 대해 아동성별, 집중 수유 기간에 따른 유의한 효과는 발견하지 못함. 최종적으로 Enet에서 선택한 6가지 유해물질을 포함시키자 p,p'-DDT만 12개월의 행동문제 위험을 증가시킴.

<표 7-16> 12개월의 행동 문제에 대한 가중 다중 선형 회귀와 다중 오염 모형에서 p-p'-DDT, p-p'-DDE 와 HCB Coef: 계수

	Coef	95%CI
p,p'-DDT (per IQR increase)	0.35	(0.08, 0.63)
p,p'-DDE (per IQR increase)	0.10	(-0.31, 0.51)
HCB (per IQR increase)	0.19	(-0.23, 0.62)
PCB156 (per IQR increase)	-0.39	(-0.80 0.02)
PBDE28 (per IQR increase)	-0.04	(-0.11, 0.05)
PBDE47 (per IQR increase)	-0.01	(-0.01, 0.03)

(Coeficient); 95%CI: 95% 신뢰구간(Confidence Interval); IQR: 사분위(Interquartile range).

- 24개월의 ITSC는 3(IQR=2) 중간값을 가짐. 변수선택 방법에서 24개월의 행동문제와 관련성이 분명한 유해물질을 제안하지 않았지만 좀 더 관련이 있어 보이는 두 개 변수(PCB170,194)를 포함시켜 다중 선형 회귀분석을 수행함. 노출과 24개월 문제 행동 사이에 유의한 관계는 없었음. 출생 후 30일 안에 모집한 모유로만 주 분석을 수행하자 p,p'-DDT와 12개월의 문제행동 사이의 부정적인 관계가 나옴(ß=0.64; 95%CI: 0.56, 0.72). 조산아를 제외시켰을 때 동일한 독성 물질이 선택되었고, 다중 회귀분석의 결과도 동일함(ß=0.62; 95%CI= 0.46, 0.89). 주 분석의 결과는 모유 수집방법 또는 실험실을 보정한 후, 산모 스트레스에 대한 보정 후에도 변하지 않음.

5) 고찰 및 결론

- 이 연구는 상관관계가 있는 노출 데이터를 적절하게 다루는 PCA와 같이 전통적인 방법과 Enet이나 BMA와 같은 현대적인 방법을 사용하여 POPs와 이른 행동문제 사이의 관계를 평가한 첫 번째 연구임. 또한 아동의 행동 문제에 대한 위험요인으로 DDT를 밝힘. 최근 연구에서 논의된 바와 같이 PCA는 상관된 노출 매트릭스에서 차원을 축소하면서 혼합물의 노출 효과를 테스트하는 유용한 방법임. PCA결과에서 나온 주성분을 회귀분석에 포함시키자 혼합물로서 POPs의 주성분그룹의 효과를 평가할 수 있었음. 그러나 PC 회귀 계수는 직접적인 해석을 할 수 없고, 특히 동일 구성요소의 일부인 노출에 대해 추가(additive)효과나 상호적인 효과(synergistic 또는 antagonistic)를 고려하지 못함

- LASSO와 RIDGE 의 포함한 Enet은 결과 잠재적으로 관련있는 POPs를 확인하기 위해 사용함.

Enet은 LASSO의 한계를 극복하여 모델에서 고차원 변수와 상관관계가 높은 변수를 모두 허용함. 또한 변수 선택 시 그룹 효과를 허용하여 고도로 상관된 변수 그룹이 비슷한 크기의 계수를 가지는 경향이 있음. 이는 그룹이 미리 알려지지 않은 경우 상관관계가 있는 그룹을 선택할 수 있음. 특히 예측치가 관찰된 수 (p>>n)보다 매우 클 때 유용함. 그러나 선형관계만 고려하고 축소 측정으로 인하여 편향된 계수를 제공한다는 약점이 있음.

- BMA는 선형모델에서 설명변수의 PIP(posterior inclusion probabilities)에 기초한 베이지안 모델 선택 방법임. 변수의 모든 조합으로 구성된 전체 모델에 대해 평균화하고, 모델의 부확실성을 설명한다는 강점이 있음. 그 결과를 해석하는 것이 어려울수 있고 선형모델이나 로지스틱 모델에 제한된다는 약점이 있음.

- 세가지 방법의 (PCA, Enet, BMA)의 약점은 화학물질간의 상호작용에 대한 평가를 명시적으로 허용하지 않음.

5. Lenters et al. Prenatal phthalate, perfluoroalkyl acid, and organochlorine exposures and term birth weight in three birth cohorts: multi-pollutant models based on elastic net regression. Environmental Health Perspectives. 2016.March; 124(3):365-72.

1) 배경

일부 오래된 물질과 새로 떠오르는 환경 오염물질은 자궁내 성장 억제에 대한 위험 물질로 의심을 받고 있음. 그러나 그 증거가 모호한데 부분적으로는 화합물의 효과를 구분하는데 어려움이 있기때문임.

2) 목적

- 환경노출과 다양한 상관관계가 있는 생물학적 노출표지자와 출생체중 사이의 관계를 평가함.

3) 연구방법

(1) 연구자료 및 연구대상자

- INUENDO cohort

- 2002.6 ~ 2004.5에 산전 케어를 방문한 사람들 중 모집된 그린란드인 513명, 폴란드인 180명, 우크라이나인 557명에서 태어난, 제태주수 37주 이상인 단일 신생아 1250명을 평가함.

- 참가자 : 18세 이상, 각 국가에서 태어난 사람

(2) 노출평가

- 폴란드인은 33주, 그린란드 25주, 우크라이나 23주에 혈액샘플 채취

- diethylhexyl, diisononyl phthalates(DEHP, DiNP), 8 perfluouoalkyl acid, 유기염소(PCB-153, p,p'-DDE)의 이차 대사산물을 모체 혈청 샘플에서 분석

(3) 건강영향평가

- 코호트 enroll시 인터뷰
- 출생 결과 데이터는 산부인과 기록으로 확인

(4) 보정변수

- 혈중 코티닌, 임신전 BMI, vitaminD, lipid(triglycerides, cholesterol), 출산경력, 산모 교육정도, 신생아 성별, 해산물 섭취, 샘플링 한 계절, 혈중 알콜

(5) 통계분석 방법

가) elastic net regression

- 다중공선성이 있는 경우, multiple linear regression 모델은 모수 추정에 적합하지 않음. 노출의 독립적 효과를 추정하기 위해 변수선택방법인 elastic net regression을 unpenalized regression 모델과 함께 사용함.

- elastic net penalty to linear regression models : lasso penalty와 ridge penalty의 혼합형, 최종모델에서 collinear predictors를 선택할 수 있음

- lasso : 필요한 변수만 채택하고 나머지는 0으로 만듦 / 가중치의 절대값의 합을 최소

- ridge : 모든 변수를 가지고 오면서 계수값을 줄임 / 가중치의 제곱의 합을 최소

$$\min_{\beta_0,\beta_j,\beta_k}\left[\frac{1}{2n}\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{e}\beta_j X_{ij}-\sum_{k=1}^{c}\beta_k Z_{ik}\right)^2 + \lambda\sum_{j=1}^{e}\left(\frac{1-\alpha}{2}\beta_j^2+\alpha|\beta_j|\right)\right], \quad [1]$$

- 람다의 공식은 조정모수이고 역할은 전체 함수를 조절해주고 고차항의 모수들을 아주 작은 값으로 만들어서 적당한 2차항 형태의 함수로 만들어주는 정규화 역할

- a값은 lasso (a=1)와 ridge(a=0) 사이의 균형을 제어하고, λ는

penalty(혹은 정규화) 매개변수를 나타내며, 람다가 증가함에 따라 수축 정도가 증가함 (즉, 람다를 크게 할수록 조정할 변수의 크기는 작아짐)

- 잠재적인 혼란 변수를 완전히 조정하기 위해 혼란변수를 unpenalized 변수로 elastic net regression모델에 포함시키고, 변수선택에 혼란변수들이 종속되지 않도록 연역적 방법으로 혼란변수를 선택함

- 최적의 penalization의 정도를 결정하기 위해 교차 검증(CV, cross-validation) 시행. 알파와 람다의 시퀀스의 격자모델을 테스트하였고, 반복된 10배(10-fold) 교차검증으로부터 예측치의 평균의 제곱의 오차(MSE)의 최소치를 산출한 조합을 선택함. 단일 교차검증보다 안정적인 선택을 하기위해서 각각 다른 partition으로 교차 검증을 100번 반복함

- 교차검증은 시뮬레이션에서 선택의 좋은 민감성(true positive rate)을 보여주지만 부풀려진 위발견율(false discovery rates)로부터 시달릴 수 있음. 그러므로 교차검증을 보완하고 결과의 통계적 추론을 돕기 위해서 Lockhart의 공분산테스트 p-value를 사용하여 더 까다롭게 추정함 ; 이는 첫 번째 모델에 포함된 다른 중요한 예측치를 조건으로, 모델에서 penalization의 정도가 감소함에 따라 그것의 포함에 대한 각 노출의 중요성을 평가함.

2) unpenalized regression model

- elastic net 모델을 통해 선택된 노출의 하위 집단 - 계수의 크기, 공분산 테스트, 적응모델에 걸친 선택 일관성에 대해 고려하여 - 중요한 것으로 추론되는 경우, 다중 노출 보통 최소 제곱(ordinary least squares, OLS) 회귀모형에 다시 적합시킴. unpenalized하고, 상호 보정된 계수의 추정값을 얻기위해.

- outlier의 영향을 감소시키기위해서 모델링 전에 노출 변수는 자연로그로 변환함.

- 회귀분석의 경우 모든 예측 변수와 연속 변수를 각각의

mean±2SD(standard deviations)로 함. penalized regressions에서 이전과 동일한 중요성을 가진 변수를 주고 이진법 예측치와 연속형 예측치에 대한 계수의 비교가능성을 향상시키기 위해서

- 비순환 그래프를 이용해서 최소로 충분한 조정 세트를 확인 ; 참여자의 지역, 임신 나이, 임신전 BMI, 출산경력.

- 출생의 제태주수는 중간변수(또는 충돌자)를 대표할 수 있어서 별도의 모델로 추가 함. - 이는 산모의 혈청 레벨에 직접적인 영향을 주진 않지만 태아의 누적 노출에 영향을 미치고, 노출에 의해 인과적 영향을 받을 수 있으므로- 문헌에서 이 변수를 조정하는 방법에 대한 의견일치가 없음.

- 추가로 조정된 (further adjusted) 모델에서 불필요한 조정 또는 과조정을 이끌어내는 잠재적 혼란변수가 더해짐 : 신생아 성별, 산모의 키, 수태기간동안 알콜섭취, 산모의 혈청코티닌, 산모 혈청의 vit.D

- 산모의 교육상태, 해산물 섭취, 혈액 샘플링한 계절 변수는 missing 데이터의 비율이 높아서 민감도 분석에서만 포함됨.

- 비교를 위해서 단일물질과 출생 체중 사이의 관계를 OLS 모델로 평가하고 위발견율을 5%로 통제함.

- 민감도 분석에서 일반화된 가산 모델(generalized additive models, GAMs)을 이용하여 보정된 단일물질과 결과와의 관계의 선형성을 평가함. penalized regression spline으로 노출 기간을 다듬음

- 층화 분석에서 신생아 성별, 임신전 BMI, 흡연에 의한 effect measure modification을 테스트함.

- 지질 표준화된 유기 염소 수준을 사용하면 지질이 중간값을 가질 때 과조절로 이어질 수 있다는 것을 인식하고 총지질을 공변량으로 포함시켜 PCD-153과 p,p-DDE의 wet weight에 대한 계수를 부가적으로 추정함. 프탈레이트와 PFAS계수의 과 보정을 피하기 위해서 1차 다중 노출 모델에는 지질이 포함되지 않음. additional model에는 DEHP와 DiNP 각각의 대사산물 대신에 두 대사산물의 몰수의 합을 포

함시켜 분석함.

4) 결과

○ Elastic net regression model

<표 7-17> 출생 체중에 대한 다중-노출 탄성 그물 회귀 모형

Table 2. Multiple-exposure elastic net penalized regression models[a] (β_{EN}) for term birth weight.

Potential predictor (increment)	Adjusted	Plus gestational age	Further adjusted
ln-MEHHP (1.70 ng/mL)	−64.67[b]	−59.43[b]	−48.61
ln-MEOHP (1.29 ng/mL)	−0.15	0	0
ln-MECPP (1.42 ng/mL)	0	0	0
ln-MHiNP (2.74 ng/mL)	0	0	0
ln-MOiNP (2.22 ng/mL)	23.81	22.26	16.31
ln-MCiOP (2.32 ng/mL)	0	0	0
ln-PFHxS (1.24 ng/mL)	−3.49	0	0
ln-PFHpA (1.84 ng/mL)	0	0	0
ln-PFOS (1.60 ng/mL)	0	0	0
ln-PFOA (1.18 ng/mL)	−11.51	−10.11	−38.82
ln-PFNA (1.03 ng/mL)	−7.05	−7.69	0
ln-PFDA (1.40 ng/mL)	0	0	0
ln-PFUnDA (2.10 ng/mL)	0	0	0
ln-PFDoDA (1.67 ng/mL)	−22.56	0	0
ln-PCB-153 (2.43 ng/g)	0	0	0
ln-*p,p*′-DDE (1.82 ng/g)	−106.39[b]	−76.63[b]	−47.02[b]

Regression coefficients (β_{EN}) represent the change in mean birth weight (g) for term infants per increment: a 2-SD increase in ln-transformed exposure biomarker levels. β_{EN} for the modeled, unpenalized covariates are not shown.
[a]The cross-validated optimum penalization was $\alpha = 1.00$, $\lambda = 3.32$ (MSE = 205,061) for the adjusted model (minimal sufficient set: study population, maternal age, prepregnancy BMI, and parity); $\alpha = 1.00$, $\lambda = 3.32$ (MSE = 177,179) for the model additionally adjusted for gestational age; and $\alpha = 0.98$, $\lambda = 2.46$ (MSE = 166,329) for the further adjusted model (plus infant sex, maternal height, alcohol consumption, serum cotinine, and vitamin D). All models, $n = 1,250$. [b]Covariance test (Lockhart et al. 2014) $p < 0.05$: p-values were 0.002, < 0.001, and 0.345 for MEHHP and < 0.001, 0.017, 0.046 for *p,p*′-DDE in the minimally adjusted, plus gestational age, and further adjusted models, respectively. $p = 0.071$ for PFOA in the further adjusted model; all other p-values were > 0.10.

- 보정 : 지역, 임신나이, 임신 전 BMI, 출산력

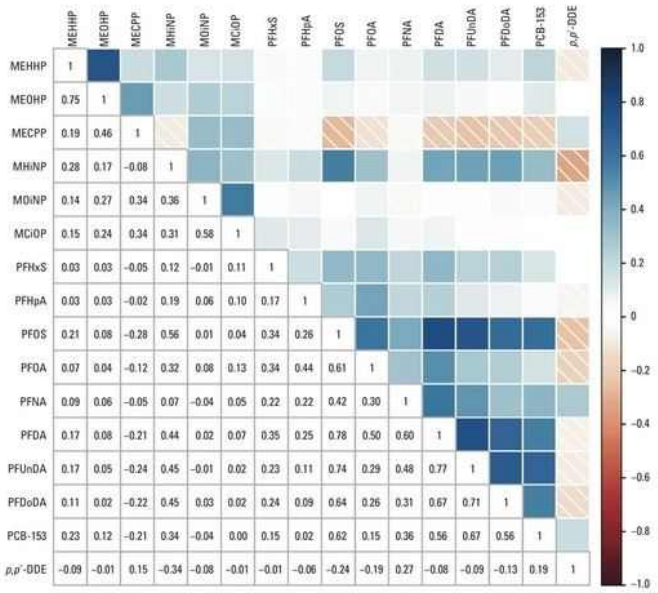
- 추가 보정 : 신생아 성별, 산모의 키, 수태기간동안 알콜섭취, 산모의 혈청코티닌, 산모 혈청의 vit.D

- 8개 변수가 선택됨. MEHHP, MOiNP, PFOA, p,p'-DDE는 3개 모델에서 지속적으로 큰 베타값을 가지면서 선택됨.

- p,p'-DDE는 세 모델에서 모두 유의했고, MEHHP는 두 모델에

서만 유의함

○ 생물학적 지표 간의 Spearman correlation coefficients



[그림 7-16] 노출 바이오마커 간의 스피어만 상관계수

음영 처리 된 상자의 색상 강도는 상관관계의 크기를 나타냄. 파란색은 양의 상관 관계를 나타내고 흰색 대각선이 있는 빨간색은 음의 상관관계를 나타냄.

○ ENR에서 선택된 물질을 이용한 multiple-exposure ordinary least squares (OLS)

<표 7-18> 탄성그물회귀로 선택된 노출과 출생 체중에 대한 다중-노출 선형 회귀 모형

Table 3. Multiple-exposure unpenalized linear regression models for the exposures selected via elastic net regression and term birth weight [β_{OLS} (95% CI)].

Predictor (increment)	Adjusted	*p*-Value	Plus gestational age	*p*-Value	Further adjusted	*p*-Value
ln-MEHHP (1.70 ng/mL)	−86.75 (−139.18, −34.32)	0.001	−83.94 (−132.68, −35.19)	0.001	−70.22 (−117.59, −22.85)	0.004
ln-MOiNP (2.22 ng/mL)	45.85 (−4.84, 96.54)	0.076	45.62 (−1.51, 92.74)	0.058	37.64 (−7.99, 83.27)	0.106
ln-PFOA (1.18 ng/mL)	−42.77 (−108.19, 22.65)	0.200	−41.02 (−101.83, 19.80)	0.186	−63.77 (−122.83, −4.71)	0.035
ln-*p,p*′-DDE[a] (1.82 ng/g)	−134.73 (−191.93, −77.53)	< 0.001	−100.75 (−154.13, −47.36)	< 0.001	−66.70 (−119.38, −14.02)	0.013
Population						
Poland	−40.24 (−133.58, 53.11)	0.398	4.84 (−82.16, 91.85)	0.913	−93.16 (−183.72, −2.60)	0.044
Ukraine	−218.89 (−300.66, −137.13)	< 0.001	−142.98 (−219.73, −66.22)	< 0.001	−256.90 (−338.77, −175.02)	< 0.001
Maternal age (years)						
27–31	88.14 (23.05, 153.24)	0.008	80.01 (19.48, 140.53)	0.010	65.44 (6.74, 124.14)	0.029
32–45	30.53 (−42.52, 103.57)	0.413	38.08 (−29.83, 105.99)	0.272	37.19 (−28.83, 103.21)	0.270
BMI (8.62 kg/m^2)	209.00 (155.53, 262.46)	< 0.001	181.37 (131.51, 231.22)	< 0.001	194.00 (145.42, 242.58)	< 0.001
Parity: multiparous	58.72 (−3.63, 121.06)	0.065	76.33 (18.32, 134.34)	0.010	85.92 (29.39, 142.44)	0.003
Gestational age (2.45 weeks)			343.78 (295.61, 391.94)	< 0.001	330.43 (283.70, 377.15)	< 0.001
Infant sex: female					−115.40 (−160.38, −70.43)	< 0.001
Maternal height (12.93 cm)					135.83 (88.38, 183.28)	< 0.001
Alcohol: ≥ 7 drinks/week					34.43 (−61.76, 130.62)	0.483
Cotinine (113.51 ng/mL)					−140.41 (−191.92, −88.89)	< 0.001
Vitamin D (22.05 ng/mL)					18.77 (−29.39, 66.93)	0.445

Regression coefficients (β_{OLS}) represent the change in mean birth weight (g) for term infants per increment: a 2-SD increase in ln-transformed exposure biomarker or untransformed continuous covariate levels, or per category for categorical covariates. Reference categories are population, Greenland; maternal age, 18–26 years; parity, nulliparous; infant sex, male; alcohol, <7 drinks/week (around the time of conception). Variance inflation factors for exposure terms ranged from 1.04 to 1.74.
[a]β_{OLS} for models including wet weight *p,p*′-DDE (ng/mL), adjusted for total lipids: −134.22 (95% CI: −191.43, −77.02), −99.91 (95% CI: −153.30, −46.52), −67.16 (95% CI: −119.80, −14.51).

- ENR에서 선택된 네가지 변수를 unpenalized 모델에 넣음. MEHHP와 p,p-DDE는 유의하나 PFOA는 음의 관계로 나오지만 유의하지 않았음. MOiNP는 양의 관계이나 유의하지 않음.

- further adjusted로 갈수록 계수값이 작아짐 PFOA제외.

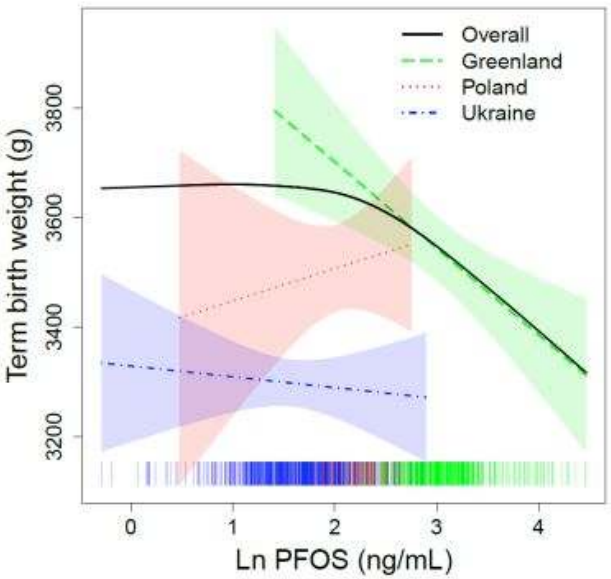
○ 단일-노출 보통 최소제곱회귀(Single-exposure unpenalized OLS-regression)

- 비교를 위해서 단일 노출 OLS 회귀 모델을 돌렸을 때 elastic net regression에서 나온 4가지 물질을 포함한 9개의 물질이 adjust 모델에서 유의하게 관련있는 것으로 나타남. gestational age에서는 4개(elastic과 동일한건 3개), further adjusted에서는 OLS 3개, elastic 4개 선택됨.

<표 7-19> 출생 체중에 대한 단일-노출 패널화하지않은 보통최소제곱(OLS) 회귀 모형

Exposure	Geometric Mean	Ln, 2-SD[a]	Adjusted β_{OLS} (95% CI)	*p*-value	Plus gestational age β_{OLS} (95% CI)	*p*-value	Further adjusted β_{OLS} (95% CI)	*p*-value
MEHHP (ng/mL)	0.559	1.700	-73.89 (-125.96, -21.81)	0.006*	-72.66 (-120.92, -24.39)	0.003*	-63.63 (-110.44, -16.83)	0.008*
MEOHP (ng/mL)	0.113	1.293	-58.72 (-109.66, -7.79)	0.024*	-48.39 (-95.65, -1.12)	0.045	-44.75 (-90.45, 0.96)	0.055
MECPP (ng/mL)	0.813	1.421	-12.15 (-65.57, 41.26)	0.656	-2.82 (-52.36, 46.73)	0.911	-12.52 (-60.44, 35.39)	0.609
ΣDEHPom (mol/mL)	5.430	1.265	-60.48 (-111.50, -9.46)	0.020*	-52.50 (-99.83, -5.17)	0.030	-53.23 (-98.98, -7.47)	0.023
MHiNP (ng/mL)	0.096	2.735	28.60 (-34.90, 92.10)	0.378	28.78 (-30.09, 87.65)	0.338	33.55 (-23.48, 90.58)	0.249
MOiNP (ng/mL)	0.016	2.219	34.93 (-15.60, 85.47)	0.176	34.04 (-12.82, 80.89)	0.155	26.11 (-19.17, 71.39)	0.259
MCiOP (ng/mL)	0.275	2.321	-10.52 (-61.09, 40.06)	0.684	-3.85 (-50.75, 43.05)	0.872	11.79 (-33.97, 57.54)	0.614
ΣDINPom (mol/mL)	1.471	2.074	-0.69 (-52.90, 51.52)	0.979	6.70 (-41.72, 55.12)	0.786	20.75 (-26.28, 67.78)	0.387
PFHxS (ng/mL)	1.842	1.241	-47.74 (-101.07, 5.59)	0.080	-11.39 (-61.15, 38.37)	0.654	-6.28 (-55.18, 42.61)	0.801
PFHpA (ng/mL)	0.047	1.837	-10.17 (-66.19, 45.84)	0.722	-11.12 (-63.05, 40.82)	0.675	-7.05 (-57.18, 43.08)	0.783
PFOS (ng/mL)	9.357	1.600	-114.36 (-206.81, -21.91)	0.015*	-69.53 (-155.59, 16.53)	0.114	-68.84 (-152.90, 15.22)	0.109
PFOA (ng/mL)	1.421	1.175	-68.94 (-134.25, -3.63)	0.039	-61.06 (-121.63, -0.49)	0.048	-78.52 (-137.01, -20.03)	0.009*
PFNA (ng/mL)	0.652	1.028	-78.62 (-130.57, -26.66)	0.003*	-60.92 (-109.20, -12.63)	0.014*	-44.67 (-92.04, 2.69)	0.065
PFDA (ng/mL)	0.245	1.397	-103.28 (-169.84, -36.72)	0.002*	-64.80 (-126.87, -2.73)	0.041	-43.93 (-104.83, 16.97)	0.158
PFUnDA (ng/mL)	0.275	2.099	-102.04 (-174.80, -29.28)	0.006*	-65.99 (-133.74, 1.77)	0.057	-37.15 (-103.86, 29.56)	0.275
PFDoDA (ng/mL)	0.068	1.672	-83.97 (-150.78, -17.16)	0.014*	-35.64 (-98.07, 26.79)	0.263	-24.40 (-86.33, 37.54)	0.440
PCB-153 (ng/g)	39.620	2.432	-143.74 (-218.92, -68.57)	0.000*	-121.09 (-190.93, -51.24)	0.001*	-78.01 (-147.15, -8.88)	0.027
p,p'-DDE (ng/g)	419.856	1.823	-137.24 (-193.92, -80.56)	0.000*	-103.11 (-156.04, -50.17)	0.000*	-73.02 (-125.21, -20.83)	0.006*
PCB-153 (ng/mL)[b]	0.299	2.331	-141.82 (-216.21, -67.43)	0.000*	-103.75 (-171.60, -35.89)	0.003*	-65.25 (-132.29, 1.80)	0.057
p,p'-DDE (ng/mL)[b]	3.144	1.774	-134.91 (-191.88, -77.95)	0.000*	-93.14 (-145.99, -40.29)	0.001*	-64.83 (-116.78, -12.89)	0.015

○ Generalized additive models



[그림 7-17] 과불화옥탄술폰산(PFOS)과 출생체중 간에 단일오염물질 노출-결과 관계를 위한 일반화 부가 모형

- PFOS와 태아 체중의 단일 노출 결과 관계에 대한 GAM

- 다른 노출과 체중 모델에서는 유의한 비선형성의 증거가 없었음. adjusted 모델에서 BMI는 평균(22.89kg/m2), 임신나이는 18-26세, 출산경력이 없는 것으로 놓으면 그림과 같이 나타남. 각 지역은 점선으로 나타나고 95% CI는 음영으로, 통합분석에서 전체 노출 결과 관계는 검정실선임. rug plot은 PFOS 생물학적 노출 지표의 밀도를 나타냄.

5) 고찰 및 결론

- correlated 된 오염원의 관계 평가의 문제를 극복하기 위해서 가장 영향력이 높은 노출 변수를 선택하는 다중 오염 penalized regression 접근법을 사용함. Elastic net regression은 가장 영향력이 적은 변수의 계수를 0으로 만듦. penalized 되지 않은 노출과 출생몸무게사이의 선형 관계의 측정을 위해 ENR에서 선택된 변수를 다중 오염 unpenalized regression 모델에 넣음

- 단일 노출 OLS와 ENR의 결과는 다소 일치했지만 OLS 모델은 각 오염원이 다른 오염원에 대해 상호간 보정이 없고 어떤 연관성은 가양성에 대한 반향을 일으킴

- 전통적인 backward-forward stepwise regression의 사용은 ENR과 비슷한 선택 결과를 보여주지만 각 스텝에서 노출 상호간의 보정이 없어서 이론상으로 선택 정확도가 낮아짐.

- ENR은 노출 수가 많은 data에서 심지어 더 큰 power를 기대할 수 있고 single-exposure environment-wide association study(EWAS) 접근법에도 상호보완적인 선택 전략을 제공할 수 있음.

- 병인 연구 (etiologic reseach question)에 적용할 때 주의해야 할 점은 높은 상관성을 가지는 두 변수가 있을 때 ENR은 약간 더 영향력이 높은 변수를 선택하고, 병인적으로 관련이 있을지도 모르는 다른 변수의 계수는 축소시킴.

6. Swartz MD, Cai Y, Chan W et al. Air toxics and birth defects: a Bayesian hierarchical approach to evaluate multiple pollutants and spina bifida. Environ Health. 2015; 14: 16. doi:10.1186/1476-069X-14-16.

1) 배경

- 산모가 벤젠에 노출되면 자손에서 이분 척추(spina bifida)와 연관이 있다는 증거가 있지만, 이 비교적 일반적인 출생 결함(spina bifida)의 위험도에 대해 여러 유해대기오염물질 (hazardous air pollutants, HAPs)의 역할을 동시에 평가한 연구는 없음.

2) 목적

- 미국 환경청(United States Environmental Protection Agency, U.S. EPA)에서 확인한 여러 유해대기오염물질(HAPs)의 모성 노출과 자손에서 이분 척추의 연관성을 확률적 탐색 변수 선택(Stochastic Search Variable Selection, SVSS)를 포함하는 계층적 베이지안 모델링(hierarchical Bayesian modeling)을 사용하여 평가하였음.

3) 방법 : 환자대조군 연구

(1) 연구자료 및 연구대상자

- the Texas Birth Defects Registry(n=1,108): 인구-기반 능동 감시 체계

- 1999년 1월 1일 - 2004년 12월 31일 까지 신경관 결손(neural tube defects, NTD; 이분 척추 포함)를 포함한 출생, 사산 및 임신종결에 대한 자료임

- 대조군(control group, n=4,132): 1999.01.01. - 2004.12.31.에 텍사스에서 태어난 층화된 무작위 표본임. 1 : 4(대조군). 출생 연도 매칭(matching)하였음.

(2) 노출 평가

- 주변 HAPs 농도에 대한 인구 통계학적(census) 추정은 U.S. EPA 의 1999 년 인구 노출 전국 평가 시스템(Assessment System for Population Exposure Nationwide, ASPEN)에서 얻었음(μg/m^3).

(3) 보정변수(Covariate, 공변량)

- 영아의 성별, 출생 년도, 모성의 인종/민족 (비 히스패닉 백인, 비 히스패닉 흑인, 히스패닉, 기타), 모성 출생지 (미국, 멕시코, 기타), 임산부 연령 (20 세 미만, 20-24 세, 25-29 세, 30-34 세, 35-39 세, 40 세 이상), 모성 교육 (<고등학교, 고등학교, 고등학교 이상), 결혼 여부 (기혼 또는 미혼), 출산력 (0, 1, 2, ≥3), 모성 흡연 (아니오 또는 예), 임신시 계절 (봄, 여름, 가을, 겨울), 사회경제적 지위 (4 분위수; 저, 중저, 중고, 고 빈곤 수준).

(4) 통계분석 방법

- 환자 대조군 중 범주형 변수에는 카이제곱검정 (chi-squared test) 수행

- HAPs 간의 상관관계는 스피어만 순위 상관관계 (Spearman's rank correlation)를 사용함.

- 확률적 탐색 변수선택(SVSS)은 변수가 상관 (즉, ρ = 0.25-0.80)되었지만 상관도가 매우 높지는 않은 경우 (즉, ρ> 0.80)에 가장 적절하기 때문에, HAPs 을 높은 상관 관계 (즉, ρ> 0.80)로 그룹화 (또는 "분류(빈, bin)")하고, 선택된 오염 물질을 기존의 과학에 기초한 주어진 빈을 나타내고 빈 내의 다른 HAPs 와의 상관 관계를 나타냈음.

- 상관 관계와 기존의 과학적 증거를 기반으로 변수를 생략하는 것은 다중공선성(multicolinearity)을 줄이기 위한 합리적인 접근법임. 유전 연관 연구에서 일반적으로 사용되는 상관관계에 기반한 알고리즘을 사용했음. 빈이 정의 된 후에는 빈을 가장 잘 나타내거나 (다른 HAPs 와 최대 상관 관계) 또는 가장 높은 상관 관계와 선천적

결함과의 연관성에 대한 기존 증거의 조합으로 빈 내의 HAPs 를 선택했음.

- 첫째, 출생 년도, 모성 교육, 모성 인종 / 민족, 모성 흡연, 인구통계 학적 빈곤 상태를 보정하여 베이지안 계층적 로지스틱 회귀분석(Bayesian hierarchical logistic regression)을 사용하여 모체가 각 HAPs 에 개별적으로 노출 된 것과 자녀의 이분 척추와의 연관성을 조사했음.

- 둘째, 동일한 공변량을 보정하면서 모든 HAPs 를 공동 분석하기 위해 확률적 탐색 변수 선택 (SSVS)과 결합된 베이지안 계층적 로지스틱 회귀 분석을 사용하여 다중 오염 물질 분석을 수행했음.

- 베이시안 계층적 로지스틱 모형은 각 HAPs 와 이분 척추의 노출 간 고정-요인(fixed-effect)을 제공하는 로지스틱 모형의 혼합-요인(mixed-effect)으로 해석할 수 있고, 또한 인구통계학적 단위 할당으로 인한 그룹 내 상관관계를 설명하기 위한 무작위 간섭(random intercept)으로 해석할 수 있음. SSVS 는 일관된 데이터 기반 확률 론적 프레임워크를 추가하여 고정-요인을 탐색하고 잠재적으로 중요한 연관성을 식별함. SSVS 와의 비교를 위해, HAPs 를 동시에 평가하면서 선택 없는 표준 계층적 베이지안 모델을 수행했음.

- HAPs 를 대조군의 분포에 기반하여 3 가지 범주로 분류했음: 고노출 (대조군의 95 백분위 수 이상), 중간 노출 (대조군의 5 백분위 수에서 95 백분위 수 사이), 낮은 노출 (5 백분위 수 아래, 참고치(reference)로 사용됨). 위 범주화에 대한 민감도 분석으로 베이지안 분석을 각 HAPs 을 연속 측정으로 반복하여 모델링하여 베이지안 분석을 수행하였음. 모든 HAPs 에서 농도 변화량이 크기 때문에, 각 오염 물질을 중심화하여 표준화했음.

- 사전변수(prior) 분포: 개별 오염 물질 분석을 위해, 이전에 확립된 방법에 기초하여 무작위 간섭(random intercept)을 하기 전에 계층적 사전변수(prior)를 사용했음. 무작위 간섭에 사용되는 사전

변수는 로지스틱 회귀에서는 정보가 없는 것으로 간주됨. SSVS 를 이용한 다중 오염 물질 모델에서 우리는 공변량에 대한 무작위 및 고정 매개변수(parameter)에 대해 동일한 사전 변수를 사용했음. HAPs 에 해당하는 매개변수에 대해서 SSVS 에 대한 사전변수의 혼합을 가정했음.

- 모형 추정과 선택: Markov chain Monte Carlo (MCMC) 방법을 사용하여 계층적 베이지안 모델의 사후(posterior) 분포를 추정했음. 단일 오염 물질 분석을 위해, 사후 평균을 사용하여 각 오염 물질에 대한 OR 를 계산하였고, 95 % 신뢰구간(credible interval, CI)은 사후 표본의 2.5th 와 97.5th 분위수를 사용하였음. SSVS 를 이용한 다중 오염 물질 모델을 위하여 각 오염 물질에 대한 한계(marginal) 베이즈 인자(Bayes factor)를 계산 하였음. 간단히 말하면, 한계 베이즈 인자는 데이터가 주어지면 각 변수의 선택에 대한 증거를 요약하는 사후 확률에 대한 사전 확률의 비율임. 따라서 베이즈 인자가 1 보다 크면 모델에 포함될 수 있는 증거가 있음을 암시하며, 1 보다 큰 값은 더 강력한 증거를 나타냄. 한계 베이즈 인자가 1 보다 큰 HAPs 를 합동(joint) 모델에 포함하고, MCMC 를 통해 선택된 모델의 합동 사후 분포를 계산하고, 사후 평균 및 베타 계수의 분위수를 사용하여 합동 모델의 OR 및 95 % CI 를 계산했음. 모든 MCMC 계산에 대해 별도의 초기 값을 갖는 2 개의 체인(사슬)을 시뮬레이션 했고, 각각은 150,000 회 반복으로 구성되었음. 각 체인의 처음 50 %를 체인을 사후 분포에 수렴하도록 버렸음. 두 체인 간에 연관된 모든 매개 변수에 대하여 사후 평균이 얼마나 잘되었는지를 통해 수렴을 평가했음. 0.95 보다 높은 상관 관계는 체인이 충분히 수렴되었다고 간주함.

- MCMC 계산은 WinBUGS 1.4 를 사용했고 사후 추론은 R (64 비트 대 3.0.2)을 사용하여 수행하였음.

4) 결과

- 상관성이 높은 HAPs 의 4 개 그룹이 있었음(Table 2). 특히 벤젠과 메틸렌 클로라이드는 아세트 알데히드, 아크롤레인, 포름알데히드, 벤젠 및 메틸렌 클로라이드로 구성된 상관성이 높은 그룹을 나타내기 위해 선택되었음. 상관성이 높은 HAPs (ρ> 0.80)의 "빈 (bin)"을 나타내기 위해 최종 분석에서 25 개의 HAPs 가 선택되었음.

<표 7-20> 상관관계가 ρ > 0.80 인 유해 공기 오염물질

Table 2 Hazardous air pollutants with correlations greater than $\rho > 0.80$, Texas, 1999

Pollutant	Acet	Acro	Form	Benz	MCl	EDbr	PD	TCE	EDcl	VCl	DPM	Ni
Acet	1.00	0.97	0.98	0.85	0.79							
Acro	0.97	1.00	0.95	0.81	0.68							
Form	0.98	0.95	1.00	0.89	0.82							
Benz[1]	0.85	0.81	0.89	1.00	0.74							
MCl[1]	0.79	0.68	0.82	0.74	1.00							
EDbr						1.00	0.98	1.00				
PD						0.98	1.00	0.99				
TCE[2]						1.00	0.99	1.00				
EDcl									1.00	0.98		
VCl[3]									0.98	1.00		
DPM[4]											1.00	0.87
Ni											0.87	1.00

*Abbreviations: *Acet* acetaldyhyde, *Acro* acrolein, *Form* formaldehyde, *Benz* benzene, *MCl* methylene chloride, *EDbr* ethylene dibromide, *PD* propylene dichloride, *TCE* 1,1,2,2-Tetrachloroethane, *EDcl* ethylene dichloride, *VCl* vinyl chloride, *DPM* diesel particulate matter, *Ni* nickel compounds.
[1]Benzene and methylene chloride selected to represent the highly correlated group including acetaldehyde, acrolein and formaldehyde.
[2]1,1,2,2-Tetrachloroethane selected to represent the highly correlated group including ethylene dibromide and propylene dichloride.
[3]Vinyl Chloride selected to represent the highly correlated group including ethylene dichloride.
[4]Diesel particulate matter selected to represent the highly correlated group including nickel compounds.

- 환자군 내 병인의 이질성을 최소화하기 위해, 염색체 이상이나 다른 증후군 (75 명), 폐쇄 결손 (예, lipomyelomeningocele, 88 명), 무뇌아 (anencephaly, 351 명)은 제외함. 모성 주소가 누락 된 사례는 제외함 (n = 61). 이러한 배제 후, 척추 이분증 533 례(case)가 분석이 가능하였음. 대조군 중 모성주소가 누락된 433 명을 제외하고 3,695 명이 분석이 가능하였음.

- 척추 이분증의 어머니는 더 히스패닉이고 대조군의 어머니와 비교하여 멕시코에서 태어났음 (각각 p = 0.003 및 p = 0.05). 또한,

케이스의 어머니는 인구조사지역 중 더 높은 빈곤 수준의 (p = 0.02) 지역에서 살 확률이 높았음(Table 3).

- 단일 오염 물질 모델에서 32 개의 HAPs 중 14 개 (44 %)는 95 % CI 가 1 을 포함하지 않았음.

- 25 개의 HAPs 중 베이즈 인자(Bayes factor)가 1 보다 큰 2 개를 확인했음 : 퀴놀린(quinoline, ORhigh = 2.06, 95 % CI : 1.11-3.87, 베이즈 인자 = 1.01), 트리클로로에틸렌(Trichloroethylene, TCE, ORmedium = 2.00, 95 % CI : 1.14-3.61, 베이즈 인자 = 3.79)(Table 4). 이러한 연관성은 공변량의 연관성보다 강했고 95% 신뢰구간은 비슷했음.

<표 7-21> 베이즈 인자가 1.00 보다 큰 확률적 검색 변수 선택(SSVS)을 사용하여 이분 척주증과 연관된 유해 대기오염물질

Table 4 Hazardous air pollutants associated with spina bifida identified using Stochastic Search Variable Selection (SSVS) with a Bayes factor greater than 1.00

Pollutant	Pollutant range (μg/m^3)	Cases/Controls	Odds ratio[1]	95% credible interval	Bayes factor
Quinoline					
Low	$<6.50\times10^{-14}$	18/176	1.00	Ref.	
Medium	6.50×10^{-14} - 1.71×10^{-5}	434/3149	1.42	(0.87, 2.42)	0.32
High	$>1.71\times10^{-5}$	39/178	2.06	(1.11, 3.87)	1.01
Trichloroethylene					
Low	<0.0524	13/176	1.00	Ref.	
Medium	0.052 – 0.16	460/3151	2.00	(1.14, 3.61)	3.79
High	>0.16	18/176	1.32	(0.61, 2.80)	0.60

[1]Adjusted for year of birth, maternal education, maternal race/ethnicity, maternal smoking, and census tract poverty status.

- 보정하지 않은 OR 은 제어되지 않은 혼란으로 인한 이러한 효과를 과대 평가하였음 (추가 파일 1 : 표 S3). 비교를 위해 SSVS 가 없는 통합 모델에서는 중간 수준의 트리클로로에틸렌만이 척추이분증과 관련있음을 확인했음 (OR = 5.72, 95 % CI : 1.44 - 24.16, 추가 파일 1 : 표 S4). 사전 확률이 0.50 인 민감도 분석은 비슷한 결과를 나타냈음 (데이터는 표시되지 않음). 연속 규모에서 HAPs 을 사용한 우리의

분석은 베이스 인자가 1 보다 큰 8 개의 HAPs 를 선택하여 연속 규모로 분석하였지만, 95 % CI 에는 모두 1.0 이 포함되었음 (데이터는 표시되지 않음).

- 역 U 자 형태의 용량 - 반응 관계는 트리클로로에틸렌과 같은 내분비 계 장애 물질로 작용하는 독성 물질에서 흔히 나타남.

