

2018년도 한국보건행정학회 가을학술대회

“공익적 보건의료 빅데이터의 가능성과 보건행정분야 활용 방안” 발표자료

건강보험 빅데이터를 이용한

# 가슴기살균제 건강피해 규명 연구

2018. 11.

김재용, 한양대학교 건강과 사회 연구소 연구교수

박종헌, 국민건강보험공단 빅데이터운영실 전문연구위원

문재인 대통령, 가슴기 살균제 피해자 면담. 국가의 존재 이유!



# 이전 정부의 문제 인식

(2011년 가습기살균제 강제 수거.. 그런데 2015년에도)



이제 가습기는 좋은데 이 청소하기가 복잡하잖아요. 때가 잘 끼고, 그래서 뭐 살균제도 넣었다가 큰일 날 뻔했고, 사고도 나고 그랬는데, 그거를 완전히 쉽게 분해해가지고 깨끗하게, 아주 쉽게 청소할 수 있는 그런 가습기가 나와서 아마 수출도 하고, 인기가 상당히 높아진다고 그래요. 같은 가습기라도 왜 이렇게 안 팔리냐, 이럴 게 아니라 소비자의 그 어떤 마음을 잘 읽어서 소비자들이 제일 귀찮게 생각하고 힘든 것이 바로 이 청소하는 거다. 몰때나 이런 게 잘 낀다

박00 국가보훈처장 : 그래서 대통령께서 말씀하신 것이 창조경제 아닙니까?

안00 경제수석 : 대통령님이 마치 홈쇼핑 호스트 같으십니다.

최00 기획재정부 장관 : 5분 내로 동이 나야 되는데...

**중양일보**

**신혼 때 가슴기 살균제 ... 37세 남성, 10년 폐질환 앓다 사망**

기사입력 2015.10.20 오전 1:57  
 최종수정 2015.10.20 오전 7:44

590 542

가- 가+

2005년 결혼식을 올린 장모(당시 27세)씨의 신혼 생활은 오래 가지 못했다. 불행은 건강했던 장씨가 어느 날 갑자기 쓰러지면서 시작됐다. 병원에서 내린 진단은 ‘간질성 폐질환’이었다. 원인이 무엇인지는 잘 모르겠지만 장씨의 폐 조직이 손상됐다는 게 주치의의 설명이었다. 장씨는 이후 산소호흡기를 달지 않고서는 생활이 불가능할 정도로 상태가 나빠졌다. 이듬해인 2006년 아내와 이혼까지 하게 됐다. 10년간 길고 외로운 투병 생활을 해온 그는 최근 증세가 급격히 악화돼 병원에 입원했다. 폐 이식에 기대를 걸었지만 이마저 여의치 않았다. 수술이 어려울 정도로 몸 상태가 좋지 않았다. 결국 폐 이식 수술을 포기하고 퇴원했고, 그로부터 한 달 뒤인 지난달 13일에