- 벤젠이 최종 모델에 없는 이유로 1) 여러 다른 HAPs 와의 높은 상관 관계 (ρ> 0.80)와 2) 최종 다중 오염 물질 모델에서 추정 효과가 다른 HAPs 만큼 강하지 않아서 등의 여러 요인이 원인일 수 있음.

- 장점 : 특히, SSVS 는 중요한 예측 인자에 관한보다 포괄적인 정보를 제공하기 위해 모델 구축 과정에서 모델 선택 불확실성을 포함함. 전통적인 단일 오염 물질 모델에서 연관된 요인이 14 가지 물질이었는데 이 중 일부는 위양성일 수 있으며, 이와 비교하여 다변수 모델은 2 가지로 검출되었음.

- 단점 : 오염 물질을 분류하는 방법(binning)에 화학적 성질, 또는 목표 원인(target source)과 같은 다른 방법을 사용할 수도 있었음. 그러나 통계 모델은 상관관계를 통해 연관성을 평가하기 때문에, 상관관계를 기반으로 HAPs 를 정의하는 것은 통계적 모델링 접근법과 가장 잘 일치하는 것으로 보임. 차원 축소에 과학적 요소를 추가했음.

- 이전의 증거들은 SSVS 가 위음성 결과를 선호하는 경향이 있음을 시사함. 베이즈 인자 임계 값을 1 보다 크게 사용하면 보통 위음성 수가 줄어듬. 그러나 위음성이 증가 된 경우에도, 평균적으로, SSVS 방법은 표준 선택 방법에 비해 올바른 연관성을 생성 할 가능성이 더 높음.

- 원래 복잡한 질병에 대한 다중 유전 변이의 역할을 평가 할 때 사용되는 방법이었음.

5) 고찰 및 결론

- 퀴놀린과 트리클로로에틸렌이 척추 이분증의 위험에 중요한 요인이 될 수 있다는 증거가 있음. 또한 SSVS 와 함께 베이지안 계층적 모델을 사용하는 것은 질병 위험에 대한 여러 환경오염 물질을 평가하는 대안 방법임.

7. Braun JM et al. Gestational exposure to endocrine-disrupting chemicals and reciprocal social, repetitive, and stereotypic behaviors in 4- and 5-year-old children: the HOME study. Environ Health Perspect. 2014 May;122(5):513-20.

1) 배경

- 내분비계교란물질(endocrine-disrupting chemicals, EDCs)은 자폐스펙트럼질환의 원인이 될 수 있다고 알려져 있으나, EDCs 혼합물 노출과 의 관련성은 잘 알려져 있지 않음.

2) 목적

- 임신 중 EDCs 노출과 자폐 행동 간의 관련성을 평가함.

3) 방법

(1) 연구자료 및 연구대상자

- The Health Outcomes and Measures of the Environment(HOME) Study: 저농도 화학물질 노출과 어린이 성장/발달 간의 관련성을 확인하기 위해 신시내티, 오하이오, 대도시 지역에서 진행된 전향적 출생코호트 연구임.

- 2003 년 3 월부터 2006 년 1 월 신시내티 지역의 3 개 병원과 관련된 산전클리닉에서 임산부를 모집함.

- 연구대상자 등록 기준: 임신 16±3 주, 18 세 이상, 1978 년 이후에 지어진 집에서 거주, HIV 과거력 없음, 발작 또는 갑상선질병을 위한 약물 복용력 없음.

(2) 노출 평가

- 임신 16 주, 26 주 쯤에 산전클리닉을 방문하여 소변시료 및 혈액시료 채취

- 출산 후 24 시간 안에 혈액시료 채취

- 소변시료 및 혈액시료에서 70 가지 EDCs 농도를 측정함.

(3) 건강영향 평가

- 어머니들이 아동이 4 세, 5 세 일 때 2 회 이상 사회반응척도(Social Responsiveness Scale, SRS) 설문지를 작성함.

- 사회반응척도(SRS)는 대인관계 행동, 의사소통, 반복적 또는 고정관념적 행동에 대한 타당성과 신뢰성이 있는 민감한 척도임.

(4) 보정변수

- 출산 시 어머니 나이, 인종, 결혼상태, 출산력, 보험상태, 고용상태, 가구소득, 산전비타민 복용력, 어머니 지능, 아동의 1 세때 양육 환경, 혈중 코티닌.

(5) 통계분석 방법

- 자녀가 4 세 또는 5 세 때 한번이라도 추적관찰한 여성과 아닌 여성에서 EDCs 농도의 기하평균을 비교함.

- 연구대상자의 EDCs 농도 중위수와 미국인을 대표하는 자료인 미국국민건강영양조사 자료에서 미국여성의 EDCs 농도 중위수를 비교함.

- 노출변수 및 보정변수에 따른 SRS 점수를 확인함.

- 복합유해물질 노출과 건강영향 간의 관련성을 평가하기 위해 반-베이즈 모형(semi-Bayesian model)를 사용함.

- semi-Bayesian model 은 복합노출물질과 전통적인 보정변수를 outlying estimates 의 영향을 줄여 다중비교를 가능하게 하고, exchangeability group(정밀하게 만들기 위한 기능 및 사전에 명시된 정규분포 residual variance parmeter(τ2)로써)의 평균으로 beta coefficient 를 축소시켜 추정값의 정확성과 타당성을 향상시킴.

- 이 통계방법은 개별 베타 계수가 동일한 교환그룹의 다른 베타 정보를 빌릴 수 있게 함.

- 모델의 첫 번째 단계에서 SRS 점수는 반복측정된 SRS 측정을 설명하기 위해 single linear mixed model 에서 52 개의 EDC biomarker 및 보정변수들을 포함하여 회귀분석 됨.

- 모델의 두 번째 단계에서 첫 번째 단계에서 계산된 베타 계수가 교환그룹 매트릭스, 교환그룹의 분산 및 τ2 에 대해서 회귀됨.

- 교환가능 그룹은 구조적으로 유사한 화학물질이 공통적인 메커니즘을 공유한다는 선험적 기대에 따라 선택됨.

- 교환가능 매트릭스는 Table S2 와 같음.

- 베타 매개변수가 -10 과 10 사이에 있다고 가정하는 τ2 를 지정함 (즉, 표준 샘플에서 평균 SRS 점수의 1 SD 이내)

<표 7-22> 반-베이지안 회귀 모형에서 교환 매트릭스 지정

Table S2. Specification of exchangeability matrix in semi-Bayesian regression model.

Chemical	Di-butyl phthalate metabolite	Di-2-ethylhexyl phthalate metabolite	Polychlorinated biphenyl	Organochlorine Pesticide	Polybrominated diphenyl ether	Perfluoralkyl Substance	Persistent Chemical
MBP	1	0	0	0	0	0	0
MiBP	1	0	0	0	0	0	0
MEP	0	0	0	0	0	0	0
MBzP	0	0	0	0	0	0	0
MCPP	0	0	0	0	0	0	0
MEHP	0	1	0	0	0	0	0
MECPP	0	1	0	0	0	0	0
MEHHP	0	1	0	0	0	0	0
BPA	0	0	0	0	0	0	0
PCB 28	0	0	1	0	0	0	1
PCB 66	0	0	1	0	0	0	1
PCB 74	0	0	1	0	0	0	1
PCB 99	0	0	1	0	0	0	1
PCB 101	0	0	1	0	0	0	1
PCB 105	0	0	1	0	0	0	1
PCB 118	0	0	1	0	0	0	1
PCB 138/158	0	0	1	0	0	0	1
PCB 146	0	0	1	0	0	0	1
PCB 153	0	0	1	0	0	0	1
PCB 156	0	0	1	0	0	0	1
PCB 157	0	0	1	0	0	0	1
PCB 167	0	0	1	0	0	0	1
PCB 170	0	0	1	0	0	0	1
PCB 172	0	0	1	0	0	0	1
PCB 177	0	0	1	0	0	0	1
PCB 178	0	0	1	0	0	0	1
PCB 183	0	0	1	0	0	0	1

Chemical	Di-butyl phthalate metabolite	Di-2-ethylhexyl phthalate metabolite	Polychlorinated biphenyl	Organochlorine Pesticide	Polybrominated diphenyl ether	Perfluoralkyl Substance	Persistent Chemical
PCB 187	0	0	1	0	0	0	1
PCB 194	0	0	1	0	0	0	1
PCB 195	0	0	1	0	0	0	1
PCB 196/203	0	0	1	0	0	0	1
PCB 199	0	0	1	0	0	0	1
PCB 206	0	0	1	0	0	0	1
PCB 209	0	0	1	0	0	0	1
β-HCH	0	0	0	1	0	0	1
HCB	0	0	0	1	0	0	1
p'p'-DDT	0	0	0	1	0	0	1
p'p'-DDE	0	0	0	1	0	0	1
Oxychlordane	0	0	0	1	0	0	1
trans-Nonachlor	0	0	0	1	0	0	1
BB 153	0	0	0	0	0	0	1
PBDE 28	0	0	0	0	1	0	1
PBDE 47	0	0	0	0	1	0	1
PBDE 85	0	0	0	0	1	0	1
PBDE 99	0	0	0	0	1	0	1
PBDE 100	0	0	0	0	1	0	1
PBDE 153	0	0	0	0	1	0	1
PBDE 154	0	0	0	0	1	0	1
PFOA	0	0	0	0	0	1	1
PFOS	0	0	0	0	0	1	1
PFNA	0	0	0	0	0	1	1
PFHxS	0	0	0	0	0	1	1

- τ2 가 작으면 교환가능 매트릭스에서 명시된 요인에 의해 대부분의 관련성이 설명되는 반면, τ2 가 크면 교환가능 매트릭스에서 명시되지 않은 요인이 더 많은 연관성을 설명한다고 가정함.

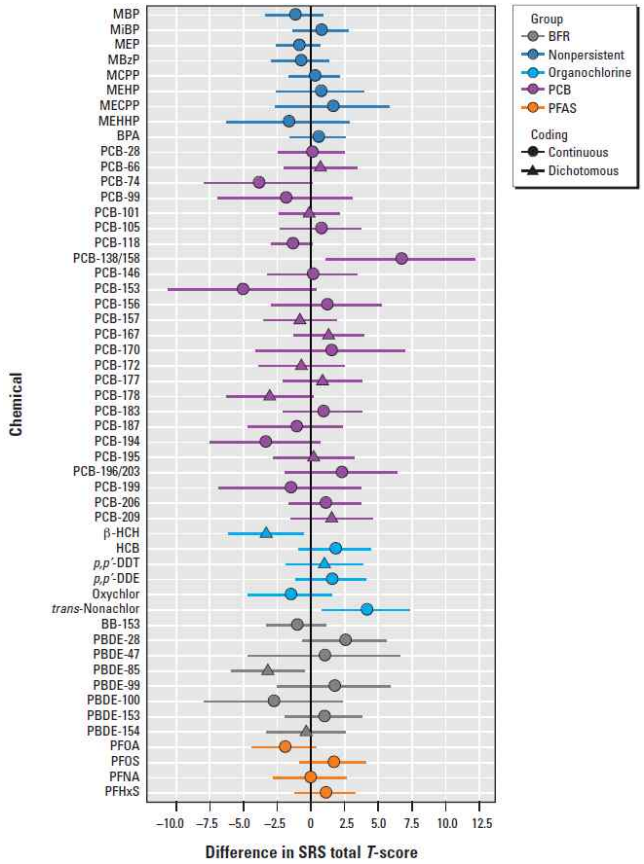
- 모델의 첫 번째 단계는 SAS PROC MIXED 를 사용하였고, 두 번째 단계는 SAS PROC IML 코드를 사용함.

4) 결과

- 대부분의 EDCs 는 무시할만한 차이인 SRS 1.5 점 이하 정도로 관련성이 있었음.

- 혈중 PBDE-138/158 농도 및 trans-nonachlor 의 2-SD 증가와 자폐행동 간에 유의한 양반응 관계가 있었음.

- 반대로 PBDE-85 는 증가할수록 자폐행동이 감소하는 관계가 관찰됨.

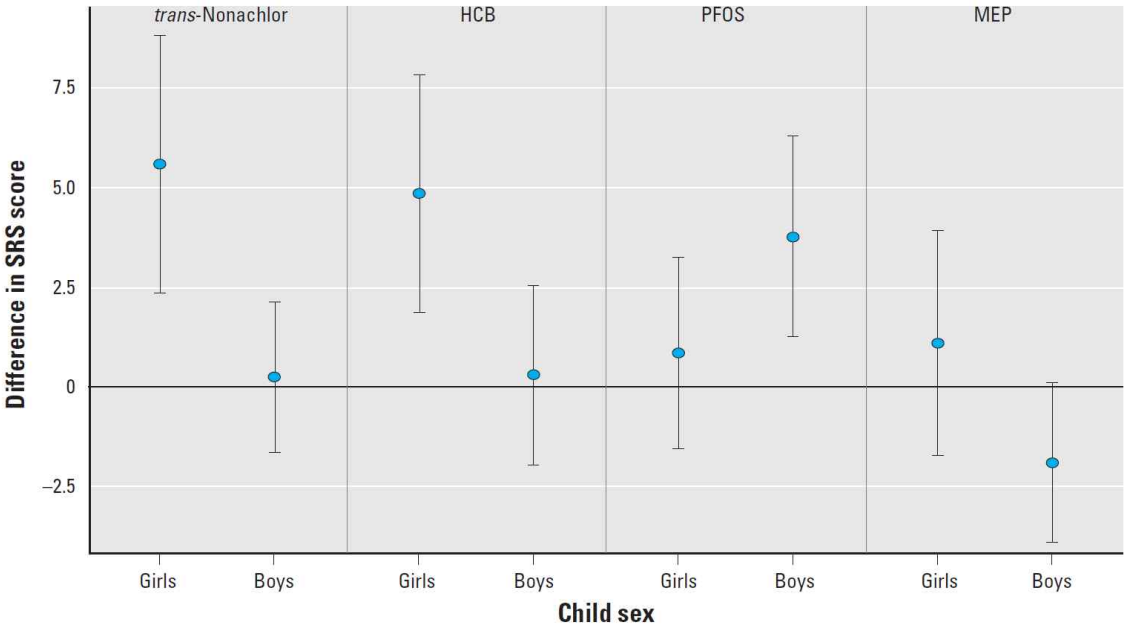


[그림 7-18] 반-베이즈 모형을 사용하여 4 세 및 5 세 신시내티 어린이의 모체 임신 소변 또는 혈청 내분비계장애 물질(EDC) 농도와 사회반응척도(SRS) 총 T-점수와의 연관성(n=175)

PCB-66, PCB 101, PCB-157, PCB-167, PCB-172, PCB-177, PCB-178, PCB-195, PCB-209, β-HCH, DDT, PBDE-85, PBDE-154, and PBDE 183 는 검출됨 vs 불검출됨 으로 코딩함. 표시된 베타는 이러한 화학 물질의 검출 가능 수준과 검출 불가능한 수준을 가진 여성에게서 태어난 어린이들 사이에서 사회반응척도(SRS) 점수의 변화임. 다른 모든 화학물질은 표준 이탈 변수와 비슷한 규모로 두기 위해 표준편차의 두 배로 나눈 연속 log10 변형 변수로 처리되었음.

공변량 보정된 반-베이지안 모형(표 2 참조)의 인구 통계 학적 요인과 임신시 우울 증상(연속), HOME 점수(연속) 및 임신 혈청 코티닌 농도(연속 log10 변환)을 보정하였음. 반-베이지안 모형의 잔차 분산(τ2)은 26.03(20 포인트 범위)으로 설정되었음. 교환 성 매트릭스는 OC, PCB, BFR, PFAS, DEHP 대사산물, DBP 대사산물이고 지속성 대비 지속성 화학 물질에 대한 절편 및 지표 (0/1) 변수를 포함함.

- 화학물질과 성별 간에 interaction 이 유의한 관련성을 보인 물질들에 대해 성별로 층화해서 분석해보았을 때 trans-nonachlor, HCB 는 여성에서만 유의하였고, PFOS 의 경우에는 남성에서만 유의하였음.



[그림 7-19] 성별로 층화한 4 세 및 5 세 신시내티 어린이의 모성 임신 소변 또는 혈청 내분비계장애물(EDC) 농도와 공변량을 보정한 사회반응척도(SRS) 총 T-점수와의 연관성(n=175)

표시된 모든 베타는 log10- 변환 된 모체 화학 물질 농도가 2-SD 증가한 아동 SRS 점수의 변화임. 임신 중 우울 증상(연속), HOME 점수(연속) 및 임신 혈청 코티닌농도(log10 변환된 연속)뿐만 아니라 인구 통계 학적 요인(표 2 참조)에 맞게 보정하였음. 중요한 화학적 × 성별 상호 작용과의 연관성만 표시되었음($p < 0.10$).

5) 고찰 및 결론

- 자폐점수와 유의한 관련성을 보인 물질들을 어린 자폐행동과 관련된 요인으로 철저히 검토할 필요가 있음.

8. Powell H, Krall JR, Wang Y et al. Ambient coarse particulate matter and hospital admissions in the medicare cohort air pollution study, 1999-2010. Environ Health Perspect. 2015; 123(11): 1152-1158. doi: 10.1289/ehp.1408720. Epub 2015 Apr 14.

1) 배경

- 최근 몇 개의 연구에서 거친 입자상 물질 (coarse particulate matter, PM10-2.5)과 사망률 및 이환율 결과 차이의 연관성을 조사했음. 그러나 이 연구들에서는 일관성이 없는 결론을 이끌어 냈음 .

2) 목적

- 1999 년에서 2010 년까지 (12 년 동안) 65 세 이상의 메디케어(Medicare) 등록자 중 PM10-2.5 와 심혈관 및 호흡기 질환 모두에 대한 응급 입원과 전국 및 지역 수준의 연관성을 추정했음.

3) 방법 : 후향적 코호트 연구

(1) 연구자료 및 연구대상자

- The Medicare Cohort Air Pollution Study(MCAPS)

- 1999 년 1 월 1 일부터 2010 년 12 월 31 일에 미국 110 개 카운티(counties, 군)이 연구에 포함될 자격이 있었음. 2000 년에 메디케어 등록자가 20,000 명 초과이고, 연구기간 중 최소 200 일 동안 공동 모니터에 PM10 과 PM2.5 가 기록된 곳이 포함될 자격이 있었음. 포아송 로그-선형 모델(Poisson log-linear model)을 사용함. 서부(n=29)와 동부(n=81)로 분류함.

(2) 노출 평가

- PM2.5 와 PM10 은 미국 환경보호국(Environmental Protection Agency, EPA)의 인구 기반 공기 질 모니터링 네트워크를 이용하여 모니터링하였음.

- 장소의 같은-날의 PM10 농도에서 매일 측정 한 PM2.5 농도를 빼서 PM10-2.5 지표를 도출했음. 나란히 놓인 모니터가 한 쌍 이상인 카운티의 경우, 연구의 각 날에 대해 PM10-2.5 의 단일 추정치를 얻고 특이점(outliers)을 방지하기 위해 10 % 정돈 된 평균(trimmed mean)을 적용했음. 나란히 놓인 모니터가 10 보다 적으면 그 날의 최소값과 최대값을 삭제했음. 두 가지 추정치만 사용 가능한 경우 평균을 사용하였음.

(3) 건강영향 평가

- 관심 건강 결과는 전국 청구 내역 파일(National Claims History Files)로 메디케어 등록자의 지불 청구로부터 심혈관계 질환과 호흡기계 질환으로 인한 응급 병원 입원의 일일 숫자(daily number)임

- 메디케어는 2012 년 해당 연령 미국 인구의 약 94%인 4,210 만 명이 등록되어있음. 각 개인에 대한 지불 청구서에는 의료 기록 외에도 인구 통계 및 거주지 카운티 정보가 포함되어 있음.

- 질병은 국제 질병 분류(International Classification of Diseases, Ninth Revision, ICD-9)으로 분류함. 카테고리 포함 기반을 병원 입원 당시의 1 차 진단(primary diagnosis)으로 사용했음.

- 심장 결과는 뇌혈관 사건 (cerebrovascular events, ICD-9 430-438), 심부전 (heart failure, HF; 428), 심장 리듬 장애 (heart rhythm disturbances, HRD; 426-427), 허혈성 심장 질환 (ischemic heart disease, IHD; 410-414 및 429), 말초 혈관 질환 (peripheral vascular disease, PVD; 440-448). 2 가지 호흡 결과는 만성 폐쇄성 폐 질환 (chronic obstructive pumonary disease, COPD; 490-492) 및 호흡기계 감염 (respiratory tract infection, RTI; 464-466 및 480-487)에 대한 입원을 평가했음.

- 각 카운티별로 매일의 시계열 입원율을 구축하였음. 고려중인 기간 동안 110 개 카운티에서 총 637 만명의 심혈관과 251 만명의 호흡기 응급 입원이 있었음.

(4) 보정변수

- National Climatic Data Center 의 각 카운티에 대한 일별 온도 및 이슬점 온도 데이터를 얻음. 이 둘은 건강 결과와 공기 오염 노출 연관성에 대하여 잠재적 교란 요인이기 때문임.

- 보정: PM2.5, 날씨, 계절, 장기간 추세

- 0, 1, 2 일의 일일 노출 지연(lags)에 더하여, 본 논문에서는 3 일의 노출 지연을 조사했음.

(5) 통계분석 방법

- 지역 (미국 동부와 서부)과 PM10-2.5 와 카운티 수준의 입원율 사이의 국가적 연관성을 알아보기 위해 2 단계 베이지안 계층 모델(two-stage Bayesian hierarchical model)을 사용했음.

- 첫 번째 단계에서는, 나이별로 시계열을 층화해서 2 개의 카테고리를 만들었음. 65 세에서 74 세 사이와 75 세 이상인 메디케어 등록자임. PM10-2.5 와 입원 간의 카운티 차원의 연관성을 추정하기 위해과 분산 된 포와송 로그 선형 모델(Poisson log-linear model)을 사용했음. 이 모델에는 해당 카운티의 주어진 날에 위험에 처한 사람들의 자연 로그의 오프셋이 포함되어 있으며, 해당 카운티의 메디케어 등록자 수를 기준으로 함; 각 연령 카테고리 (65-74 세 및 75 세 이상)의 각기 다른 기준 입원을 설명하기 위한 별도 간섭; 요일 지표; 75 세 이상의 사람들을 나타내는 지표; 현재의 기온과 지난 3 일간의 평균 기온의 부드러운 함수(smooth functions). 이 두 기온 모두 자유도 6 도를 사용했음; 현재의 이슬점 온도와 지난 3 일간의 이슬점 온도의 평균값의 부드러운 함수. 둘 다 자유도 3 도를 사용했음; 달력 시간의 부드러운 함수. 심혈관 입원시 연간 자유도 8 도, 호흡기 입원시 연간 자유도 12 도 (호흡기 입원은 심혈관 입원보다 더 강하게

계절적임); 변화하는 장기간의 추세를 포착하기 위해 나이에 대한 지표로 하여 시간 상호작용의 부드러운 함수. 연간 자유도가 1 도임; 주어진 지연에서 매일의 PM10-2.5 농도.

- 시간의 부드러운 함수에 대한 연간 자유도는 PM10-2.5 와 입원의 장기간과 계절적 변동이 제거되어, 건강 영향을 추정하기 위해 단기 변동만 남기고 선택되었음. 온도와 이슬점 온도 부드러운 함수에 대한 자유도는 이들 인자와 건강 결과 사이의 비선형 관계를 수용하기 위해 선택되었음.

- 또한, 위에 나열된 변수 외에도 PM2.5 농도를 포함한 두 번째 모델도 역시 적합했음. 동일한 지연 시간에 PM10-2.5 와 PM2.5 를 동시에 포함하여 PM2.5 의 잠재적 영향을 보정하였음.

- R 통계 소프트웨어, 버전 3.2.1 (R Core Team 2015)의 glm 함수를 사용하였음.

- 두 번째 단계에서는, tlnise 패키지 내에 있는 베이지안 계층 모델을 사용하여 PM10-2.5 와 병원 입원 간의 전국 및 지역 수준의 연관성을 추정했음. 이를 통해 카운티 내 상대 위험 추정치를 결합하는 동시에, 카운티 내 통계상의 오류 및 실제 상대 위험의 카운티 간 변동성을 고려할 수 있음. 110 개 카운티 전체에 걸친 카운티-특정 상대 위험도가 결합되어 국가-차원의 추정치를 산출했음. 마찬가지로 미국 서부 및 동부 지역에 대한 지역 추정치는 서쪽에 위치한 29 개 카운티와 동쪽에 위치한 81 개 카운티에 대한 카운티-특정 상대 위험도를 결합하여 산출하였음.

- 통계적 유의성은 0 의 값을 포함하지 않는 95 % 후방 구간 (posterior intervals, PI)에 의해 평가되었음.

- 민감도 분석: 시간, 온도 및 이슬점 온도의 부드러운 함수에 사용 된 자유도와 관련하여 심혈관 및 호흡기 입원에 대한 당일 국가 추정치의 민감도를 평가했음. 달력 시간과 온도 모두에 대해 2 에서 20 까지 자유도의 범위를 고려했음; 이슬점 온도에 대해서는 1 에서

10 까지의 범위를 고려했음. 각 부드러운 함수의 자유도에 대한 다른 선택이 다른 부드러운 함수와 독립적으로 고려되었음.

4) 결과

- 연구 기간 동안 PM10-2.5 농도의 중앙값은 서부에 비해 동부가 낮았으며, PM2.5 에 대해서는 반대의 패턴이 관찰되었음 (그림 1, 표 2). 온도와 이슬점 온도는 모두 미국 동부에서 더 높았음.

- 전국 (그림 2 및 4) 및 지역 (그림 3) 수준의 연관성은 거친 입자상 물질의 10-μg/m^3 증가에 대한 응급 입원 횟수의 변화율(95% PI 와 함께)로 표시함. 단일 오염 물질 (PM10-2.5 만 포함) 및 2 가지 오염 물질 (PM10-2.5 및 PM2.5 가 공동으로 포함됨) 모델 모두에 대하여 지연 0, 1, 2, 3 일 지연을 표시했음.

- PM10-2.5 의 10-μg/m^3 증가는 같은 날 심혈관 질환 입원의 유의한 증가와 연관이 있었음[0.69 %; 95 % 후방 구간(posterior interval, PI) : 0.45, 0.92]. PM2.5 를 보정 한 후에 이 연관성이 유의하게 유지되었음(0.63 %; 95 % PI : 0.38, 0.88). PM10-2.5 의 10-μg/m^3 증가는 호흡 관련 입원의 유의 한 증가와 관련이 없었음(그림 2).

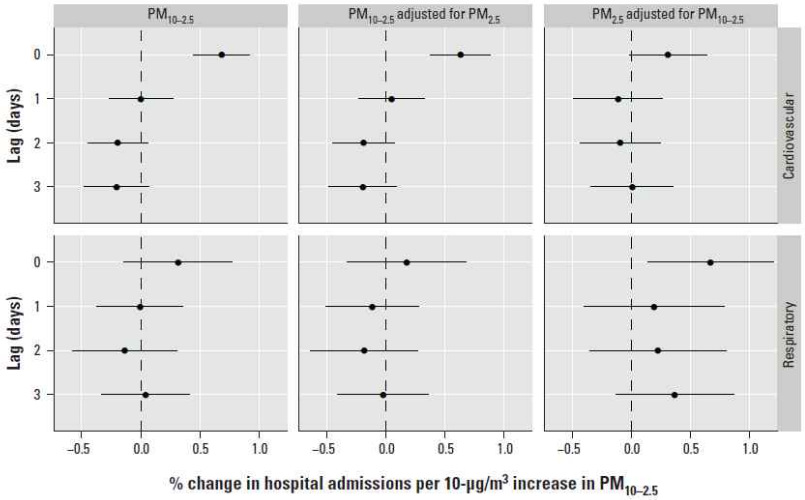


Figure 2. Estimated national-level associations and 95% PIs between cardiovascular and respiratory disease admissions and a 10-µg/m^3 increase in $PM_{10-2.5}$.

[그림 7-20] 국가 수준의 심혈관 및 호흡기 질환 입원과 PM10-2.5 의 10-μg/m^3 증가의 연관성 및 95 % PI 추정

- 동부 (0.84 %, 95 % PI : 0.51, 1.17) 및 서부 (0.44 %, 95 % PI : 0.04, 0.85)에서 PM10-2.5 의 10-µg/m^3 증가는 심혈관 입원과 양의 상관관계가 있었음(그림 3). PM2.5 를 보정 한 후, 동부 (0.74 %, 95 % PI : 0.38, 1.10)는 연관관계가 유지되었지만 서부 지역은 그렇지 않았음 (0.41 %; 95 % PI : -0.03, 0.85).

- 서부에서 노출 지연 2 일에서, 심혈 관계 입원 (단일 오염 물질 모델, -0.52 %; 95 % PI : -0.97, -0.08 및 2 가지 오염 물질 모델, -0.54 %, 95 % PI : -1.00, -0.08)에서 음성적 연관성이 유의했음.

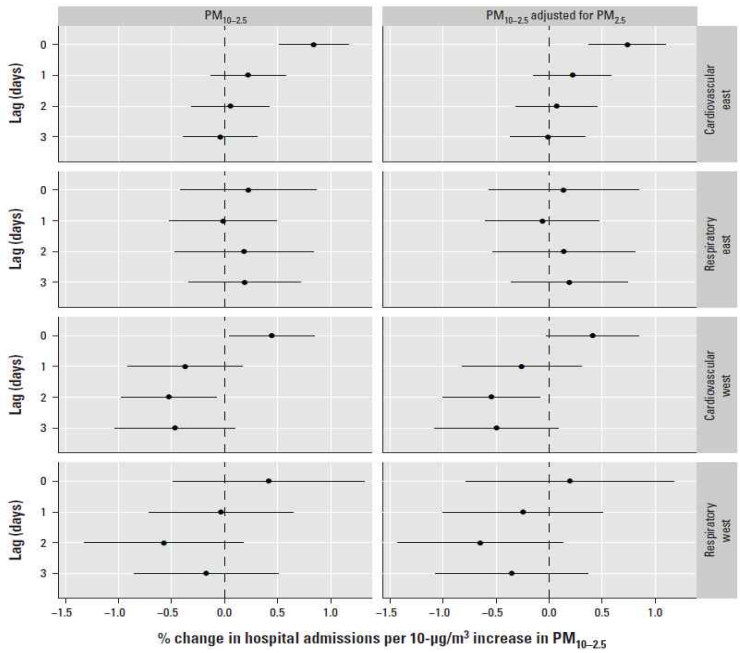


Figure 3. Estimated regional-level associations and 95% PIs between cardiovascular and respiratory disease admissions and a 10-μg/m^3 increase in $PM_{10-2.5}$.

[그림 7-21] 지역 수준의 심혈관 및 호흡기 질환 입원과 PM10-2.5 의 10-μg/m^3 증가의 연관성 및 95 % PI 추정

- 단일 오염 물질 하에서 시간, 온도, 이슬점 온도의 부드러운 기능에 사용 된 자유도와 관련하여 당일 전국 평균 추정치의 민감도를 독립적으로 평가했음(그림 5). 심혈관계나 호흡기계 입원 추산치가 온도나 이슬점 온도의 부드러운 함수에 대한 자유도 선택에 실질적인 민감성을 보이지 않음을 알 수 있음. 추정치는 시간의 부드러운 함수, 특히 호흡기 입원 관련에 대한 연간 자유도 선택에 약간 더 민감함. 그러나 우리가 고려한 매우 적은 수의 자유도와 95 % PI 의 증가 폭을 감안할 때 이 작은 양의 민감도가 예상됨.

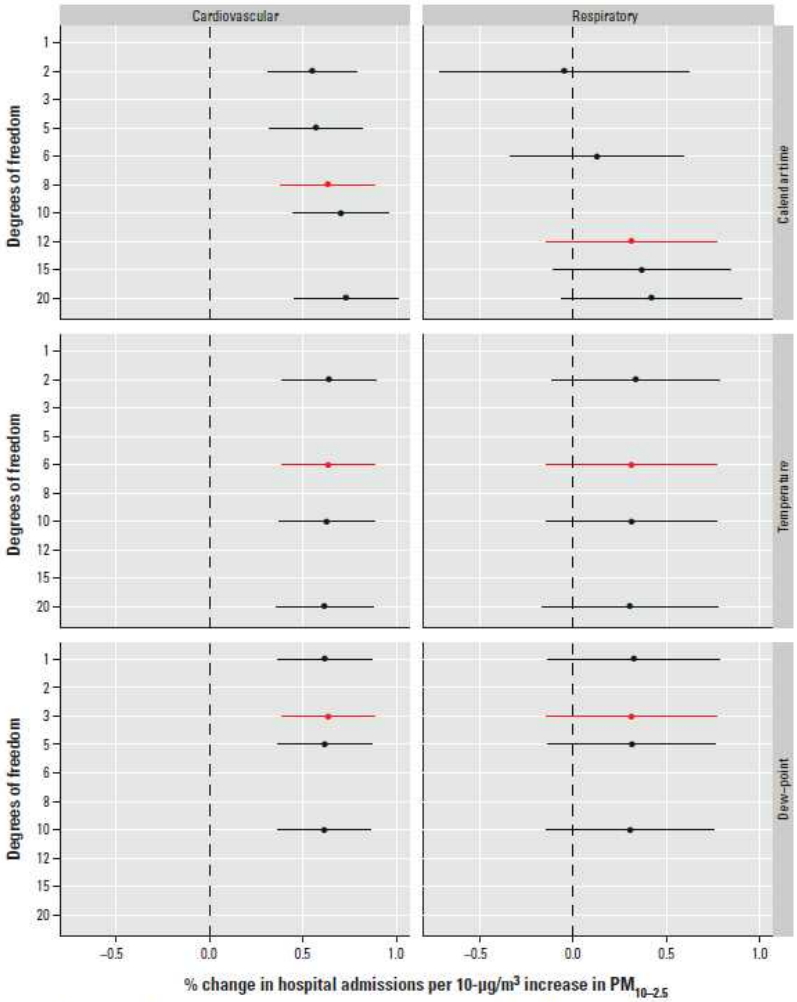


Figure 5. Estimated same-day national-level associations and 95% PIs between cardiovascular and respiratory disease admissions and a 10-μg/m^3 increase in $PM_{10-2.5}$ under varying degrees of freedom for the smooth functions of time, temperature, and dew point temperature. The estimates and PIs shown in red represent the degrees of freedom used in the final model.

[그림 7−22] 당일 국가 수준에서 시간, 온도 및 이슬점 온도의 부드러운 함수를 위한 다양한 자유도 하에서 PM10−2.5 에서 10−μg/m^3 증가와 심혈관 및 호흡기 질환 입원 사이의 연관성 및 95% PI 추정

빨간색으로 표시된 추정치 및 PI는 최종 모델에서 사용 된 자유도를 나타냄.

- 단점 : 외부 모니터 값의 영향을 줄이기 위해, 농도를 추정하기 위해 10 % 정돈된 평균을 사용했음. 하지만 이전의 PM10-2.5 연구에서 PM10-2.5에 대해 10 % 정돈된 평균값과보다 복잡한 측정 오차 모델을 사용하는 것 사이에 추정 된 건강 영향이 실질적으로 다르다는 것을 발견하지 못했음.

5) 고찰 및 결론

- PM10-2.5의 일중 변동은 65 세 이상 Medicare 가입자의 심혈관 질환 응급 입원과 유의한 연관성이 있었음. 이 연관성은 PM2.5 농도 보정에도 견고했음.

9. Horton MK, Hsu L, Claus Henn B et al. Dentine biomarkers of prenatal and early childhood exposure to manganese, zinc and lead and childhood behavior. Environ Int. 2018; 21(Pt 1): 148-158. doi: 10.1016/j.envint.2018.08.045. Epub 2018 Sep 8.

1) 배경

- 금속 노출은 신경발달적 결과를 변화시킴. 하지만 노출이 가장 큰 영향을 발휘할 때 감수성의 결정적인 창(critical windows)에 대해서는 거의 알려져 있지 않음. 누적 및 생애 초기 특이적 효과와 유년기 행동 결과에 대한 연구도 제한적이며, 신경독성 금속에 co-exposure 를 조사한 역학 연구도 제한적임

2) 목적

- 치아 생체 표지자로 망간(Mn), 아연(Zn), 납(Pb)과 아동기 후기의 행동의 연관성을 조사함.

3) 방법 : 코호트 연구

(1) 연구자료 및 연구대상자

- The Early Life Exposures in Mexico to Environmental Toxicants(ELEMENT) Study: 멕시코시티에서 수행된 어머니-자녀 쌍으로 구성된 전향적 출생코호트 연구임. 산전 환경 요인과 소아의 발달과 관련된 장기간의 결과를 조사하기 위해 구축됨

- 1994 년부터 2001 년에 구축되었으며 18-39 세의 저소득, 또는 중간소득 어머니를 모집함

- 제외기준: 5 분 Apgar score 6 점 이하(신생아 중환자실 치료가 필요한 상태)

- 133 명의 치아(상아질) 금속 측정을 완료하였음. 8-11 세에 BASC-2(Behavioral Assessment Systsem for Children, 2nd edition) 행동 결과를 측정함.

(2) 노출 평가

- 생태 표지자, 신경발달적 결과, 공변량 자료
- 임신 주수 4 개월(-4 mo.) - 출산 후 12 개월(12 mo.) 사이에 중금속 농도 측정
- 치아의 금속 분석을 위해 자연히 빠진 치아를 가져오거나 우편으로 보냈음
- laser ablation-inductively coupled plasma-mass spectrometry(LA-ICPMS)로 치아의 금속을 분석함
- 검출 한계: 0.05 ug/g

(3) 건강영향 평가

- 8-11 세에 BASC-2(Behavioral Assessment Systsem for Children, 2nd edition) 행동 결과를 측정함.
- 어머니들이 아동이 8-11 세일 때 행동을 평가함
- 전반적인 행동 기술 지수(overall behavioral skills index, BSI) 종합 점수, 내부화 및 외부화 종합 점수, 개인적 하위 항목으로 불안, 우울, 주의력, 과다행동 측정. 점수가 높을수록 행동 문제가 증가함.

(4) 보정변수

- 출산 시 어머니 나이, 어머니 교육 수준, 어머니의 임신 시 흡연, 사회경제적 상태, 어머니 IQ
- 재태연령, 출생 체중, 출생 키

(5) 통계분석 방법

- 표준 단변량 또는 이변량 탐사를 하였으며, 필요시 자연로그 변환을 하였음.
- 변화된 분포된 래그 모델(distributed lag model, DLM)으로 신경발달적 결과에 대해 민감한 창(window)을 확인하였음. DLM 은

시계열 데이터에 대한 회귀 모형으로, 특정 시간 창에서 발생하는 노출의 영향을 추정하는 동시에 다른 시간 및 공변량의 노출을 조정함. 노출의 효과가 시간 경과에 따라 부드럽게 변하는 것으로 가정함. 환경보건연구에 광범위하게 사용되고 있는 방법임. 본 연구에서는 역 분포된 래그 모델(rDLM)에서 시간 계수를 갖는 기능 스플라인 모델을 사용하였음.

- Bonferroni 가 조정 한 신뢰 구간과 Holm-Bonferroni 조정 신뢰 구간을 일정한 간격으로 겹쳐서 시간에 따라 변하는 연관성 매개 변수가 그려지게 됨. 민감한 창은 노출과 결과 변수 사이의 연관 매개 변수에 대한 추정 된 포인트와 95 % 신뢰 구간이 0 을 포함하지 않을 때 식별됨 (Chiu et al., 2016). 결과는 노출의 단위 변화 당 결과의 단위 변화 (1 단위 = 1 표준 편차 (BASC-2 점수의 1 표준 편차 (SD))로 표시됨.

- 본 연구에서는 rDLM 을 확장하여 지연된 가중 분위 합(Lagged Weighted Quantile Sum, WQS) 회귀 (Bello et al., 2017)를 사용하여 세 가지 금속에 대한 동시 노출의 시간 의존 효과를 모델링했음. 결과는 노출의 단위 변화 당 결과의 단위 변화 (1 단위 = BASC-2 점수의 1 표준 편차 (SD))로 표시됨 (1 단위 = 로그 변환 된 치아 금속 농도 1 SD).

- 가중 분위 합(WQS) 회귀는 관련 변수 (즉, 화학 물질)를 일차원 인덱스로 그룹화하는 음수가 아닌 단위 합계 제약 조건을 사용함. 지수는 결과 변수와의 상관관계를 결정하고 각 화학 물질의 상대적 또는 중요성, 즉 비례 무게에 관한 일반화된 추론을 하기 위해 결과 변수에 대해 회귀됨 (Carrico et al., 2014; Czarnota et al. , 2015). 지연된(lagged) WQS 는 가중치 색인을 사용하여 혼합 효과를 추정하기 위해, WQS 의 불이익을 회피하는 회귀선 확장으로 가장 크고 잠재적으로 생물학적으로 관련이있는 혼합 효과가 있는 기간을 확인 할 수 있음.

- 민감도 분석에서 누적 상아질 금속 농도와 BASC-2 의 연관성을 조사하기 위해서 각 금속에 대해 50 점을 합산하여 치아 금속 농도를 통합하여 계산하고, BASC-2 결과와 단일 집계 된 노출 측정 간의 연관성을 조사하기 위해 공변량을 보정한 후 선형 회귀 모델을 사용했음. 또한, 특정 창문 (출생 ± 2 주, 출생 후 3-5 개월)을 종합하여 이 창에서 수집된 금속 농도와 BASC-2 결과간의 연관성을 조사했음.

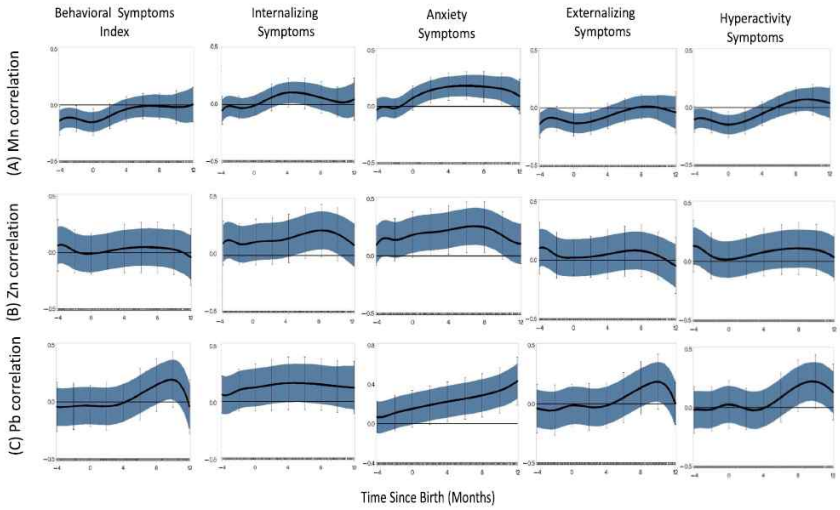
4) 결과

- 개별 증상 점수는 전반적인 BASC-2 Behavioral Composite 점수와 유의 한 양의 상관 관계를 보였음 ($r = 0.4-0.8$, $p < 0.001$).

- 모든 Mn, Zn 및 Pb 의 치아 농도는 아이의 성별, 재태 기간, 출생 체중 및 모성 빈혈 상태와 관련이 없었음.

- 이 연구에 포함 된 133 명의 피험자 중 출생 전 및 출생 후의 치아 Mn 은 아동기 행동 문제, 특히 과다 활동 및 주의력에 대해 보호적인 것으로 나타났음. 출생 후의 치아 Mn 은 내면화 문제, 특히 불안의 증가 된 보고와 관련이 있었음. 6 개월 째에서, 1 단위 Mn 의 증가 (단위 = 로그 농도의 1SD)는 내면화 증상 점수가 0.18 단위 (단위 = BASC-2 점수의 1SD) 증가하고 불안이 0.25 단위 증가한 것과 관련이 있었음.

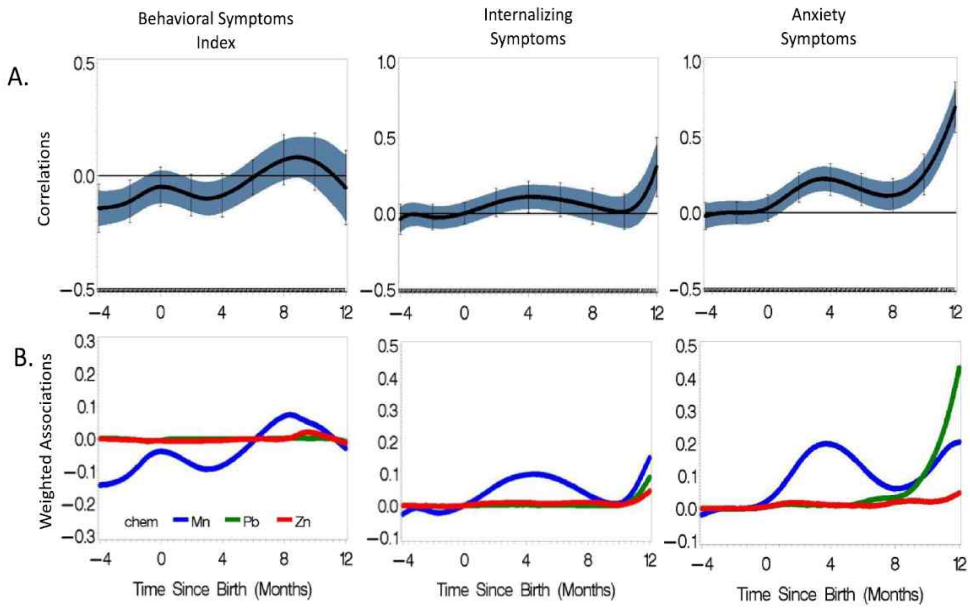
- 출생 후의 Pb 는 불안 증상의 증가와 관련이 있었음. 12 개월에 Pb 가 1 단위 증가했을 때 불안 증상이 0.4 단위 증가했음.



[그림 7-23] 개별 역 분포된 래그 모델(rDLM)은 미세한 규모 분해의 치아 금속 수준과 BASC-2 행동 증상 간의 연관성을 보여줌

산모 교육 및 재태 연령을 보정한 (A) 망간 (위), (B) 아연 (중간), (C) 납 (아래)) (n=133); Y 축 = 노출과 결과 사이의 시간 변화 상관; X 축 = 치아 수집의 시점을 나타내는 출생 이후 시간; 중심 검은 선은 = 상관 관계를 나타냄; 파란색 막대는 95 % 신뢰 구간을 나타내고 세로 막대는 95 % Holms-Bonferroni family-wise 신뢰 구간을 나타냅니다. 중요한 창을 나타내는 유의한 연관성은 0 을 포함하지 않는 rDLM 의 영역으로 정의됨. (이 그림의 범례에서 색상에 대한 참조를 해석하려면 이 논문의 웹 버전을 참조하십시오.)

- 금속 혼합체로 조사했을 때, 불안 증상이 증가 할 가능성이 있는 두 가지 창문을 관찰했음 : 첫 번째 창 (0-8 개월)은 Mn 에 의해 유도된 것처럼 보였고, 두 번째 창 (8-12 개월)은 Pb 이 주도적인 금속 혼합체로 인한 것임. 혼합 지수의 1 단위 증가는 불안 증상의 SD 에서 0.7 단위 증가와 연관이 있었음.



[그림 7-24] ELEMENT133 명을 대상으로 8-11 세에 평가한 BASC-2 행동 기술 점수, 내재화 증상, 불안 증상을 사용하여 치아 망간, 아연, 납에 대한 공동 노출의 시간 변화 연관성을 나타내는 지연된 가중 분위 합(WQS).

(A) 95 % 신뢰 구간 (파란색 띠)과 95 % Holm-Bonferroni 가족 단위의 신뢰 구간 (수직 막대)을 갖는 지연된 WQS 그래프. Y 축은 공동 노출과 결과 사이의 시간 변화 상관 관계를 나타냄.

(B) 플롯 A 에 표시된 관찰 된 혼합 효과에 대한 개별 금속의 가중치 관계. 두 그림 모두에서 X 축은 치아 샘플링의시기 (개월)를 나타내는 출생 이후의 시간을 나타냄.

- 장점 : 우리의 결과는 각각의 개별 금속뿐만 아니라 금속을 측정할 때 시간 창에 따라 달라지는 금속의 조합에 대해서도 상대적으로 작고 통계적으로 유의미한 관계가 있음을 나타냄.

- 단점 : 지연된 WQS 방법은 시간에 따라 변하는 금속 혼합물의 포괄적인 프로파일을 제공하지만, 잠재적으로 승수(multiplicative) 효과(즉, 금속 상호 작용)의 추정을 허용하지 않음

5) 고찰 및 결론

- 태아기(prenatal)의 치아 Mn 은 보호적일 수 있으며, 출생 후 초기의 과도한 Mn 은 부정적 행동에 대한 위험을 증가시킬 수 있음. Mn, Zn 및 Pb 의 농도가 높을수록 행동에 나쁜 영향을 줄 수 있음.

<Method 참고>
- Mn 의 확인을 위해 임신 2 기 (Arora et al., 2012)에서 샘플링 한 바닥 먼지(floor dust)에서 Mn loading 을 가진 치아 Mn 사이에 유의한 양의 상관 관계가 있었음. 납은 임산부 혈청 Pb 농도를 사용하여 검증 하였음. 그 외 금속은 (Andra et al., 2016) 참고.
- 낙엽(deciduous)의 치아는 금속을 축적하고 산후와 조기 출산시기(앞니의 경우 임신 주 13-16, 대구치는 출생 후 10-11 개월때)에 광물이 점진적 패턴 (성장 고리에 가까움)으로 진행됨 (Ash and Neslon, 2002; Hillson, 1996). 신생아 라인 (출생시 에나멜과 상아질에서 형성된 조직 학적 특징)과 샘플링 포인트에 시간 정보를 할당하기위한 일일 성장 점진적 표시를 사용하여 임신 3 개월부터 임신 1 년까지의 약 50 개의 시점에서 각 치아를 샘플링했습니다. 약 7-10 일마다 샘플링 빈도. 이것은 대략 매주 해상도에서의 금속 섭취량을 반영합니다. 검출 한계는 0.05 µg / g
각 금속에 대해. 미네랄 밀도가 치아 내에서 그리고 치아 유형과 개인 사이에서 다양하기 때문에 각 시료 채취 지점에서의 금속 강도는 분석물 농도를 칼슘으로 나누고 비율 (예 : 55Mn : 43Ca, 66Zn : 43Ca, 또는 208Pb : 43Ca).

10. Romano ME, Eliot MN, Zoeller RT et al. Maternal urinary phthalate metabolites during pregnancy and thyroid hormone concentrations in maternal and cord sera: The HOME study. In J Hyg Environ Health. 2018; 221(4): 623-631. doi: 10.1016/j.ijheh.2018.03.010. Epub 2018 Mar 23.

1) 배경

- 소비자 제품에 흔히 발견되는 내분비 교란 물질인 프탈레이트(phthalate)는 갑상선 호르몬에 악영향을 미칠 수 있지만 이전의 역학 연구 결과는 상충되고 있음.

2) 목적

- 전향적 코호트 연구로 임신 중에 측정 된 모체 소변 프탈산 대사물질 농도와 프탈레이트 혼합물이 임산부와 신생아에서 갑상선 호르몬과 관련이 있는지 조사했음.

3) 방법 : 코호트 연구

(1) 연구자료 및 연구대상자

- The Health Outcomes and Measures of the Environment(Home) Study: 오하이오주 신시내티에서 수행된 어머니-자녀 쌍으로 구성된 전향적 출생코호트 연구임. 일반환경화학물질 노출이 어린이 건강에 미치는 영향을 평가하기 위해 고안됨

- 구축 시 임신 16±3 주, 18 세 이상, 영어 가능, 1978 년 이전 지어진 집에 거주자를 2003 년 3 월부터 2006 년 1 월까지 모집함

- 제외기준: HIV 감염, 발작(seizure) 이나 갑상선 질환으로 약물 치료중인 여성

- 1,263 명 중 468 명(37%)가 모집되었으며 389(83%)는 한 명(singleton) 영아 출산을 통해 추적 하였음.

(2) 노출 평가

- 모체의 소변 내 프탈레이트 대사체(metabolites)

- 어머니들은 임신 16 주 (범위 : 10-23)와 임신 26 주 (범위 : 19-35)에 2 개의 소변 샘플을 제공했음. 이 중 16 주의 요중 프탈레이트를 사용함.

- 9 종의 프탈레이트 모노에스테르 대사물 [모노에틸 프탈레이트 (monoethyl phthalate, MEP), 모노 n-부틸 프탈레이트 (mono-n-butyl phthalate, MnBP), 모노이소부틸 프탈레이트 (mono-isobutyl phthalate, MiBP), 모노벤질 프탈레이트 (monobenzyl phthalate, MBzP), 모노-3-카복시프로필 프탈레이트 (mono-3-carboxypropyl phthalate, MCPP), 모노(2-에틸헥실) 프탈레이트 (mono(2-ethylhexyl) phthalate, MEHP), 모노(2-에틸-5-하이드록시헥실) 프탈레이트 (mono(2-ethyl-5-hydroxyhexyl) phthalate, MEHHP), 모노(2-에틸-5-옥소헥실) 프탈레이트 (mono-(2-ethyl-5-oxohexyl) phthalate, MEOHP), 모노(2-에틸-5-카복시펜틸) 프탈레이트 (mono(2-ethyl-5-carboxypentyl) phthalate, MECPP)] 는 적어도 6 개의 모 (parent) 프탈레이트 노출을 반영함.

- cf 모 프탈레이트): diethyl phthalate, di-n-butyl phthalate, di-isobutyl phthalate, benzylbutyl phthalate, di-n-octylphthalate, di(2-ethylhexyl) phthalate (DEHP)

- 프탈레이트 대사산물 농도를 크레아티닌으로 표준화하였음.

- MECPP 의 분자량에 표준화 된 DEHP (ΣDEHP)의 대사 산물을 각각의 대사 산물의 요중 농도를 분자량으로 나누어 4 개의 측정 된 DEHP 대사 산물 [MEHP, MEHHP, MEOHP 및 MECPP] 대사 산물 농도 및 MECPP (308 g / mol)의 분자량을 곱하였음.

- 검출 한계: 0.2 - 1.2 ng/ml

(3) 건강영향 평가

- 혈청 갑상선 호르몬 농도
- 임신 16 주와, 분만시 정맥 제대혈(venous cord blood) 측정.
- 갑상선 자극 호르몬(thyroid stimulating hormone, TSH), 전체와 유리 티록신(total thyroxine, TT4 and free thyroxine, FT4), 트리이오도사이로닌(triiodothyronine, TT3 and FT3)

(4) 보정변수

- 임신 2 분기에 설문조사를 실시함
- 생식 및 병력, 인구학적, 사회경제적, 주산기, 행동적 요인. 분만 방법과 출생아 성별. 요중 코티닌(흡연 및 간접 흡연), 모성 총 비스페놀 A(bisphenol A)
- 77%에서는 요중 이오다인(iodine) 측정
- 80%에서 산모 혈청 다염화 비페닐(polychlhorinated biphenyls, PCB)-153, 다브롬화 비페닐 에스터 (polybrominated diphenyl ethers, PBDE) 동족체 (BDE-28, BDE-47)

(5) 통계분석 방법

- 로그 10 변환을 한 요중 프탈레이트 대사체 농도 평균을 회귀 모형에 사용함. TSH 는 자연로그 변환함.
- 다변량 선형 회귀 분석(multivariate linear regression)을 사용하여 개별 갑상선 호르몬과, 모체의 소변 프탈레이트 대사 산물 농도의 각각 10 배 증가의 보정된(adjusted) 차이를 추정했음.
- 모체 소변 MEP, MnBP, MiBP, MBzP, MCPP, ΣDEHP 의 10 배 증가와 TSH 의 퍼센트 차이 [%diff= e^{β}-1*100, β 는 관심의 추정된 회귀 계수], T3 또는 T4 의 평균 차이를 구하였음
- 보정: 출산 시 모성의 나이, 인종, 교육, 결혼 상태, 가구 소득, 출산력(parity), 체질량지수, 임신 중 혈청 코티닌, 산전 비타민 사용
- 모성 혈청 갑상선 호르몬은 임신 16 주 log10-모성 요중 BPA 농도를 보정함

- 제대혈 갑상선 호르몬의 추가 보정은 임신 16 주, 26 주의 log10-모성 요중 BPA 농도, 영아의 성별, 분만시 임신 주수, 분만 방법

- 가중 분위 합(Weighted Quantile Sum, WQS) 회귀를 사용하여 모성 프탈레이트 대사체 혼합물과 모성 및 제대 갑상선 호르몬의 연관성을 조사하였음. WQS 회귀 분석을 통해 상관도가 높은 노출을 지표로 결합한 다음 화학 혼합물과 관심 결과 간의 연관성을 추정하는 데 사용할 수 있음 (Carrico et al., 2015). 이전에 HOME 연구 참가자에서 모성 소변에서 반복적으로 측정 한 프탈레이트와 모체의 요오드 프탈레이트 대사 산물 간의 상관 관계는 약하거나 중간 정도이었음 (Shoaff et al., 2016). 산모의 phthalate 대사 산물의 평균 농도 사이의 Spearman 상관 관계는 본 연구의 분석 표본에서 0.03-0.44 의 범위였음. 따라서 본 연구에서 WQS 를 사용하여 단일 프탈레이트 지수를 만들었음. 이 프탈레이트 지수는 개별 프탈레이트 대사 산물의 사 분위수와 각 갑상선 호르몬의 연관성을 사용하여 지수에 포함 된 각 프탈레이트 대사 산물의 경험적 가중치를 추정함. 경험적 가중치는 혼합 성분 (즉, 개별 프탈레이트 대사 산물)의 전체 결합에 대한 상대 기여를 나타내므로, 관심 결과에 더 큰 영향을 미치는 프탈레이트 대사 산물은 약한 효과를 갖는 프탈레이트 대사 산물보다 더 높은 가중치를 가짐. 본 연구에서는 프탈레이트 지수를 다변수 선형 회귀 모델(multivariable linear regression)에 통합하여 위에서 설명한 사항을 보정하고 총체적인 영향을 추정해 보았음. 이 문맥에서 프탈레이트 지수에 대한 베타는 모체 또는 신생아 갑상선 호르몬에 대한 프탈레이트 혼합물의 효과를 추정하는 것으로 해석 할 수 있음. WQS 는 단일 방향으로 추론을하기 때문에 (Carrico et al., 2015), 지수에 직접적으로 포함되지 않은 모든 프탈레이트 대사 산물은 다변수 WQS 회귀 모형에 조정 변수로 추가됨. 개별 프탈레이트 대사 산물에 대한 다변수 회귀 분석의 결과는 어떤 프탈레이트가 각 갑상선

호르몬에 대한 프탈레이트 지수에 포함될 수 있는지 알려주는 데 사용되었음.

- R 패키지 wqs (0.0.1)를 사용함.

- 상호 작용에 대한 p-value는 0.10 미만으로 효과 측정 수정 (effect measure modification, EMM)을 나타냈음.

* 프탈레이트 / 요오드 및 기타물질 / 성별(EMM)

4) 결과

- 연구에 참여한 어머니는 주로 25-35 세 (64%), 비 히스패닉 백인 (66%), 결혼 (71%)이었음.

- 제대혈 불충분한 갑상선 기능 기준: TSH > 20 mIU/L, 모체의 무증상 갑상선 저하 기준 TSH > 3.0 mIU/L

- 개별 모체 요중 프탈레이트 대사 산물은 일반적으로 모성 갑상선 호르몬 변화와 관련이 없었음. 그러나 16주 모체 요중 MEP가 10배 증가하면 모체 혈청 총 thyroxine (TT4)가 0.52 ug/dL (95% CI : -1.01, -0.03) 감소했음.

- 평균 (16 주 및 26 주) 모체 요중 MBzP의 10 배 증가에 대해, 제대 혈청 TSH는 19% (95% CI : -33.1, -1.9) 감소했음.

- 모체 혈청 갑상선 호르몬을 평가하기 위해 WQS 회귀 분석을 사용하였고, 프탈레이트 지수는 모체 혈청 TT4와 역 상관관계가 있었음 (WQS β= -.060; 95% CI: -1.01, -0.18). MEP (0.39)와 MCPP (0.37)가 가장 강한 개별 가중치를 가지고 있었음.

- 신생아 중 프탈레이트 지수는 제대혈 TSH (WQS β = -0.11; 95% CI : -0.20, -0.03)와 TT4 (WQS β = -0.53; 95% CI : -0.90, -0.16)와 역의 상관관계가 있었음. TSH의 경우 가중치가 큰 순서로 MBzP (0.43)와 MiBP (0.30)이었고, TT4의 경우는 MEP (0.32), MiBP (0.28), MBzP (0.23) 순이었음.

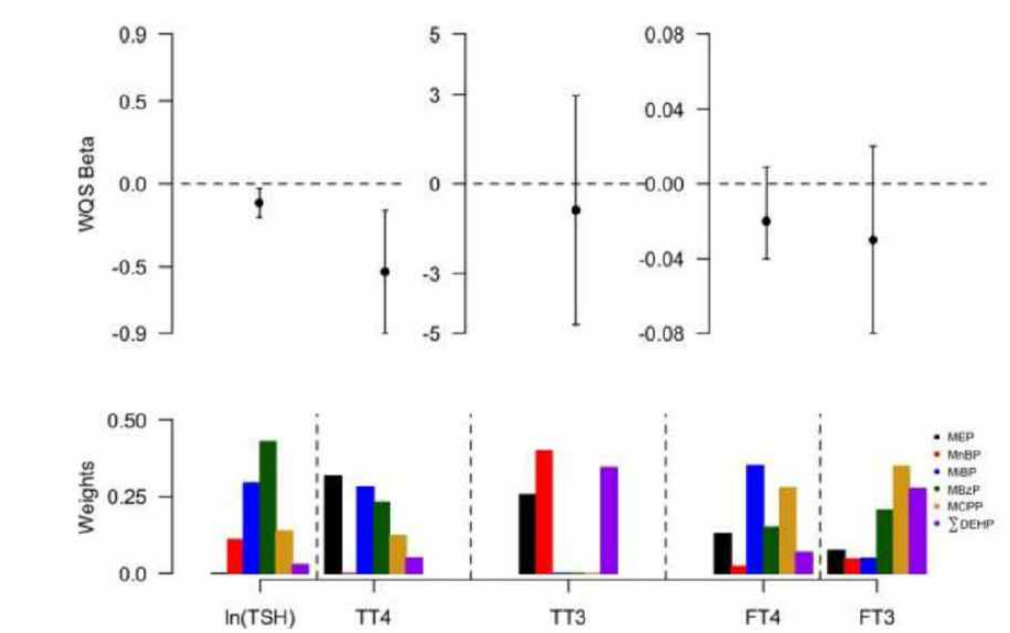


Figure 1.
Beta coefficients, 95% confidence intervals, and weights from weighted quantile sum regression of maternal urinary phthalate index and thyroid hormones in cord serum

[그림 7-25] 모성 요중 프탈레이트 지수와 제대 혈청의 갑상선 호르몬의 베타 계수, 95% 신뢰구간, 가중분위합(WQS) 회귀에서의 가중치

<표 7-23> 모성 요중 프탈레이트와 모성 및 제대 혈청에서 갑상선 호르몬 농도의 가중 분위합(WQS)으로부터 가중치와 계수

Table S3: Weights and coefficients from Weighted Quantile Sum regression of maternal urinary phthalates during pregnancy with thyroid hormone concentrations in maternal and cord sera

	Weights						WQS beta (95% CI)[a]	p-value
Maternal Serum Thyroid Hormones [a]	MEP	MnBP	MiBP	MBzP	MCPP	∑DEHP		
Ln(TSH)	0.311	---	0.217	---	0.216	0.256	0.12 (-0.05, 0.30)	0.17
TT_4	0.388	---	0.134	---	0.367	0.111	-0.60 (-1.01, -0.18)	0.005
TT_3	---	---	0.417	---	0.418	0.165	-3.44 (-9.50, 2.62)	0.27
FT_4	---	0.174	0.204	0.426	0.196	---	0.001 (-0.02, 0.02)	0.96
FT_3	0.145	0.219	0.056	0.173	0.115	0.292	-0.03 (-0.09, 0.03)	0.38
Cord Serum Thyroid Hormones [b]								
Ln(TSH)	---	0.110	0.295	0.428	0.139	0.029	-0.11 (-0.20, -0.03)	0.01
TT_4	0.316	---	0.281	0.231	0.122	0.050	-0.53 (-0.90, -0.16)	0.005
TT_3	0.257	0.400	---	---	---	0.344	-0.88 (-4.72, 2.96)	0.65
FT_4	0.130	0.024	0.350	0.150	0.278	0.069	-0.02 (-0.04, 0.01)	0.21
FT_3	0.075	0.046	0.049	0.207	0.348	0.275	-0.03 (-0.08, 0.02)	0.18

[a] Models for maternal serum thyroid hormone concentrations were adjusted for maternal age at delivery, race, education, marital status, household income, parity, serum cotinine, body mass index, prenatal vitamin use, $\log_{10}$-maternal urinary bisphenol A concentration at 16 weeks' gestation.

[b] Models for cord serum thyroid hormone concentrations were adjusted for maternal age at delivery, race, education, marital status, household income, parity, serum cotinine, body mass index, prenatal vitamin use, infant sex, average of $\log_{10}$-maternal urinary bisphenol A concentrations from 16 and 26 weeks' gestation, gestational age at delivery, mode of delivery

- 장점 : 우리의 분석은 여러 프탈레이트가 갑상선 축에 어떻게 영향을 미칠 수 있는지 이해하는 중요한 단계임.

5) 고찰 및 결론

- 출생 코호트에서 여러 프탈레이트에 대한 공동 노출이 특정 갑상선 호르몬 (임산부와 신생아의 TT4, 신생아의 TSH)과 반비례하였음.

11. Lenters V, Portengen L, Smit LA, Jönsson BA, Giwercman A, Rylander L, Lindh CH, Spanò M, Pedersen HS, Ludwicki JK, Chumak L, Piersma AH, Toft G, Bonde JP, Heederik D, Vermeulen R. Phthalates, perfluoroalkyl acids, metals and organochlorines and reproductive function: a multipollutant assessment in Greenlandic, Polish and Ukrainian men. Occup Environ Med. 2015 Jun;72(6):385-93.

1) 배경

- 수많은 환경유해물질이 생식 건강과 관련되어 있으나, 노출의 복잡한 상호 관계 구조와 다중검정(multiple testing) 문제가 현재 존재하는 근거들의 해석을 제한함.

2) 목적

- 여러 가지 환경유해물질 노출에서 남성 생식기능의 생물학적 지표와 관련된 노출 물질을 확인하는 것임.

3) 방법

(1) 연구자료 및 연구대상자

- 본 연구는 단면연구 디자인으로 수행되었음.

- 폴란드 바르샤바에 있는 큰 중앙 병원, 우크라이나 카르키프에 있는 병원 3곳과 산전 클리닉 8곳, 그린란드 전역의 19개 시에 있는 지역 병원에서, 2002년부터 2004년까지 산전 관리를 위해 방문한 임신부와 남성 파트너를 모집하였음.

- 1710명의 커플(45% 참여율) 중, 모집 지역 3군데에서 각각 약 200명의 남성 파트너가 정액 연구에 참여한다고 동의할 때까지(38% 참여율) 계속해서 남성 파트너를 모집하였음.

- 참가자들은 정액 샘플, 정맥 혈액 샘플 및 생활 습관에 대한 정보를 제공함.

(2) 노출 평가

- 프탈레이트 대사체와 과불화알킬산(perfluoroalkyl acids, PFAAs)은 혈청 샘플에서 liquid chromatography-tandem mass spectrometry 로 분석하였음.

- 프탈레이트 대사체는 다음과 같은 물질을 포함함: di(2-ethylhexyl) phthalate (DEHP)의 이차 산화 대사체 [mono(2-ethyl-5-hydroxyhexyl) phthalate (MEHHP, alternatively 5OH-MEHP), mono(2-ethyl-5-oxohexyl) phthalate (MEOHP, alternatively 5oxo-MEHP) 및 mono(2-ethyl-5-carboxypentyl) phthalate (MECPP, alternatively 5cx-MEPP)], 그리고 diisononyl phthalate (DiNP)의 이차 산화 대사체 [mono(4-methyl-7-hydroxyoctyl) phthalate (MHiNP, alternatively 7OH-MMeOP), mono(4-methyl-7-oxooctyl) phthalate (MOiNP, alternatively 7oxo-MMeOP) 및 mono(4-methyl-7-carboxyheptyl) phthalate (MOiCP, alternatively 7cx-MMeHP)].

- 과불화화합물의 일종인 과불화알킬산의 분석은 perfluorooctane sulfonic acid (PFOS), perfluorooctanoic acid (PFOA), perfluorhexane sulfonic acid (PFHxS), perfluorononanoic acid (PFNA), perfluorodecanoic acid (PFDA), perfluoroundecanoic acid (PFUnDA) 및 perfluorododecanoic acid (PFDoDA)을 포함함.

(3) 건강영향 평가

- 표 1에 열거 된 22 개의 생식 표지자를 표준화 된 프로토콜에 따라 평가 함. : 생식 호르몬, 정액 특성, 정자 염색질 온전성 및 세포자연사의 표지자, 부고환 및 부속기 성 기능의 표지자 및 정자의 Y:X 비율.

<표 7-24> 임산부의 남성 배우자에 대하여 생식 기능과 가능한 교란요인에 대한 바이오마커

Table 1 Summary statistics for biomarkers of reproductive function and potential confounders among 602 male partners of pregnant women

Outcomes and covariates	n*	Greenland (n=199) GM (5, 95 P)	Warsaw, Poland (n=197) GM (5, 95 P)	Kharkiv, Ukraine (n=206) GM (5, 95 P)	p Value†
Male reproductive hormones in serum‡§¶					
Follicle-stimulating hormone (IU/L)	456	4.38 (1.70, 9.20)	3.46 (1.40, 7.40)	3.67 (1.60, 9.40)	0.001
Luteinising hormone (IU/L)	456	4.05 (1.80, 8.50)	3.81 (1.90, 7.10)	3.79 (1.90, 7.60)	0.34
Inhibin B (ng/L)**	456	184.21 (83.00, 314.00)	157.94 (71.00, 275.00)	194.74 (98.00, 321.00)	<0.001
Sex hormone-binding globulin (nmol/L)	455	26.90 (15.60, 46.70)	21.61 (9.45, 40.90)	26.19 (15.50, 43.20)	<0.001
Total testosterone (nmol/L)**	456	14.81 (7.27, 24.41)	13.08 (6.83, 20.76)	18.01 (10.26, 26.89)	<0.001
Free testosterone (nmol/L)**	455	0.31 (0.18, 0.48)	0.29 (0.16, 0.42)	0.39 (0.22, 0.58)	<0.001
Estradiol (pmol/L)	454	63.44 (40.50, 98.70)	71.50 (45.80, 112.60)	80.73 (52.50, 137.50)	<0.001
Conventional semen characteristics					
Semen volume (mL)††,‡‡	535	3.21 (1.20, 7.20)	3.37 (1.25, 7.19)	3.29 (1.59, 7.57)	0.68
Sperm concentration (10^6/mL)††	600	51.12 (10.75, 199.00)	55.95 (7.15, 257.50)	53.07 (10.15, 185.80)	0.65
Total sperm count (10^6/ejaculate)††,‡‡	533	160.15 (31.50, 692.30)	186.68 (18.82, 1061.96)	179.46 (34.20, 804.80)	0.39
Morphologically normal (%)**	598	6.97 (2.00, 13.00)	6.66 (2.00, 14.00)	7.37 (1.00, 15.00)	0.20
Progressive motility (% A+B)**,††,§§	565	55.29 (22.00, 81.00)	60.98 (19.00, 87.00)	55.08 (14.50, 87.00)	0.007
Sperm chromatin integrity§,††					
SCSA DNA fragmentation index (%)	546	7.66 (3.32, 20.28)	10.07 (4.16, 31.50)	10.83 (3.54, 31.56)	<0.001
High DNA stainability (%)	546	11.37 (4.37, 29.88)	8.66 (3.60, 25.14)	10.29 (4.35, 26.55)	<0.001
TUNEL DNA fragmentation index (%)	462	2.95 (0.79, 13.37)	11.64 (2.57, 38.28)	6.46 (1.90, 30.18)	<0.001
Apoptotic markers§,††					
Fas positivity (%)**	454	25.00 (2.22, 70.60)	48.57 (3.13, 96.42)	28.24 (0.00, 90.60)	<0.001
Bcl-xL positivity (%)**	286	26.54 (0.44, 90.65)	18.05 (0.81, 69.64)	66.01 (9.54, 99.17)	<0.001
Epididymal and accessory sex gland function§,††,‡‡					
Neutral α-glucosidase (mU/ejaculate)	448	16.32 (4.80, 52.00)	21.81 (6.55, 71.52)	17.51 (4.86, 52.56)	0.001
Prostate-specific antigen (μg/ejaculate)	507	3187.74 (1078.00, 8330.85)	3708.59 (1312.37, 11127.89)	2497.94 (784.08, 7707.70)	<0.001
Zinc (μmol/ejaculate)	497	4.68 (1.06, 12.63)	6.31 (1.50, 25.23)	4.22 (1.11, 16.56)	<0.001
Fructose (μmol/ejaculate)	506	47.00 (12.67, 153.85)	42.24 (6.10, 161.65)	39.36 (10.69, 117.35)	0.18
Y chromosome sperm cells (%)**	411	51.26 (48.52, 55.49)	50.30 (48.79, 52.11)	50.76 (48.72, 53.61)	<0.001
Age (years)**	597	31.34 (20.83, 43.15)	30.34 (25.36, 36.80)	27.88 (20.67, 38.75)	<0.001
BMI (kg/m^2)**	592	25.99 (20.56, 31.77)	25.81 (20.43, 31.93)	24.20 (19.71, 29.41)	<0.001
Abstinence period (days)	557	2.86 (0.50, 7.10)	4.27 (1.00, 30.00)	3.41 (1.50, 7.00)	<0.001
Current smoker, n (%)	597	144 (72.7%)	53 (27.2%)	136 (66.7%)	<0.001
Current smokers: cotinine (ng/mL)	589	113.73 (0.96, 465.23)	23.95 (<LOD, 358.42)	33.59 (<LOD, 447.36)	<0.001
Time of blood sampling, <12:00, n (%)	504	17 (16.8%)	187 (94.9%)	183 (88.8%)	<0.001

*Number available out of 602.
†Test for difference in levels between the three study populations: ANOVA for means and χ^2 test for proportions.
‡In adjusted analyses, adjusted for time of blood sampling.
§In adjusted analyses, adjusted for age.
¶In adjusted analyses, adjusted for BMI.
**Outcome was not ln-transformed; arithmetic mean presented.
††In adjusted analyses, adjusted for ln-abstinence period.
‡‡Samples with spillage were excluded (n=67 for semen volume and sperm count; n=41–52 for the 4 epididymal and accessory sex gland function markers).
§§Samples with a delay of >1 h from collection to semen analysis were excluded (n=28).
BMI, body mass index; GM, geometric mean; LOD, limit of detection; P, percentile, SCSA, sperm chromatin structure assay; TUNEL, terminal deoxynucleotidyl transferase dUTP nick end-labelling.

(4) 통계분석 방법

- 노출 프로파일과 각 결과 간의 연관성을 평가하기 위해 희소 부분 최소 제곱(sparse partial least squares, sPLS) 회귀 모델링을 사용함.

- 부분최소제곱(PLS) 회귀 분석에서 노출의 X 행렬(X matrix)과 Y 결과 벡터(Y outcome vector)는 동시에 잠재 변수(latent variable)로 분해되고 X 와 Y 사이의 공분산(covariance)을 최대화하는 방식으로 회귀됨.

- 구성 요소(component)라고 하는 이러한 잠재 변수(latent varialbe)는 입력 변수의 전체 세트를 직각으로 선형 조합 한 것임.

- Chun 과 Keleş 는 희소부분최소제곱(sPLS)를 제안했는데, penalisation(shirankage 라고도 부름)을 통해 '유익하지 않은(uninformative)' X 변수를 제거함으로써 변수 선택을 PLS 알고리즘에 포함시킴.

- 차원 축소 중에 페널티 기간(penalty term) (L_1)이 적용되고 축소된 X 변수 집합에 대한 PLS 를 통해 회귀 계수가 계산됨.

- 모델 복잡성은 모델을 구성하는 데 사용되는 구성 요소의 수(K)와 희소(sparsity) 정도(η)의 함수임.

- 최적의 모델을 결정하기 위해 (overfitting 을 피하기 위해) 우리는 Monte Carlo 교차 검증(cross validation)을 수행하고 교차 검증 루프(cross validation loop)에 null (절편 만 포함한 모델) 모델을 추가하여 중요성을 평가함.

- 우리는 K 값 1-5 와 0.01 의 단계에서 0.01 에서 0.99 사이의 η 값을 갖는 모델을 5 배 교차 검증을 사용하여 500 번 모델을 (데이터의 서로 다른 임의의 파티션과 함께) 반복해서 테스트함.

- 전체 최소 평균 제곱 오차 예측 (MSEP, minimum mean squared error of prediction) 중 한 표준오차 (SE, standard error) 내에서 가장 간략한 모델을 선택함.

- sPLS 분석에 착수하기 전에 몇 가지 데이터 전처리 및 탐색 단계를 수행함.

- 검출 한계 (LOD, limit of detection) 이하 노출 데이터 (0-18 %) 및 노출 매트릭스 에서 무작위로 누락 된 값 (missing-at-random) (2-16 %)에 대해 결측값 대체 (imputation)를 수행함.

- 1 차 분석에서 HCB, PCB-153 및 p, p′-DDE는 지질 보정 (lipid adjusted) 되었으며 시료의 70 % 이상이 검출 한계보다 높은 농도를 갖는 4 가지 오염 물질은 분석에서 제외됨.

- 정규성에 근사하기 위해, 모든 노출 변수에 대해 로그(ln) 변환을 하였고, 22 개의 결과 변수 중 14 개 결과 변수를 로그 변환함.

- 잠재적 혼란을 해결하기 위해 2 단계 회귀 접근 방식을 사용함: 첫째, 각 결과 및 각 노출은 잠재적인 혼동 요인에 대해 개별적으로 회귀되었으며 둘째, 계산된 잔차를 입력하여 sPLS 회귀 모델을 적합(fitting)하였음.

- 본 연구에서는 결과 변수에 따라 선행적으로 (연구 집단 및 혈청 코티닌 수준, 가변 연령, 체질량 지수, 금욕 기간 및 혈액 채혈 시간을 포함한) 잠재적인 혼란변수들을 선택하였음.

- 누락 된 공변량 데이터 (0.8-16 %)에 대해서 결측값 대체 함.

- 과보정을 평가하기 위해, 2 차 분석에서는 인구 집단에 대해 선행 표준화 된 노출과 결과 변수만을 포함한 sPLS 모델을 평가함.

- 또한 평균 중심 맞춤 (mean centring)은 기준선 오프셋을 제거했으며, 로그 변환을 통한 변수 범위의 표준화는 X 변수 간의 분산 차이를 줄였으며 차원 축소 중에 상대적 중요성을 감소시킴.

- 추가 선별 접근법으로, 단일 오염 물질, 공변량 보정 OLS 선형 회귀 모형을 15 개의 노출 (> 70 % 검출 빈도)과 22 개의 결과 사이의 330 개 연관성에 대해 테스트함.

- 다중 비교를 보정하기 위해 위발견율 (FDR, False Discovery Rate)을 계산하고 위발견율 <10 %로 유의 수준을 설정함.

- 민감도 분석에서, 검출 및 비 검출 변수로 이분화 된 70 % 미만 검출 빈도를 가진 4 가지 오염 물질 노출을 테스트하였고, 결측값

대체된 교란 변수와 (무작위로 누락 된) 노출 데이터가 있는 참가자를 제외한 모델을 테스트 함.

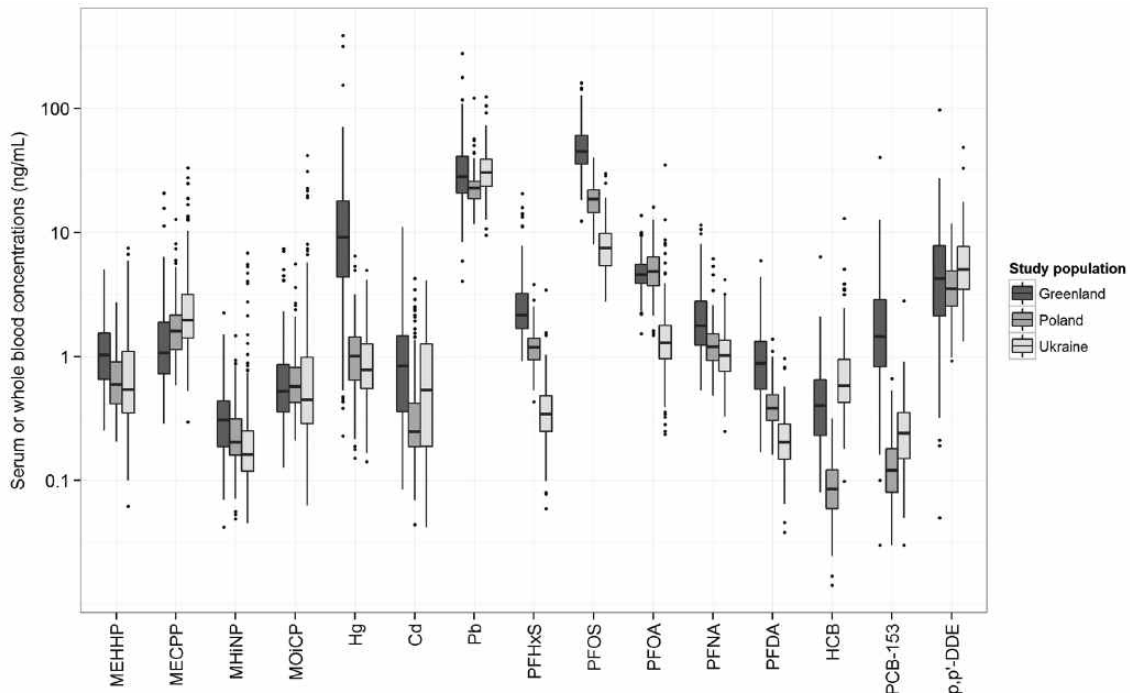
- 또한 공변량으로서 총 지질을 포함하여 지질 비 조정 유기 염소 (lipid-unajusted organochlorines)가 있는 모델을 평가함.

- 마지막으로 인구집단 층화 회귀의 플롯을 평가하여, 인구집단 변수에 대해 상호 작용 (interaction term)이 있는 연구 집단 사이의 효과 추정치에서 이질성 (heterogeneity)을 테스트 함.

- sPLS 및 OLS 회귀 계수는 노출량의 ln 단위 증가 당으로 나타내며, 해석 가능성을 위해서는 노출의 IQR 증가 당 결과의 퍼센트 변화로 나타남; 로그 변환된 결과변수의 경우 비례 변화(proportional change)로, 로그 변화되지 않은 결과변수를 위해서는 평균 결과 수준과 관련된 절대 변화로 나타냄.

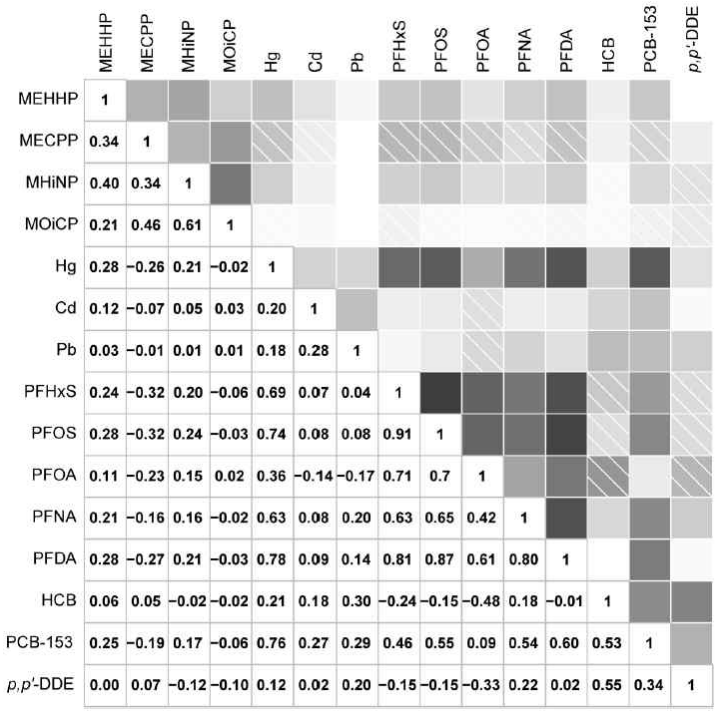
- 통계 분석은 R v.3.0.1 (R for Statistical Computing, Vienna, Austria)을 사용하여 sPLS 모델링을 위한 spls 패키지로 수행됨.

4) 결과



[그림 7-26] 연구 대상자 당 노출 바이오마커 농도 분포의 박스 플롯

MEHHP, mono(2-ethyl-5-hydroxyhexyl) phthalate; MECPP, mono(2-ethyl-5-carboxypentyl) phthalate; MHiNP, mono(4-methyl-7-hydroxyoctyl) phthalate; MOiCP, mono(4-methyl-7- carboxyheptyl) phthalate; Hg, mercury; PFHxS, perfluorhexane sulfonic acid; PFOS, perfluorooctane sulfonic acid; PFOA, perfluorooctanoic acid; PFNA, perfluorononanoic acid; PFDA, perfluorodecanoic acid; HCB, hexachlorobenzene; PCB, polychlorinated biphenyl; DDE, dichlorodiphenyldichloroethylene.