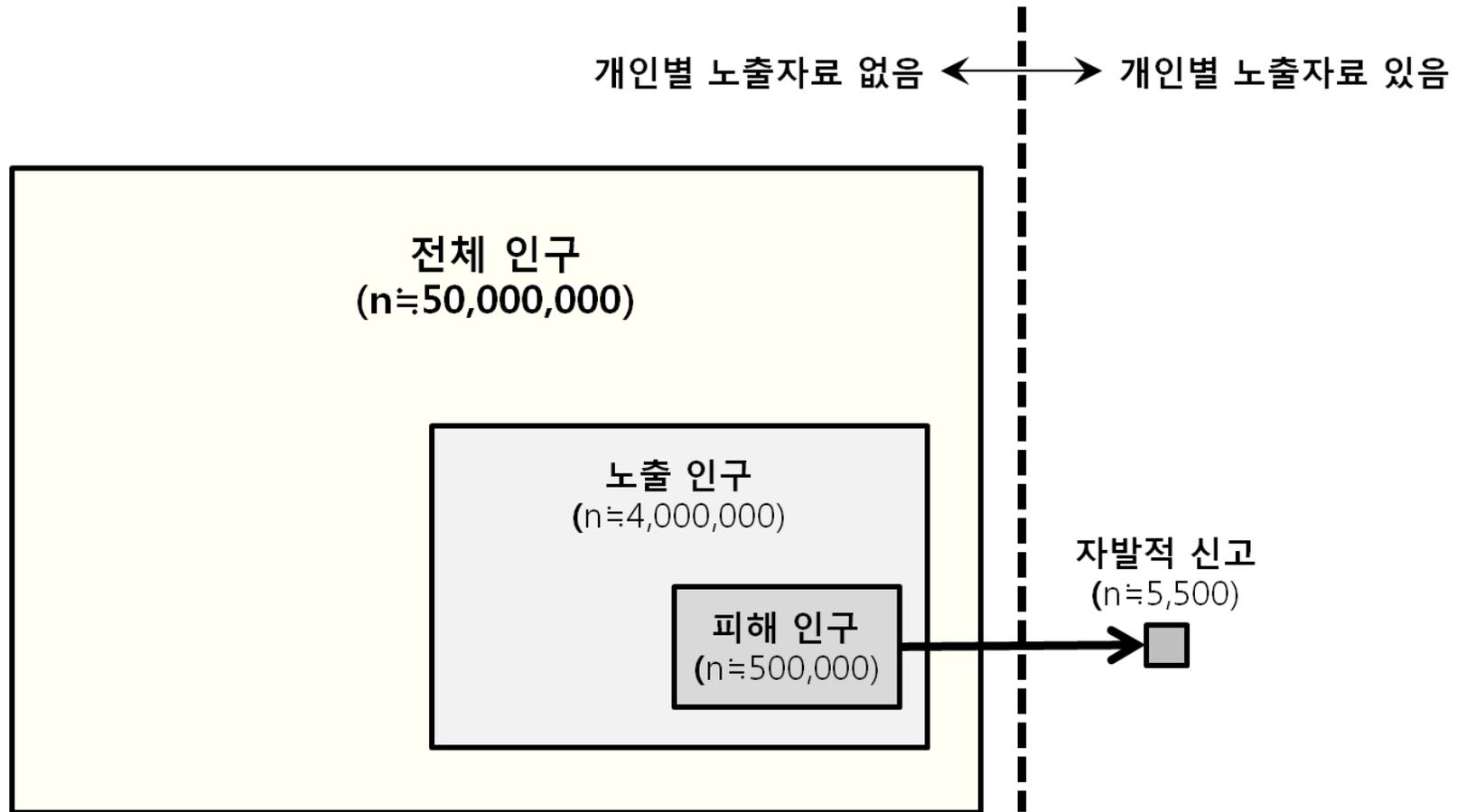
# 배경

- 2011년에 발견된 가슴기살균제 사건은 전세계적으로도 유례가 없는 인위적 대규모 환경성 재앙 사건이었음.
- 최소 400만명 이상이 노출되었고 50만명 전후가 건강문제를 겪었을 것으로 추정됨.
- 그러나, 당시까지 정부 정책은 자발적인 피해신고자 약 6천여 명에게만 초점을 맞춤. 이 중에서도 이미 1300명이 사망함.

# 연구 목표

- 가습기살균제 노출의 역학적 특성을 규명
- 가습기살균제로 인한 초과 건강피해를 규명하고 그 관련성을 밝히는 것
- 가습기살균제로 인한 기여위험율과 상대위험도를 규명하는 것 → 피해보상과 관련된 법률적 의미

# 연구 분석의 기본 틀

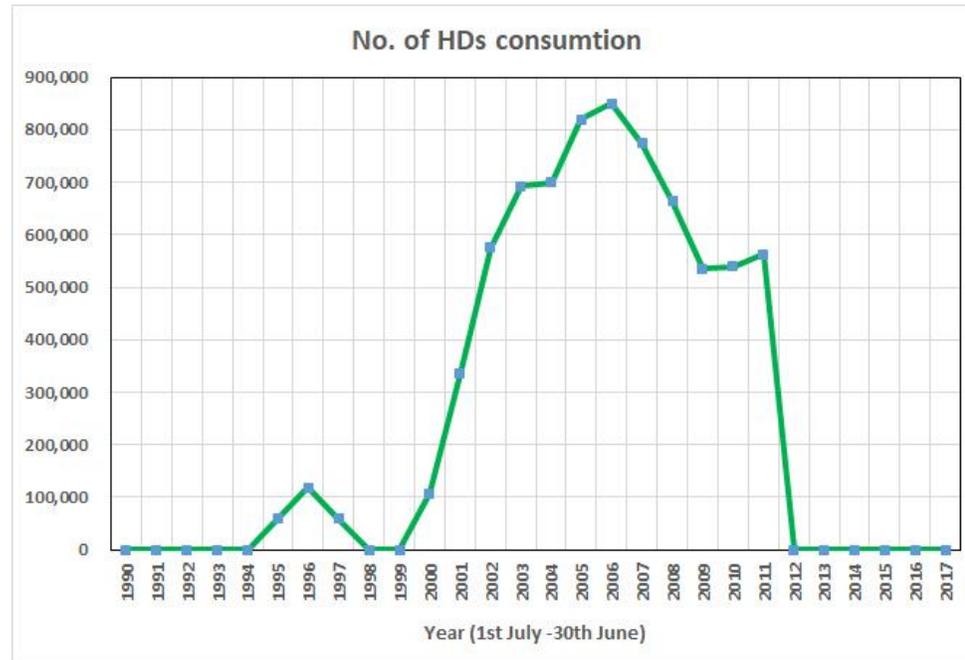


- 대규모 환경재앙에 대한 초동대응의 기본원칙인 노출 인구 파악이 미흡
- 전국 단위 판매정보를 이용한 전국민 코호트 방식의 접근법이 불가피

# 가용 자료가 무엇이었나?

- 한국 갤럽이 시행하는 전국 표본조사

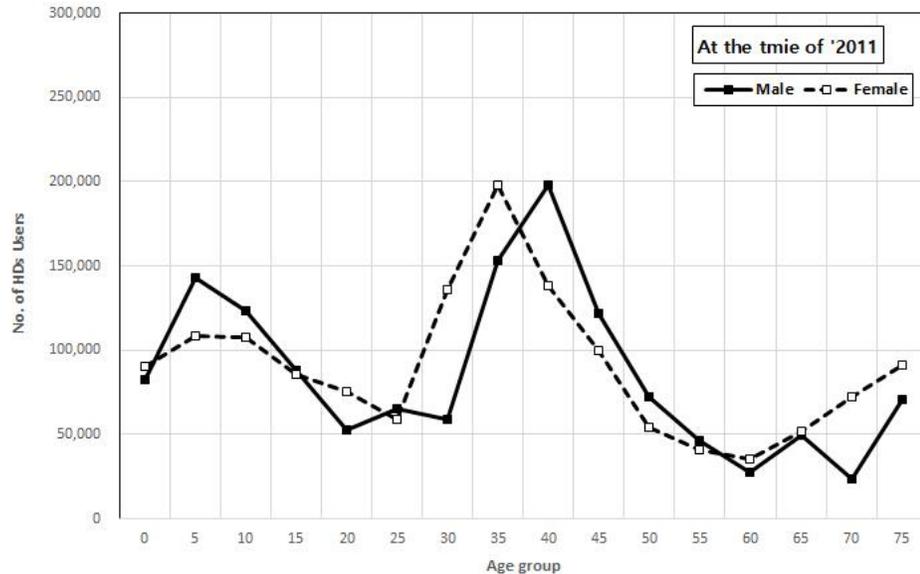
- 환경부가 조사한 연도별  
가습기살균제 판매량



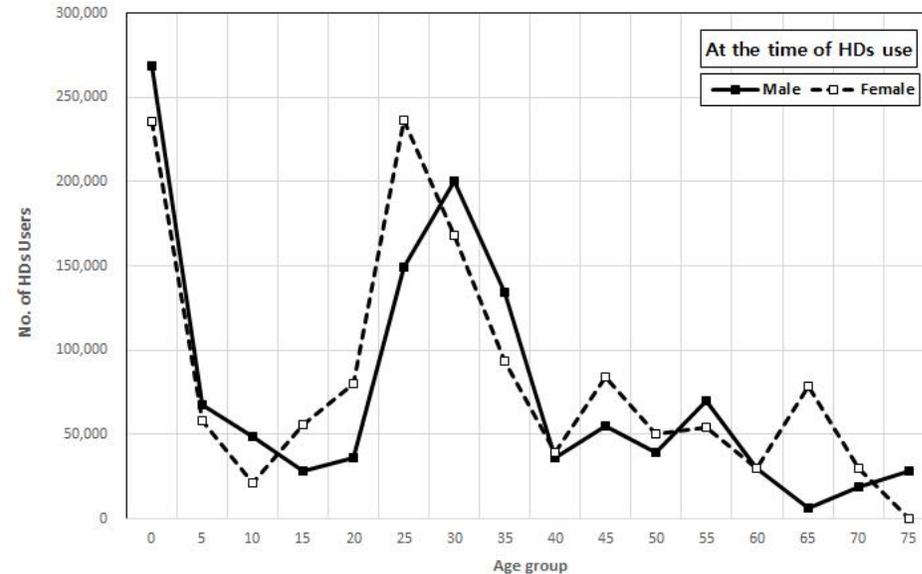
- 그리고, 국민건강보험 빅데이터...

# 전국 표본 설문 조사

- 설문조사의 구조를 변경함  
(응답자 개인 → 배우자, 소아, 노인 등 가족구성원 전체)
- 사용 시점 중심의 재조사 실시 및 가중치 변경
- 최초 사용 시점 중심으로 재분석



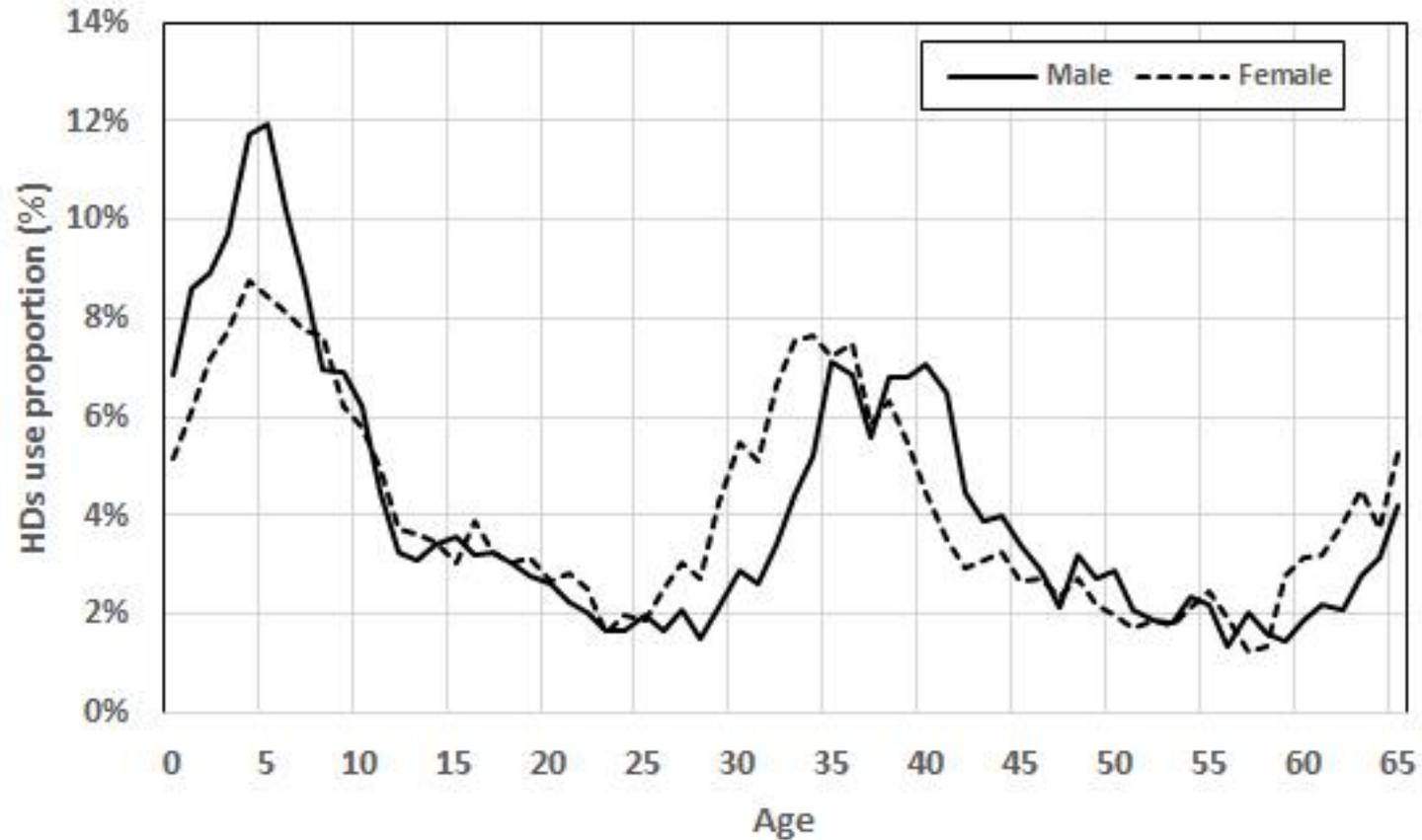
변경 전 가슴기살균제 연령별 사용자 수



변경 후 가슴기살균제 연령별 사용자 수

# 전국 표본 설문 조사

- 인-년(person-year) 단위로 사용률을 재계산한 결과



# 국민건강보험 빅데이터

- 연인원 5천만명 이상의 전국민을 하나의 코호트로 구축
- 행위별 수가제에서 상세한 의료이용 정보 활용 가능(16년-)
- 지역, 소득수준 등 사회경제적 변수들도 활용 가능