	MEHHP	MECPP	MHiNP	MOiCP	Hg	Cd	Pb	PFHxS	PFOS	PFOA	PFNA	PFDA	HCB	PCB-153	*p,p'*-DDE
MEHHP	1														
MECPP	0.34	1													
MHiNP	0.40	0.34	1												
MOiCP	0.21	0.46	0.61	1											
Hg	0.28	−0.26	0.21	−0.02	1										
Cd	0.12	−0.07	0.05	0.03	0.20	1									
Pb	0.03	−0.01	0.01	0.01	0.18	0.28	1								
PFHxS	0.24	−0.32	0.20	−0.06	0.69	0.07	0.04	1							
PFOS	0.28	−0.32	0.24	−0.03	0.74	0.08	0.08	0.91	1						
PFOA	0.11	−0.23	0.15	0.02	0.36	−0.14	−0.17	0.71	0.7	1					
PFNA	0.21	−0.16	0.16	−0.02	0.63	0.08	0.20	0.63	0.65	0.42	1				
PFDA	0.28	−0.27	0.21	−0.03	0.78	0.09	0.14	0.81	0.87	0.61	0.80	1			
HCB	0.06	0.05	−0.02	−0.02	0.21	0.18	0.30	−0.24	−0.15	−0.48	0.18	−0.01	1		
PCB-153	0.25	−0.19	0.17	−0.06	0.76	0.27	0.29	0.46	0.55	0.09	0.54	0.60	0.53	1	
p,p'-DDE	0.00	0.07	−0.12	−0.10	0.12	0.02	0.20	−0.15	−0.15	−0.33	0.22	0.02	0.55	0.34	1

[그림 7−27] 노출 바이오마크의 짝지은 피어슨 상관계수, 또한 열지도로 표시함.

MEHHP, mono(2-ethyl-5-hydroxyhexyl) phthalate; MECPP, mono(2-ethyl-5-carboxypentyl) phthalate; MHiNP, mono(4-methyl-7-hydroxyoctyl) phthalate; MOiCP, mono(4-methyl-7-carboxyheptyl) phthalate; Hg, mercury; PFHxS, perfluorhexane sulfonic acid; PFOS, perfluorooctane sulfonic acid; PFOA, perfluorooctanoic acid; PFNA, perfluorononanoic acid; PFDA, perfluorodecanoic acid; HCB, hexachlorobenzene; PCB, polychlorinated biphenyl; DDE, dichlorodiphenyldichloroethylene.

<표 7-25> 다중 오염 물질 희소부분최소제곱(sPLS) 회귀 모델링 및 이와 대응하여 단일 오염 물질 보통최소제곱(OLS) 회귀 계수에서 선택한 노출-결과 연관성 및 회귀 계수

Table 2 Exposure–outcome associations and regression coefficients selected from multipollutant sPLS regression modelling and corresponding single-pollutant OLS regression coefficients

		sPLS		OLS		
Outcome	Exposure*	K, η†	β_{sPLS}	β_{OLS}	95% CI	p Value
LH* (IU/L)	*p,p'*-DDE (ng/g)	1, 0.99	0.083	0.083‡,§	(0.031 to 0.135)	0.002¶
Inhibin B (ng/L)	Hg (ng/mL)	1, 0.99	10.79	10.82	(3.90 to 17.73)	0.002¶
Total testosterone (nmol/L)	MHiNP (ng/mL)	1, 0.90	−1.14	−1.15‡	(−1.74 to −0.57)	0.0001¶
Free testosterone (nmol/L)	MHiNP (ng/mL)	1, 0.64	−0.0113	−0.019‡	(−0.032 to −0.007)	0.002¶
	MOiCP (ng/mL)		−0.0091	−0.010‡	(−0.020 to 0.000)	0.043
	Cd (ng/mL)		0.0091	0.012	(0.001 to 0.023)	0.029
Semen volume* (mL)	MEHHP (ng/mL)	1, 0.99	−0.106	−0.106	(−0.167 to −0.045)	0.0008¶
Progressive sperm (%)	PCB-153 (ng/g)	1, 0.99	−3.37	−3.37§	(−5.48 to −1.25)	0.002¶
TUNEL DFI* (%)	MHiNP (ng/mL)	1, 0.99	−0.218	−0.218‡	(−0.332 to −0.104)	0.0002¶
NAG* (mU/ejaculate)	MEHHP (ng/mL)	1, 0.77	−0.163	−0.164	(−0.255 to −0.073)	0.0005¶

Regression coefficients (β) are expressed per ln-unit change in exposure. Refer to online supplementary table S3 for per cent changes and coefficients from models only adjusted for study population. sPLS and OLS models were adjusted for study population, cotinine and for additional potential confounders as specified in table 1; sPLS models by inputting exposures and outcomes 'prestandardised' for covariates (residuals).
*Variable ln-transformed in statistical analyses.
†sPLS tuning parameters: K, the number of components used and η, the degree of sparsity.
‡Interaction p value <0.10 for the cross-product term between exposure and study population (see online supplementary figure S3 for population-stratified regression plots).
§Sensitivity analysis: β_{OLS} (95% CI) for models with organochlorines unadjusted for lipids (ng/mL), and with total lipids (g/L) included as an additional covariate: LH and *p,p'*-DDE, 0.070 (0.017, 0.123); progressive sperm and PCB-153, −3.592 (−5.871 to −1.313).
¶Significant after adjustment for multiple comparisons in single-pollutant OLS regression modelling (FDR <10%).
DDE, dichlorodiphenyldichloroethylene; DFI, DNA fragmentation index; FDR, false discovery rate; Hg, mercury; LH, luteinising hormone; MHiNP, mono(4-methyl-7-hydroxyoctyl) phthalate; MOiCP, mono(4-methyl-7-carboxyheptyl) phthalate; MEHHP, mono(2-ethyl-5-hydroxyhexyl) phthalate; NAG, neutral α-glucosidase; OLS, ordinary least squares; PCB, polychlorinated biphenyl; sPLS, sparse partial least squares; TUNEL, terminal deoxynucleotidyl transferase dUTP nick end-labelling.

- 희소부분최소제곱(sPLS) 모델에서 테스트 된 300 개 이상의 노출-결과 연관 관계 중 8 개 결과를 포함하는 10 개의 연관 관계를 발견함. 다음의 몇몇 연관 관계는 세 연구 집단에서 방향성이 두드러지게 일관되었음: 수은과 inhibin B, 그리고 카드뮴과 테스토스테론 사이의 양방향 연관성; DiNP 대사 산물과 테스토스테론, 폴리염화바이페닐-153 과 진행성 정자 운동성, 그리고 DEHP 대사 산물과 부고환 기능의 지표 인 중성 α-글루코시다제 사이의 역상관 관계.

5) 고찰 및 결론

- 이러한 환경오염물질의 혼합물에 대한 종합 평가는 일부 유기염소 및 프탈레이트가 남성 생식 건강의 일부 지표에 유해한 영향을 미친다는 것을 나타냄.

12. Porru S, Pavanello S, Carta A, Arici C, Simeone C, Izzotti A, Mastrangelo G. Complex relationships between occupation, environment, DNA adducts, genetic polymorphisms and bladder cancer in a case-control study using a structural equation modeling. PLoS One. 2014 Apr 10;9(4):e94566.

1) 배경

- DNA adduct(DNA 부가물)는 발암물질 노출과 발암 초기 단계를 통합하는 측정치로 간주됨. 방광조직 보다 더 접근하기 쉬운 말초혈액 림프구(pheripheral blood lymphocytes, PBL) 는 방광조직의 림프구와 비슷한 수치를 보여줌.

2) 목적

- 이 연구에서 우리는 말초혈액 림프구 DNA 부가물의 형성과 방광암의 위험이 관련이 있는지 여부와 이러한 관계가 유전적 다형성(genetic polymorphisms), 방광암의 환경 및 직업 위험요인들에 의해 어떻게 조절되는지를 조사한다.

3) 방법

(1) 연구자료 및 연구대상자

- 병원기반 환자대조군(hospital-based case-control study) 연구임.

- 연구대상자 포함 기준은 20 세에서 80 세 사이의 Bresia(북부 이탈리아)에 거주하는 남성임.

- 1997 년 7 월부터 2000 년 12 월까지 Bresia 의 두 주요 병원의 비뇨기과에서 새로이 진단받고 조직학적으로 확인된 방광암 환자가 입원함.

- 대조군은 다양한 비뇨기과적 비종양질환(urological non-neoplastic diseases)에 영향을 받은 214 명의 환자들이었으며,

연령(±5 세), 입원 기간 및 입원한 병원을 환자군과 빈도 매칭(frequency matching)하였음.

(2) 노출 평가

- 모든 피험자는 인구통계학적 변수 및 흡연, 커피 및 기타 액체 소비, 식습관, 직업에 대한 정보를 수집하기 위해 입원 중에 설문 조사에 참여함.

- 다환방향족 탄화수소(polycyclic aromatic hydrocarbon, PAHs) 및 방향족 아민(aromatic amines, AAs)에 대한 작업장 노출은 이전 공보에 설명 된 방법론에 따라 산정되었음.

- AAs 및 PAHs 에 대한 누적 노출 지수는 각각 각 직업에서 노출의 길이 (l), 강도 (i) 및 빈도 (f)의 곱 (i × f × l)으로 계산하여, 모든 직업의 노출지수를 더하여 계산하였음.

- 평생 흡연한 담배의 양은 갑년(packyears) 단위로 계산하였음.

- 평생 마신 커피의 시간 가중 평균은 잔/일(cups/day) 단위로 0(전혀 마시지 않는 사람), ≤3, 4, ≥5 잔/일로 기록하였음.

- PAHs 가 포함된 음식, 과일, 큰 잎 채소 및 기타 채소 섭취는 네 가지 범주 (1 회 미만/달, 1 회 미만/주, 1-3 회/주, 3 회 초과/주)로 분류되었음.

- GSTM1, GSTT1, GSTP1, NAT1, NAT2, SULT1A1, XRCC1-3, XPD, CYP1A2, MPO, COMT, MnSOD 및 NQO1 의 유전자형(genotyping)은 Amplification Refractory Mutation System assay 와 GeneAmp PCR System 9700(Applied Biosystems, Italy)을 사용하여 평가하였음.

- PCR 은 효소 소화 및 PCR-RFLP 분석에 의해 수행되었음.

- 단변량 로지스틱 회귀 분석에서 방광암 위험과 관련이 있는 것으로 판명 된 모든 변수는 다변량 비조건부 로지스틱 회귀 모델(multivariate unconditional logistic regression model)에서 강제

실행 된 후 단계별 선택법(backwards stepwise selection) ($p < 0.05$)을 통해 이 중 일부가 보정변수로 선택되었음.

- 방광암 위험은 흡연량, 커피 과다섭취, MnSOD(Val/Val 유전자형)와 함께 유의하게 증가하였고, 큰 잎 채소 소비량 및 MPO(G-463A 동형 접합체 변이 형)로 인해 유의하게 감소하였음.

- AAs 의 누적 노출량은 통계적으로 유의하지 않았지만 실질적인 교란변수 이 때문에 유지되었음.

(3) 건강영향 평가: PBLs 에서 DNA 추출

- 혈액 샘플은 입원 기간 동안 모든 피험자로부터 수집되었으며, 혈액샘플이 수집된 당일에 원심 분리를 한 다음 말초 혈액 림프구(PBL)를 채취하였음.

- 자동화 된 DNA 추출을 위한 프로토콜은 이전에 기술 된 제조자의 지시에 따라 Extragen kit (Talent 사의 Extragen BC)로 수행하였음.

- 특히 최대 10ml 까지 준비된 전혈로 부터 버피 코트(buffy coats) 2.5ml 를 DNA 추출을 위해 사용하였음.

- 일반적으로 정상기증자에서 150-400 μg 의 DNA/extraction 을 추출하였음.

- DNA 부가물은 ^{32}P-Post-labeling 방법으로 분석하였음.

(5) 통계분석 방법

- 구조방정식(structural equation modeling, SEM)에서, 갑년 단위의 흡연 량, 커피 및 채소 소비량, MnSOD 유전자형, MPO 유전자형, 누적 AAs 노출량 (이전 연구에 따른 방광암 위험과 관련된 변수) 및 XRCC1 유전자형(아래 참조)은 외생 변수들(exogenous variables; 회귀 기반 기술에서 예측 변수에 해당)로 사용되었음.

- 가설 1: 방광암 위험과 DNA 부가물은 하나 또는 그 이상의 외생 변수에 의해 영향을 받는 내생 변수(endogenous variables; 결과 변수에 해당)가 될 수 있음.

- 가설 2: 방광암 위험은 DNA 부가물 형성을 통해 간접적으로 영향을 받을 수 있음.

- 두 개의 경쟁 가설은 두 모델의 구조 방정식으로 변환되어 어떤 모델이 관측 된 데이터에 가장 잘 맞는지를 찾고자 함.

- SEM 구조 방정식에는 "점근분포무관법(asymptotic distribution free)"이 적용되었음.

- 왜냐하면 점근분포무관법은 모든 변수들이 정규성 가정을 만족하지 않아도 되므로 주어진 변수(특히, DNA 부가물)를 사용하는 것이 허용되었기 때문임.

- 본 연구는 SEM 의 두 가지 goodness-of-fit statistics(적합도 검정 통계)를 사용함.

(1) "모델 대 포화"(포화 모델은 공분산에 완벽하게 맞는 모델)을 위한 카이 제곱 검정.

(2) 동시 방정식 시스템(simultaneous equation system)의 분석으로부터 얻어진 안정성 지수(stability index).

- SEM 에 필요한 샘플 크기는 모델의 복잡성, 사용 된 평가 방법 및 관측 된 변수의 분포 특성에 따라 달라짐.

- 가장 좋은 방법은 모델의 복잡성 (즉, 외생 변수의 수)과 다음의 경험 법칙을 고려하는 것임 : 최소 비율 5:1; 추천 비율 10:1; 정규 분포가 없는 데이터의 경우 권장 비율은 15:1 임.

- SEM 모델에 8 개의 외생 변수가 사용 된 경우, 120 (= 15 × 8)명의 피험자가 있어야 하지만, 본 연구에서는 실제로 412 명의 피험자를 모집하여 본 연구에서 SEM 분석에 필요한 피험자 수를 만족함.

- 본 연구는 STATA 12로 분석함.

4) 결과

- PBLs DNA 부가물과 방광암 간의 간접적인 경로는 발견되지 않아 가설 2는 입증하지 못함.

- DNA 부가물은 AAs (p = 0.028)에 대한 직업 누적 노출과 양의 상관관계가 있었고, XRCC1 Arg 399 (p <0.006)는 감소된 부가물 수준과 관련이 있었으나 방광암 위험에는 영향을 미치지 않음.

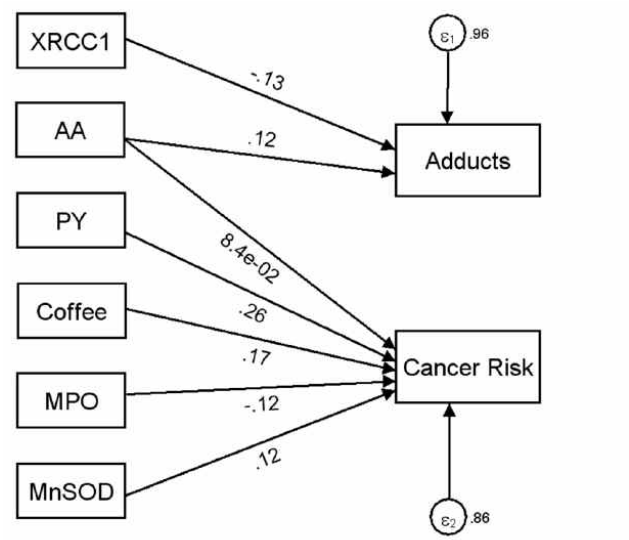
- 이전 연구결과에서 관찰되었던 흡연량 (p <0.001), 커피 섭취 (p <0.001), 누적 AAs 노출 (p = 0.041) 및 MnSOD (p = 0.009)에 의한 방광암 위험 증가와 MPO에 의한 방광암 위험 감소 (p <0.008)가 SEM 분석을 통해 다시 확인되었음.

<표 7-26> 구조방정식 결과: 베타 계수(마이너스는 역 상관관계를 나타냄), 표준 오차, z-검정과 관련 p-value와 95% 신뢰구간

Table 3. SEM results: beta coefficients (with "minus" sign indicating inverse relationship), standard errors, z tests and the corresponding p-values, along with 95% confidence intervals for endogenous variables of structural equations; variances and covariance.

	Endogenous variable	Exogenous variables	Beta Coef.	Std. Err.	z	P>|z|	95% CI	
							Lower	Upper
Structural Equations	Adducts	Packyears	0.064	0.048	1.35	0.177	−0.029	0.157
		Coffee	0.015	0.047	0.31	0.757	−0.078	0.107
		Occupational exposure to AAs	0.117	0.048	2.46	0.014	0.024	0.210
		XRCC1	−0.129	0.047	−2.75	0.006	−0.221	−0.037
		MPO	−0.024	0.039	−0.63	0.531	−0.100	0.052
		MnSOD	−0.053	0.045	−1.17	0.244	−0.142	0.036
	Cancer Risk	Packyears	0.256	0.041	6.19	0.000	0.175	0.337
		Coffee	0.166	0.043	3.83	0.000	0.081	0.250
		AA	0.084	0.034	2.45	0.014	0.017	0.152
		XRCC1	−0.028	0.044	−0.65	0.519	−0.114	0.058
		MPO	−0.115	0.036	−3.21	0.001	−0.185	−0.045
		MnSOD	0.120	0.044	2.71	0.007	0.033	0.206
Variances	Adducts		0.962	0.018			0.927	0.998
	Cancer Risk		0.858	0.025			0.811	0.908
Covariance	Adducts × Cancer Risk		−0.038	0.047	−0.80	0.421	−0.131	0.055

Packyears = Life-long consumption of cigarettes.
Coffee = Life-long time-weighted average of cups/day of coffee.
AA = Occupational cumulative exposure to aromatic amines.
XRCC1 = X-ray repair cross-complementing protein 1.
MPO = Myeloperoxidase (*G-463A* homozygous variant).
MnSOD = Manganese Superoxide Dismutase (*Val/Val* genotype).
doi:10.1371/journal.pone.0094566.t003



[그림 7-28] 결과에 대한 경로 다이어그램(path diagram)

변수(사각형); 변수(원); 원인 흐름(화살표); 경로(화살표 노선). 추정된 베타 계수는 경로를 따라 나타났음.

Packyears = 평생 담배 소비량; Coffee = 평생 시간 가중 평균 컵/일 커피; AA = 방향족 아민에 대한 직업적 누적 노출; XRCC1 = X-ray repair cross-complementing protein 1; MPO = Myeloperoxidase (G-463A homozygous variant); MnSOD = Manganese Superoxide Dismutase (Val/Val genotype).

5) 고찰 및 결론

- 본 연구 결과는 처음으로 DNA 부가물 및 방광암 위험과 방향족 아민(AAs)에 대한 직업 누적 노출 간의 관련성에 대한 근거를 제공하였고, 방광암 발생에 있어서 방향족 아민(AAs)의 중심적인 역할을 강화시켰음.

13. Arlinghaus A, Lombardi DA, Willetts JL, Folkard S, Christiani DC. A structural equation modeling approach to fatigue-related risk factors for occupational injury. Am J Epidemiol. 2012 Oct 1;176(7):597-607.

1) 배경

- 직업 상 부상은 긴 근무 시간 및 짧은 수면 기간을 포함하여 수많은 개인 및 업무 관련 위험 요소와 관련되어 있음.
- 그러나 이러한 부상을 일으키는 복잡한 기전은 아직 완전히 이해되지 않았음.

2) 목적

- 구조 방정식 모델링(structural equation modelin, SEM)을 사용하여 피로와 관련된 직접 및 간접 잠재적 부상(injury) 위험 요인을 조사함.

3) 방법

(1) 연구자료 및 연구대상자

- 연구 표본은 미국 인구를 대표하는 단면적, 다단계, 군집, 계층화된 가중치 표본에 대한 조사 인 미국국민건강면접조사(national health interview survey, NHIS)에서 추출되었음.
- NHIS 에는 미국 인구의 인구 통계 특성, 건강 상태 및 건강 행태에 대한 정보가 포함되어 있으며, 매년 대중에 자료를 공개함.
- 본 연구는 지난 6 년간의 조사 자료 (2004-2009)를 모아 큰 연구 표본을 확보하여 부상을 당한 사례의 수를 늘렸고, 이 6 년간 NHIS 에서 제공된 표본 가중치를 보정하여 유효 인구 추정치(valid population estimates)를 만듦.

- 18 세에서 74 세 사이의 모든 면접 대상자 중 (n = 341,281), 지난 주 동안 임금을 받고 일하였다고 보고한 213,506 명은 주당 근무 시간에 대한 정보를 제공함.

- 이 중 성인 중심 NHIS(Adult Core of NHIS)에서 수면 시간, 건강 상태 및 체질량지수에 대한 정보를 수집하여 샘플 크기가 89,366 개로 줌.

- SEM 분석을 위해 보정변수에 결측 값이 있는 3,049 명의 대상자를 제외하고, 최종적으로 86,137 명을 대상으로 연구를 수행함.

(2) SEM 를 위해 사용된 변수

- 직업적 부상: 참가자들은 의료 처치가 필요한 면담 3 개월 전에 부상이나 중독의 경험이 있었는지 질문 받았음. 직업상의 부상은 분석에서 이분형 종속 변수로 사용되었고 부상 당했을 시점에 그 사람이 어떤 활동에 관여했는지 묻는 질문에 “급여를 받는 직업의 업무 중”이라는 답으로 확인하였음.

- 고용 상태 및 주 중 노동 시간: 각 취업자(면접 이전 주간에 "급여를 받는 직장에서 일하는 중"이라고 대답한 사람을 뜻함)는 지난 주 동안 근무한 총 시간을 보고하였음. 주당 노동 시간은 선행 <20, 21-30, 31-40, 41-50, 51-60 및> 60 시간 범주로 분류되었으며, 분석 시 독립 변수로 사용되었음.

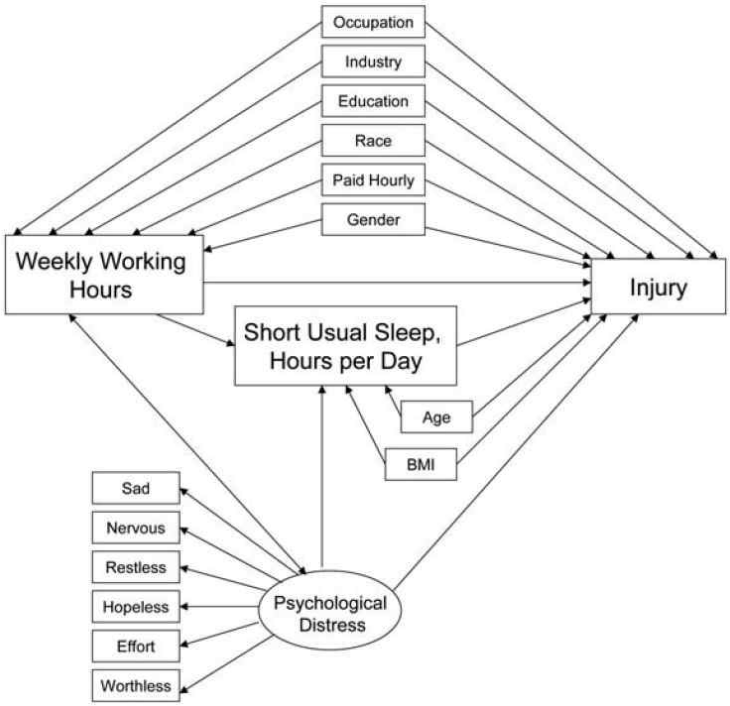
- 수면 시간: 일상적인 수면 지속 시간은 "평균적으로 24 시간 동안 몇 시간의 수면을 취합니까?"라는 질문에 대한 응답으로 측정되었음. 이 변수는 부상 위험에 영향을 주는 두 번째 예측 변수로 포함되었음. 수면 시간이 짧을수록 부상 위험이 증가하는 방향이 되게 추정하도록 다음과 같이 반대로 코딩함; 10-9-9.99, 8-8.99, 7-7.99, 6-6.99, 5-5.99 또는 <5 시간/일. 동시에 이 변수는 종속 변수로 지정되었음.

- 인구 통계 요인, 업무 관련 변수 및 건강 정보: 사회인구학적 요인과 직업 관련 요인들이 주당 업무 시간과 부상 위험 모두에 영향을

주는 통제 변수로 포함됨(성별, 교육, 인종/민족, 직업, 업종, 급여 유형). 산업과 직업은 각각 북미 산업 분류 체계와 표준 직업 분류를 사용하여 NHIS 에 코드화 됨. SEM 추정을 위해, 우리는 각 응답을 이분형 더미 변수(아니오/예)로 변환하였고, 다음 명목 변수의 의미있는 해석을 위해 참조 범주(reference category)로 "아니오"를 사용함: 산업, 직업, 교육 수준, 인종/민족.

- 정신 건강과 수면 및 부상 위험과의 관계를 테스트하기 위해, 심리적 고통(psychological distress)을 구성하기 위한 K6 척도로 측정한 지표를 포함시킴. 이 척도에는 6 가지 질문이 포함되었고, 다음과 같은 질문들로 구성됨. "지난 30 일 동안 당신은 다음과 같은 기분을 얼마나 자주 느꼈나요?": "너무 슬퍼서 기운이 나게 할 수 있는 것이 없었나요?", "초조했나요?", "가만히 있을 수 없었나요?", "희망이 없었나요?", "모든 일이 노력이었나요?", "가치가 없었나요?". 위 설문은 5 점 범위로 "전혀 없었다"부터 "항상 그랬다"까지로 답변되었음. 이러한 질문은 확인 요인 분석(cirfirmatory factor analysis)의 평균을 통해 "심리적 고통"이라는 잠재 변수(latent variable)로 합해짐. 잠재 변수는 관측된 변수(observed variable)로 측정 된 정보를 요약하고 각 관측된 변수가 기본 구성에 미친 영향(요인 가중치)을 설명함. 최근의 프로그래밍 개발로 인해 연구원은 잠재(연속 또는 범주형)변수와 SEM 의 다른 연속 변수들 간의 상관관계를 지정할 수 있음.

- 이전 연구 결과와 이론적 가정을 바탕으로, 본 연구에서는 체질량지수와 연령이 일상적인 수면 기간을 경유하여 부상 위험에 영향을 주는 것으로 가정하였음. 또한 심리적 고통은 직업상 부상과 짧은 수면 기간 모두에 직접적인 위험 요인으로 테스트되어 수면을 통해 직간접적으로 부상 위험에 영향을 미침. SEM 에 지정된 모델은 SEM 변수에 대한 표준 표기법을 사용하여 그림 1 과 같이 표현함.



[그림 7-29] 주당 긴 근무 시간과 짧은 수면 시간으로 인한 산업 재해 위험도를 예측하기 위한 구조 방정식 모델의 상술

인구 통계, 건강 및 업무 관련 특성의 중재(mediating) 및 혼란(confounding) 효과 포함. 이 모델은 6 년(2004-2009)에 걸쳐 국가 건강 면담 조사(National Health Interview Survey)에서 수집 된 고용 노동자의 자체보고 데이터임. 사각형은 관찰된 변수를 나타내며 타원은 잠복변수/구조를 나타냄. 단일 화살표는 방향 효과를 나타내며, 양방향 화살표는 상관 관계를 나타냄.

(3) 통계분석 방법

- 분석 된 모든 변수에 대한 가중 빈도, 백분율 및 평균값은 설문 조사 데이터를 위한 SAS(버전 9.2; SAS Institute Inc., Cary, North Carolina)를 사용하여 계산됨.

- **구조 방정식 모델링:** 로지스틱 회귀 분석을 사용하여 SEM 은 이분형 결과변수인 부상 위험에 대한 모든 예측 변수의 직접 효과(direct effect)를 추정함. 연속 예측 변수들 간의 관계는 선형 회귀 기법을 사용하여 조사됨. 모든 SEM 분석은 강한 표준 오류(robust standard errors)가 있는 가중 최대우도추정(weighted maximum likelihood estimation)을 사용하여, Mplus, 버전 6.1 (Múthen 및 Múthen, Los Angeles, California)으로 수행함. P <0.05 (양측 검정)를 통계적 유의 수준으로 사용 함. 사용된 모델은 NHIS 데이터에 제공된 샘플링 가중치, 클러스터링 및 층화를 포함하여 "Type = Complex"문으로 지정된 NHIS 의 복잡한 조사 디자인을 모델링 한 것임. 간접 효과는 예측변수, 매개변수 및 결과변수 간의 직접 효과를 곱하여 계산함. 결과 변수에 대한 예측 변수의 총 효과를 추정하기 위해, 간접 효과에 대한 매개 변수 추정(parameter estimate)을 결과 변수에 대한 예측 변수의 직접 효과에 더함. 총 효과에 대한 오즈비는 회귀 변수의 거듭제곱에 의해 구해짐. 간접 효과의 표준 오류(통계적 유의성)는 손으로 계산할 수 있지만 Mplus 는 최대 우도 추정을 사용하여 복잡한 샘플에서 총 영향의 통계적 유의성을 예측하는 방법을 아직 제공하지 않음. 따라서, 우리는 가중 최소 제곱 접근법 (즉, 2 항 회귀 분석)으로 두 번째 추정을 사용하여 총 효과(양측 P <0.05)에 대한 통계적 시험을 수행함.

- 모델 적합(model fit): 모델 적합성은 χ^2 검정 또는 표준 적합 지수(normed fit index), 비교 적합 지수(compartive fit index, CFI), 근사 평균 제곱근 오차(root mean square error of approximation)와 같은 모델 적합 지표들을 사용하여 평가할 수 있음. χ^2검정은 큰 표본에서 모형 거부를 산출 할 가능성이 있지만, 위와 같은 모형 적합 지표는 표본 크기와 무관한 정보를 제공함. 그러나 이러한 "전통적"모델 적합 지표는 최대 우도 추정과 함께 로지스틱 회귀 분석을 포함한 모델에 적용 할 수 없음. 이 모델에 사용할 수있는

유일한 적합 지표는 AIC (Akaike Information Criterion) 및 BIC (Bayesian Information Criterion)이며 모델의 매개 변수 개수와 모델의 우도 함수에 따라 달라짐. 따라서 결과는 "좋음" 또는 "나쁨"과 같이 모델 적합을 평가하는 결과로 해석 될 수 없음. 따라서 모델 적합성에 대한 전반적인 테스트는 그러한 모델에 사용할 수 없지만 AIC 및 BIC 를 사용하여 서로 다른 모델을 테스트 할 수 있음. 이 경우 가장 작은 AIC 및 BIC 가있는 모델을 선호하는 것이 좋음. 이러한 제한점을 극복하기 위해서 2 가지 다른 방법을 사용하여 모델을 적합(fitting)시킴: 이분형 회귀분석(binomial regression)을 수행하기 위해 견고한 표준오차(robust standard error)를 가진 가중 최소제곱 근사법(weighted least squares approximation)을 사용하였는데, 이 방법은 근사값의 평균제곱근오차(mean square eror of approximation), CFI, Tucker-Lewis 지표(TLI)와 같은 전통적인 모델 적합 지표를 산출하였음. 로지스틱 회귀분석을 포함한 SEM 의 매개변수 계수(parameter coefficient)를 얻기위해, 견고한 표준오차(robust standard error)를 가진 최대 우도 추정(maximum likelihood estimation)을 사용하는 두 번째 방법을 사용함.

4) 결과

<표 7-27> 2004-2009 년국가 건강 면담 조사(n=89,366)에서 근로자 표본의 인구학적 특징

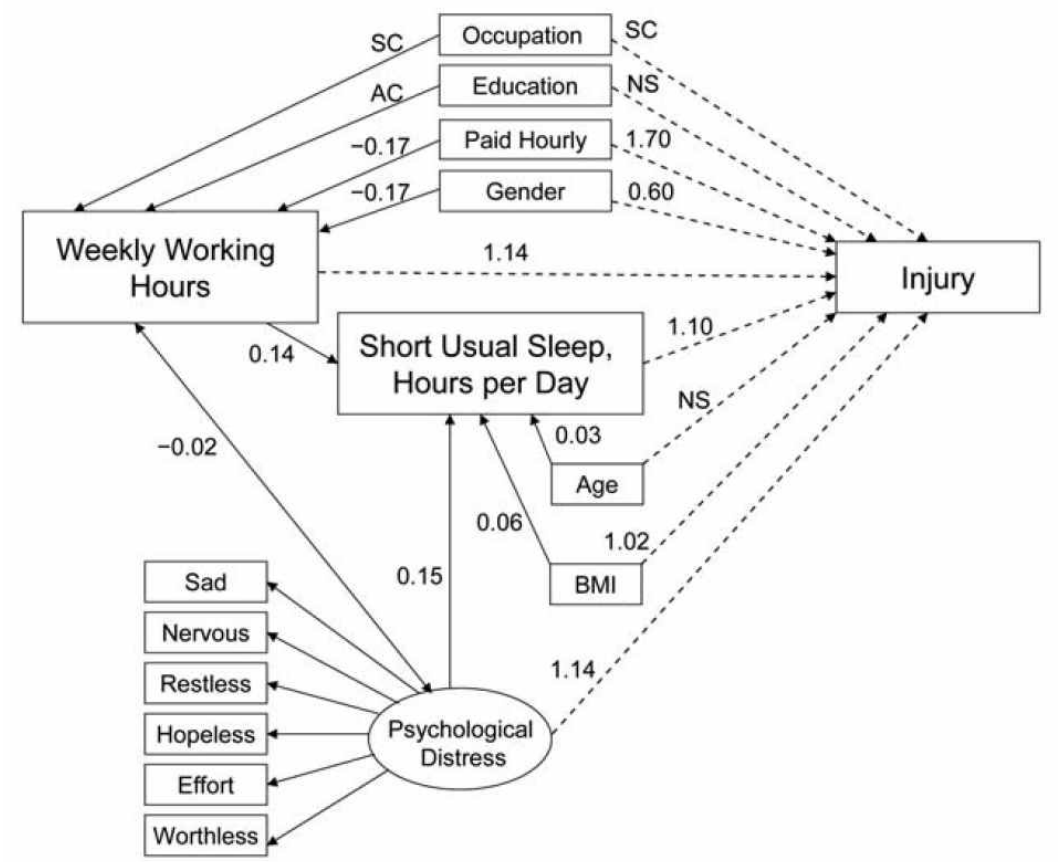
Characteristic	No. of Subjects Interviewed	Weighted %
Gender		
Men	44,473	50.8
Women	44,893	49.2
Highest level of education		
Did not complete high school	11,088	9.8
High school graduate	23,285	25.8
Some college	27,611	31.9
Bachelor's degree	17,865	21.6
Master's degree	6,664	8.0
Professional school degree or doctoral degree	2,405	2.9
Race/ethnicity		
Hispanic	16,893	11.9
Non-Hispanic white	54,339	71.3
Non-Hispanic black	13,133	12.1
Non-Hispanic other	5,001	4.7

[a] Numbers may not sum to total because of missing data.

<표 7-28> 2004-2009 년국가 건강 면담 조사(n=89,366)에서 근로자 표본의 작업-관련 요인, 산업 부상 발생, 수면 시간

Characteristic	No. of Subjects Interviewed	Weighted %
Hours worked per week		
≤20	8,119	9.4
21–30	7,362	8.3
31–40	48,633	52.5
41–50	14,980	17.7
51–60	6,989	8.2
>60	3,283	3.9
Paid hourly		
Yes	52,549	57.4
No	36,196	42.6
Occupation		
Management	7,665	9.5
Education, training, and library	5,165	6.3
Health-care practitioners and technical service	4,582	5.5
Food preparation and serving-related jobs	4,549	5.0
Building and grounds cleaning and maintenance	3,889	3.8
Sales and related jobs	9,011	10.8
Office and administrative support	12,263	14.1
Construction and extraction	5,130	5.6
Production	6,022	6.5
Transportation and material moving	4,970	5.5
Other[a]	23,268	27.0
Injury at work		
Yes	639	0.7
No	88,727	99.3
Daily duration of sleep, hours		
≥10	1,475	1.6
9–9.99	2,815	3.1
8–8.99	26,695	29.3
7–7.99	30,965	35.2
6–6.99	20,537	23.1
5–5.99	5,194	5.8
<5	1,685	1.9

[a] Other occupations included business and financial operations; installation, maintenance, and repair; personal care and service; computer and mathematical work; health-care support; protective services; arts design, entertainment, sports, and media; architecture and engineering; community and social services; legal; life, physical, and social sciences; farming, fishing, and forestry; and military.



[그림 7-30] 주당 긴 근무 시간과 짧은 수면 시간으로 인한 산업 재해 위험을 예측하기 위한 구조 방정식 모델 결과

인구학적, 건강, 작업-관련 특징의 매개와 교란 영향을 포함함. 모형은 자기-보고 자료로 2004-2009 년(6 년)동안 국가 건강 면담 조사임. 실선은 선형회귀분석을 나타냄(모수 계수는 베타 계수); 점선은 로지스틱 회귀 분석(모수 계수는 오즈비)임. “주간 근무 시간”은 <20, 21-30, 31-40, 41-50, 51-60 또는 >60 으로 분류되었음; “짧은 평소 수면” 하루 10 시간이상부터 <5 시간으로 1 시간 간격으로 분류함; “직업”에는 관리, 교육/훈련/도서관, 건강관리 실무자/기술 서비스, 음식 준비/서빙 관련 작업, 건물 및 부지 청소 및 유지 보수, 판매 및 관련

작업, 사무실 및 행정 지원, 건설/추출, 생산 및 운송/자재 이동(참조 카테고리: “아니오”); “교육”에는 고등학교 이하, 고등학교 졸업, 전문대학, 학사, 석사, 박사 학위(참조 카테고리: “아니오”)가 포함되었음. “성별”에 대한 참조 범주는 남성이었음. “시간당 지불”에 대한 참조 범주는 “시간당 지불되지 않음”이었음; 나이와 체질량지수는 연속 변수임. (AC, 모든 범주가 유의함, NS, 유의하지 않음, SC, 선택된 변수가 유의함)

<표 7-29> 짧은 수면 시간 및 주당 근무시간과 산업 재해 위험도에 대한 관계를 예측하기 위해 구조 방정식 모델에서 선형 회귀를 사용하여 추정된 경로 계수

Model	Parameter Estimate (*b*)	Standard Error for *b*	β	95% Confidence Interval for β
Sleep Duration (Hours, Ranging From >10 Hours/Day to <5 Hours/Day)				
Weekly working hours (10-hour intervals ranging from <30 hours/week to >60 hours/week)	0.13***	<0.01	0.14	0.13, 0.14
Age, years	<0.01***	<0.01	0.03	0.02, 0.03
Body mass index[a] (continuous)	0.01***	<0.01	0.06	0.05, 0.07
Psychological distress[b]	0.07***	<0.01	0.15	0.14, 0.16
Weekly Working Hours (10-Hour Intervals Ranging From <30 Hours/Week to >60 Hours/Week)				
Occupation (yes vs. no)				
Management	0.30***	0.02	0.08	0.07, 0.09
Education, training, and library	−0.22***	0.03	−0.05	−0.06, −0.04
Health-care practitioners and technical service	0.14***	0.02	0.03	0.02, 0.04
Food preparation and serving-related jobs	−0.28***	0.02	−0.05	−0.06, −0.05
Building and grounds cleaning and maintenance	−0.28***	0.02	−0.05	−0.05, −0.04
Sales and related jobs	−0.02	0.02	−0.01	−0.01, 0.01
Office and administrative support	<0.01	0.01	<0.01	−0.01, 0.01
Construction and extraction	−0.01	0.02	−0.01	−0.01, 0.01
Production	0.28***	0.02	0.06	0.05, 0.07
Transportation and material moving	0.16***	0.02	0.03	0.02, 0.04
Paid hourly (yes vs. no)	−0.39***	<0.01	−0.17	−0.18, −0.16
Education (yes vs. no)				
Did not complete high school	0.08**	0.07	0.02	−0.01, 0.05
High school graduate	0.19**	0.06	0.07	0.03, 0.11
Some college	0.14*	0.06	0.06	0.01, 0.10
Bachelor's degree	0.18**	0.06	0.07	0.03, 0.10
Master's degree	0.19**	0.07	0.05	0.02, 0.07
Professional school degree or doctoral degree	0.39***	0.07	0.06	0.04, 0.08
Gender (female vs. male)	−0.39***	0.01	−0.17	−0.18, −0.16

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.
[a] Weight (kg)/height $(m)^2$.
[b] Latent variable measured by means of the observed variables “sad,” “nervous,” “restless,” “hopeless,” “effort,” and “worthless” (see text).

국가건강면담조사(National Health Interview Survey)(n=86,317), 2004 - 2009

<표 7-30> 산업 재해 위험도와 다양한 요소의 직접, 간접, 총 효과를 예측하기 위해 구조방정식 모델에서 로지스틱 회귀를 사용하여 추정된 경로 계수와 오즈 비

Model	Direct Effect				Indirect Effect[a]		Total Effect[a]	
	Parameter Estimate (*b*)	Standard Error for *b*	OR	95% Confidence Interval for OR	Parameter Estimate (*b*)	OR	Parameter Estimate (*b*)	OR
Sleep duration (hours, ranging from >10 hours/day to <5 hours/day)	0.10*	0.05	1.10	1.02, 1.19				
Weekly working hours (10-hour intervals ranging from <30 hours/week to >60 hours/week)	0.13**	0.04	1.14	1.06, 1.22	0.01*	1.01	0.14*	1.15
Gender (female vs. male)	−0.52***	0.11	0.60	0.50, 0.71	−0.05*	0.95	−0.56*	0.57
Age, years	−0.01	<0.01	1.00	0.99, 1.00	0.01*	1.00	−0.01	0.99
Education (yes vs. no)								
Did not complete high school	0.57	1.02	1.76	0.33, 9.41	0.01	1.01	0.58	1.79
High school graduate	0.58	1.02	1.79	0.33, 9.65	0.02*	1.02	0.60	1.82
Some college	0.77	1.01	2.15	0.41, 11.3	0.02	1.02	0.79	2.20
Bachelor's degree	0.28	1.01	1.33	0.25, 7.03	0.02*	1.02	0.30	1.35
Master's degree	−0.25	1.04	0.78	0.14, 4.31	0.02*	1.02	−0.23	0.79
Professional school degree or doctoral degree	−0.11	1.10	0.90	0.15, 5.51	0.05*	1.05	−0.06	0.94
Occupation (yes vs. no)								
Management	−0.31	0.20	0.74	0.53, 1.03	0.04*	1.04	−0.27	0.76
Education, training, and library	0.43*	0.23	1.54	1.06, 2.24	−0.03*	0.97	0.40	1.49
Health-care practitioners and technical service	0.50*	0.21	1.65	1.16, 2.35	0.02*	1.02	0.52*	1.68
Food preparation and serving-related jobs	0.07	0.24	1.07	0.72, 1.59	−0.04*	0.96	0.03	1.03
Building and grounds cleaning and maintenance	0.42*	0.22	1.52	1.05, 2.20	−0.04*	0.96	0.38	1.46
Sales and related jobs	−0.21	0.20	0.81	0.58, 1.12	0.00	1.00	−0.21	0.81
Office and administrative support	−0.28	0.18	0.75	0.56, 1.01	0.00	1.00	−0.28	0.76
Construction and extraction	0.69***	0.16	1.99	1.52, 2.60	0.00	1.00	0.69	1.99
Production	0.48**	0.17	1.62	1.22, 2.15	0.04*	1.04	0.52*	1.68
Transportation and material moving	0.29	0.18	1.34	0.99, 1.80	0.02*	1.02	0.31	1.36
Paid hourly (yes vs. no)	0.53***	0.11	1.70	1.41, 2.04	−0.02*	0.98	0.51*	1.67
Body mass index[b] (continuous)	0.02**	0.01	1.02	1.01, 1.03	0.01*	1.00	0.02*	1.02
Psychological distress[c]	0.14***	0.02	1.14	1.10, 1.19	0.01*	1.01	0.15*	1.16

Abbreviation: OR, odds ratio.
* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.
[a] Statistical significance was calculated using a structural equation model with weighted least squares estimation.
[b] Weight (kg)/height (m)^2.
[c] Latent variable measured by means of the observed variables "sad," "nervous," "restless," "hopeless," "effort," and "worthless" (see text).