→ 유일한 가슴기살균제 노출정보(전국단위 판매량)에 대응하는  
 → 유일한 전국민 코호트 연구 수행

2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

↓ Re-arrange annual period considering HDs use

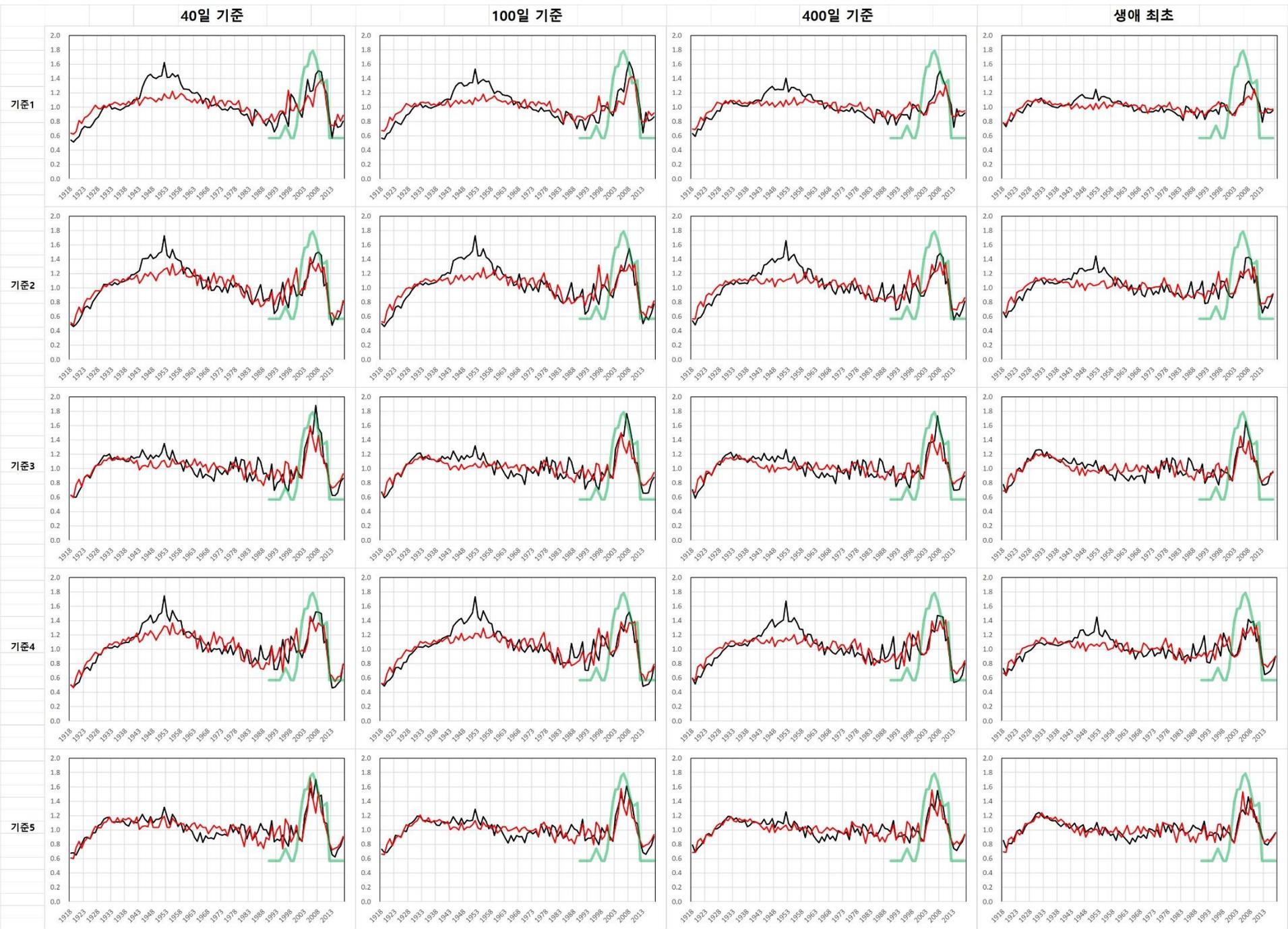
02-03 =2003	03-04 =2004	04-05 =2005	05-06 =2006	06-07 =2007	07-08 =2008	08-09 =2009	09-10 =2010	10-11 =2011	11-12 =2012	12-13 =2013	13-14 =2014	14-15 =2015	15-16 =2016	16-17 =2017
----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------

# 국민건강보험 빅데이터

- 행정자료를 질병역학 연구자료로 전환(Episode of Care)
- 미지의 연구 → 4종의 에피소드 기준과 5종의 판단기준 → 민감도 분석

Window periods for episode of care	Purpose
Within 40 days	to catch the prevalence of diseases of short-term & recurrent events
Within 100 days	to catch the prevalence of diseases of seasonal & recurrent events
Within 400 days	to catch the prevalence of diseases with yearly following-up events
life-long	to catch the incidence of diseases with irreversible & chronic events

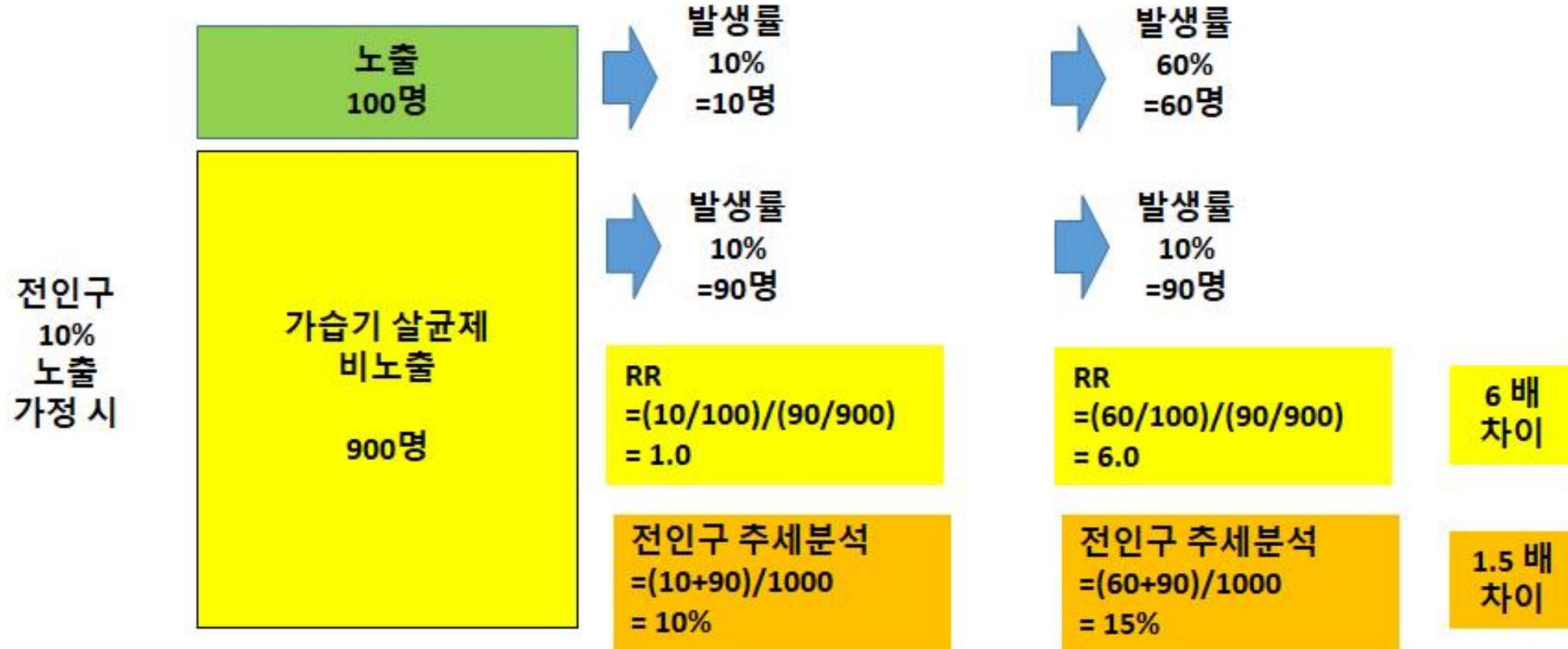
Principal or second diagnostic criteria	Purpose
All episodes of care with 'J84' code	to test all relevant disease events
Within 1 digit classification 'J' and all patients	to test similar disease events
Within 1 digit classification 'J' and in-patients	to test similar disease admissions
Within 3 digit classification 'J84' and all patients	to test specific disease events
Within 3 digit classification 'J84' and in-patients	to test specific disease admissions



# 연령-시기-코호트(Age-period-cohort , APC) 및 이중차분(Difference in difference, DID) 분석

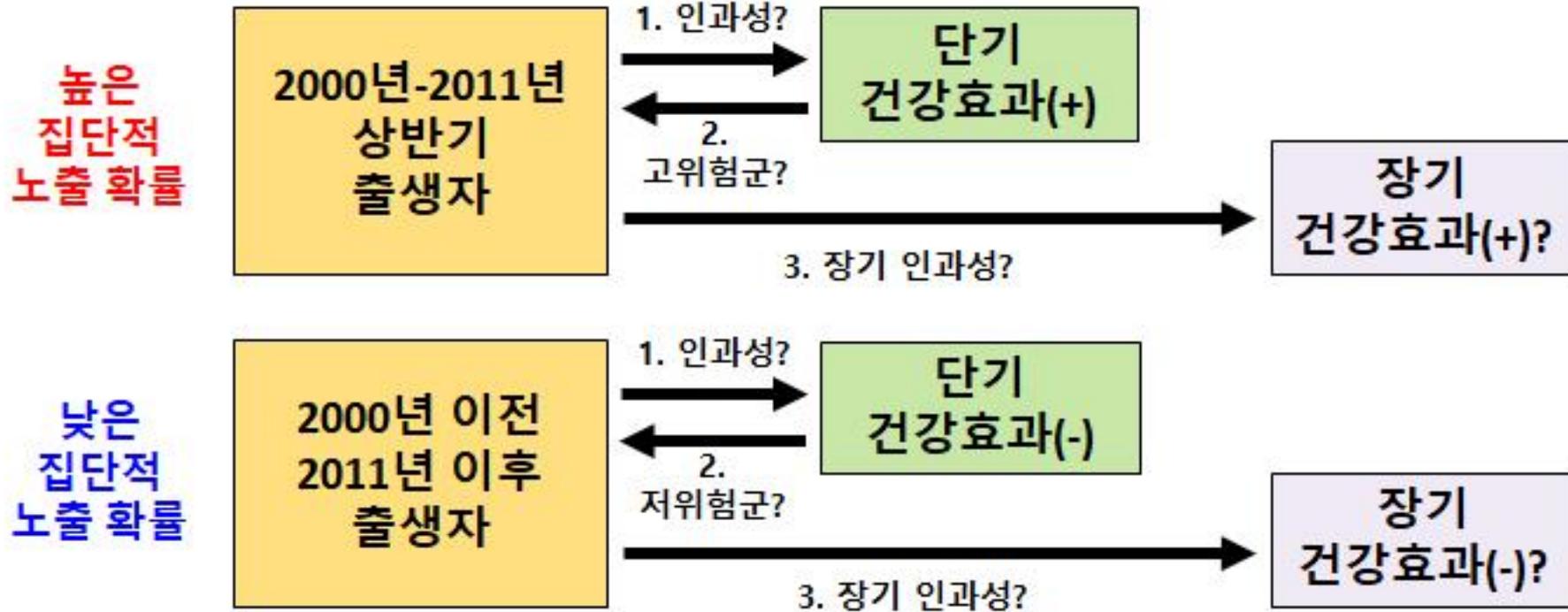
- APC analysis using SAS Glimmix poisson dist. model
  - separate analysis by sex
  - fixed : age, income, district / random : period, cohort
- DID analysis to estimate period X age interaction
  - HDs impact period (2003-2012) vs. Non-impact period(2013-2017)
- Population Attributable Fraction =  $\frac{RR \text{ in total population} - 1}{RR \text{ in total population}}$
- RR by Exposure =  $\frac{RR \text{ in total population} - (1 - \% \text{ of exposed})}{\% \text{ of exposed}}$

# 추세 분석 결과해석 시 유의사항



- 노출분율은 연령에 따라 5-15% 전후로 추정되고 있음(임종한 등, 2017).
- 개인별 노출여부 정보가 없는 추세분석에서 상대위험도 1.5배는 10% 노출분율을 가정하면 노출인구 집단에서 6배의 상대위험도를 의미함.