국가건강면담조사(National Health Interview Survey)(n=86,317), 2004 - 2009

- 긴 근무 시간과 짧은 수면 시간은 독립적으로 부상의 위험을 증가시켰음.
- 추가적인 직접 위험 요소로는 성별, 직업, 급여 유형 및 체질량지수가 있었음.
- 동시에, 긴 노동 시간, 높은 심리적 고통 및 높은 체질량지수는 수면 지속 기간을 경유하여 부상의 위험을 증가시킴.

5) 고찰 및 결론

- 이러한 결과는 구조방정식 모형(SEM)이 복잡한 표본에서 이분형 결과 및 간접 효과를 조사 할 때 유용한 접근법임을 나타내며, 본 연구 결과는 포괄적이고 새로운 부상 예측 모델을 제공함.

14. Valeri L, Mazumdar MM, Bobb JF et al. The joint effect of prenatal exposure to metal mixtures on neurodevelopmental outcomes at 20-40 months of age: Evidence from Rural Bangladesh. Environ Health Perspect. 2017; 125(6): 067015. doi: 10.1289/EHP614.

1) 배경

- 화학 혼합물에 대한 노출은 복잡한 영향을 평가할 수 있는 새로운 통계적 방법을 필요로 하며, 모든 인구집단에서 실제-삶의 시나리오로 인식됨. 이전 연구는 단면연구가 대부분임

2) 목적

- 아동의 신경 발달에 대한 비소(arsenic), 망간(manganese) 및 납(lead)에 대한 자궁 노출의 공동 효과를 평가하고자 함.

- 상호작용, 비선형효과 고려

3) 방법 : 전향적 코호트 연구

(1) 연구자료 및 연구대상자

- 2008 년과 2011 년 사이에 임신 한 여성을 대상으로 전향적 연구에 참여하여 만성 저농도 비소 노출이 생식 건강 결과에 미치는 영향을 조사했던 연구이고, 2010-2013 년에 후속 연구에 참여하도록 권유했음. 참가자는 방글라데시의 Sirajdikhan 과 Pabna Sadar upazilas 의 Dhaka Community Hospital Trust (DCH)가 운영하는 두 개의 농촌 보건 진료소에서 모집되었음.

- 이 분석의 연구 기반은 생식 코호트 연구 당시 태어난 모든 20-40 개월의 어린이들로 구성되었음.

- 1,613 명 여성 중 재대혈 및 신경발달점수가 모두 가능한 표본은 825 명이었음

(2) 노출 평가

- 제대혈 비소, 망간 및 납 농도를 임신 후기의 생체 표지 물질로 사용했음

- 동적 반응 세포-유도 결합 플라스마 질량분석기(dynamic reaction cell-inductively coupled plasma mass spectrometer) (Perkin Elmer)로 금속 농도를 분석함.

(3) 건강영향 평가

- 영유아 발달 베일리 척도(Bayley Scales of Infant and Toddler Development, BSID-III TM)(Bayley 2006)를 벵갈어(Bengali)로 번역하여 방글라데시 농촌 지역에서 사용할 수 있도록 했음.

- 두 가지 주요 결과는 각 참가자에 대하여 인지 및 언어 발달의 원시 점수를 합산함으로써 유도되었음. a) 원시 인지 발달 점수 (cognitive development score, CS); b) 원시 언어 발달 종합 점수 (language development composite score, LCS; 표현 및 수용 척도 원시 점수의 합).

- 질관리를 위해 비디오로 신경발달 평가를 측정함.

(4) 보정변수

- 어머니 연령, 교육, 흡연력, 사회경제적 지위(HOME score), 산모의 IQ, 임신 중 단백질 섭취량

- 아기의 성별, 신경발달 검사 시 연령, 출생 체중, 키, 머리둘레, 영양상태

- 2 차(quadratic) 효과(effect): 아이의 검사 시 연령, 산모의 IQ, 어머니 연령, HOME score

(5) 통계분석 방법

- 망간, 납, 비소는 자연로그변환하였고 연속변수로 모델링했음. 이는 오른쪽으로 왜곡되어 있음(right-skewedness)을 의미함.

- 다변수 선형 회귀:

$$Y_i = \beta_0 + \beta_1 As_i + \beta_2 Mn_i + \beta_3 Pb_i + \boldsymbol{\beta}^T \mathbf{Z_i} + e_i, \qquad (1)$$

Y 는 신경발달 점수; Mn, As, Pb; $Z=Z_1,..., Z_p$ 는 잠재적인 교란요인임.

- 베이지안 커널 기계 회귀(Bayesian kernel machine regression, BKMR): 커널 함수를 사용하는 방법임. BKMR 의 핵심 특징은 노출-반응 함수의 형태가 유연하게 모델링되어서, 사전변수(prior)가 (예: 2 차 항을 사용하여) 지정 될 필요가 없다는 점임.

$$Y_i = h(As_i, Mn_i, Pb_i) + \boldsymbol{\beta}^T \mathbf{Z_i} + e_i \qquad (2)$$

- 함수 h()는 비선형성 및/또는 혼합물 구성 요소간의 상호 작용을 수용하는 노출-반응 함수임. 일반적으로 M 금속의 경우 $(1 + DF)^M - 1$ 매개 변수임. (cf. 자유도, degree of freedom, DF=3). BKMR 은 h()에 대한 커널 노출-반응 기계 표현을 사용하여 노출-반응 함수의 잠재적으로 고차원적인 매개 변수 공간을 처리함. 커널 머신 표현은 추정 된 노출-반응 함수를 정규화하고 몇 가지 계산 이점을 제공하는 선형 혼합 모델 (Liu et al., 2007)로 모델 (2)을 공식화 함 (Bobb et al. 2015). 직관적으로, 결과적인 선형 혼합 모델은 비슷한 노출 프로파일 (예 : z_i = (As_i, Mn_i, Pb_i)의 유사한 값)을 갖는 두 각각이 유사한 건강 영향을 예측한다고 가정함 (예 : h_i = h (z_i) 이고 각 i 와 j 에 대해 h_j = h (z_j)에 가까울 것임). 여기서 각각의 노출 프로파일 간의 유사성 (또는 거리)은 커널 함수를 사용하여 측정됨.

- BKMR 모델 피팅(fitting)은 커널 함수의 선택에 달려있음. 우리는 가우시안(Gaussian) 커널을 사용했는데, 이는 h()에 대한 다양한 기본 함수 형태를 유연하게 포착하고 현실적인 노출-반응 시나리오에 기반한 시뮬레이션 연구에서 잘 작동하는 것으로 나타났음 (Bobb et al., 2015). 피팅을 하면 노출-반응 함수와 불확실성의 추정치가 산출됨.

- 일반화된 부가 모델(generalized additive models, GAMs): BKMR 결과의 재현성을 평가하고 혼합물 성분들 간의 잠재적인

상호작용(interaction)을 시험하기 위해, 이어서 다변수 일반화된 부가 모델을 피팅하였음. GAM 은 용량-반응 곡선을 유연하게 모델링하는 대안적인 접근법이며 저-차원 설정에서 BKMR 의 사전 가정의 유효성을 평가하는 데 유용할 수 있음. 우리는 GAM 의 두 사양을 비교했음. 하나는 중금속 각각이며, 두 번째는 비소와 망간의 통합 부드러운 효과임.

- 3 가지 민감도 분석: a) 이상치가 있거나 없는 모델을 피팅하여 극한 관측치의 영향을 평가했음. b) 또한 부드러움(smoothness, 평탄도)의 다른 정도를 허용하는 2 가지의 선행 가정을 사용하여 BKMR 에 대한 선행 특정화의 영향을 평가했음 (보충 자료, BKMR 의 민감도 분석 참조). c) 마지막으로 Rodrigues et al.(2016)에서 고려한 20-40 개월의 중금속에 대한 어린이 노출을 추가로 보정하여 결과의 견고성을 평가했음. 그룹화된 변수 선택을 하였음. 이는 노출의 가장 중요한 시간 창 (산전과 산후)을 구분할 수 있게 함.

$$Y_i = h[\text{Group}_1 = (As_{i,\text{pre-natal}}, Mn_{i\,\text{pre-natal}}, Pb_{i\,\text{pre-natal}}), \text{Group}_2 = (As_{i\,\text{post-natal}}, Mn_{i\,\text{post-natal}}, Pb_{i\,\text{post-natal}})] + \boldsymbol{\beta}^T \mathbf{Z_i} + e_i$$

- R(version 3.1.3; R foundation for Statistical Computing)

4) 결과

- Sirajdikhan 과 비교해서 Pabna 어린이는 출생 체중이 낮고 CS 가 낮았음. 제대혈 망간과 비소 농도도 Pabna 에서 유의하게 높았음. Pabna 클리닉의 참가자들 사이에서만 금속 농도가 유의미한 상관관계를 보였으므로 금속 원천이 두 장소에서 다를 수 있음을 나타냄(Figure 1). 제대혈 납 농도는 Sirajdikhan 에서 높았음. 2 차 분석을 위해 클리닉 별로 층화하였음.

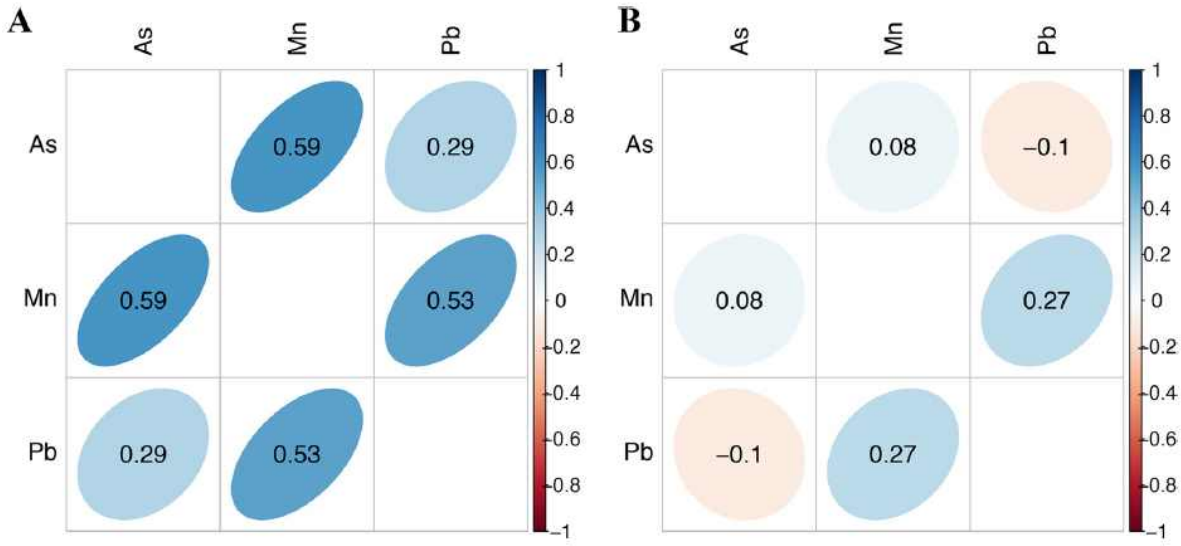


Figure 1. Correlation (Pearson) matrix for maternal and cord blood metals. (*A*) Pabna (*B*) Sirajdikhan.

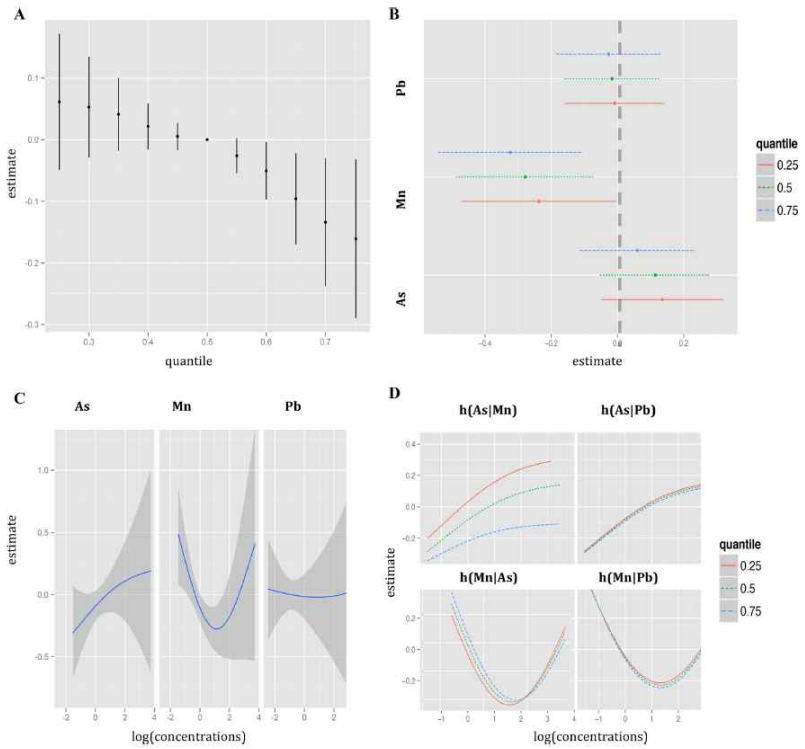
[그림 7-31] 모성과 제대혈 금속 상관관계 매트릭스

(A) Pabna (B) Sirajdikhan

- 잠재적 혼란 요인 중 아동의 연령, 클리닉, 모성 교육은 신경 발달 결과와 가장 강하게 연관되어 있었음. 단백질 섭취는 CS 와 LCS 와 유의한 연관성이 있었음. HOME 척도는 CS 와 유의미하게 관련이 있었음. 보정 후 Pabna 의 망간 농도의 베타만 유의하게 연관있었음.

- Pabna 에서 비소와 납을 보정하였을 때 CS 점수는 로그 망간 농도의 사분위 범위 (interquartile range, IQR) 당 0.2 SD 씩 감소 할 것으로 예상됨(data not shown).

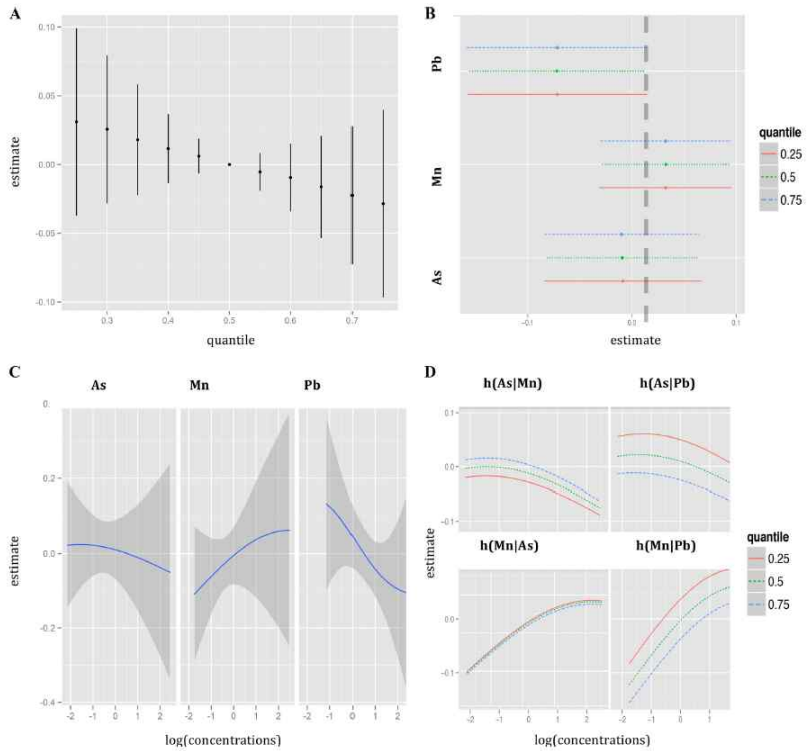
- 베이지안 커널 기계 회귀 분석: 세 금속의 공동 노출-반응 함수를 추정함. 혼합 효과에 대한 추론은 각 혼합 성분의 수준 변화와 연관된 CS 의 변화에 대한 사후 평균 추정치 (및 95 % 사후 신뢰 구간(posterior credible interval))를 계산하여 얻었음. (그림 2-3 및 그림 S1-S4; 표 3 이러한 결과를 요약함)



[그림 7-32] 베이지안 커널 기계 회귀(Bayesian Kernel Machine Regression, BKMR)로 추정한 Pabna 클리닉에서 인지 점수 (CS)에 대한 혼합물의 공동 효과.

아동 성, 모성 IQ, 모성 교육, 모성 단백질 섭취량, 흡연 환경, 시험시 연령 및 모성 나이를 보정하였음. (A) 혼합물의 전반적인 효과 (추정치와 95 % credible interval). 이 그림은 노출이 모두 50 백분위 수에 해당 할 때와 비교하여 노출이 특정 백분위 수 (x 축)에 있을 때 신경 발달 점수의 예상 변화를 나타냄. (B) 단일 오염물질 연관성 (추정치와 95 % credible interval, 회색 점선은 null 에 있음). 이 플롯은 다른 모든 노출이 25, 50 또는 75 백분위 수로 고정 된 경우 단일 오염 물질이 75 번째 대(vs) 25 번째 백분위 수에 있을 때 신경 발달 점수를 비교함. - 망간이 중요한 영향을 미치는 금속임 (C) 단변량 노출-반응

함수 및 각 오염 물질의 중앙값에 고정 된 각 금속에 대한 95 % 신뢰 대역(confidence bands). - 망간과 비소의 비선형성 (D) 이변 량 노출-반응 함수: 망간이 25, 50 또는 75 백분위 수에 고정되고 납은 중앙값 (상단 왼쪽 패널)에 고정되는 경우의 비소; 납이 25 번째, 50 번째 또는 75 번째 백분위 수에 고정되고 망간이 중앙값 (오른쪽 상단 패널)에 고정되는 경우의 비소; 비소가 25, 50 또는 75 백분위 수에 고정되고 납이 중앙값 (하단 왼쪽 패널)에 고정되는 경우의 망간; 납이 25, 50 또는 75 백분위 수에 고정되고 비소가 중앙값 (오른쪽 하단 패널)에 고정되는 경우의 망간. [percentiles: (0.25,0.5,0.75); As=(0.56,0.88,1.60), Mn=(4.80,7.12,17.5), Pb=(1.15, 1.63, 2.42)]. - 비소와 망간의 상호작용이 추정되지만 통계적으로 유의하지는 않았음.



[그림 7-33] Sirajdikhan 클리닉에서 인지 점수 (CS)에 대한 혼합물의 공동 효과.

그림 31 과 같음. [percentiles: (0.25,0.5,0.75); 비소(As)=(0.26,0.40,0.60), 망간(Mn)=(3.93,5.14,6.66), 납(Pb)=(3.87,6.12,9.67)]. - 선형이고 비상호작용인 것으로 보임

<표 7-31> 금속 혼합물 (비소, 망간, 납)과 인지 점수(CS)의 공동 연관에 대한 결과 요약

Table 3. Descriptive summary of results on the joint association of the metal mixture (arsenic, manganese, lead) and cognitive score.

Site	Arsenic	Manganese	Lead	Overall mixture
Pabna	Suggestive evidence of protective main effect[††]; suggestive evidence of interaction with manganese when lead is at the median (Figures 2B and 2D)	Significant nonlinear toxic main effect when lead and arsenic above 60th percentile (Figure 2B)	No main effect (Figure 2B)	Significant toxic effect (Figure 2A)
*Sirajdikhan	No main effect (Figure 3B)	No main effect (Figure 3B)	Suggestive evidence of linear main effect at all levels of arsenic and manganese (Figure 3B)	Not significant (Figure 3A)
Combined Sites	Protective main effect when lead and arsenic are below the median[††] (Figure S1b); potentiates the toxic effect of manganese (Figure S1d)	Significant nonlinear effect at all levels of arsenic and lead (Figure S1b)	Significant linear effect when manganese and arsenic are at the median (Figure S1b)	Suggestive evidence of toxic effect (Figure S1a)

*Mean Blood Pb = 6μg/dl, which may have overwhelmed mixture in stratified analysis.
††This result might be due to residual unmeasured confounding.

- 클리닉 별로 보정한 회귀 분석: Pabna 지역이 대상자 수가 더 많아서 영향을 미치는 것으로 추정되며, GAM 모델에서도 비슷한 결과를 보였음.

- 노출 프로파일의 차이로 인해 클리닉 별로 계층화하여 분석하였음. 높은 망간 농도 (4 분위수 범위 (interquartile range, IQR) : 4.8, 18 μg/dl)를 보이는 Pabna 지역에서 비소, 납 및 망간의 혼합물이 인지 점수에 미치는 통계적으로 유의한 부정적 영향을 보였음. 제대혈 금속 농도가 중위수와 비교하여 (As = 0.5 μg/dl, Mn = 5.8 μg/dl, Pb = 3.1 μg/dl) 모든 값이 60 번째 백분위 수 (As ≥ 0.7 μg/dl, Mn ≥ 6.6 μg/dl, Pb ≥ 4.2 μg/dl)이상인 경우임. 망간의 비선형 효과에 대한 증거가 발견되었음(Figure 2). 비소와 망간이 중앙값에 있을때 로그 망간의 25 번째에서 75 번째 백분위 수의 변화는 인지 점수가 -0.3 (-0.5, -0.1) 표준편차 감소와 연관이 있었음. 연구결과 비소가 망간 독성을 증대시킬 수 있음을 시사하였음.

- 민감도 분석: 비소-망간 상호작용이 여전히 유의했음. 민감도 분석은 제대혈 금속 노출이 20-40 개월의 신경발달에 가장 중요한 동인임을 나타냄.

- 장점 : 비선형성, 부가 효과를 정량화하고 시각화함. 많은 회귀 모델을 피팅할 때 단일 금속 효과 추정, 모델 오정렬 및 잘못된 발견 증가와 같은 기본 분석의 중요한 한계를 극복했음.

- 단점 : 남아있는 잠재적 혼란이 편견을 유발할 수 있음. 하지만 클리닉 별로 계층화된 분석은 이질성을 감소시켰음.

5) 고찰 및 결론

- 화학 혼합물의 건강 영향 연구를 위한 새로운 통계 방법을 사용하여, 비소와 망간 사이의 잠재적인 상승 작용뿐만 아니라 혼합물의 신경 독성의 증거를 발견했음.

15. 이슬비. 임신 중 복합 환경유해물질 노출이 6개월 영유아 아토피 피부염 발생에 미치는 영향. 이화여자대학교. 2019

1) 배경

- 환경유해물질 노출에 가장 민감성이 높은 시기를 확인하는 연구디자인이 필수적임

2) 목적

- '산모·영유아의 환경유해인자 노출 및 건강영향 연구'에 참여한 산모아 출생아를 대상으로 6개월 시기 아토피 피부염 발생에 중요한 영향을 미치는 환경유해물질을 시기별로 확인하고, 이들의 복합 노출이 출생 후 6개월 시기의 아토피 피부염 발생에 미치는 요인을 확인하고자 함

3) 방법 : 전향적 코호트 연구

(1) 연구자료 및 대상자

- Mothers and Children's Environmental Health(MOCEH) 연구
- 302쌍의 엄마와 아이로 구성되어있음.

(2) 노출평가

- 환경유해물질로는 중금속 3종(납, 수은, 카드뮴), 비스페놀 A(BPA), 프탈레이트 대사체 3종(ME-HHP, MEOHP, MnBP), 대기오염물질(NO2, PM10, PM2.5)을 포함하여 총 10가지 물질

(3) 건강영향평가

- 6개월 시기의 아토피 피부염 여부

(4) 보정변수(Covariate, 공변량)

- 성별

(5) 통계분석 방법

- 먼저 환경유해물질의 단일 노출과 아토피 피부염과의 연관성을

확인한다. 통계 분석 모형은 중요한 물질 선택과 연관성의 크기를 동시에 평가하기 위해 탄성그물회귀(Elastic-net regression, ENR)를 이용하였음. 또한, 선택된 유해물질들의 누적 노출 영향을 평가하기 위해 부분최소제곱-판별 분석(Partial least square-discriminant analysis, PLS-DA)을 이용하였으며, 비선형 연관성과 교호 작용 효과를 반영한 누적 노출 영향을 평가하기 위해 베이지안 커널 기계 회귀(Bayesian kernel machine regression, BKMR)를 이용하였음.

4) 결과

- 환경유해물질의 단일 노출과 6개월 시기 아토피 피부염 발생의 연관성을 확인하였을 때 임신 말기 납, MEHHP, MEOHP, PM10 노출이 6개월 시기 아토피 피부염과 유의한 연관성을 나타냈음.

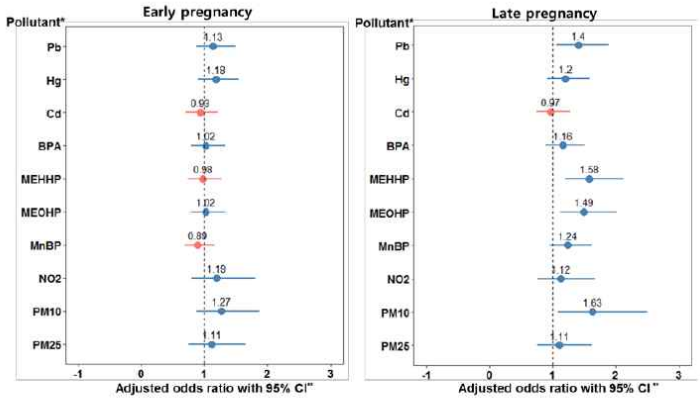


Figure 5. Association between prenatal exposures and atopic dermatitis in 6-month-infants based on multiple logistic regression.

[그림 7-34] 다중 로지스틱 회귀분석을 기반으로 한 산전 노출과 6개월 유아의 아토피 피부염의 연관성

- 다음으로 총 10가지 환경유해물질을 이용하여 ENR 분석을 한 결과, 임신 초기에는 납, 수은, MnBP, PM10 노출이 선택되었으며 임신 말기에는 납, 수은, MEHHP, MEOHP, PM10이 선택되었음

<표 7-32> 탄성그물회귀(ENR)로 노출을 선택하고 6개월 유아의 아토피 피부염에 대한 다중노출 로지스틱 회귀 모형

Table 8. Multiple-exposure unpenalized logistic regression models for the selected exposures via ENR and atopic dermatitis in 6-month-infants

Pollutant	Heavy metal	BPA and phthalate	Air pollutant	All exposure
		Early pregnancy		
Pb	1.148 (0.862,1.538)			1.112 (0.844,1.471)
Hg	1.168 (0.889,1.537)			1.163 (0.885,1.532)
Cd	0.890 (0.667,1.173)			–
BPA		1.046 (0.789,1.386)		–
MEHHP		–		–
MEOHP		1.151 (0.839,1.583)		–
MnBP		0.813 (0.591,1.117)		0.931 (0.709,1.228)
NO_2			–	–
PM_{10}			1.268 (0.869,1.864)	1.227 (0.882,1.844)
$PM_{2.5}$			–	–
		Late pregnancy		
Pb	**1.389 (1.047,1.869)**			**1.389 (1.037, 1.887)**
Hg	0.884 (0.634,1.566)			1.251 (0.932, 1.684)
Cd	–			–
BPA		–		–
MEHHP		1.768 (0.939,3.412)		1.909 (0.978, 3.836)
MEOHP		0.879 (0.458,1.693)		0.853 (0.428, 1.705)
MnBP		–		–
NO_2			1.627 (1.083,2.500)	–
PM_{10}				**1.616 (1.068, 2.507)**
$PM_{2.5}$				–

Data are expressed as odds ratio (95% confidence interval)

The models were adjusted for mother's education level, mother's pre-pregnancy BMI, parity, mother's atopic dermatitis, mother's urinary cotinine level, and season at birth

- 또한, 임신 말기에 선택된 물질 중 다른 노출들을 공동유해인자로써 보정하였을 때 납 (오즈비 [OR]: 1.389, 95% CI: 1.037, 1.887)과 PM10 (OR: 1.616, 95% CI: 1.068, 2.507)의 개별 노출이 6개월 시기 아토피 피부염을 유의하게 증가시키는 것으로 나타났음.

<표 7-33> 산전 노출과 6개월 유아의 아토피 피부염에 대한 부분최소제곱-판별 분석(PLS-DA) 모형

Table 9. PLS-DA models for the prenatal exposures and atopic dermatitis in 6-month-infants

Chemical set	Pollutant	Odds Ratio (95% CI)[a]	
		Unadjusted model	Adjusted model[b]
Early pregnancy			
Heavy metal	Pb, Hg, Cd	1.260 (0.960,1.667)	1.271 (0.947,1.720)
BPA and phthalates	BPA, MEHHP, MEOHP, MnBP	1.160 (0.830, 1.625)	1.212 (0.848,1.740)
Air pollutant	NO2, PM10, PM2.5	**1.192 (1.002,1.423)**	1.185 (0.886,1.596)
All exposures	Pb, Hg, Cd, BPA, MEHHP, MEOHP, MnBP, NO2, PM10, PM2.5	**1.262 (1.047,1.528)**	1.324 (0.989,1.793)
Late pregnancy			
Heavy metal	Pb, Hg, Cd	1.249 (0.968, 1.623)	**1.479 (1.111, 1.998)**
BPA and phthalates	BPA, MEHHP, MEOHP, MnBP	**1.218 (1.026, 1.457)**	**1.331 (1.105, 1.621)**
Air pollutant	NO2, PM10, PM2.5	1.224 (0.805, 1.903)	1.169 (0.728, 1.907)
All exposures	Pb, Hg, Cd, BPA, MEHHP, MEOHP, MnBP, NO2, PM10, PM2.5	**1.311 (1.083, 1.604)**	**1.492 (1.209, 1.871)**

Abbreviations: PLS-DA, partial least square-discriminant analysis; CI, confidence interval;
[a] The estimated odds ratio and 95% CI for an increase in the score of each component with the 1-unit
[b] The models were adjusted for mother's education level, mother's pre-pregnancy BMI, parity, mother's atopic dermatitis, mother's urinary cotinine level, and season at birth.

- PLS-DA를 통해 모든 환경유해물질의 누적 노출 점수를 산출하고 6개월 시기 아토피 피부염 발생과의 연관성을 살펴본 결과 임신 말기 누적 노출 점수가 증가할수록 아토피 피부염의 발생 위험이 유의하게 증가하였음(OR:1.492, 95% CI: 1.209, 1.871). 또한, BKMR을 통해 환경유해물질 노출과 아토피 피부염 발생 사이의 비선형 연관성 및 유해 물질 사이의 교호작용 효과가 존재하는 것을 확인하였으며 이들을 반영하여 누적 노출의 영향을 확인하였을 때도 임신 말기의 누적 노출과 아토피 피부염 발생 사이의 유의한 연관성이 나타났음.

<표 7-34> 베이지안커널기계회귀(BKMR)로 노출에 대한 후부포함확률(PIP)

Table 10. Posterior inclusion probabilities (PIP) for exposures in BKMR[a]

Exposure	Early pregnancy		Late pregnancy	
	Group PIP	Conditional PIP	Group PIP	Conditional PIP
Pb	0.52	0.46	0.85	0.98
Hg	0.52	0.37	0.85	0.01
Cd	0.52	0.15	0.85	0
BPA	0.25	0.19	0.97	0.01
MEHHP	0.25	0.24	0.97	0.80
MEOHP	0.25	0.13	0.97	0.17
MnBP	0.25	0.41	0.97	0
NO_2	0.36	0.15	0.44	0.22
PM_{10}	0.36	0.53	0.44	0.55
$PM_{2.5}$	0.36	0.31	0.44	0.22

[a]The models were adjusted for mother's education level, mother's pre-pregnancy BMI, parity, mother's atopic dermatitis, mother's urinary cotinine level, and season at birth.

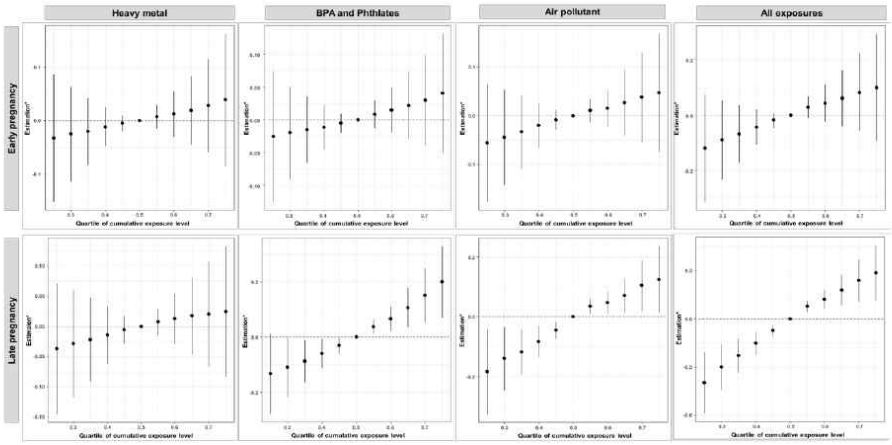


Figure 6. Overall effect of prenatal exposures for atopic dermatitis in 6-month-infants.

*The estimated effect and 95% confidence interval were shown by black circle and lines, respectively; The models were adjusted for mother's education level, mother's pre-pregnancy BMI, parity, mother's atopic dermatitis, mother's urinary cotinine level, and season at birth

[그림 7-35] 6개월 유아에서 산전 노출과 아토피피부염의 전체적인 효과

5) 고찰 및 결론

- 임신 중 환경유해물질에 복합적으로 노출될 경우 각각의 물질들이 함께 작용하여 영유아의 아토피 피부염 발생 위험을 유의하게 증가시킬 수 있음을 시사하며, 특히 임신 말기가 환경유해물질의 복합노출에 감수성이 큰 시기일 가능성을 시사하고 있음.

16. Luo L, Hudson LG, Lewis J et al. Two-step approach for assessing the health effects of environmental chemical mixtures: application to simulated datasets and real data from the Navajo Birth Cohort Study. Environ Health. 2019; 18(1): 46. doi: 10.1186/s12940-019-0482-6.

1) 배경

- 다양한 방법으로 다양한 상관관계를 갖는 고차원 데이터 세트에서 수집된 화학 혼합물 노출을 분석 할 때, 기존 방법이 효과적으로 데이터 차수(dimension)를 줄이고 상호작용 효과를 감지하는 데 한계가 있음을 인식하였음.

2) 목적

- 이 연구에서 Navajo Birth Cohort Study 의 기존 데이터 세트와 두 개의 시뮬레이션 된 데이터 세트를 사용하여 화학 혼합물 노출의 분석 과제를 해결하는 2 단계 통계 접근법의 성능을 조사하고자 함. biomarker: 산화 스트레스 바이오마커

3) 방법 : 코호트 연구

(1) 연구자료 및 연구대상자

- the Navajo Birth Cohort Study(NBCS)

- NBCS 는 미국 남서부의 Four Corners 지역에 있는 나바호족 국가(Navajo Nation, 애리조나, 뉴멕시코, 유타 포함, 아메리카 원주민 부족 최대 거주)의 버려진 우라늄 광산(abandoned uarnium mine AUM)-관련 금속 노출, 출생 결과 및 초기 발달 지연 사이의 관계를 조사하기 위해 고안된 전향적 역학 연구임. 금속 혼합물 노출 분석임. 약 160,000 명의 나바호족.

- 2013 년 NBCS 는 나바호족 국가에서 최소 5 년간 살았으며 참여하는 NAIHS 병원에서 출산한 후 출산 후 1 년 동안 추적관찰을 한 14 세- 45 세의 임산부 600 명 이상을 모집하기 시작했음.

(2) 노출 평가

- 등록 시 혈액, 혈청, 소변 이렇게 세 가지 샘플에서 측정된 금속 노출로 구성됨

- 우라늄, 총 비소, 비소(arsenous III) 산(arsenite, AsIII), 디메틸아르시닉산(dimethylarsinic acid, DMA), 아연

- 노출의 정도가 다른 금속을 공정하게 비교하기 위해, 먼저 변수 분포의 왜도를 줄이기 위해 로그 변환을 수행한 다음 통계 모델링에서 z 점수 변환을 사용하여 표준화를 수행했음.

(3) 건강영향 평가

- 주요 건강 결과는 산화 스트레스의 바이오마커이며, 이는 다양한 질병의 중요한 구성 요소로 간주됨. 8-iso-PGF2α 와 prostaglandinF2α 비(ratio).

- 세계 보건기구(WHO) 혈청 아연 농도 기준 70 μg/dL 이상과 이하로, 두 그룹에서 각각 66 명의 참가자로 132 명의 참가자를 무작위로 선택했음.

(4) 보정변수

- 연령, 흡연, 고용, 임신 3 기 등록, BMI

(5) 통계분석 방법

- 랜덤 포레스트(random forest, RF)와 화학 혼합물 데이터를 분석하기 위한 적응형 lasso(올가미) 접근법을 결합하는 2 단계 절차를 검토할 것을 제안함.

- 시뮬레이션 된 데이터 세트(NIEHS 워크샵에서 제공 한 데이터 세트 # 1 및 다양한 상관관계를 갖는 많은 수의 상관된 노출 변수를 나타내도록 시뮬레이션 된 데이터 세트)와 NBCS 실제 데이터 세트를 사용하여 제안 된 2 단계 방법의 성능을 평가했음. 제안 된 방법을 (1)

단일 단계 적응형 lasso; (2) CART 를 이용하고 이후 적응형 lasso 접근법의 2 단계, 이는 NIEHS 워크숍에서도 널리 활용되었음

- 분류 및 회귀 나무(Classification and regressino trees, CART): 데이터 공간을 재귀 적으로 분할하고 각 파티션 내에 간단한 예측 모델을 적용하여 얻은 데이터로부터 예측 모델을 구성하는 머신 러닝 방법임. 연속형 종속 변수의 경우 회귀 나무에 대한 예측 오차는 관측값과 예측값 사이의 제곱 차로 추정되며, 예측 가능성이 가장 높은 변수를 선택하는 데 사용할 수 있음.

- 무작위 숲(Random forest, RF): 많은 나무를 키우는 앙상블 기반 접근 방식임. 각 나무는 재귀적 분할 방법을 사용하여 구축되어, 모든 예측 변수에 걸쳐있는 특정 공간을 예측 변수와 결과 변수 간에 유사한 연관 패턴을 가진 대상 그룹으로 분할함. 특히, 각 나무는 임의로 추출된 데이터의 붓스트랩(bootstrap) 샘플을 사용하여 성장하게 됨. 변수의 무작위로 선택된 하위 집합을 기반으로, 평균 제곱 오류를 기반으로 나무 노드를 분할하는 데 사용됨. 나무의 앙상블(전체)을 평균하여 예측함. 변수 중요도(variable importance measure to pre-select important variables, VIMP) 측정 값을 사용하여 중요한 변수를 2 단계 분석에 사전 선택했음. VIMP 는 변수가 얼마나 중요한지에 대한 측정값으로, 변수가 분석에서 제거되면 예측 오류의 변화를 추정함. VIMP 값이 클수록 변수의 예측 가능성이 높아짐.

- 적응형 lasso(adaptive lassl, alasso): Tibshirani 가 제안한 lasso(올가미) 방법은 변수 선택하는데 광범위하게 연구 된 L1 규범 페널티를 부과하여 많은 계수를 0 으로 축소시킴. 단점으로 여러 개 중 한 개만 선택. 적응형 lasso 는 lasso 의 한계를 해결하기 위해 Zou 가 개발한 수정된 버전임. 융통성 있는 가중치 체계를 사용하여, 적응형 lasso 는 다른 계수에 다른 수축량을 적용하는 반면, lasso 는 잠재적 인 편향을 유발할 수 있는 모든 회귀 계수에 동일한 페널티를 적용함.

적응형 lasso는 하위 집합 선택과 능선 회귀 모두의 우수한 기능을 결합하여 다중공선성이 존재할 때 잘 수행되는 것으로 보임.

- 성과 평가(performance evaluation): 세 가지 방법중 성과를 비교했음: (1) 단일 단계 적응형 lasso; (2) CART를 이용하고 이후 적응형 lasso 접근법의 2 단계; (3) 무작위 숲을 이용하고 이후 적응형 lasso 접근법의 2단계. 각각의 방법에 대해 노출 변수에 대한 모든 주요 효과를 포함하는 모델과 노출 변수 사이의 모든 쌍별 상호 작용 항을 추가로 포함하는 두 가지 모델을 적용함. R2, 보정된 R2, 평균 제곱 오차(mean squared error, MSE) 및 평균 제곱 예측 오차(mean squared prediction error, MSPE)를 포함한 여러 성능 메트릭을 계산하여 세 가지 방법(상호작용 항의 유무에 각각)의 성능을 측정했음.

- R 통계 소프트웨어, 버전 3.4.2 (R Core Team, 2017?)의 glmnet(적응형 lasso)패키지, randomForestSRC(무작위 나무) 패키지를 사용하였음.

4) 결과

시뮬레이션 데이터 세트 1: NIEHS 워크숍 데이터세트

- 노출 변수의 양반응관계, 혼란변수는 정규분포된 무작위 노이즈(noise)임. 노출 X1 및 X2는 X1이 X2보다 2배 더 강력한 결과에 긍정적으로 기여하도록 시뮬레이션을 설정하였음; X4 및 X5는 X4보다 4.5배 강력한 X5의 결과에 부정적인 영향을 미쳤으며; X7은 결과에 긍정적으로 기여했음. n=500 개의 샘플이 합성 데이터 세트에 포함되었음.

- Fig 1, Table 2, Table S2은 세 가지 방법을 사용하여 노출 변수와 결과 (상자)와 95 % 신뢰 구간 (CI) (선) 간의 관계를 모델링하는 회귀 계수를 설명하는 산림도(forest plot)임. 범례에 표시된 대로 색상으로 구분됨. 단일 단계 적응성 lasso 방법은 X1과 X7 사이의 긍정적인 연관성과 결과, 그리고 X4와 X5 사이의 부정적인

연관성으로 X5 가 X4 보다 4.1 배 강력함을 확인했음. 최종 선택된 모델에는 X2 및 X3 의 두 변수가 포함되지만 계수는 통계적으로 0 과 유의하게 다르지 않았음. RF 와 그에 따른 적응 lasso 방법은 X1, X5 와 X7 이 결과와 유의하게 연관되어 있음을 식별했으며, 계수 추정치는 단일 단계 적응형 lasso 접근법을 사용하여 확인된 것과 유사함. 더 작은 크기의 연관 효과를 갖는 X2 및 X4 는 확인되지 않았지만, 선택된 모델에 의해 설명 된 총 분산은 단일 단계 적응형 lasso 방법이 더 복잡한 모델에 가까움(조정 된 R2 는 각각 91.9 및 92.2%). MSPE 는 거의 동일함(9.47 vs. 9.55). CART 다음에 적응형 lasso 방법은 X1 과 X7 의 연관성을 올바르게 식별했으며, 이는 전체 결과 분산의 81 % 만 차지함. 조정 된 R2 는 다른 두 가지 방법에 비해 약 10 % 더 적으며 MSPE 는 훨씬 더 큼.

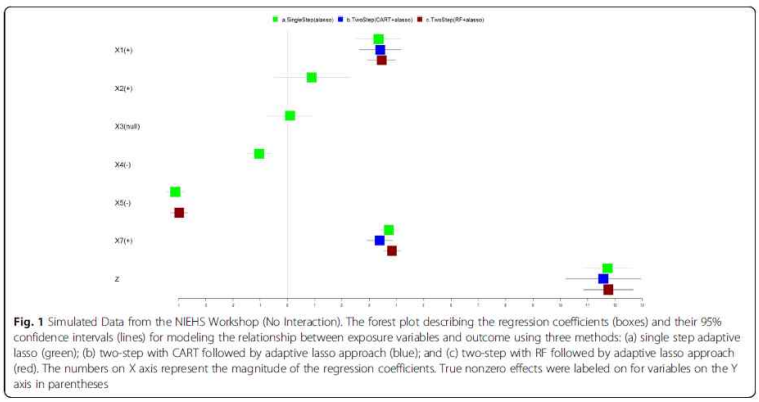


Fig. 1 Simulated Data from the NIEHS Workshop (No Interaction). The forest plot describing the regression coefficients (boxes) and their 95% confidence intervals (lines) for modeling the relationship between exposure variables and outcome using three methods: (a) single step adaptive lasso (green); (b) two-step with CART followed by adaptive lasso approach (blue); and (c) two-step with RF followed by adaptive lasso approach (red). The numbers on X axis represent the magnitude of the regression coefficients. True nonzero effects were labeled on for variables on the Y axis in parentheses

[그림 7-36] NIEHS 워크숍에서 시뮬레이션 된 데이터(상호 작용 없음)

3 가지 방법을 사용하여 노출 변수와 결과 간의 관계를 모델링하기 위한 회귀 계수(상자)와 95 % 신뢰 구간(선)을 설명하는 산림도(forest plot): (a) 단일 단계 적응형 lasso(녹색); (b) 2 단계 CART 및 적응형 lasso(파란색); (c) 2 단계 RF 및 적응형 lasso(빨간색). X 축의 숫자는

회귀 계수의 크기를 나타냄. Y 축의 변수에 대해 괄호 안에 0 이 아닌 실제 효과가 표시되었음.

<표 7-35> NIEHS 데이터셋에서 각 접근에 따른 성과 평가(상호작용 없음)

Table 2 Performance Evaluation for Each Approach for NIEHS Dataset [a] (No Interaction)

	Adaptive lasso	CART+ Adaptive lasso	RF+ Adaptive lasso
R^2	92.3%	81.5%	91.9%
Adjusted R^2	92.2%	81.4%	91.9%
MSE	9.080	21.73	9.50
MSE.CV	9.082	21.72	9.076
MSPE.CV	9.47	22.00	9.55

[a]The dataset includes 500 samples and 8 predictor variables with Pearson's correlations ranging from −0.13 to 0.89. *MSE* Mean square error, MSE. *CV* Cross validation, *MSPE* Mean square predictive error

[a] 피어슨 상관지수는 -0.13-0.89. *CV* Cross validation

- Fig 2, Table 3, Table S3 은 노출 변수의 주요 및 쌍별 상호작용 효과를 모델링하는 세 가지 방법의 선택된 공변량 및 모델 성능을 나타냄.

- NIEHS 에 의한 시뮬레이션 데이터 셋 적용에서 주요 발견은 2 단계 접근법이 추정 된 연관 효과와 상호작용 효과를 덜 식별하는 반면, 단일 단계 접근법과 비교할 때 비슷한 성능을 가진 보다 포용적인 모델을 발견했다는 것임. RF 는 중요한 변수를 효과적으로 식별하고 데이터 차원 축소를 수행 할 수 있는 도구이지만 워크샵에서 제공되는 것과 같은 작은 차원 데이터 집합을 분석하는 데에는 큰 도움이 되지 않을 수 있음.

시뮬레이션 데이터 세트 2: 실제 노출 프로파일을 모방하는 변수와 상관 된 데이터 세트

- m=20 이상인 노출 변수를 갖는 데이터세트, 결과 변수가 연속형인 데이터세트를 시뮬레이션했음. 시뮬레이션은 노출 X1, X2, X12 및

X15 의 긍정적 연관 효과, X9 및 X16 의 부정적인 효과뿐만 아니라 X1 과 X12 사이의 상승적 상호 작용을 포함하도록 설정되었음.

- Fig 3, Table 4, Table S4 는 실제 모델 계수와 주 효과 모델로 세 가지 방법을 사용하여 식별 된 계수를 나타냄.

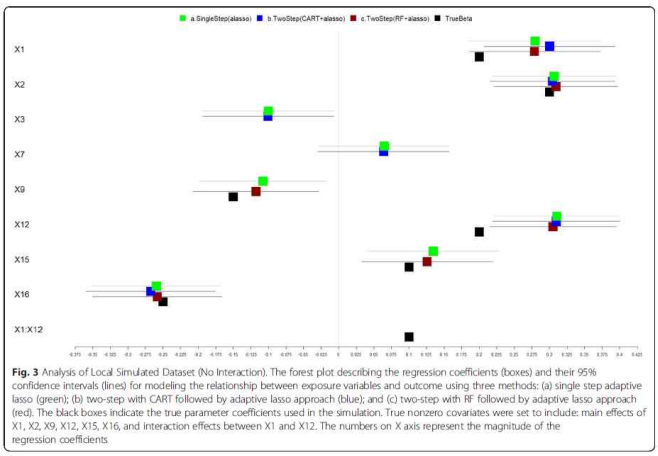


Fig. 3 Analysis of Local Simulated Dataset (No Interaction). The forest plot describing the regression coefficients (boxes) and their 95% confidence intervals (lines) for modeling the relationship between exposure variables and outcome using three methods: (a) single step adaptive lasso (green); (b) two-step with CART followed by adaptive lasso approach (blue); and (c) two-step with RF followed by adaptive lasso approach (red). The black boxes indicate the true parameter coefficients used in the simulation. True nonzero covariates were set to include: main effects of X1, X2, X9, X12, X15, X16, and interaction effects between X1 and X12. The numbers on X axis represent the magnitude of the regression coefficients

[그림 7-37] 국소 시뮬레이션 데이터셋 분석(상호작용 없음)

그림 35 와 같음. 검은색 박스는 시뮬레이션에 사용된 실제 파라미터 계수를 나타냄. 0 이 아닌 공변량은 X1, X2, X9, X12, X15, X16 의 주요 효과 및 X1 과 X12 간의 상호작용 효과를 포함하도록 설정되었음. X 축의 숫자는 회귀 계수의 크기를 나타냄.

- Fig 4, Table 5, Table S5 는 쌍별 상호 작용 효과를 포함한 모델의 해당 결과를 나타냄. 2 단계 RF + 적응형 lasso 방법은 0 이 아닌 계수의 주요 효과를 정확하게 식별하고 나머지는 0 으로 축소하였음. 단일 단계 적응형 lasso 방법은 (X1, X2, X9, X12, X15 및 X16)에 대한 실제 연관 효과를 감지했지만, X3 및 X7 에 대해서는 널 효과를 잘못 감지했음. 단일 단계 적응형 lasso 및 2 단계 RF + 적응형 lasso 의 성능은 교차 검증을 사용하여 유사한 보정 된 R2 및 더 작은 MSPE 로 측정한 것과 유사함. 다시 한 번, 2 단계 RF + 적응형 lasso 접근 방식은 복잡성이 덜한 수준 높은 모델을 만들었음. 2 단계 RF +

적응형 lasso 방법에 쌍별 교호작용 항을 추가하면 X1 과 X12 사이의 교호작용 효과가 검출됨.

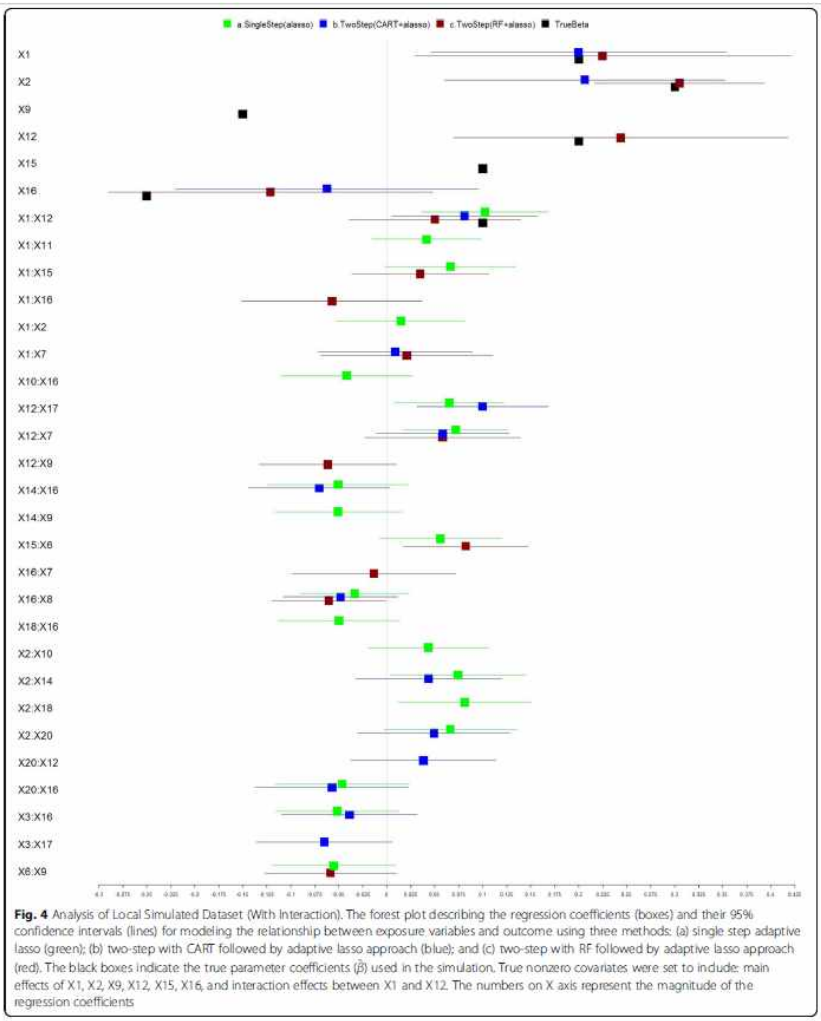


Fig. 4 Analysis of Local Simulated Dataset (With Interaction). The forest plot describing the regression coefficients (boxes) and their 95% confidence intervals (lines) for modeling the relationship between exposure variables and outcome using three methods: (a) single step adaptive lasso (green); (b) two-step with CART followed by adaptive lasso approach (blue); and (c) two-step with RF followed by adaptive lasso approach (red). The black boxes indicate the true parameter coefficients (β) used in the simulation. True nonzero covariates were set to include: main effects of X1, X2, X9, X12, X15, X16, and interaction effects between X1 and X12. The numbers on X axis represent the magnitude of the regression coefficients

[그림 7-38] 국소 시뮬레이션 데이터셋 분석(상호작용 있음)

그림 38 과 같음. 검은색 박스는 시뮬레이션에 사용 된 실제 파라미터 계수(^β)를 나타냄.

- 예측 변수 사이의 다양한 상관 관계로 모델 성능을 결정하기 위해, 시뮬레이션 데이터셋 2 와 동일한 수의 변수(m=20)와 샘플(n=500)로 예측 변수(r=0.3, 0.5, 0.7 및 0.8) 사이에서 다양한 증가 상관 관계로 시뮬레이션을 수행했음. 각 상관 설정에 대해 추정 회귀 계수를 실제 시뮬레이션 계수와 비교하고 단일 단계 적응형 lasso, CART + 적응형 lasso 및 RF + 적응형 lasso, 또는 각각 상호 작용 효과가 있고 없고로 구했음.

단일 단계 적응형 lasso 는 실제 모델보다 더 많은 변수를 식별함. 2 단계 방법은 중간 정도에서 보통 큰 상관 예측 변수 (r <0.8)를 가진 시뮬레이션 된 데이터 세트에 대해 개선된 결과가 일관되게 관찰됨. 그러나 상관관계가 매우 큰 경우(예: r> 0.8), 세 가지 접근 방식 중 어느 것도 잘 수행되지 않으며, 각 모델은 실제 null 공변량의 하위 집합을 최종 모델에 잘못 포함함. 이러한 경우 생물학적 입력을 기반으로 고도로 상관 된 쌍에서 하나의 변수를 제거하여 분석 전에 데이터를 사전 필터링하면 모델 성능을 향상시키는 데 도움이 될 수 있음.

NBCS 의 적용

- Fig 5, Table 6, Table S14(상호작용 없음) / Fig 6, Table 7, Table S15(상호작용 있음)는 NBCS 에서 수집 된 산화스트레스 바이오마커에 대한 금속 혼합물 노출의 영향 분석에 세 가지 방법을 적용한 비교 결과를 나타냄. 흡연, 임신 3 기, 연령 보정함. 제안 된 2 단계 RF + 적응형 lasso 방법은 주요 효과 모델링에 대한 교차 검증을 사용하여 가장 큰 보정된 R2 및 더 작은 MSPE 를 가진 모델을 산출했음(Table 6). 산화 스트레스 바이오 마커와 소변에서 세슘과 디메틸아르신산(DMA)에 대한 노출은 양의 연관성이 있었고, 소변에서 바륨과 탈륨에 대한 음의 연관성이 관찰되었음(Fig 5).

<표 7-36> 나바호 출생코호트(NBCS)의 분석에 사용된 각 접근법의 성능 평가(상호작용 없음)

Table 6 Performance Evaluation of each approach used in analyses of NBCS [a] (No Interaction)

	Adaptive lasso	CART+ Adaptive lasso	RF+ Adaptive lasso
R^2	37.0%	21.5%	36.7%
Adjusted R^2	32.2%	15.5%	32.6%
MSE	0.0048	0.0061	0.0049
MSE.CV	0.0047	0.0061	0.0047
MSPE.CV	0.0078	0.0078	0.0062

[a]The dataset includes 132 samples and 30 predictor variables including 25 metals and 5 demographic variables (age, smoking, employment, trimester and BMI). The Pearson's correlations between the metals range from −0.32 to 0.76

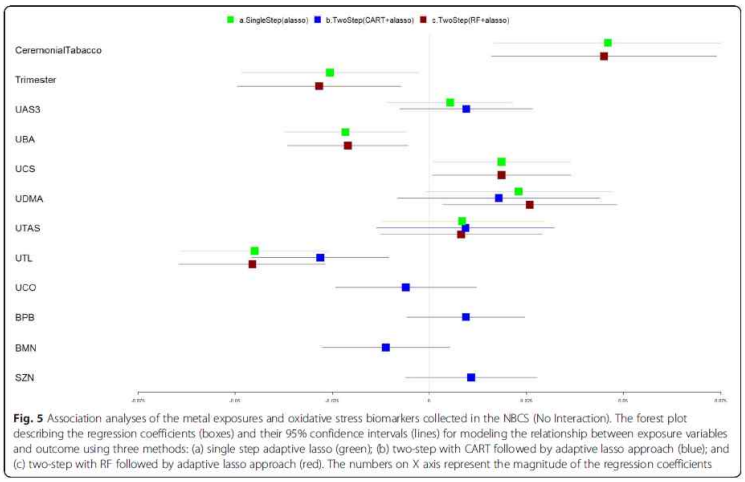


Fig. 5 Association analyses of the metal exposures and oxidative stress biomarkers collected in the NBCS (No Interaction). The forest plot describing the regression coefficients (boxes) and their 95% confidence intervals (lines) for modeling the relationship between exposure variables and outcome using three methods: (a) single step adaptive lasso (green); (b) two-step with CART followed by adaptive lasso approach (blue); and (c) two-step with RF followed by adaptive lasso approach (red). The numbers on X axis represent the magnitude of the regression coefficients

[그림 7-39] 나바호 출생코호트(NBCS)에서 금속 노출과 산화 스트레스 바이오마커의 연관성 분석(상호작용 없음)
X 축의 숫자는 회귀 계수의 크기를 나타냄.

- 쌍별 상호작용 항을 포함하는 모델의 경우, 제안 된 2 단계 RF + 적응형 lasso 방법은 산화 스트레스 결과와 관련하여 소변에서 바륨,

세슘 및 탈륨의 주요 효과, 아연과 소변과 혈청 코발트, 혈청에서 아연과 소변에서 세슘, 그리고 소변 중 비소 (arsenous III)산과 소변에서 탈륨 사이의 상호작용을 감지했음. 제안된 2 단계 방법은 가장 작은 MSPE 를 산출했음. 단일 단계 적응형 lasso 는 가장 큰 보정된 R2 를 가진 모델을 산출했으며 다른 모델보다 더 많은 상호작용 효과 항을 포함했음.

그러나 이것은 교차 검증을 사용하여 더 큰 MSPE 를 야기한 가능한 과적합(overfitting)을 설명 할 수 있음. NBCS 데이터세트에대한 이 응용에서, 2 단계 CART + 적응형 lasso 방법은 다른 두 방법에 비해 성능이 떨어졌음(작은 보정된 R2 및 더 큰 MSPE).

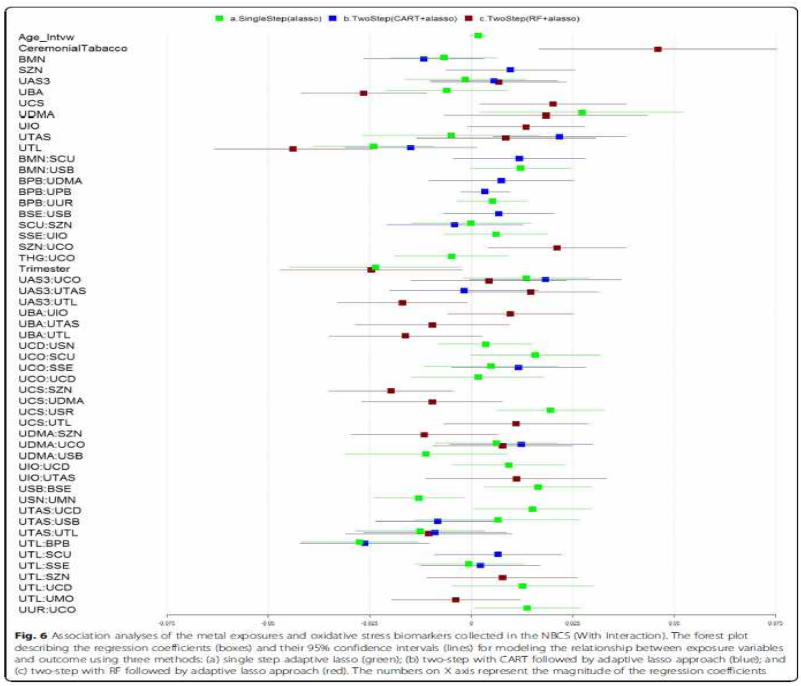


Fig. 6 Association analyses of the metal exposures and oxidative stress biomarkers collected in the NBCS (With Interaction). The forest plot describing the regression coefficients (boxes) and their 95% confidence intervals (lines) for modeling the relationship between exposure variables and outcome using three methods: (a) single step adaptive lasso (green); (b) two-step with CART followed by adaptive lasso approach (blue); and (c) two-step with RF followed by adaptive lasso approach (red). The numbers on X axis represent the magnitude of the regression coefficients

[그림 7-40] 나바호 출생 코호트(NBCS)에서 금속 노출과 산화 스트레스 바이오마커의 연관성 분석(상호작용 있음)

X 축의 숫자는 회귀 계수의 크기를 나타냄.

- 단점 : 이 모델링 방식에는 몇 가지 제한점이 있음. RF 방법은 노출과 결과 변수 사이의 비선형 관계를 설명하는 중요한 변수를 선택하는 유틸리티를 제공하지만, 적응형 lasso 접근법을 사용하여 두 번째 단계에서 결정된 최종 모델은 결과와 노출 변수 사이의 선형 관계만 캡처함(독립적인 금속 또는 상호작용하는 금속). 예측 변수 그룹간에 과도한 상관관계가 존재하는 경우 lasso 접근 방식은 해당 그룹에서 하나의 변수만 선택하고 편향된 효과 추정치를 제공 할 수 있음. -> ENR 제안

- 요약: 시뮬레이션 된 데이터 세트를 활용하고 이 방법을 Navajo Birth Cohort Study 의 실제 데이터 세트에 적용하여 제안된 2 단계 접근법의 우수한 성능을 입증했음. 시뮬레이션 데이터 세트의 결과는 다른 방법과 비교하여 변수 차원 축소의 효과와 parsimonious(지극히 검소한) 모델을 신뢰할 수 있는 확인을 나타냄: 1 단계 적응 lasso 또는 2 단계 CART 방법임.

5) 고찰 및 결론

- 2 단계 접근 방식은 변수 선택 및 적응형 수축 전략의 장점을 결합하여, 고-처리량 화학 혼합물 노출이 건강 결과에 미치는 영향을 분석하는 강력한 방법을 제공함.

RF + 적응형 lasso 2-단계 접근

- RF 스크리닝 단계는 관심 결과를 선형 또는 비선형 관계로 유지하는 효과적인 차원 축소를 수행함.
- 표준 올가미 접근 방식이 유사한 예측 유틸리티를 생성하는 여러 상호 작용 항으로 불안정한 모델을 식별한다는 제한을 극복함
- 표준 접근법이 0 이 아닌 실제 공변량보다 더 많은 공변량을 포함하는 모델을 추정하는 경향이 있는 과적합 문제를 해결함

- 단일 단계 접근 방식보다 더 나은 예측 성능을 제공하는 실제 모델에 더 가까운 parsimonious 또는 sparse 모델을 제공함.

이 2 단계 접근은 다음 상황에 적절함

- 데이터 세트가 10-50 예측 변수로 구성된 경우
- 상호작용 효과의 추정 및 효과 크기의 정량화에 관심이 있는 경우
- 예측 변수 사이에 중간 정도에서 조금 큰 상관관계 (r <0.8)가 존재하는 경우

17. Casas M, Valvi D, Ballesteros-Gomez A, Gascon M, Fernández MF, Garcia-Esteban R, Iñiguez C, Martínez D, Murcia M, Monfort N, Luque N, Rubio S, Ventura R, Sunyer J, Vrijheid M. Exposure to Bisphenol A and Phthalates during Pregnancy and Ultrasound Measures of Fetal Growth in the INMA-Sabadell Cohort. Environ Health Perspect. 2016 Apr;124(4):521-8.

1) 배경

- bisphenol A(BPA)와 프탈레이트(phthalates)에 산전 노출된 경우, 태아의 성장에 영향을 끼칠 수 있으나 이전 연구 결과는 일관성이 없고, 연구가 부족함.

2) 목적

- 스페인 출생 코호트의 모-자녀 488 쌍에 대하여 BPA 와 프탈레이트에 대한 태아기 노출이 태아성장과 관련이 있는지 평가함.

3) 방법

(1) 연구대상자

- INMA (INfancia y Medio Ambiente; Childhood and Environment)는 스페인 Sabadell 지역에서 2004 년에서 2006 년 사이 657 명의 임산부를 대상으로 한 인구 기반 출생 코호트 연구임.

- 산전 진찰 첫 방문에 16 세 이상, 거점병원에서 출산희망, 외동임신, 자연임신, 의사소통에 문제가 없는 경우 포함시켰는데 평균 13.4±1.7 재태기였음.

(2) 산전 BPA, 프탈레이트 노출

- 모체의 소변 샘플은 12±1.7 주와 32±1.4 주에 수집되어 -20℃에 보관하였고 소변 크레아티닌 농도를 통해 희석을 고려하여 액체 크로마토 그라피-질량 분광법으로 측정함.

- BPA 는 479 명의 산모에서 측정하고, 프탈레이트 대사산물은 391 명에서 측정함.

(3) 태아성장과 출생결과

- 태아 초음파 검사는 12, 20, 34 주에 시행하였고, 대퇴골 길이(FL, femur length), 머리 둘레(HC, head circumference), 복부 둘레(AC, abdominal circumference), 양쪽마루뼈지름(BPD, biparietal diameter)를 매개변수로 측정함.

- 머리엉덩길이(CRL, crown-rump length)가 자가보고한 생리주기와 7 일 이상 차이가 있을 경우 CRL 을 통해 제태기간을 결정하였고, 태반을 수집해 자궁 내 성장의 척도로 태반 무게를 분석함.

- 5 개의 태아 매개변수에 대한 종적 성장 곡선을 얻기 위해 선형혼합모형(Linear mixed models)을 사용하였음.

- 태아성장에 영향을 끼치는 것으로 알려진 생리학적 요인(모 연령, 출산력, 출신국가, 모 신장, 임신전 체중 혹은 BMI, 부 신장, 부 체중 또는 BMI, 자녀의 성별)을 우도비 검정(likelihood ratio test, $p<0.05$)를 통해 재태연령과의 상호작용을 검정함.

- 출생결과는 재태연령에 크게 의존하기 때문에 Box-cox power exponential method 를 이용하여 출생 체중, 길이, HC, 태반 무게에 대한 성장곡선을 성별로 보정하여 40 주까지 표준화하였음.

(4) 공변량

- 모 연령, 교육, 사회계층, 출신국가, 체중 및 신장, 출산력, 과거력, 결혼상태, 식생활, 가정용 세제 사용, 직접흡연, 소변 수집 시간, 출생시기에 대한 정보는 12 주와 32 주에 훈련된 면담자가 설문지를 통해 얻음.

- 코티닌(Cotinine)은 32 주에 소변을 통해서 측정함.

(5) 통계분석

- 제 1 석달(trimester)과 제 3 석달 사이의 소변 BPA 와 프탈레이트 대사산물의 재현성은 확률일원절편 선형혼합모형(random one-way intercept linear mixed models)을 사용하여 급내상관계수(ICC, intraclass correlation coefficient)를 계산하여 얻음.

- 로그-변환된 성분농도의 Pearson correlations 를 계산함.

- 소변농도의 근사 로그정규분포 때문에 평균 크레아티닌에 보정된 오염물질의 농도는 $\log_2$ 로 변환하였음.

- 일반화 부가 모형(generalized additive models, GEM)을 사용하여 노출-결과 관계의 선형성을 평가함.

- 태아크기/성장, 출생 매개변수에 대한 BPA/프탈레이트와 표준점수(SD scores) 사이의 연관성은 일변량 및 다변량 선형회귀모형으로 검정함.

- 다변량모형에 포함될 공변량(covariates)를 결정하기 위해 유향 비순환 그래프(DAGs, directed acyclic graphs)를 적용함.

- 과거 연구에서 노출 혹은 결과와 관련이 있다고 기술된 경우 공변량은 DAGs 에 포함하였고, 이변량 분석을 통해 이번 연구에서도 연관성을 확인하였음 ($p \leq 0.1$).

- DAGs 에 의거하여 최종 다변량모형은 어머니의 교육수준(1 차, 2 차, 대학), 임신 중 흡연(비흡연자, 임신 중에만 비흡연, 임신 중 흡연), 출산력(초산, 경산)에 대하여 보정되었음.

- 다변량모형에서 시작하여, 잠재적 교란요인을 포함하여 전진 단계별 선택법(forward stepwise selection)을 이용하였고, 출생시기, 임신 중 소변 코티닌 수준, 임신 중 통조림생선 섭취, 소변 수집 시각, 가정용 세제 사용에 대하여 계수가 10% 초과하여 변경되는지 검사하여 출생시기(겨울, 봄, 여름, 가을)과 소변 코티닌 수준은 대부분의 계수를 10% 초과하여 변경하여 최종 모형에 포함하였음.

- 태아의 HC 성장모형은 이미 출산력에 대하여 보정되어 태아 HC 다변량모형에 포함되지 않았고, 신생아의 HC 모형도 출산유형(질식분만, 기구분만, 제왕절개)에 따라 보정되었음.

- 성별에 의한 유의미한 변화($p \leq 0.1$)를 확인하기 위하여 노출과 성별과의 상호작용에 대하여 우도비 검정을 시행함.

- 결과의 로버스트성(robustness)을 평가하기 위하여 다양한 민감도 분석(sensitivity analyses)를 하였고 다른 모형에서 일관된 결과를 보이는지를 고려하였음.

- 이를 위해, 소변 크레아티닌 농도를 보정하는 모형을 반복하고, “희석효과”를 방지하기 위해 크레아티닌 값이 극한치인 값을 제외하였고, 4 가지 주요 노출 변수(BPA, ΣDEHPm, MBzP, ΣLMWPm)가 포함된 다중 오염물질 모형을 수행하여 데이터를 전체 관측치로 제한하여 단일 오염물질 모형과 계수의 변화가 있는지 비교함.

4) 결과

- 임신 중 BPA 혹은 DEHPm 의 노출이 태아성장 매개변수와 관련을 보이지 않음. 산전 MBzP 노출은 20-34 주에 FL 과 양의 상관관계를 보였음. MBzP 는 남아에서는 출생 체중과 양의 상관관계를 보였으나 여아에서는 보이지 않음.

- LMWPm mono-n-butyl phthalate (MnBP)는 HC 와 임신 12-20 주에 음의 상관관계를 보임.

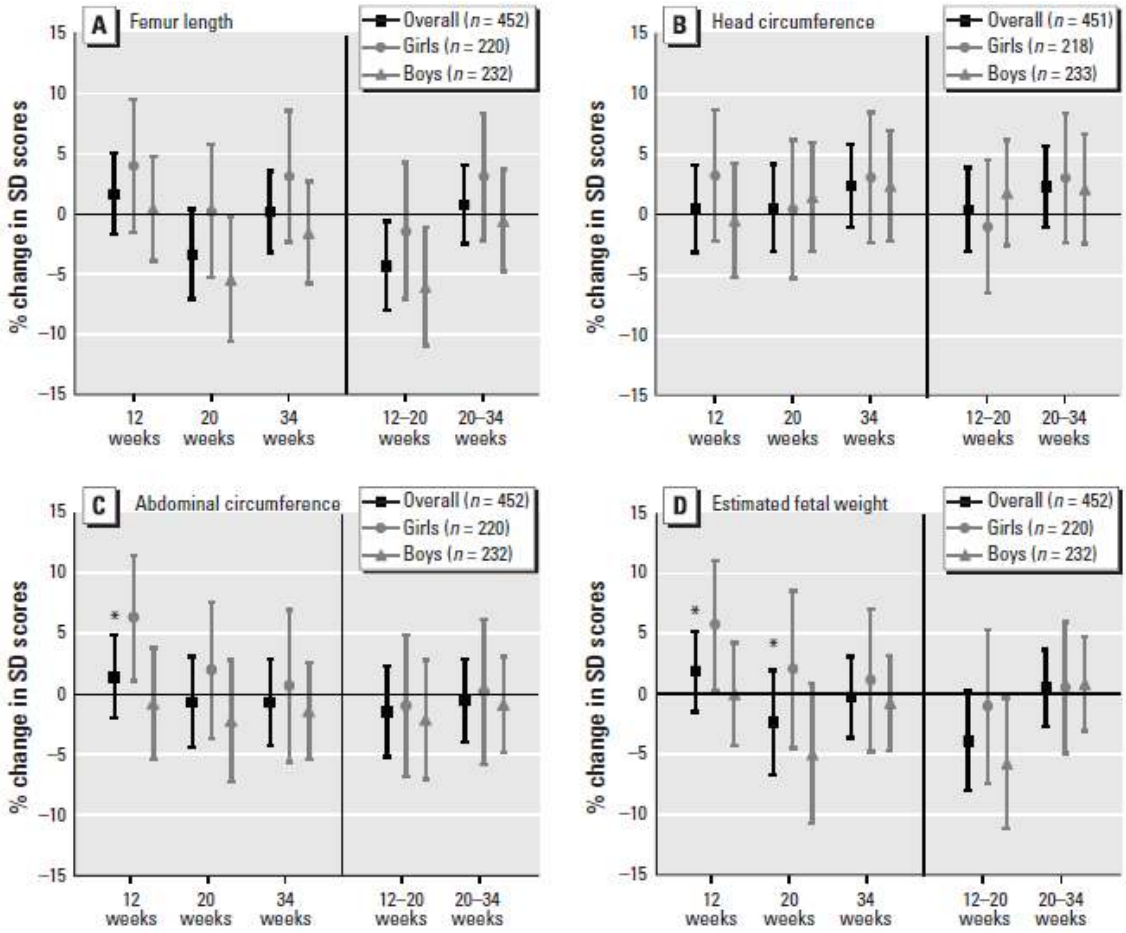


Figure 1. Adjusted associations between maternal urinary BPA levels (μg/g creatinine) and fetal size and growth parameters in the overall population, in girls, and in boys: femur length (*A*), head circumference (*B*), abdominal circumference (*C*), and estimated fetal weight (*D*).

BPA, bisphenol A. Mean percent difference in standard deviation scores per doubling of BPA levels (levels were $\log_2$-transformed). Femur length model was adjusted for maternal age, maternal height, paternal weight, maternal country of origin, maternal education, smoking during pregnancy, parity, birth season, and urinary cotinine levels during pregnancy. Head circumference model was adjusted for maternal age, maternal height/weight, paternal weight, maternal education, smoking during pregnancy, parity, birth season, and urinary cotinine levels during pregnancy. Abdominal circumference model was adjusted for maternal age, maternal height/weight, paternal weight, maternal country of origin, maternal education, smoking during pregnancy, parity, birth season, and urinary cotinine levels during pregnancy. Estimated fetal weight model was adjusted for maternal age, maternal height, paternal weight, maternal country of origin, maternal education, smoking during pregnancy, parity, birth season, and urinary cotinine levels during pregnancy. *p-Interaction for sex ≤ 0.1.

[그림 7-41] 모성 소변 BPA 수준(㎍/g 크레아티닌)과 전체 성별, 소녀, 소년의 태아 크기 및 성장 매개변수 사이의 보정된 연관성

<표 7-37> 모성 소변 BPA와 프탈레이트 대사체 수준(㎍/g 크레아티닌)과 전체 성별, 소녀, 소년의 출생 결과 사이의 보정된 연관성

Table 2. Adjusted associations between maternal urinary BPA and phthalate metabolites levels (µg/g creatinine) and birth outcomes in the overall population, in girls, and in boys.

	Overall		Girls		Boys		
Compounds	*n*	β (95% CI)	*n*	β (95% CI)	*n*	β (95% CI)	*p* Interaction
BPA							
Weight (g)	448	−16.24 (−54.10, 21.62)	219	−21.56 (−82.93, 39.82)	229	−11.01 (−57.39, 35.37)	0.99
Placental weight (g)	398	0.53 (−10.79, 11.86)	193	−8.01 (−28.03, 12.01)	205	6.98 (−6.47, 20.44)	0.28
Length (mm)	441	−0.37 (−2.00, 1.25)	212	−0.64 (−3.21, 1.93)	229	−0.27 (−2.31, 1.77)	0.93
Head circumference (mm)	435	0.47 (−0.66, 1.60)	208	0.38 (−1.47, 2.23)	227	0.54 (−0.77, 1.85)	0.77
Gestational age (weeks)	453	1.19 (−0.19, 2.58)	220	1.65 (−0.56, 0.39)	233	0.63 (−1.21, 2.46)	0.54
ΣDEHPm							
Weight (g)	371	15.56 (−28.75, 59.87)	179	5.92 (−60.59, 72.42)	192	14.68 (−44.11, 73.48)	0.82
Placental weight (g)	325	1.31 (−12.30, 14.93)	154	−14.25 (−36.73, 8.24)	171	13.94 (−2.85, 30.72)	0.04
Length (mm)	364	0.38 (−1.59, 2.35)	172	0.78 (−2.06, 3.61)	192	−0.25 (−2.91, 2.40)	0.84
Head circumference (mm)	359	0.16 (−1.15, 1.47)	169	0.00 (−1.98, 1.97)	190	0.17 (−1.46, 1.80)	0.56
Gestational age (weeks)	375	−0.13 (−1.72, 1.46)	180	0.20 (−2.23, 2.63)	195	−0.48 (−2.65, 1.69)	0.71
MBzP							
Weight (g)	371	14.11 (−19.08, 47.29)	179	−27.30 (−79.41, 24.80)	192	47.78 (5.78, 89.78)	0.04
Placental weight (g)	325	0.75 (−9.79, 11.30)	154	−20.57 (−39.22, −1.92)	171	12.53 (0.27, 24.79)	0.02
Length (mm)	364	−0.42 (−1.90, 1.07)	172	−1.42 (−3.67, 0.84)	192	0.43 (−1.49, 2.35)	0.24
Head circumference (mm)	359	−0.07 (−1.05, 0.92)	169	−0.92 (−2.47, 0.62)	190	0.70 (−0.48, 1.88)	0.07
Gestational age (weeks)	375	−0.39 (−1.61, 0.83)	180	−0.43 (−2.34, 1.48)	195	−0.43 (−2.06, 1.20)	0.97
ΣLMWPm							
Weight (g)	371	−10.02 (−42.84, 22.80)	179	−11.55 (−62.08, 38.97)	192	6.16 (−37.64, 49.95)	0.47
Placental weight (g)	325	4.20 (−6.05, 14.46)	154	−1.65 (−19.31, 16.01)	171	9.18 (−3.42, 21.78)	0.18
Length (mm)	364	−0.74 (−2.20, 0.72)	172	−0.24 (−2.39, 1.92)	192	−0.77 (−2.75, 1.20)	0.97
Head circumference (mm)	359	−0.54 (−1.51, 0.43)	169	−0.13 (−1.62, 1.36)	190	−0.62 (−1.85, 0.61)	0.89
Gestational age (weeks)	375	0.27 (−0.94, 1.48)	180	0.38 (−1.48, 2.24)	195	−0.29 (−1.96, 1.39)	0.42

Abbreviations: BPA, bisphenol A; CI, confidence interval; DEHPm, di(2-ethylhexyl) phthalate metabolites; MBzP, mono-benzyl phthalate; LMWPm, low-molecular-weight phthalate metabolites. Betas represent the estimated difference in each outcome associated per doubling of exposure levels (levels were log_2-transformed). All models were adjusted for maternal education, smoking during pregnancy, parity, birth season, and urinary cotinine levels during pregnancy. Head circumference models also adjusted for type of delivery.

5) 고찰 및 결론

- BPA 또는 프탈레이트 노출과 태아성장의 연관성에 대한 증거는 거의 없으나, 프탈레이트 대사물인 MBzP와 MnBP는 일부 태아성장 매개변수와 관련이 있어 더 많은 연구가 필요함.

18. Gascon M, Casas M, Morales E, Valvi D, Ballesteros-Gómez A, Luque N, Rubio S, Monfort N, Ventura R, Martínez D, Sunyer J, Vrijheid M. Prenatal exposure to bisphenol A and phthalates and childhood respiratory tract infections and allergy. J Allergy Clin Immunol. 2015 Feb;135(2):370-8.

1) 배경

- bisphenol A(BPA)와 프탈레이트(phthalates)에 산전 노출된 경우, 어린이의 감염에 대한 감수성이나, 알레르기, 천식에 영향을 줄 수 있을 것으로 생각되지만, 이에 대한 전향적 연구가 부족함.

2) 목적

- bisphenol A(BPA)와 프탈레이트(phthalates)에 대한 산전 노출이 출생 시부터 7세 까지의 어린이에게 호흡기계 및 알레르기의 위험을 증가시킬 수 있는지 여부를 평가하려고 하였음.

3) 방법

(1) 연구대상자

- INMA (INfancia y Medio Ambiente; Childhood and Environment)는 스페인 Sabadell 지역에서 2004년에서 2008년 사이 657명의 임산부를 대상으로 한 출생 코호트 연구임.

- 산전 진찰 첫 방문에 16세 이상, 거점병원에서 출산희망, 외동임신, 자연임신, 의사소통에 문제가 없는 경우 포함시켰음.

(2) 호흡기계 및 알레르기 결과

- 조사자는 자녀가 6개월, 14개월, 4세, 7세가 되던 해에 천명, 흉부 감염 및 습진에 대한 정보를 설문지를 통해 수집함. 기관지염에 대한 정보는 6개월, 14개월, 4세에 수집함.

- 4세가 되던 해에 RAST를 이용하여 특이 IgE를 측정하였고, Dermatophagoides pteronyssinus, 고양이 상피세포, Phleum pratense에 2 kU/L 인 경우, 아토피로 분류되었음.

(3) 노출 변수

- 산모의 소변 샘플은 12주 및 32주에 수집하여 보관하였음.

- 총 BPA를 BPA 노출의 바이오마커로 사용하였고, 프탈레이트 대사산물 8가지(mono-(2-ethyl-5-hydroxyhexyl) phthalate (MEHHP); mono-(2-ethyl-hexyl) phthalate (MEHP); mono-(2-ethyl-5-oxohexyl) phthalate (MEOHP); mono-(2-ethyl-5-carboxypentyl) phthalate (MECPP); mono-benzyl phthalate (MBzP); mono-ethyl phthalate (MEP); monoisobutyl phthalate (MiBP); and mono-n-butyl phthalate (MnBP))를 측정하였음.

- BPA와 프탈레이트 농도는 소변 희석을 조절하기 위해 크레아티닌으로 보정하였고, 제1석달(trimester)과 제3석달의 농도의 평균을 노출 변수로 사용하였음.

- 프탈레이트 대사산물은 서로 다른 물리화학적 성질을 가지기 때문에 원물질(parent of metabolites)에 따라 $\sum_4$DEHP 대사산물(MEHP, MEHHP, MEOHP, MECCP), MBzP 대사산물, $\sum_3$LMWP 대사산물(MEP, MiBP, MnBP)로 나누었음.

(4) 공변량

- 제1석달과 제3석달과 자녀가 14개월이 되었을 때, 설문지를 통하여 공변량에 대한 정보를 수집함(산모 나이, 교육, 출신국가, 임신 중 흡연, 임신 중 간접흡연, 임신 중 애완동물 유무, 나이 많은 형제자매의 수, .생후 첫 1년간 보육원 출석 여부, 완전모유수유 기간, 산모의 통조림 참치 섭취량, 어머니와 아버지의 알레르기, 천식 증상 과거력).