# 생태학적 연구의 분석 전략과 예시



- 노출 오분류로 인한 저평가 경향은 불가피 → 과평가 가능성은 낮을 수 있음.
- 고위험군 선정 및 건강효과 검증과정에 구성타당도 검토가 필수적.

# 고위험 집단의 구성과 추적 분석 → 장기적 건강영향 모니터링



## 특정 세대(cohort)의 구성요건

1. 가슴기살균제와 특이적 관련성
2. 단기 건강영향의 특성(예: 고평로)
3. 교란요인들의 R/O (구성타당도)
4. 고위험 인구집단의 구성 및 장기추적
5. 타 집단과의 지속적 비교분석 결과  
→ 가슴기살균제 세대?

# 연구 결과

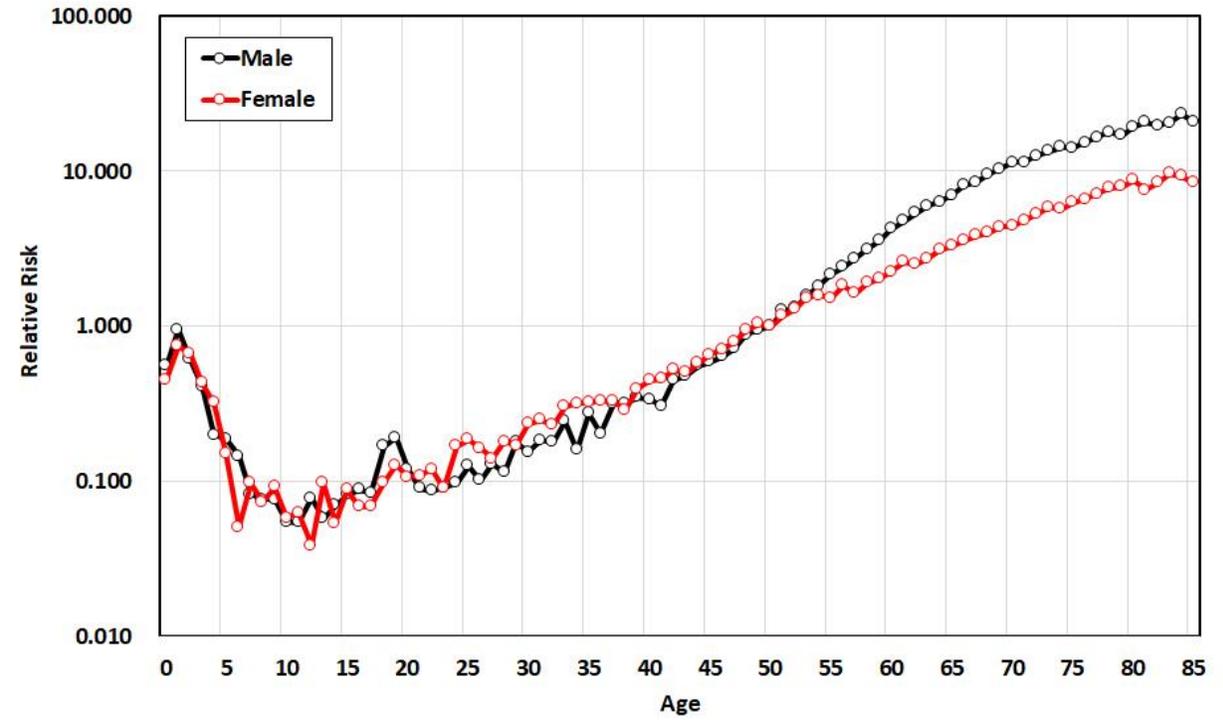
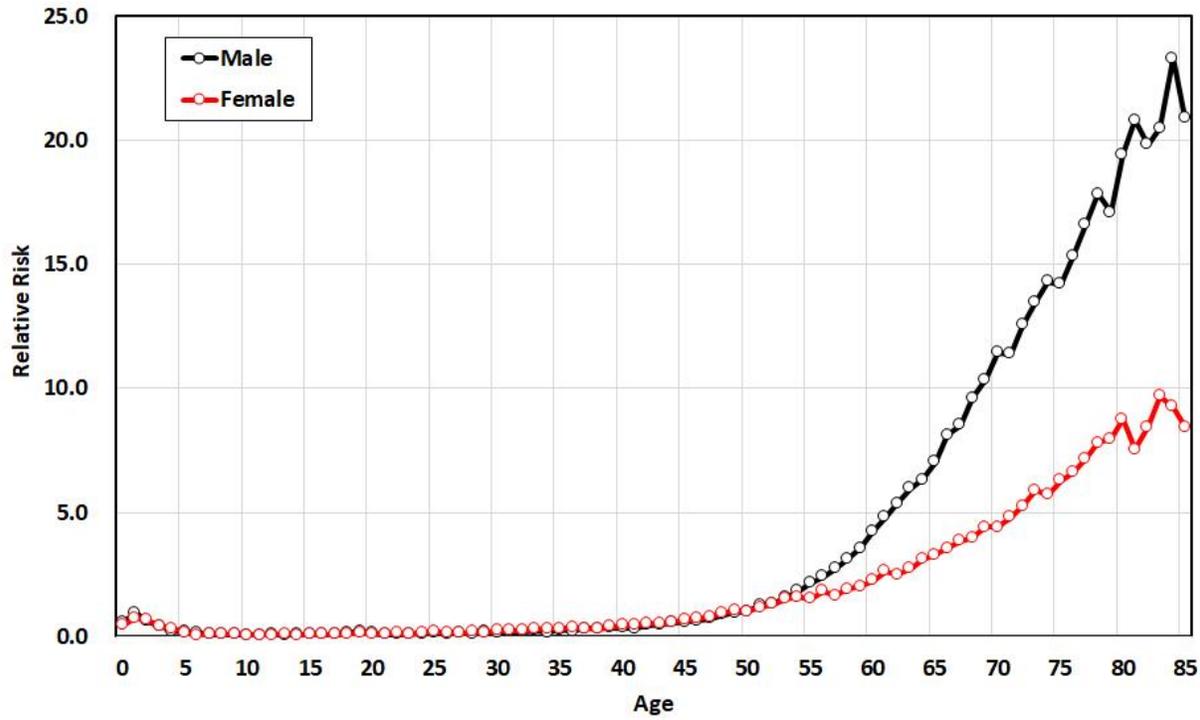
- 호흡기계 질병군은 전체 38개 분류군 중에서 25개 군에서 특이적인 기간효과 또는 코호트 효과 또는 전후(DID) 효과를 보임.
- 피해신고자들이 다양한 호흡기계 질병을 호소한 것과 유사한 양상을 보임.