- 산모의 임신 전 체질량지수(BMI), 재태연령, 출생 체중, 출생 계절, 자녀의 성별을 수집함.

(5) 통계분석

- 공변량의 누락값(0%~0.8% 사이)은 다중대체법(multiple imputation methods)을 이용하여 처리하였고, 100 번 이상의 대체를 시행하여 평균을 사용함.

- 657 명의 산모 중 608 명은 관찰기간 동안 적어도 1 번 이상 천명, 흉부 감염, 기관지염, 습진에 대한 정보를 제공하였고, 이 중 462 명은 산전 BPA 노출, 391 명은 산전 프탈레이트 노출 정보가 있었음.

- 474 명의 자녀에서 7 세 때의 천식에 대한 정보가 있었으며, 이 중 361 명은 산전 BPA 혹은 프탈레이트 노출에 대한 정보가 있었음.

- 311 명의 자녀에서 특이 IgE 를 측정하였고, 175 명의 자녀는 BPA 에 대한 정보를, 176 명의 자녀는 프탈레이트 노출에 대한 정보가 있었음.

- 오염물질 농도에 대한 정보가 치우쳐(skewed) 있었기 때문에 노출변수는 $\log_2$ 로 변환 되었음.

- BPA 와 프탈레이트 노출과 7 세까지의 호흡기계 면역의 위험을 보기 위하여 일반화추정방정식(generalized estimating equations)과 비구조적상관행렬(unstructure correlation matrix)과 추적관찰 시의 자녀 나이와 노출변수 사이의 상호작용항(interaction term)를 사용하였음.

- 이 상호작용항은 추적관찰 시의 나이의 차이에 대한 노출 효과의 차이점을 검정하는데 사용하였음.

- 종단연구에서 일반화추정방정식은 서로 다른 시간 지점에서 수집된 결과의 가능한 상관관계(correlation)와 일반화선형모형(generalized linear model)의 매개변수의 추정을 가능하게 하고, 어느 시점에서 불완전한 정보를 가진 대상을 포함시킬 수 있음.

- 관심노출과 단일 시점의 이분결과변수(dichotomous outcome variables)의 연관성을 평가하기 위해 로지스틱회귀모형(Logistic regression models)을 사용하였음.

- 다변량모형(multivariate models)에 포함될 공변량(covariates)를 결정하기 위해 유향 비순환 그래프(DAGs, directed acyclic graphs)를 적용함.

- 과거 연구에서 노출 혹은 결과와 관련이 있다고 기술된 경우 공변량은 retrospective DAGs 에 포함하였고, 이변량 분석을 통해 연관성을 확인하였음 ($p \leq 0.1$).

- DAGs 에 따라 BPA 에 대한 최종 다변량모형은 산모의 교육수준, 자녀의 형제자매 수, 임신 중 흡연을 포함하였고, 프탈레이트 대사산물에 대한 다변량모형은 산모의 임신 전 체질량지수(BMI), 산모의 천식 혹은 알레르기 과거력을 추가로 포함하였음.

- 다양한 노출변수와 결과의 연관성의 선형성(linearity of the association)은 일반화가법모형(GAM, generalized additive models)를 이용하여 평가하였음.

- 비선형성의 증거가 없었기 때문에, BPA 와 프탈레이트의 농도는 연속변수로 취급하였음.

- 민감도 분석(sensitivity analyses)는 각 프탈레이트 대사산물(MEHP, MEHHP, MEOHP, MECPP, MBzP, MEP, MiBP, MnBP)에 대해 수행함.

- 각 노출평가의 역할과 중요성을 차별화하기 위하여 노출변수 BPA, Σ_4DEHP, MBzP, Σ_3LMWP 를 포함한 다중 오염물질 모형(multipollutant model)을 시행하였음.

- 과거 1 개의 연구만이 BPA 노출이 각 임신 석달(trimester)에 미치는 영향을 평가하여 취약기(windows of susceptibility)를 확인했기 때문에 이번 연구에서 2 개의 석달별로 민감도 분석을 시행하였음.

- 통계분석 소프트웨어는 STATA 12.0 과 R 3.0.2.를 사용함.

4) 결과

- 산모의 소변 BPA 농도가 2배가 될 때마다, 천명에 대한 비교위험도는 RR=1.20 (95%CI=1.03-1.40), 흉부 감염은 RR=1.15 (95%CI=1.00-1.32), 기관지염은 RR=1.18 (95%CI=1.01-1.37)이었음.
- Σ_4DEHP 대사산물의 경우, 천명에 대한 비교위험도는 RR=1.25 (95%CI=1.04-1.50), 흉부 감염은 RR=1.15 (95%CI=0.97-1.35), 기관지염은 RR=1.20 (95%CI=1.01-1.43)이었음.
- MBzP의 경우, 천명에 대한 비교위험도는 RR=1.15 (95%CI=1.00-1.33)이었음.

<표 7-38> 모성 소변 BPA와 프탈레이트 대사체 수준과 소아기의 호흡기계 및 알레르기 발현 간의 연관성

TABLE III. Associations between maternal urinary BPA and phthalate metabolite levels* and occurrence of respiratory and allergy outcomes during childhood

	BPA, RR (95% CI)	*P* value	Σ_4DEHP,† RR (95% CI)	*P* value	MBzP, RR (95% CI)	*P* value	Σ_3LMWP,‡ RR (95% CI)	*P* value
Wheeze, chest infections, bronchitis, and eczema from birth until age 7 y§								
No.	462		391		391		391	
Unadjusted								
Wheeze	1.16 (1.00-1.35)	.05	1.29 (1.08-1.54)	.01	1.15 (1.00-1.32)	.05	0.99 (0.86-1.14)	.92
Chest infections	1.10 (0.96-1.26)	.18	1.18 (1.00-1.39)	.04	1.07 (0.95-1.22)	.27	1.02 (0.90-1.16)	.79
Bronchitis	1.12 (0.97-1.30)	.13	1.22 (1.03-1.46)	.02	1.06 (0.92-1.22)	.40	0.96 (0.84-1.11)	.61
Eczema	0.99 (0.84-1.16)	.91	1.05 (0.88-1.26)	.60	1.07 (0.93-1.23)	.37	0.94 (0.82-1.08)	.94
Adjusted								
Wheeze	1.20 (1.03-1.40)	.02	1.25 (1.04-1.50)	.02	1.15 (1.00-1.33)	.05	0.95 (0.82-1.10)	.49
Chest infections	1.15 (1.00-1.32)	.05	1.14 (0.97-1.35)	.11	1.08 (0.95-1.23)	.24	1.00 (0.87-1.13)	.96
Bronchitis	1.18 (1.01-1.37)	.04	1.20 (1.01-1.43)	.04	1.06 (0.92-1.22)	.43	0.95 (0.82-1.09)	.45
Eczema	1.00 (0.85-1.18)	.99	1.00 (0.83-1.20)	.99	1.05 (0.91-1.21)	.51	0.91 (0.79-1.05)	.21
Asthma at age 7 y‖								
No.	361		361		361		361	
Unadjusted	1.22 (0.95-1.56)	.12	1.38 (1.05-1.81)	.02	1.25 (1.02-1.55)	.04	1.07 (0.85-1.35)	.57
Adjusted	1.21 (0.94-1.57)	.14	1.38 (1.05-1.82)	.02	1.26 (1.01-1.82)	.02	1.06 (0.83-1.35)	.63
Atopy at age 4 y¶								
No.	175		176		176		176	
Unadjusted	1.04 (0.64-1.70)	.86	1.07 (0.60-1.90)	.81	0.97 (0.61-1.54)	.89	1.08 (0.69-1.68)	.74
Adjusted	1.07 (0.65-1.77)	.78	1.13 (0.60-2.11)	.71	0.97 (0.61-1.56)	.91	1.11 (0.68-1.81)	.66

BPA models were adjusted for maternal education, number of siblings, and maternal smoking during pregnancy, and phthalate models were additionally adjusted for maternal history of asthma/allergy and maternal body mass index.

*RR per doubling concentration (levels were $\log_2$ transformed).

†The Σ_4DEHP metabolites include MEHHP, MEHP, MEOHP, and MECPP.

‡The Σ_3LMWP metabolites include MEP, MiBP, and MnBP.

§Wheeze, chest infections, and eczema were assessed at ages 6 and 14 months and 4 and 7 years, and bronchitis was assessed at ages 6 and 14 months and 4 years.

‖Children were classified as asthmatic if the mother reported ever doctor-diagnosed asthma at age 7 years, asthma treatment in the last 12 months (at age 7 years), or wheeze in the last 12 months at the age of 7 years plus wheeze in at least 1 of the other previous follow-ups.

¶Children were classified as atopic if they had IgE levels of 2 kU/L or greater to any of the following common allergens: *Dermatophagoides pteronyssinus*, cat epithelium, and *Phleum pratense*.

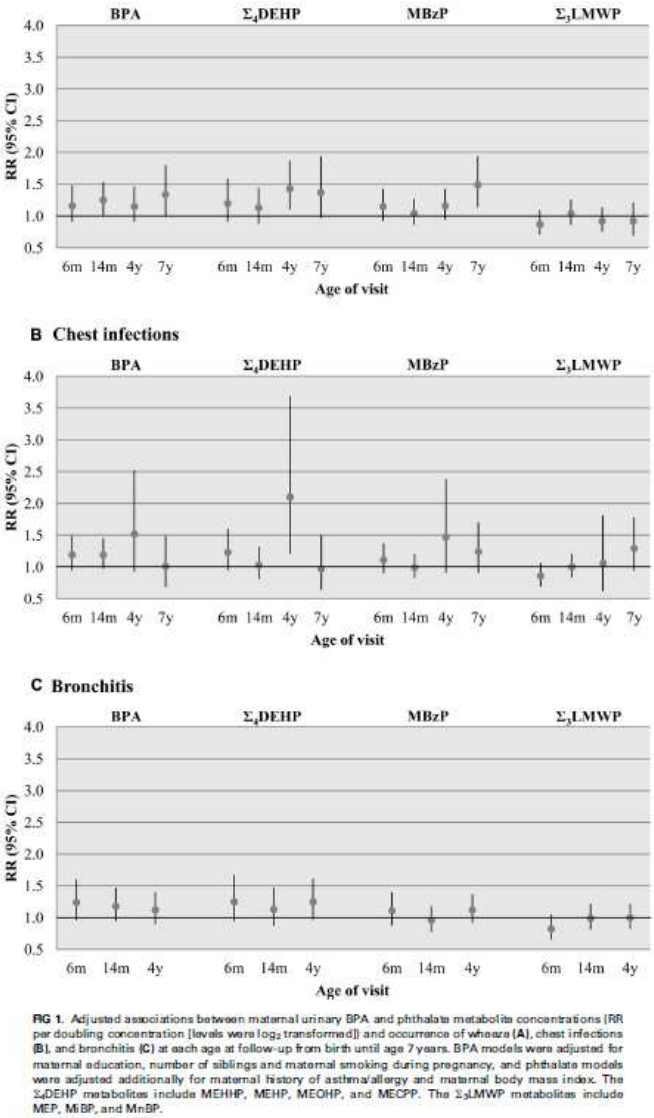


FIG 1. Adjusted associations between maternal urinary BPA and phthalate metabolite concentrations (RR per doubling concentration [levels were log_2 transformed]) and occurrence of wheeze (**A**), chest infections (**B**), and bronchitis (**C**) at each age at follow-up from birth until age 7 years. BPA models were adjusted for maternal education, number of siblings and maternal smoking during pregnancy, and phthalate models were adjusted additionally for maternal history of asthma/allergy and maternal body mass index. The Σ4DEHP metabolites include MEHHP, MEHP, MEOHP, and MECPP. The Σ3LMWP metabolites include MEP, MiBP, and MnBP.

[그림 7-42] 모성 소변 BPA와 프탈레이트 대사체 농도와 천명, 흉부 감염, 기관지염간의 보정된 연관성

5) 고찰 및 결론

- 산전 BPA와 고분자 프탈레이트는 유아기의 천식과 호흡기계 감염의 위험을 상승시킬 수 있음.

19. Maresca MM, Hoepner LA, Hassoun A, Oberfield SE, Mooney SJ, Calafat AM, Ramirez J, Freyer G, Perera FP, Whyatt RM, Rundle AG. Prenatal Exposure to Phthalates and Childhood Body Size in an Urban Cohort. Environ Health Perspect. 2016 Apr;124(4):514-20.

1) 배경

- 프탈레이트 노출은 비만을 증가시킬 것으로 추정 되나, 과거의 연구는 단면연구가 대부분임.

2) 목적

- 산전 프탈레이트 노출과 자녀의 5세, 7세 시점의 체질량지수(BMI)의 연관성을 평가함..

3) 방법

(1) 연구대상자

- CCCEH (Columbia Center for Children's Environmental Health) 종적 출생 코호트의 비만 프로젝트(Obesity Project)에 등록된 어린이들을 대상으로 하였으며, 1998년에서 2006년까지 맨하탄 북부의 뉴욕장로병원(New-York Presbyterian Hospital), 할렘병원센터(Harlem Hospital Center)에서 18세에서 35세 사이의 임산부를 대상으로 모집하였음.

- 현재흡연, 불법약물사용, HIV 환자, 고혈압, 당뇨인 경우는 제외하였으며 제3석달(trimester) 동안 참가자들에게서 사회인구학적 정보를 수집하고 출산 의무기록에서 유아의 성별과 출생 체중을 수집하였음.

- 비만은 연령과 성별에 따른 체질량지수(BMI) 95% 백분위수를 초과한 경우로 정의되었음.

(2) 소변 프탈레이트 측정

- 제 3 석달의 산모와 3 세 및 5 세 때 자녀에서 소변 샘플을 채취하여 di(2-ethylhexyl) phthalate (DEHP)의 대사산물인 MEHP, MEHHP, MEOHP, MECPP 와 di-isobutyl phthalate (DiBP), di-n-butyl phthalate (DnBP), butylbenzyl phthalate (BBzP), diethyl phthalate (DEP), and di-n-octyl phthalate (DOP)의 각 대사산물인 MiBP, MBP, MBzP, MEP, MCPP 에 대하여 분석함.

(3) 데이터분석

- 모든 분석은 SPSS (PASW® Statistics 21)와 SAS 9.3 을 이용하여 수행하였음.

- CDC 의 권고에 따라 MEP 와 MBzP 에는 각 0.66 과 0.72 를 곱하여 보정하였고 소변 프탈레이트 대사산물 농도는 자연로그(ln)를 취해 분석함.

- 로그변환된 소변 대사산물의 농도의 상관관계는 피어슨의 상관관계 분석(Pearson product moment correlation analyses)을 사용하여 평가하였고, 소변 프탈레이트 대사산물 농도가 중간에서 강한 상관관계(0.34-0.99)를 보여서 다중 비교(multiple comparisons)를 줄이기 위한 프탈레이트 노출 패턴의 요약값(summary measures)을 만들기 위해 주성분분석(PCA, principal component analysis)를 시행하였음.

- 복합화학혼합물로부터 노출패턴을 확인하기 위하여 산모 및 자녀(3 세, 5 세)의 소변 대사산물 농도를 로그변환하여 베리맥스 회전(varimax rotation)을 이용하여 주성분분석함.

- 성분점수(component scores)는 산모 및 자녀(3 세, 5 세)에 계산되었으며 대상의 대사산물이 대사산물 패턴과 얼마나 일치하는지를 나타냄.

- PCA 결과는 상관구조(correlation structure)에 의존하기 때문에 PCA 및 상관관계를 2003 년과 2004 년 NHANES 프탈레이트 소변 대사산물 정보에 적용하여 국가 데이터와 일치하는지 확인하였다.

- 2003-2004 NHANES (National Health and Nutritional Examination Survey)에 참여한 15 세-45 세 여성의 소변 프탈레이트 대사산물 농도(MEHP, MEHHP, MECPP, MEOHP, MCPP, MiBP, MBP, MBzP, MEP)를 얻어 로그변환하여 베리맥스 회전하여 주성분분석하였고, NHANES 복합설문조사 가중치가 분석에 포함되었음.

- 산모 소변 프탈레이트 농도의 성분점수가 자녀(5 세, 7 세)의 BMI z-score 와 연관성이 있는지 여부를 확인하기 위해 일반화추정방정식(GEE, generalized estimating equation model)를 선형연결함수(linear link function), 비구조적 상관구조(unstructured correlation structure)를 이용하여 수행함.

- 산모 소변 프탈레이트 농도 성분점수가 자녀(7 세)의 체지방률, 체지방지수(FMI, fat mass index), 허리둘레와 연관성이 있는지 확인하기 위해 선형회귀분석(linear regression analyses)을 사용하였음.

- 모든 모형은 성분점수에 의거하여 산모의 인종/민족, 임신 중 공적부조(receipt of public assistance) 여부, 산전 비만상태, 출생 체중, 자녀 나이, 산모 요비중(urine specific gravity) z-score 에 대하여 보정되었음.

- 분석은 남, 녀따로 수행되었고, 이차분석(seconday analyses)은 계층화된 분석에서 유의한 연관성($p < 0.05$)이 확인되었을 때 이루어졌으며 인체측정결과와 개인의 프탈레이트 대사산물 간의 관계를 평가하였음.

- 3 세, 5 세의 자녀에서 프탈레이트 노출에 의한 잠재적 교란효과(potential confounding effects)를 확인하기 위해 민감도 분석(sensitivity analyses)을 시행하였음.

- 산모와 자녀의 프탈레이트 노출 사이의 가능한 교란을 평가하기 위해 어머니의 성분점수와 자녀(3 세 및 5 세)의 성분점수의 상관관계를 평가하여 GEE 모형에 추가함.

- 산모 소변 프탈레이트 성분점수와 5 세, 7 세의 추적관찰 사이의 가능한 상호작용은 성분점수 상호작용항(interaction term)을 모형에 포함시켜 평가하였음.

- 불완전한 추적관찰로 인한 비뚤림(bias)를 평가하기 위해 역확률가중치(inverse probability weights)를 계산하여 GEE 모형에 포함시킴.

4) 결과

- 주성분분석을 통해 DEHP 와 non-DEHP 를 확인하였고, 남아에서 높은 산모의 non-DEHP 성분점수는 낮은 BMI z-score, 낮은 체지방률, 작은 허리 둘레와 관련이 있었으나, 여아에서는 연관성을 보이지 않음. 성별과 non-DEHP 의 연관성의 상호작용은 통계적으로 유의했음($p<0.01$).

<표 7-39> 모성 산전 프탈레이트 대사체 요소 점수와 5 세와 7 세 소아의 체질량지수(BMI) z-점수 간의 연관성

Table 3. Associations between maternal prenatal urinary phthalate metabolite component scores and child BMI *z*-scores at ages 5 and 7 years.

	Model 1[a]			Model 2[b]			Model 3[c]		
Phthalate component score	Girls (n = 181) β (95% CI) *p*-value	Boys (n = 156) β (95% CI) *p*-value	*p*-int[d]	Girls (n = 127) β (95% CI) *p*-value	Boys (n = 113) β (95% CI) *p*-value	*p*-int[d]	Girls (n = 154) β (95% CI) *p*-value	Boys (n = 124) β (95% CI) *p*-value	*p*-int[d]
Prenatal DEHP component score	−0.10 (−0.29, 0.09) 0.30	−0.08 (−0.30, 0.13) 0.45	ND[e]	−0.16 (−0.42, 0.09) 0.21	−0.06 (−0.34, 0.22) 0.66	ND[e]	−0.12 (−0.31, 0.08) 0.25	0.00 (−0.24, 0.24) 0.99	ND[e]
Prenatal non-DEHP component score	0.07 (−0.18, 0.31) 0.58	−0.30 (−0.50, −0.10) 0.003	0.003	0.16 (−0.17, 0.48) 0.35	−0.31 (−0.56, −0.07) 0.01	0.003	0.13 (−0.14, 0.41) 0.34	−0.30 (−0.54, −0.06) 0.01	0.001
Age 3 phthalate component score	NA	NA	NA	0.02 (−0.34, 0.37) 0.92	0.00 (−0.29, 0.29) 0.99	ND[e]	NA	NA	NA
Age 5 DEHP component score	NA	NA	NA	NA	NA	NA	−0.01 (−0.20, 0.23) 0.91	−0.13 (−0.38, 0.13) 0.33	ND[e]
Age 5 non-DEHP component score	NA	NA	NA	NA	NA	NA	−0.12 (−0.37, 0.14) 0.36	−0.34 (−0.58, −0.10) 0.01	ND[e]

Abbreviations: DEHP, di-(2-ethylhexyl) phthalate; NA, not applicable; ND, not done.
[a]GEE-based analyses were used with BMI *z*-score data from ages 5 and 7 years combined; the β coefficients reflect the average difference in BMI *z*-score per unit difference in component score. Model 1 adjusts for age (in months) at time of measurement, maternal prepregnancy obesity, birth weight, maternal race/ethnicity, maternal receipt of public assistance during pregnancy, and urinary specific gravity. The DEHP and non-DEHP component score variables are entered into the same regression model and thus the results are mutually adjusted. [b]Model 2 additionally adjusts for child age 3 urinary metabolite concentration component score. [c]Model 3 additionally adjusts for child age 5 urinary metabolite concentration DEHP and non-DEHP component scores. [d]*p* for interaction with sex. [e]Interaction test not performed because stratified analyses did not suggest heterogeneity of effect sizes by child sex.

<표 7-40> 모성 산전 소변 프탈레이트 요소 지수와 7 세 소아의 신체 조성과 허리 둘레 간의 연관성

Table 4. Associations[a] between maternal prenatal urinary phthalate component scores and child body composition and waist circumference outcomes at age 7 years.

	Percent body fat			Fat mass index			Waist circumference		
Phthalate component	Girls (n = 164) β (95% CI) p-value	Boys (n = 142) β (95% CI) p-value	p-int[b]	Girls (n = 164) β (95% CI) p-value	Boys (n = 142) β (95% CI) p-value	p-int[b]	Girls (n = 154) β (95% CI) p-value	Boys (n = 124) β (95% CI) p-value	p-int[b]
Prenatal DEHP component score	−0.14 (−1.19, 0.92) 0.80	−0.31 (−1.56, 0.94) 0.63	ND[c]	−0.01 (−0.42, 0.39) 0.96	−0.11 (−0.56, 0.34) 0.63	ND[c]	−0.11 (−1.46, 1.25) 0.88	−0.65 (−2.26, 0.97) 0.43	ND[c]
Prenatal non-DEHP component score	0.62 (−0.64, 1.88) 0.33	−1.62 (−2.91, −0.34) 0.01	0.002	0.34 (−0.15, 0.82) 0.17	−0.50 (−0.96, −0.04) 0.03	0.003	1.20 (−0.43, 2.84) 0.15	−2.02 (−3.71, −0.32) 0.02	0.0002

Abbreviations: DEHP, di(2-ethylhexyl) phthalate; ND, not done.
[a]All analyses adjust for age (in months) at time of measurement, maternal prepregnancy obesity, birth weight, maternal race/ethnicity, receipt of public assistance during pregnancy, and urinary specific gravity. The DEHP and non-DEHP component score variables are entered into the same regression model and thus the results are mutually adjusted. The β coefficient reflects the average difference in the outcome variable per unit difference in component score. [b]p for interaction with sex. [c]Interaction test not performed because stratified analyses did not suggest heterogeneity of effect sizes by child sex.

<표 7-41> 모성 소변 프탈레이트 대사체와 소아 신체계측 결과 간의 연관성

Table 5. Associations[a] between maternal urinary phthalate metabolites and child anthropometric outcomes.

	BMI z-score at ages 5 and 7[b]		Percent body fat at age 7		FMI at age 7		Waist circumference at age 7	
Maternal phthalate urinary metabolite	Girls (n = 181) β (95% CI) p-value	Boys (n = 156) β (95% CI) p-value	Girls (n = 164) β (95% CI) p-value	Boys (n = 142) β (95% CI) p-value	Girls (n = 164) β (95% CI) p-value	Boys (n = 142) β (95% CI) p-value	Girls (n = 154) β (95% CI) p-value	Boys (n = 124) β (95% CI) p-value
Molar sum of DEHP metabolites	−0.08 (−0.25, 0.09) 0.37	−0.09 (−0.28, 0.11) 0.39	−0.13 (−1.09, 0.84) 0.80	−0.39 (−1.57, 0.79) 0.51	−0.03 (−0.40, 0.34) 0.88	−0.13 (−0.55, 0.29) 0.54	−0.13 (−1.37, 1.12) 0.84	−0.65 (−2.16, 0.87) 0.40
MCPP	0.10 (−0.17, 0.37) 0.46	−0.24 (−0.47, −0.01) 0.04	0.08 (−1.07, 1.23) 0.89	−0.63 (−1.86, 0.60) 0.32	0.14 (−0.30, 0.58) 0.52	−0.19 (−0.63, 0.25) 0.40	0.70 (−0.76, 2.16) 0.34	−0.88 (−2.51, 0.76) 0.29
MiBP	0.02 (−0.18, 0.22) 0.86	−0.17 (−0.36, 0.01) 0.06	0.78 (−0.33, 1.88) 0.17	−0.97 (−2.10, 0.17) 0.09	0.35 (−0.08, 0.77) 0.11	−0.29 (−0.70, 0.11) 0.16	1.12 (−0.34, 2.58) 0.13	−1.18 (−2.66, 0.31) 0.12
MBP	0.11 (−0.12, 0.34) 0.33	−0.22 (−0.44, 0.00) 0.05	−0.52 (−0.72, 1.76) 0.41	−1.05 (−2.26, 0.15) 0.09	0.32 (−0.15, 0.80) 0.18	−0.34 (−0.78, 0.09) 0.12	0.85 (−0.76, 2.47) 0.30	−1.34 (−2.91, 0.23) 0.10
MBzP	−0.04 (−0.18, 0.10) 0.56	−0.16 (−0.31, −0.01) 0.04	0.03 (−0.81, 0.87) 0.95	−0.75 (−1.64, 0.15) 0.10	0.03 (−0.30, 0.35) 0.87	−0.18 (−0.50, 0.14) 0.26	0.10 (−1.03, 1.19) 0.86	−0.70 (−1.87, 0.48) 0.24
MEP	0.02 (−0.15, 0.19) 0.83	−0.04 (−0.19, 0.10) 0.56	0.33 (−0.61, 1.27) 0.49	−0.86 (−1.78, 0.07) 0.07	0.12 (−0.24, 0.49) 0.50	−0.32 (−0.65, 0.01) 0.06	0.63 (−0.56, 1.83) 0.30	−1.27 (−2.46, −0.08) 0.04

Abbreviations: BMI, body mass index; FMI, fat mass index; MBP, mono-n-butyl phthalate; MBzP, monobenzyl phthalate; MCPP, mono(3-carboxypropyl) phthalate; MEP, monoethyl phthalate; MiBP, mono-isobutyl phthalate.
[a]All analyses adjust for age (in months) at time of measurement, maternal prepregnancy obesity, birth weight, maternal race/ethnicity, receipt of public assistance during pregnancy, and urinary specific gravity. Metabolite concentrations were ln-transformed for analyses. The β coefficient reflects the average difference to in the outcome variable per unit difference in the ln-transformed metabolite concentration. [b]GEE-based analyses were used with BMI z-score data from ages 5 and 7 years combined.

5) 고찰 및 결론

- 가설과는 반대로, 산전 non-DEHP 프탈레이트 노출은 유아기의 남아에서 낮은 BMI z-score, 허리 둘레, FMI 와 관련이 있었음.

20. Talbott EO, Marshall LP, Rager JR, Arena VC, Sharma RK, Stacy SL. Air toxics and the risk of autism spectrum disorder: the results of a population based case-control study in southwestern Pennsylvania. Environ Health. 2015 Oct 6;14:80.

1) 배경

- 자폐스펙트럼장애(ASD, autism spectrum disorders)는 68 명의 어린이 중 한명에게 영향을 미치는 중요한 공중보건 문제임.

2) 목적

- ASD 의 원인에 대한 이해가 부족하여 대기오염물질 중의 신경독소와의 연관성을 평가하기 위하여 남서부 펜실베니아주에서 환자-대조군 연구를 수행하였음. .

3) 방법

(1) 환자의 선정

- 펜실베니아 남서부의 Allergheny, Armstrong, Beaver, Butler, Washington, Westmoreland 자치주에서 2005 년 1 월 1 일에서 2009 년 12 월 31 일 사이 태어난 ASD 환아가 연구에 포함되었음.

- 2007 년 6 개 자치주에서 23,399 명의 출생을 기준으로 하여, 유병률이 1,000 명당 6 명이므로, 연간 연구지역에서 140-141 명의 ASD 환자가 진단할 것으로 추정하였고, 이중 절반이 연구에 참여할 것으로 예상하였음.

- 펜실베니아에는 자폐증 등록기관이 없어 ASD 환자의 학부모에게 연락할 수 있는 허가를 받을 수 있는 중앙기관이 없어, ASD 환자를 모집하기 위하여 광범위한 봉사활동 캠페인을 사용함.

- 자녀가 입양되었거나, 부모가 영어 혹은 면담이 불가능한 경우는 연구에 포함되지 않았고, 총 217 명의 환자가 동의를 하여 연구를 위해 면담하였음.

(2) 대조군의 선정

- 대조군은 2 종류로 설계되었음. 첫 번째 대조군(Interviewed controls)은 2005 년에서 2009 년 사이 PADOH (Pennsylvania Department of Health)에서 5,007 명(남 4:여 1 비율)을 모집하여 출생년도, 성별, 인종에 따라 매칭하여 226 명을 선발하였음.

- 설문지를 통해 훈련된 면담자가 부모의 인구학적 및 사회경제적 정보, 상세주거력, 부모의 직업력, ASD 가족력, 흡연력, 산모의 생식 및 임식력, 자녀의 병력을 수집하였고, 마지막 생리기간 3 개월 전부터 자녀가 2 세가 된 해까지 거주한 주소에 대한 정보를 수집함.

- 두 번째 대조군(BC, birth certificate controls)은 2005 년에서 2009 년 사이 출생증명서를 통해 5,007 명(남 4:여 1 비율)을 모집하였으나 16 명이 ASD 환자로 분류되어 제외되었고, 출생증명서를 통해 출생 시 거주지, 산모 나이, 흡연력, 산모의 교육수준, 인종 및 기타 특성에 대한 정보를 수집하였음.

(3) 노출평가

- 2005 년 NATA (National Air Toxics Assessment) 평가에서 얻은 모형데이터를 이용하여 대기 중 유해 대기오염 농도를 추정하였음.

- NATA 를 통해 입수할 수 있는 177 개의 공기 독성물질 중 미국 환경보호청(US EPA)에서 신경학적 혹은 발달, 내분비교란 효과가 있는 37 개의 독성 물질의 분포, 가변성, 상관관계를 조사하였음.

- 37 개의 독성 물질 중 7 개(carbon tetrachloride, chloroform, ethylene dibromide, ethylene dichloride, hexachlorbenzene, methyl chloride, PCBs)는 분포가 다양하지 않아 분석에서 제외하여 총 30 개 독성 물질에 대하여 분석하였음.

- 환자 대조군의 분석을 위해 거주지 주소를 ArcGIS (ver. 10.1; ESRI Inc., Redlands, CA)를 통해 X, Y 좌표로 지역코딩하였고 수동으로 확인하였고, 성공적인 코딩이 안 된 경우, MapQuest Latitude/Lonitude Finder를 포함한 다른 방법을 이용함.

- 2000년 인구조사 보고서(11자리 FIPS 코드)는 ArcGIS 10.1을 이용하여 할당되었고 2000년 미국 인구조사에 대한 2009 Tiger Line 파일에 연결되었음.

- 거주지의 위치와 변화 및 거주지에서 보낸 시간을 고려하여 공기 독성물질에 각각에 대한 개인 별 노출 추정치를 계산하였고, 평균 노출 추정치는 임신기간 및 생후 첫해 및 2번째 해에 대하여 계산되었음.

- 외국 거주로 인해 NATA 정보를 사용할 수 없었던 2명의 참가자는 분석에서 제외되어 217명의 환자와 224명의 대조군이 분석 그룹에 포함됨.

- 출생증명서 데이터 분석을 위해 NATA 농도를 출생 시 거주지의 인구조사 보고서와 연결하였음.

- BC 대조군의 분석에서는 16명이 환자였고, 20명이 NATA 노출이 할당이 되지 않았고, 환자군에서 1명이 출생증명서와 연결이 되지 않아 총 216명의 환자와 4,971명의 대조군이 최종 분석 대상이 됨.

(4) 통계분석

- NATA 대기오염물질 노출과 ASD에 대한 연관성을 조사하기 위해 로지스틱 회귀(logistic regression)를 사용하였고, 개인 교차비를 계산하기 위해 30개의 NATA 오염물질에 대하여 사분위 절단점을 계산하여 상위 사분위와 하위 사분위와 비교하였음.

- 환자군과 면담 대조군의 경우, 각 오염물질에 대하여 임신기간에 로지스틱회귀모형(seperate logistic regression models)을 수행하였고, 생후 첫해 및 2번째 해에 대하여 이차적으로 수행하였음.

- 환자군과 BC 대조군의 경우, 출생 시의 거주지 정보만 활용 가능하였음.

- 모든 분석은 산모 나이, 교육수준, 인종, 흡연, 자녀의 출생연도 및 성별에 따라 보정되었음.

- 화합물을 개별적으로 검사하는 것 외에도, 구조적 특성에 따라 3종류로 분류하였음: 셀레늄을 제외한 금속(비소, 카드뮴, 크로뮴, 납, 수은, 니켈), 방향족 용제(벤젠, 에틸벤젠, 스티렌, 톨루엔, 크실렌), 염화용제(methylene chloride, perchloroethyle, trichloroethylene, trichloroethane, vinyl chloride)

- 금속, 방향족 용제, 염화용제의 각 그룹에 따라 색인점수(index scores)를 계산하여 면담 대조군의 사분위 절단점을 계산함.

- 면담 대조군과 BC 대조군에서 어머니의 나이, 교육수준, 인종, 흡연, 자녀의 출생연도 및 성별을 통제하면서 각 지수별로 사분위와 낮은 사분위를 비교하기 위해 로지스틱회귀모형을 IBM SPSS Statistics 20, 22 를 사용하여 수행함.

- 대조군에 비해 환자군에서 다태분만(multiple births)이 많았는데, 다태분만은 조산 및 다른 문제와 관련이 높으므로, 다태분만을 포함 혹은 제외하고 민감도 분석(sensitivity analysis)을 시행하였음.

- 이전의 개별 오염물질에 대하여 로지스틱회귀모형에서 사용된 동일 공변량을 제어하면서 NATA 화합물의 가장 유의미한 효과를 고려하기 위해, 어머니의 나이, 교육수준, 인종, 흡연, 자녀의 출생연도 및 성별을 보정한 환자-대조군 비교에서 유의하다고 확인된 모든 물질에 대하여 후진다중 로지스틱분석(backward multiple logistic analysis)을 시행하였음.

- 공기 독성물질은 종종 상관관계가 있고, 사람들은 보통 복합혼합 대기 오염물질에 노출되는데, Spearman 상관행렬(correlation matrix)은 많은 공기 독성물질이 높은 상관관계가 있음을 보여 주었음.

- von Ehrenstein 등이 기술한 방법론과 마찬가지로, 30 개 공기 독성물질의 상관구조를 분석하기 위해 인자분석(factor analysis)을 수행하였고, 인자는 주성분분석(PCA, principal component analysis)과 베리맥스 회전(varimax rotation)을 이용하여 추출되었고, 고유값(eigenvalue) >1 에 따라 유지를 결정하였음.

4) 결과

- 4 사분위와 1 사분위 노출을 비교할 때, 면담 대조군에서 aOR (adjusted odds ratios)=2.04 (95%CI=1.17-3.58), BC 대조군에서 aOR=1.61 (95%CI=1.08-2.40)이었음.

- 크로뮴의 경우 BC 대조군에서 OR=1.60 (95%CI=1.08-2.38), 면담 대조군에서 OR=1.52 (95%CI=0.87-2.66)이었음.

- methylene chloride (OR=1.41, 95%CI=0.96-2.07)와 PAHs (OR=1.44, 95%CI=0.98-2.11)는 통계적 유의미의 경계선에 있었음.

<표 7-42> 임신 중 국가 공기 독성물질(NATA) 노출과 연관된 자폐스팩트럼장애의 보정된 오즈비

Table 4 Adjusted[a] OR (95 % CI) for ASD associated with exposure during pregnancy to NATA air toxics (fourth vs first quartile)

Pollutant	Interviewed (217 Cases and 224 Controls)		Birth Certificate (215 Cases and 4,856 Controls)	
	All Births	Singleton Births	All Births	Singleton Births
Arsenic	1.28 (0.71–2.28)	1.38 (0.75–2.52)	1.33 (0.87–2.02)	1.62 (1.03–2.54)**
Chromium	1.52 (0.87–2.66)	1.58 (0.88–2.81)	1.60 (1.08–2.38)**	1.65 (1.09–2.50)**
Methylene chloride	1.07 (0.61–1.85)	1.33 (0.74–2.36)	1.41 (0.96–2.07)*	1.63 (1.08–2.45)**
PAHs	1.33 (0.76–2.32)	1.56 (0.87–2.77)	1.44 (0.98–2.11)*	1.64 (1.09–2.46)**
Styrene	2.04 (1.17–3.58)**	2.23 (1.24–4.00)**	1.61 (1.08–2.40)**	1.67 (1.09–2.54)**
Lead	1.10 (0.63–1.94)	1.21 (0.67–2.18)	1.33 (0.90–1.97)	1.52 (1.00–2.30)**
Cyanide	1.19 (0.67–2.12)	1.32 (0.73–2.36)	1.36 (0.92–2.01)	1.52 (1.01–2.30)**

[a]Models adjusted for maternal smoking, age, education, race, child's birth year, and sex of the child
**Significant (p-value < 0.05) *$p < .10$

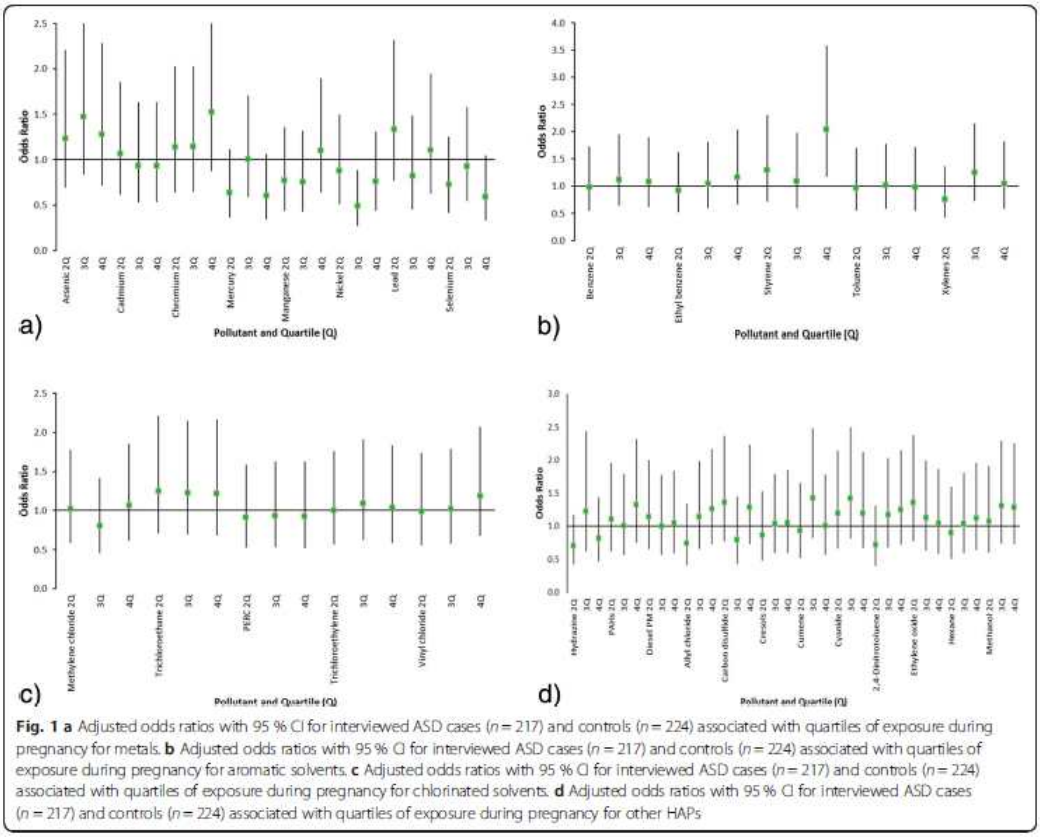


Fig. 1 **a** Adjusted odds ratios with 95 % CI for interviewed ASD cases ($n = 217$) and controls ($n = 224$) associated with quartiles of exposure during pregnancy for metals. **b** Adjusted odds ratios with 95 % CI for interviewed ASD cases ($n = 217$) and controls ($n = 224$) associated with quartiles of exposure during pregnancy for aromatic solvents. **c** Adjusted odds ratios with 95 % CI for interviewed ASD cases ($n = 217$) and controls ($n = 224$) associated with quartiles of exposure during pregnancy for chlorinated solvents. **d** Adjusted odds ratios with 95 % CI for interviewed ASD cases ($n = 217$) and controls ($n = 224$) associated with quartiles of exposure during pregnancy for other HAPs

[그림 7-43] 임신중 금속, 방향족 용제, 염화 용제, 유해공기오염물질과 연관된 면담한 자폐스펙트럼장애 환자군(n=217)과 대조군(n=224)의 보정된 오즈비

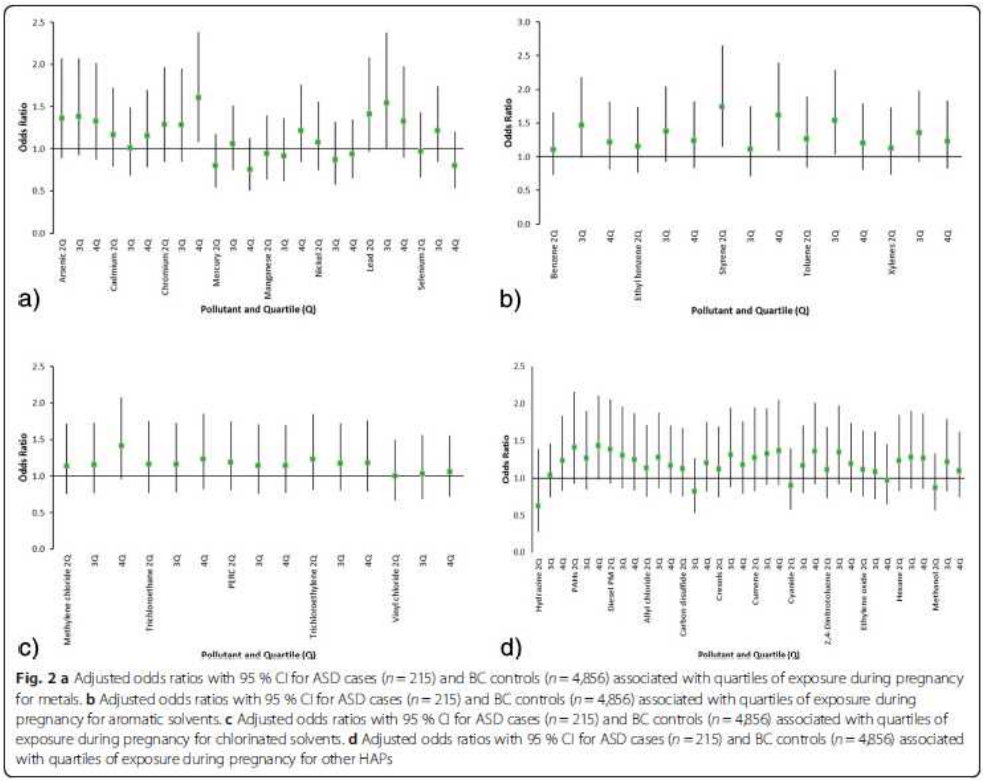


Fig. 2 a Adjusted odds ratios with 95 % CI for ASD cases ($n = 215$) and BC controls ($n = 4,856$) associated with quartiles of exposure during pregnancy for metals. **b** Adjusted odds ratios with 95 % CI for ASD cases ($n = 215$) and BC controls ($n = 4,856$) associated with quartiles of exposure during pregnancy for aromatic solvents. **c** Adjusted odds ratios with 95 % CI for ASD cases ($n = 215$) and BC controls ($n = 4,856$) associated with quartiles of exposure during pregnancy for chlorinated solvents. **d** Adjusted odds ratios with 95 % CI for ASD cases ($n = 215$) and BC controls ($n = 4,856$) associated with quartiles of exposure during pregnancy for other HAPs

[그림 7-44] 임신중 금속, 방향족 용제, 염화 용제, 유해공기오염물질과 연관된 면담한 자폐스펙트럼장애 환자군(n=215)과 출생증명서(BC) 대조군(n=4,856)의 보정된 오즈비

5) 고찰 및 결론

- 임신기간 동안 스티렌과 크로뮴의 농도가 높은 지역에서 사는 것은 ASD 위험 증가와 관련이 있고, PAHs 와 methylene chlroide 도 위험 증가의 경계선에 있음. 이번 결과는 과거의 연구와 같은 결과를 보이지만, 각 화학물질이 각기 위험 인자인지, 혹은 오염물질의 혼합 효과에 의한 것인지는 확실하지 않음. 향후의 연구에서는 개인의 일상생활 중의 이동을 고려하여 공기 독성물질의 시공간적 평가가 향상되어야 함.