급성 비인두염[감기] (J00)	만성 후두염 및 후두기관염(J37)
급성 부비동염 (J01)	달리 분류되지 않은 성대 및 후두의 질환(J38)
급성 인두염(J02)	상기도의 기타 질환(J39)
급성 편도염(J03)	급성인지 만성인지 명시되지 않은 기관지염(J40)
급성 후두염 및 기관염(J04)	단순성 및 점액농성 만성 기관지염(J41)
급성 폐색성 후두염[크루프] 및 후두개염(J05)	상세불명의 만성 기관지염(J42)
다발성 및 상세불명 부위의 급성 상기도 감염(J06)	폐기종(J43)
인플루엔자(J09-J11)	기타 만성 폐색성 폐질환(J44)
폐렴(J12-J18)	천식(J45-J46)
급성 기관지염(J20)	기관지확장증(J47)
급성 세기관지염(J21)	진폐증(J60)
상세불명의 급성 하기도감염(J22)	진폐증을 제외한 외부요인에 의한 폐질환(J66-J70)
혈관운동성 및 알레르기성 비염(J30)	성인호흡곤란증후군(J80)
만성 비염, 비인두염 및 인두염 (J31)	폐부종(J81)
만성 부비동염(J32)	달리 분류되지 않은 폐호산구증가증(J82)
코폴립(J33)	기타 간질성 폐질환(J84)
코 및 부비동의 기타 장애(J34)	폐 및 종격의 농양(J85)
편도 및 아데노이드의 만성 질환(J35)	달리 분류되지 않은 흉막삼출액(J90)
편도주위 농양(J36)	기흉(J93)

\* 음영 표시된 질병군들은 기간-코호트-전후 효과 모두에서 특이 소견이 보이지 않은 질병군들임.

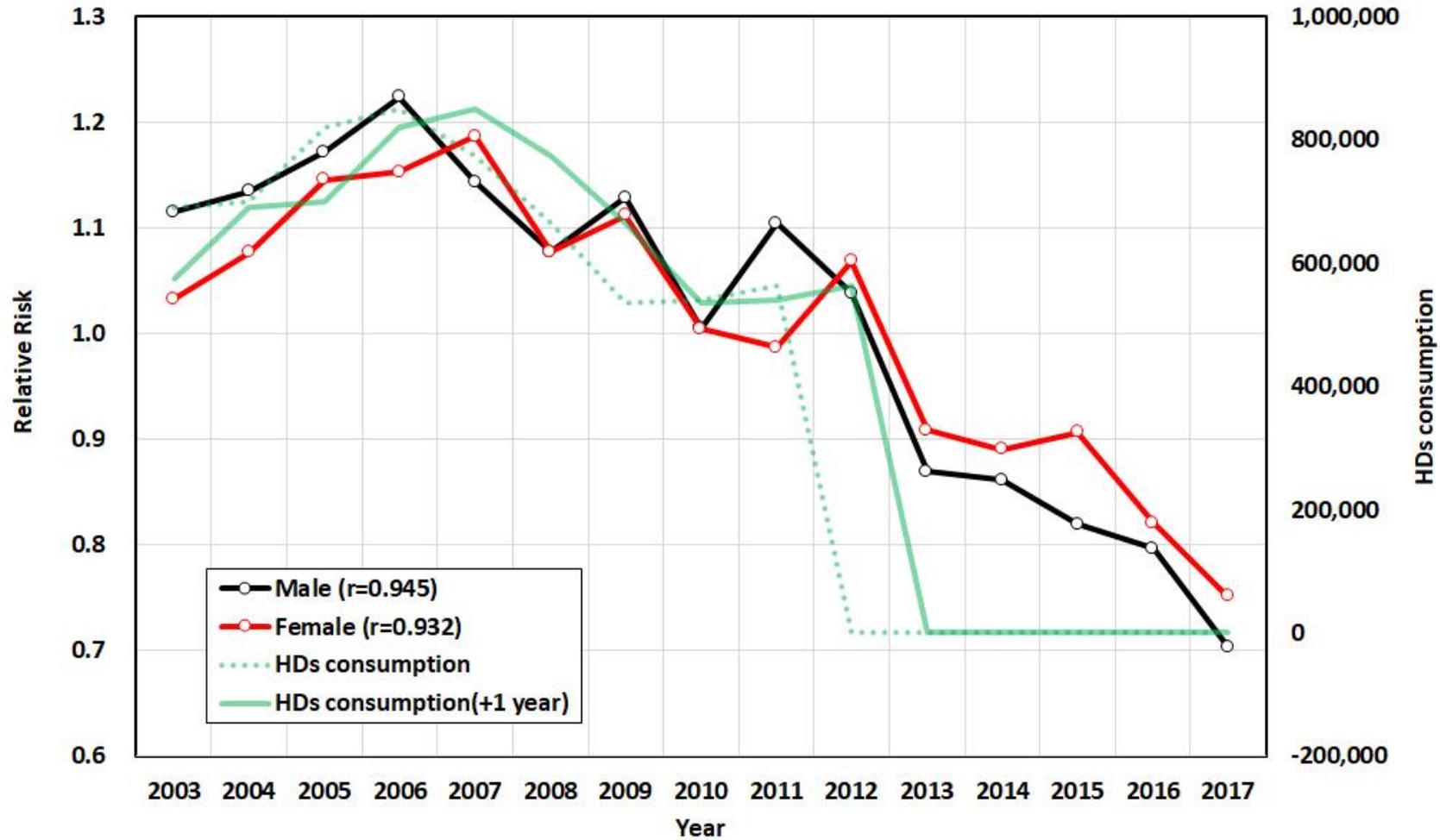
- 악성 신생물에서 특이적인 양상이 관찰되지 않음. 다만, 소화기관의 악성 신생물(C15-C26), 호흡기 및 흉곽내기관의 악성 신생물(C30-C39), 백혈병(C91-C95)에서 영아에 속하는 한 두 연령에서 전후(DID) 효과가 미미하게 나타나 이것이 노이즈인지 여부를 파악하기 위한 추적 관찰 필요.
- 그 외 질병군 중에서 특이적인 소견이 관찰된 질병군들은 결막염(H10), 중이염(H65-H67), 독성 간질환(K71), 아토피 피부염(I20)이었음. 또한 여성 심부전(I50)과 남성 15-25세에서 급성심근경색(I21-I24)은 추적 관찰 필요.
- 혈액 및 조혈기관의 기타 질환(D70-D77)은 4-7세에서 유의하지 않으나 연속적인 증가 경향이 보여 추적 관찰 필요.
- 뇌졸중(I60-I64)은 1-10세 사이에 유의하지 않으나 연속적인 증가 경향이 보여 추적 관찰 필요.