21. Kim TW, Koh DH and Park CY. Decision Tree of Occupational Lung Cancer Using Classification and Regression Analysis. Safety Health Work 2010;1:140-148.

1) 배경

- 전 세계적으로 폐암은 암에 의한 사망원인 중 가장 높고 직업과 관련된 암중 가장 흔함.
- 직업관련 폐암이 다른 암과 조직학적이나 병리적 형태, 임상징후, 결과에는 차이가 없고 직업적 노출에 의한 발생이라는 것 밖에 없어 직업성과 비직업성 폐암을 구분하는 것은 매우 어려움.
- 한국에서는 직업병으로 승인된 근로자만 근로복지공단으로부터 보상을 받기 때문에 질병의 직업 관련성을 결정하는 것은 근로자와 근로복지공단에 매우 중요한 문제임. 이 중요성에도 불구하고 의사결정과정에 대한 포괄적인 연구가 부족함.

2) 목적

- 폐암의 업무관련성에 영향을 미치는 변수를 찾고, 업무관련성을 객관적으로 결정하는데 도움을 주기 위하여 직업성 폐암의 결정 tree를 도출 하고자 함

3) 연구방법

(1) 연구자료 및 연구대상자

- 1992년부터 2007년까지 폐암 160건이 근로복지공단에서 업무관련성의 결정을 위해 요청됨. 그 중 6건은 연기되었고 1건은 역학조사 보고서가 포함되어 있지 않아서 제외함 총 153건이 연구에 포함됨.

(2) 노출평가

- 산업안전보건연구원의 폐암에 대한 각 역학조사 보고서를 리뷰하고 직업성 폐암에 영향을 미칠 수 있는 결정요인 15개의 변수를 찾아

냈으나 그 중 8개의 변수만 포함시킴. 이 변수들은 폐암의 업무관련성 (직업성 vs 비직업성 폐암), 인구학적 특성 (성별, 폐암 진단시기의 나이), 건강습관 (흡연력), 임상요인 (조직학적 분류), 직업적 요인 (산업분류, 노출되었거나 노출 가능성 있는 물질, 잠재기, 업무기간) 임.

- 직업적 폐암은 직업과 관련 있는 일차성 폐암으로 정의함. 산업분류는 국제표준 산업 분류(ISIC Rev 4.)를 이용하여 제조업과 비제조업으로 구분함. 여러 노출 물질 혹은 노출 되었을 물질들 중 주요한 한 가지 폐 발암성물질을 노출 물질로 선택함. 국제암 연구소(IARC)의 분류법에 따라 이러한 노출 물질을 폐에 대한 발암성이 알려진 물질(발암성, IARC 그룹 1)과 발암성이 알려지지 않은 물질(상당한(probably), 가능성 있는(possibly), 발암성 없는, IARC 그룹 2A, 2B, 4)로 구분함. 잠재기는 근로자가 처음 노출물질에 노출된 날부터 진단 받은 날까지의 기간으로 정의함. 노출 기간이 정확하지 않다면 잠재기는 처음 일한 날부터 진단 받은 날까지의 기간으로 정의함. 업무 기간은 근로자가 알려지거나 알려지지 않은 발암물질에 노출된 사업장에 고용된 기간으로 산정함.

(3) 통계분석 방법

- 진단 시 나이, 흡연력, 잠재기, 일한 기간은 t-test로 분석하고, 성별, 조직학적 분류, 산업분류는 chi-square test로 분석함.

- 직업적 폐암 결정 tree를 도출하기 위해서 분류와 회귀 나무 분석(Classification and Regression Tree, CART)을 사용함. CART는 나무 구축 기술이고 임상 의사결정 규칙을 생성하는데 적합함. 범주형 예측변수 뿐만 아니라, 심하게 치우치(highly skewed)거나 다중 모드의 수치 데이터(multi-modal numerical data)를 다룰 수 있는 비모수 기술임. 설명된 결과변수 결정에 가장 중요한 예측변수들 사이의 복잡한 상호작용을 발견할 수 있음. CART 분석법은 비교적 자동 머신러닝 기법으로, 해석을 쉽게 할 수 있는 의사결정 나무를 생산함.

- CART 분석에서 직업성폐암의 결정은 target 변수로 선택되었고

7개의 변수는 혼란변수로 포함됨. tree의 검증을 위하여 대상을 무작위로 표본자료(122건, 80%)와 검증자료(31건, 20%)로 나눔. 의사결정 tree는 표본자료로부터 도출되었고, 검증자료에 의해 평가됨. 다변량 로지스틱 회귀분석은 표본자료로 구성되었고, 검증자료로 테스트함. CART와 로지스틱 회귀분석의 정확성 비교를 위하여 두 모델의 receiver operating characteristic curves(ROC) 아래 면적(AUCs)이 표본자료와 검증자료에서 사용됨.

4) 결과

○ 일반적 특성

- 153건 중 57건(37.3%)이 직업성폐암으로 결정된 건이었고, 96건(62.7%)이 비직업성 폐암으로 결정된 건이었음. 두 그룹의 평균 나이(50.8 vs 50.3)는 비슷함. 남성은 비직업성 암보다 직업성 암의 큰 비율을 차지하지만 유의하지 않음. 비직업성 폐암의 흡연력은 직업성 폐암보다 5갑년 더 많았지만 유의하지 않았음. 직업성 폐암의 잠재기는 비직업성폐암보다 겨의 7년 더 길었고, 직업성 폐암의 업무기간은 6년 더 길었음. 산업분류에 따라 비제조업에서 직업성폐암의 비율은 제조업에서보다 유의하게 더 높았음. 조직학적 분류에 따르면 adenocarcinoma가 가장 흔했고, SCC(squamous cell carcinoma), SCLC(small cell lung carcinoma) 순이었음.

<표 7-43> 연구 대상자의 직업적 및 비-직업적 폐암에 따른 기술적 특성

Table 1. Descriptive characteristics of the survey populations by occupational and non-occupational lung cancer

Variables	Occupational lung cancer (n = 57)	Non-occupational lung cancer (n = 96)	p-value
Age (Mean ± SD, years)	50.8±7.7	50.3±8.7	0.728
Gender (%)			0.620
Male	55 (39.6)	84 (60.4)	
Female	2 (14.3)	12 (85.7)	
Smoking (pack-years)	12.2 ± 14.7	17.3 ± 17.1	0.089
Latent period‡ (years)	22.8 ± 8.7	15.5 ± 9.5	< 0.001
Working period§ (years)	21.0 ± 7.3	14.7 ± 9.1	< 0.001
Industry (%)			0.003
Manufacturing	30 (29.8)	73 (70.2)	
Non-manufacturing	26 (53.1)	20 (46.9)	
Histological type (%)			0.754
Adenocarcinoma	22 (37.9)	36 (62.1)	
SCC*	16 (37.2)	27 (62.8)	
SCLC†	5 (33.3)	10 (66.7)	
Others	2 (66.7)	1 (33.3)	

○ 노출물질

- 본 연구에서 24종의 노출물질이 있었음. 석면(45건, 29.4%)이 가장 많았고, 결정형 실리카 (21건, 13.7%), 먼지(18건, 11.8%) 순임. 알려진 폐 발암성물질은 석면, 결정형실리카, 6가크롬, PAHs, 라돈, 코크스 오븐 배출물질(COE) 이었음. 알려진 폐 발암성물질에 노출된 케이스는 92건 이었음. 6가 크롬에 노출된 케이스는 직업성폐암의 가장 높은 비율(71.4%)을 차지했고, PAHs에 노출된 근로자는 가장 낮은 비율(50%) 차지함. 라돈과 COE에 노출된 케이스는 각각 1건 임.

<표 7-44> 직장에서 일하는 중 노출되는 물질

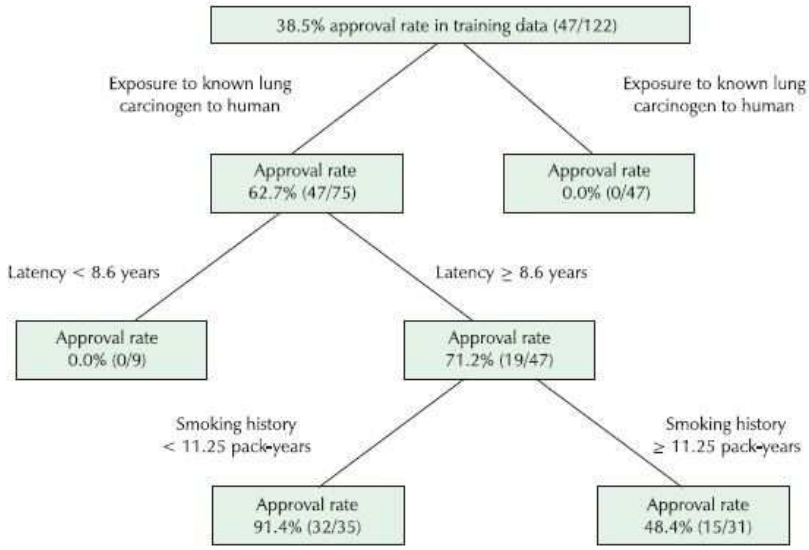
Table 2. Exposed material during work in the workplace ( ) = %

Exposure material	IARC group	Carcinogen to lung	Occupational lung cancer (n = 57)	Non-occupational lung cancer (n = 96)	Total
Asbestos	1	Known	29 (64.4)	16 (35.6)	45 (29.4)
Crystalline silica			12 (57.1)	9 (42.9)	21 (13.7)
PAHs*			5 (50.0)	5 (50.0)	10 (6.5)
Cr(IV)†			8 (80.0)	2 (20.0)	10 (6.6)
Lead Chromate			2 (50.0)	2 (50.0)	4 (2.6)
Radon			0 (0.0)	1 (100)	1 (0.7)
COE‡			1 (100)	0 (0.0)	1 (0.7)
VCM§		Unknown	0 (0.0)	2 (100)	2 (1.3)
Radiation			0 (0.0)	2 (100)	2 (1.3)
Formaldehyde			0 (0.0)	1 (100)	1 (0.7)
Lead	2A		0 (0.0)	2 (100)	2 (1.3)
MMVF‖	2B		0 (0.0)	2 (100)	2 (1.3)
Hydrochloric acid			0 (0.0)	1 (100)	1 (0.7)
Carbon black			0 (0.0)	1 (100)	1 (0.7)
Dust	3		0 (0.0)	18 (100)	18 (11.8)
Organic solvents			0 (0.0)	8 (100)	8 (5.2)
Welding fume			0 (0.0)	3 (100)	3 (2.0)
Paints			0 (0.0)	2 (100)	2 (1.3)
Dye			0 (0.0)	1 (100)	1 (0.7)
Harmful gas			0 (0.0)	1 (100)	1 (0.7)
MWF¶			0 (0.0)	1 (100)	1 (0.7)
Tar			0 (0.0)	1 (100)	1 (0.7)
Cr(III)**			0 (0.0)	11 (100)	11 (7.1)
Missing			0 (0.0)	4 (100)	4 (2.6)
Total			57 (37.3)	96 (62.7)	153 (100)

*PAHs: polycyclic aromatic hydrocarbons.
†Cr(IV): hexavalent chromium.
‡COE: coke oven emission.
§VCM: vinyl chloride monomer.
‖MMVF: man-made vitreous fiber.
¶MWF: metal working fluid.
**Cr(III): trivalent chromium.

○ CART모델

- 표본모델에서 직업성폐암 결정의 첫 번째 예측변수는 노출 물질의 폐에 대한 발암성이었음. 알려진 폐 발암성물질에 노출된 75건의 케이스 중 47건(62.7%)이 직업성 폐암으로 결정됨. 알려지지 않은 폐 발암성물질에 노출된 모든 케이스는 비직업적 요인에 의해 폐암으로 발전된 것으로 결정됨. 알려진 폐 발암성물질에 노출된 케이스 중 두 번째 예측변수는 8.6년 이상의 잠재기였음. 8.6년 이상의 잠재기를 가진 66건 중 47건(71.2%)이 직업성 폐암으로 결정됨. 8.6년 미만의 잠재기를 가진 모든 케이스는 비직업성으로 결정됨. 알려진 폐 발암성물질에 노출되었고, 8.6년 이상의잠재기를 가진 케이스 중 11.25 갑년 이하의 흡연력은 추가 예측 값을 제공함. 35건 중 32(91.4%)건이 직업성 폐암으로 결정됨. 11.25갑년 이상의 흡연력을 가진 31건의 케이스 중 15건(48.4%)이 직업성 노출에 의해 폐암이 발전된 것으로 결정됨.

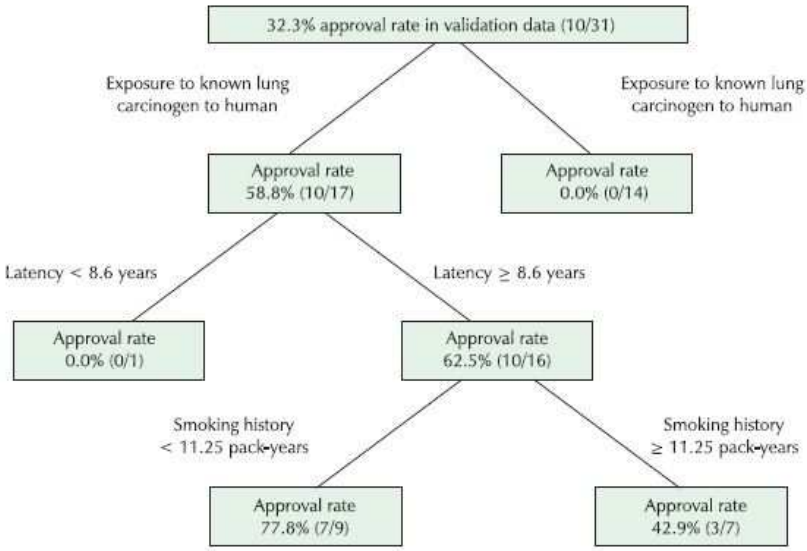


[그림 7-45] 훈련 데이터에서 직업성 폐암의 예측인자

각 노드는 각 예측 변수에 사용 가능한 데이터를 기반으로 함. 각 예측 변수에 대한 각 승인 비율은 상자에 표시되어 있음. 각 예측 변수는 줄

옆에 작성되었음.

- 검증 tree에서 표본 자료의 규칙을 적용하면, 알려진 폐 발암물질에 노출된 31건 중 17건(58.8%)이 직업성 폐암으로 결정됨. 알려진 폐 발암성물질에 노출되고 8.6년 이상의 잠재기를 가진 16건 중 10건(62.5%)이 직업성 폐암으로 결정됨. 폐 발암성물질에 노출되고, 8.6년 이상의 잠재기를 가지며, 11.25갑년 이하의 흡연력을 가진 9건 중 7건(77.8%)가 직업성 암으로 결정됨.



[그림 7-46] 유효성 검사 자료에 적용되는 직업성 폐암 결정 나무

각 노드는 각 예측 변수에 사용 가능한 데이터를 기반으로 함. 각 예측 변수에 대한 각 승인 비율은 상자에 표시되어 있음. 각 예측 변수는 줄 옆에 작성되었음.

- 로지스틱 회귀분석에서 노출 물질의 폐 발암성, 흡연력, 업무기간이 중요한 변수로 분석됨. 1갑년이 증가하면 직업성 폐암에 대한 업무관련성이 0.94배 감소함. 업무기간이 1년 증가하면 업무관련성이 1.15배 증가함.

- AUCs에 근거하면 CART모델(0.914)의 정확성은 로지스틱 회귀 모델(0.824) 보다 더 높았음.

5) 고찰 및 결론

- 일반적으로 고형암의 잠재기는 10-12년임. 본 연구에서는 잠재기의 컷 포인트가 9년 동안 석면에 노출된 근로자가 직업성 폐암으로 결정되었기 때문에 짧음. 이 한 건을 제외하면 잠재기가 10년 보다 더 높음. CART의 목적은 정확한 예측변수와 분류에 있음. 그러므로 8.6년 이상이라는 잠재기는 비직업성 폐암의 9건을 제외하고 계산되어짐. 이 결과는 변수 구간의 분할 지점에서 예측 오류가 발생할 수 있다는 CART의 단점을 보여줌.

- 산업분류 별 chi-square test에서 제조업과 비교했을 때 비제조업의 직업성 폐암이 유의하게 증가함. CART모델에서는 산업분류가 포함되지 않았는데 첫 번째 예측변수(알려진 폐 발암성 물질에 의한 직업적 노출)에 따라 분류하자 산업분류의 유의성이 사라짐. 이 결과는 패턴 분류에 대한 의사결정 프로세스를 제시하는 CART의 특성을 나타냄.

- 직업성 폐암을 승인하기 위한 예측 변수로 폐 발암성물질의 노출, 잠재기, 흡연력을 발견함. 업무관련성을 결정하는 의사는 각 사례의 복잡성을 고려해야하기 때문에 이 의사결정 tree는 폐암의 업무관련성의 최소한의 결정 표준으로 사용되어야만 함. 직업성폐암에 대한 정확한 결정 표준을 만들기 위하여 검증력 향상을 위한 부가적인 연구가 수행되어야함.

22. Buckley JP et al. Prenatal phthalate Exposures and Childhood Fat Mass in a New York City Cohort. Environmental Health Perspectives. 2016.April; 124(4):507-13.

1) 배경

- 동물 실험연구와 제한적인 역학적 증거는 프탈레이트에 노출된 태아가 안드로겐 비활성화 능력을 가진 프탈레이트의 잠재적인 성 특이적 효과로 비만이 될 수 있음을 시사함.

2) 목적

- 전향적 코호트에서 태아의 프탈레이트 노출과 소아 체지방량 사이의 관계를 평가하고자 함

3) 연구방법

(1) 연구자료 및 연구대상자

- The Mount Sinai Children's Environmental Health Study는 1998년부터 2002년 사이에 뉴욕에 있는 Mount Sinai Diagnostic and Treatment Center와 인접한 2개의 개인 의원으로부터 단일아를 임신한 479명의 초산부를 enroll시킴. 여성들은 Mount Sinai Medical Center에서 출산하였고, 75명의 여성은 제외됨. 최종 출생 코호트는 출생 데이터가 있는 404명의 신생아로 구성됨. 아동들은 4-5.5세(평균 4.9, 첫 번째 방문), 6세(평균 6.2, 두 번째 방문), 7-9세(평균 7.8, 세 번째 방문)때 3번의 follow-up방문을 하도록 함.

- 출생 코호트에 포함된 404명의 신생아중 382명이 산모의 소변에서 측정된 태아 프탈레이트 대사 산물 농도를 가지고 있었음. 이전연구에 따라 부정확한 바이오 마커의 측정 가능성 때문에 매우 희석된 소변에서 얻은 두 대상을 제외시킴. 380명의 신생아 중에 한번이라도 follow-up 결과가 있는 180명이 본 연구에 포함됨.

(2) 노출평가

- 산모들은 25-40주(평균 31.5주) 사이의 소변 샘플을 제공함.

- monoethyl phthalate(MEP), mono-n-butyl phthalate(MnBP), monoiso-butyl phthalate(MiBP), mono phthalate(MCPP), monobenzyl phthalate(MBzP), DEHP의 대사산물 4개(MEHP, MEHHP, MEOHP, MECPP)

- 태아의 프탈레이트 대사 산물 농도의 분포는 추적 방문이 있든 없든 간에 유사함.

(3) 건강영향평가

- Tanita scale(TBF-300)을 이용하여 생체 전기 임피던스 분석을 통하여 체중과 체성분을 평가함. stadiometer를 사용하여 키를 중복 측정하였고, 두 측정의 차이가 2.0cm를 초과하면 세 번째 측정을 함.

- Tanita scale로 보고된 체지방 추정치를 이용하여 (체지방/키)*100로 체지방률(%)을 계산함. 체지방에 대한 Tanita 방정식은 7세 미만에서 유효하지 않기 때문에 아동에서 생체전기 임피던스 측정으로부터 체지방을 추정하는 유효한 두가지 대안 방정식을 테스트 함. 그 결과 Tanita 독점 방정식을 사용하기로 함. 보충 분석을 위하여 몸무게/키*키 (Kg/㎡)로 BMI를 계산하였고, CDC SAS매크로를 이용하여 성별 및 연령별 표준화된 BMI z-score를 결정함.

(4) 보정변수

- 산모와의 인터뷰를 통해 데이터 수집. Mount Sinai Hospital의 주산기 데이터 베이스에서 임신과 분만 특성을 확인함.

- 임신 중 체중 증가의 적정성을 <2009 의학연구소의 권고수준*100>에 기초하여 평가함. 권고보다 적은 체중증가(<86%), 권고 수준(86-120%), 권고보다 많은 체중증가(>120%)로 분류함.

- 아동의 신체활동 정도를 각 follow-up 방문 때 부모나 보호자가 “거의 대부분 활동적”이라고 보고하면 활동적으로, “가끔 활동적”이거나 “거의 활동하지 않음”으로 보고하면 비활동적으로 분류함.

(5) 통계분석 방법

- 태아의 프탈레이트 대사산물 농도와 체지방률(%) 사이의 관계를 평가하기 위하여 베이지안 모델링 프레임워크를 사용함.

a) LOD 미만의 대사산물 대치(imputing), b) 다중 대사산물 모델로부터 추정 안정화, c) 누락된 공변량 데이터의 대치 d) follow-up에 대한 잠재적인 정보 누락 검토와 같은 잠재적인 바이어스를 고려함과 동시에 관계를 추정하기 위해서 이 접근방법을 선택함.

먼저 꼬리를 자른(truncated) 정규분포에서 값을 대치하여 LOD 미만의 프탈레이트 대사산물 농도에 대하여 처리함. Markov chain Monte Carlo (MCMC) 알고리즘을 반복할 때마다 관찰된 분포(하한 값은 0, 상한 값은 LOD인)의 평균과 표준편차로 정의된 매개변수의 끝이 잘린 정규분포로부터 각 대사산물의 자연로그 값을 대치함.

- DEHP 대사산물은 같은 물질에서의 노출을 나타내고 높은 상관성을 가지기 때문에 MCMC 반복할 때 물질의 대사산물 값을 사용하여 계산된 몰의 합으로 평가함.

- 연구집단의 분포와 관련하여 프탈레이트의 상대적 효과 크기에 대한 용이한 비교를 위해서 각 프탈레이트 대사산물의 자연로그 또는 합을 평균과 표준편차로 표준화 함.

- 프탈레이트 대사산물 농도는 공통 모 프탈레이트 및 근원물질이 같기 때문에 상관관계가 있는데 체지방률과 단일 프탈레이트 대사산물 사이의 관계에 대한 추정은 다른 대사산물에 의해 혼동될 수 있음을 시사함. 따라서 베이지안 프레임워크 내에서 다중 대사산물의 계층적 모델을 이용하여 연관성을 추정함. 각 표준화된 프탈레이트 대사산물의 베타 계수를 0을 평균으로 하고 $1/\tau^2$을 편차로 하는 독립 정규 사전 분포에 줌. 자연로그 한 프탈레이트 대사산물 농도에서 SD 차이의 효과의 95%는 체지방률 평균의 ±1SD 안에 있을 것이라는 사전 믿음을 반영한 τ 값을 선택함 (τ=1/16). 민감도 분석에서 프탈레이트의 베타 계수에 대한 동일한 사전 분포를 지정했지만 한번에 한 대사산물만 포함하

는 단일 대사산물 모델을 사용하여 관계를 추정함.

- 방향성 비순환 그래프를 사용하여 식별된 잠재적인 혼란변수를 보정함. 인구통계학적 및 사회경제적 특성에는 산모의 인종/민족(비 히스패닉 백인/ 비 히스패닉 흑인/ 히스패닉), 나이, 교육수준(대학 미만 / 대학 이상), 임신 중 업무(피고용인/ 학생 혹은 주부)가 포함됨. 아동의 과체중 상태와 관련 있는 조기 요인을 설명하기 위하여 산모의 신체 특성(임신 전 BMI와 임신 중 체중 증가의 적절성), 임신 중 흡연(예/아니오), 모유수유(한 적 있음/ 한번도 하지 않음)를 보정함. 프탈레이트 노출의 일시적 트렌드와 아동 비만의 유병률을 설명하기 위하여 소변 수집 일정을 보정함. 부가적으로 자연로그 한 크레아티닌(소변 희석을 설명하기 위해), 아동의 성별, follow-up 한 개월수, 아동 성별과 follow-up 개월 수의 곱(product term)을 보정함.

- 정확도 향상을 위하여 산모의 키와 아동의 활동성(활동적/비활동적)을 포함시켜 결과에 대한 강한 예측변수를 포함시킴. 성별과 각 대사산물의 곱(product term)을 포함시켜 효과 측정 층화변수(modifier)로 아동의 성별을 평가함. 이분성 공변량과 추정 효과 크기의 용이한 비교와, MCMC 수렴을 개선하기 위해서, 연속 공변량(산모 나이, 임신전 BMI, 산모의 키, 임신 중 체중 증가의 적절성, 소변 수집 날짜, 자연 로그 한 크레아티닌, follow-up 개월 수)를 $(x-y_x)/(2\delta_x)$로 표준화함. 모델 적합성 향상을 평가하는 likelihood ratio test와 GAM(generalized additive models)를 사용하여 체지방률과 공변량의 연관성을 시각적으로 평가하는데 기초한 모델링을 위하여 고차 다항식을 사용함 (산모 나이(3차), 임신 전 BMI(2차), 임신중 체중 증가 적절성(3차), 자연로그한 크레아티닌(2차)). 이분법 공변량의 효과의 95%가 연구집단에서 체지방률의 평균±2SD안에 있을 것이라는 사전 믿음을 반영하여 각 공변량에 대한 $\tau=1/64$로 하여 독립적 null-centered priors를 지정함.

- 다중 대치에 대한 베이지안 프레임워크의 이점을 활용하여 MCMC 절차 내에서 값이 무작위로 결측되었다는 가정 하에 임신중 체

중증가의 적정성, 모유수유, follow-up시 활동성에대한 결측치를 치환함. 임신중 체중증가의 적절성에 대한 결측은 마지막 임신중 체중의 결측에 의한 것이기 때문에 산모 교육수준, 인종 및 민족, 아동 성별, 임신중 흡연, 임신 중 업무상태, 출산 시 산모 나이, 출산시 체중, 산모 키, 제태수, 산모의 임신 중 첫 번째 몸무게, 임신전 산모 BMI의 조건에 따라 정규 분포된 무작위 변수로 마지막 체중을 모델링함. MCMC 알고리즘을 반복 할 때 임신 중 체중증가의 적정성은 치환된 임신 중 마지막 몸무게를 사용하여 계산됨. 산모의 교육수준, 인종 및 민족, 아동의 성별, 임신 중 흡연, 임신중 근무상태, 출산시 나이, 임신중 체중증가의 적정성, 임신전 BMI, 출생시 체중에 따라 조건부 로지스틱 모델을 사용하여 모유수유를 모델링함. 인종 및 민족, 임신전 BMI, 임신중 흡연, follow-up시 개월 수, 출생시 몸무게, 아동의 성별에 따라 로지스틱 혼합 효과 모델을 사용하여 각 follow-up 방문 시 신체 활동성을 모델링함. 이분성 공변량의 효과의 95%가 임신중 마지막 몸무게 평균의 ±2SD 안에 있거나 모유수유와 신체활동성에 대한 OR값이 0.14-7.1안에 있을 것이라는 사전믿음을 반영하여 τ=1이고, null-centered, 독립적 치환 모델에 베타 계수를 부여함.

- 각 아동별 다중 관측치를 설명하기 위해서 무작위 간섭(random intercepts)와 선형 혼합효과 회귀모델을 사용하여 체지방률과 임신 3기 산모의 요중 프탈레이트 대사산물 농도의 연관성을 평가함. 자연로그한 프탈레이트 대사산물 농도에서 사후평균 베타계수와 표준편차 증가 당 95% credible intervals (CI)를 추정함. credible intervals은 전통적인 분석의 confidence interval보다 더 직관적임. 특히 주어진 데이터와 모델에서 참값이 이 구간안에 있을 확률은 95%임. 아동의 성별에 따른 층화 추정 모델에서 성별과 각 대사산물의 곱과 성별과 크레아티닌 자연대수 값의 곱을 포함시킴. 80% CI가 null 값을 넘지 않는다면 의미있는 보정인 것으로 간주함. 잠재적인 비선형 양반응관계를 평가하기 위해서 각 대사산물의 두 번째, 세 번째 삼분위수에 대한 지표변수를 모델에

적용함. 소변 희석정도를 설명하기 위해서 크레아티닌 보정농도를 사용하여 노출을 분류하고 공변량으로 자연대수 크레아티닌 값을 포함시키지는 않음.

- 아동기의 지방과다증의 두 번째 측정을 평가하고 체성분 데이터가 없는 연구와의 용이한 비교를 위하여 프탈레이트 대사산물의 농도와 BMI z-score와의 연관성을 추정함. 최초의 10000번 반복 후에 각각의 최종 베이지안 모델을 50000번 반복하였고 표준진단방법을 통하여 모델 수렴도를 평가함. DAGitty ver.2.2를 사용하여 인과관계 다이어그램를 그렸고, SAS ver 9.3으로 기술 분석을 수행함. WinBUGS ver.1.4.3으로 베이지안 모델을 수행함.

□ follow-up loss에 대한 민감도분석

- 적어도 한번 follow-up 방문을 한 아동의 최초 분석은 반응이 관찰된 결과와 공변량에 조건부로 결측되었을 것이라는 가정을 함. 출생 코호트와 follow-up샘플 사이에 관출된 변수의 분포의 차이가 있었으며, 체지방 측정이 누락될 가능성이 실제(누락된)값에 의존될수 있기 때문에 우리는 잠재적인 선택 바이어스에 대한 결과의 민감도를 평가함. 태아의 프탈레이트 대사 산물을 측정한 모든 380명의 아동의 체지방률과 프탈레이트 노출 사이의 연관성에 대한 모델링을 위하여 선택 모델 접근법을 사용함. 이 분석을 위해 각 follow-up때 결과값이 누락되었는지 여부에 대한 이진 지표에 대한 로지스틱 모델과 체지방율 모델을 함께 맞춤. 결측 데이터 지표 모델은 누락된 체지방률뿐만 아니라 follow-up loss와 관련이 있을 것으로 예상되었거나 관찰된 공변량 (출산시 산모나이, 인종/민족, 임신전 BMI, 임신중 체중증가의 적절성, 아동의 성별, follow-up 시기)에 의존함. 선택모델의 결과는 가정된 결측 데이터 메터니즘에 민감하기 때문에 결측 데이터 표시 모델(임신중 흡연, 모유수유, 산모 교육수준, 임신중 노동상태, 출생무게)을 탐색하고, 연속변수의 매개변수를 달리함. 이진 공변량(혹은 연속 공변량에서

2-SD 변화)의 영향의 95%가 0.25-4 오즈비 내에 있다는 신념을 반영하기 위해 베타계수에 null-centered, τ=2, 독립 사전값(priors)을 부여함.

4) 결과

- 모든 프탈레이트 대사산물은 3기 산모 소변 샘플의 90%이상에서 검출됨. 기하평균 농도는 비슷한 시기의 미국 여성의 대표적인 샘플인 국민건강영양조사(200-2002)에서 6세 이상 여성에서 보고된 것 보다 다소 높게 나타남.

- 아동의 21%는 적어도 한 번의 추후 방문 동안 비만으로 분류 됨. 체지방률은 아동의 연령에 따라 증가하였고, 성별에 따라 다름. 전 연령에서 소년들 간의 SD보다 소녀들 사이의 SD가 더 큼.

- 체지방률과 BMI의 z-score 사이의 상관관계는 연령에 따라 증가하였고 spearman R^2 값은 0.75, 0.82, 0.91(첫 번째, 두 번째, 세 번째 방문)이었음.

- 자연로그 표준화한 프탈레이트 대사산물 농도와 체지방률 사이의 상관관계를 추정한 결과 95% CI는 모두 null값을 포함하고 있고, 성별에 의한 보정에도 관계가 없었음.

<표 7-45> 3삼분기 모성 소변 프탈레이트 대사체 농도와 4-9세 소아의 지방 질량 백분위수 간의 보정된 연관성

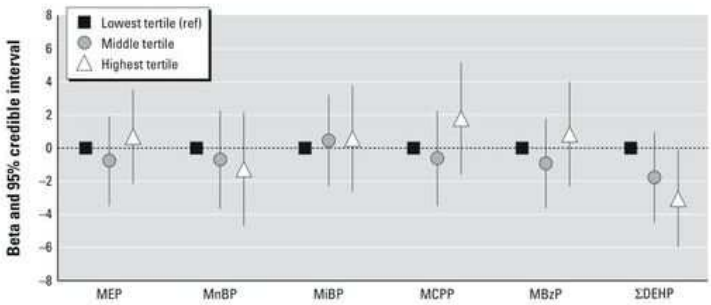
Table 4 Adjusted associations between third trimester maternal urinary phthalate metabolite concentrations and percent fat mass among children age 4–9 years in the Mount Sinai Children's Environmental Health Study.

Metabolite	Overall	Girls	Boys[a]
MEP	0.12 (−1.34, 1.58)	−0.35 (−2.43, 1.75)	0.75 (−1.31, 2.80)
MnBP	−0.86 (−3.07, 1.36)	−0.34 (−3.71, 3.05)	−0.86 (−3.46, 1.74)
MiBP	0.34 (−1.54, 2.20)	1.04 (−1.36, 3.44)	−0.88 (−3.44, 1.68)
MCPP	0.63 (−1.55, 2.82)	1.21 (−1.44, 3.87)	−0.08 (−3.22, 3.03)
MBzP	0.67 (−1.31, 2.65)	0.62 (−1.77, 3.02)	0.98 (−2.17, 4.14)
∑DEHP	−0.89 (−2.24, 0.47)	−0.80 (−2.81, 1.23)	−0.64 (−2.46, 1.16)

Beta coefficients (95% credible intervals) per standard deviation increase in natural log phthalate metabolite concentrations were estimated in a multiple metabolite linear mixed-effects regression model. Sex-specific estimates were modeled by including product terms for phthalate × child's sex. Estimates are adjusted for urine dilution and collection date; maternal race/ethnicity, age, education, work status, and smoking during pregnancy; maternal height and prepregnancy BMI; adequacy of gestational weight gain; breastfeeding; months of age and physical activity at follow-up; and, for overall models, child's sex. [a]Sex-specific estimates did not meet criteria for heterogeneity (i.e., 80% credible intervals for all phthalate × sex product terms included the null value).

- 대사산물 삼분위를 평가하는 모델에서 DEHP의 합과 체지방률의 연관성은 역상관관계로 나타남. DEHP 합의 가장 낮은 삼분위와 비교하면 middle tertile에서 체지방률은 1.77%(95%CI -4.48,0.97) 더 낮고, 가장높은 삼분위 수에서 3.06%(95%CI -5.99,-0.09) 더 낮음. 다른 대사산물의 삼분위수와 체지방률 사이에 관련이 없었음. CI는 넓지만, 아동의 성별에 따른 체지방률과 프탈레이트 삼분위수의 연관성의 차이도 발견하지 못함.

<표 7-46> 3삼분기 모성 소변 프탈레이트 대사체 농도의 3분위수와 4-9세 소아의 지방 질량 백분위수 간의 보정된 연관성



- 민감도 분석결과 다중 대사산물에서 추정된 연관성은 단일물질과 비슷하지만 다중 모델에서 베타 값은 null값에서 약간 멀어지고 CI는 덜 정확함.

- 추적관찰 손실에 대한 민감도 분석의 결과는 누락된 체지방량에 대한 데이터가 무시할 수 없음을 나타냄 (즉, 추적관찰의 가능성은 잠재적으로 아동의 측정되지 않은 지방률에 의존함). 지방량이 더 많은 아동은 추적관찰에 누락될 가능성이 적었음 : 지방량 1% 증가 당 추적관찰이 누락될 가능성에 대한 OR 값은 0.81(95%CI : 0.69, 0.94)임. 그러나 무시할 수 없는 누락 결과를 설명하는 선택 모델에서 추정된 프탈레이트 대사산물과 지방률 사이의 연관성은 일차분석과 유사함.

<표 7-47> 3삼분기 모성 소변 프탈레이트 대사체 농도와 4-9세 소아의 지방 질량 백분위수 간의 연관성에 대한 민감도 분석

Table 5 Sensitivity analyses for associations between third trimester maternal urinary phthalate metabolite concentrations and percent fat mass among children 4–9 years of age in the Mount Sinai Children's Environmental Health Study.

Metabolite	Primary analysis[a]	Single metabolite models[b]	Loss to follow-up[c]
MEP	0.12 (−1.34, 1.58)	0.08 (−1.14, 1.30)	−0.12 (−1.62, 1.36)
MnBP	−0.86 (−3.07, 1.36)	−0.07 (−1.41, 1.27)	−0.93 (−3.15, 1.30)
MiBP	0.34 (−1.54, 2.20)	0.23 (−1.12, 1.52)	1.03 (−0.91, 2.97)
MCPP	0.63 (−1.55, 2.82)	0.30 (−1.05, 1.66)	0.40 (−1.81, 2.60)
MBzP	0.67 (−1.31, 2.65)	0.40 (−0.93, 1.74)	0.48 (−1.47, 2.47)
∑DEHP	−0.89 (−2.24, 0.47)	−0.57 (−1.79, 0.57)	−0.62 (−2.01, 0.81)

Beta coefficients (95% credible intervals) per SD increase in natural log phthalate metabolite concentrations, adjusted for urine dilution and collection date; maternal race/ethnicity, age, education, work status, and smoking during pregnancy; maternal height and prepregnancy BMI; adequacy of gestational weight gain; breastfeeding; months of age and physical activity at follow-up; and child's sex. [a]Associations among children with at least one follow-up visit (n = 180) estimated in a multiple metabolite linear mixed-effects regression model. [b]Associations among children with at least one follow-up visit (n = 180) estimated in a separate linear mixed-effects regression model for each metabolite. [c]Associations among all children in the birth cohort with measured phthalate metabolite concentrations (n = 380) estimated in a multiple metabolite linear mixed-effects regression model using a selection model for potentially nonignorable missing outcome data.

- 누락 데이터를 표시하는 모델에서 다양한 매개변수들이 다른 결과를 가져오지는 않았음.

- 프탈레이트 대사산물 농도와 BMI z-score 사이의 관계패턴은 추정치가 null을 향해 감쇠했지만, 체지방률의 분석결과와 일관됨. 특히 DEHP 대사산물의 합과 BMI z-score 간의 연관성은 연속모델이든 삼분위 모

델이든 관찰되지 않음.

5) 고찰 및 결론

- 본 연구에서 프탈레이트에 대한 태아 노출은 4-9세 아동의 체지방 증가와 관련이 없었음. 4분위 수 모델에서 3기 산모의 요중 DEHP 대사산물의 합의 농도가 가장 높은 삼분위수에 해당하는 아동은 가장 낮은 삼분위수에 해당하는 아동과 비교하여 체지방률이 더 낮았음. 프탈레이트 대사산물 농도와 비만과 관련 있는 결과를 가진 상당한 횡단면 연구들이 있지만 이러한 연구 결과를 반복하기 위해서는 태아 노출을 조사하는 대규모 전향적 연구가 필요함.

〈〈연　구　진〉〉

연 구 기 관 : 산업안전보건연구원

연구총괄자 :	이지혜	직업건강연구실 역학조사부 중부권역학조사팀 팀장
연　구　원 :	예신희	직업건강연구실 역학조사부 중부권역학조사팀 선임연구위원
	이슬비	이화여자대학교 박사
	성정민	직업건강연구실 역학조사부 중부권역학조사팀 과장
	신무영	직업건강연구실 역학조사부 전공의

〈〈연 구 기 간〉〉

2019. 01. 01 ~ 2019. 11. 30

본 연구보고서의 내용은 연구책임자의 개인적 견해이며, 우리 연구원의 공식견해와 다를 수도 있음을 알려드립니다.

산업안전보건연구원장

유해물질 복합노출로 인한 건강영향에 대한 통계적 분석 방법

(2019-연구원-1514)

발 행 일 : 2019년 11월 30일
발 행 인 : 산업안전보건연구원 원장 고 재 철
연 구 책 임 자 : 직업건강연구실 역학조사부 중부권역학조사팀
팀장 이 지 혜
발 행 처 : 안전보건공단 산업안전보건연구원
주 소 : (44429) 울산광역시 중구 종가로 400
전 화 : (032) 510-0752
F A X : (032) 510-0759
Homepage : http://oshri.kosha.or.kr