# 간질성 폐질환 (J84)



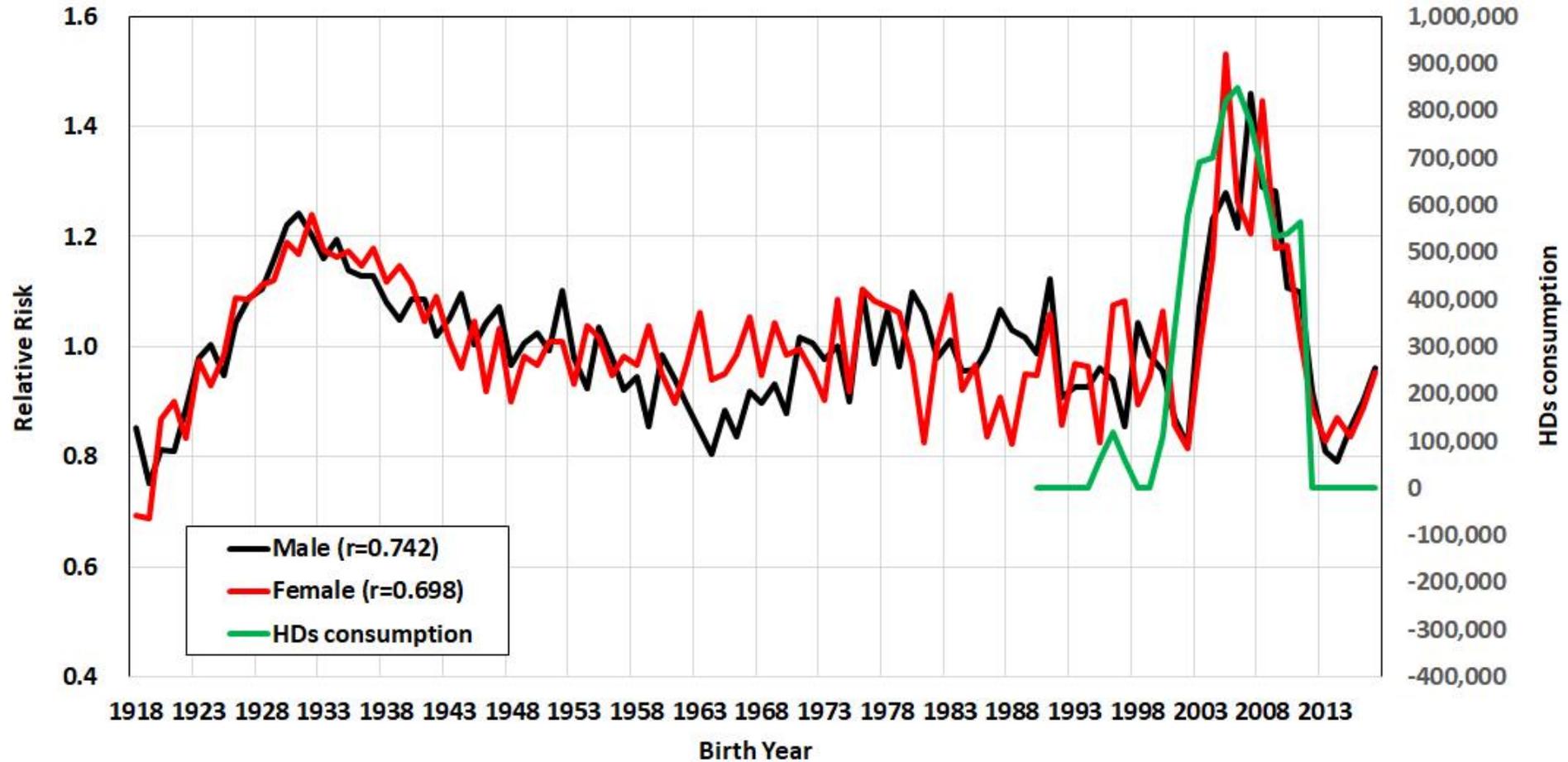
연령 효과

# 간질성 폐질환 (J84)



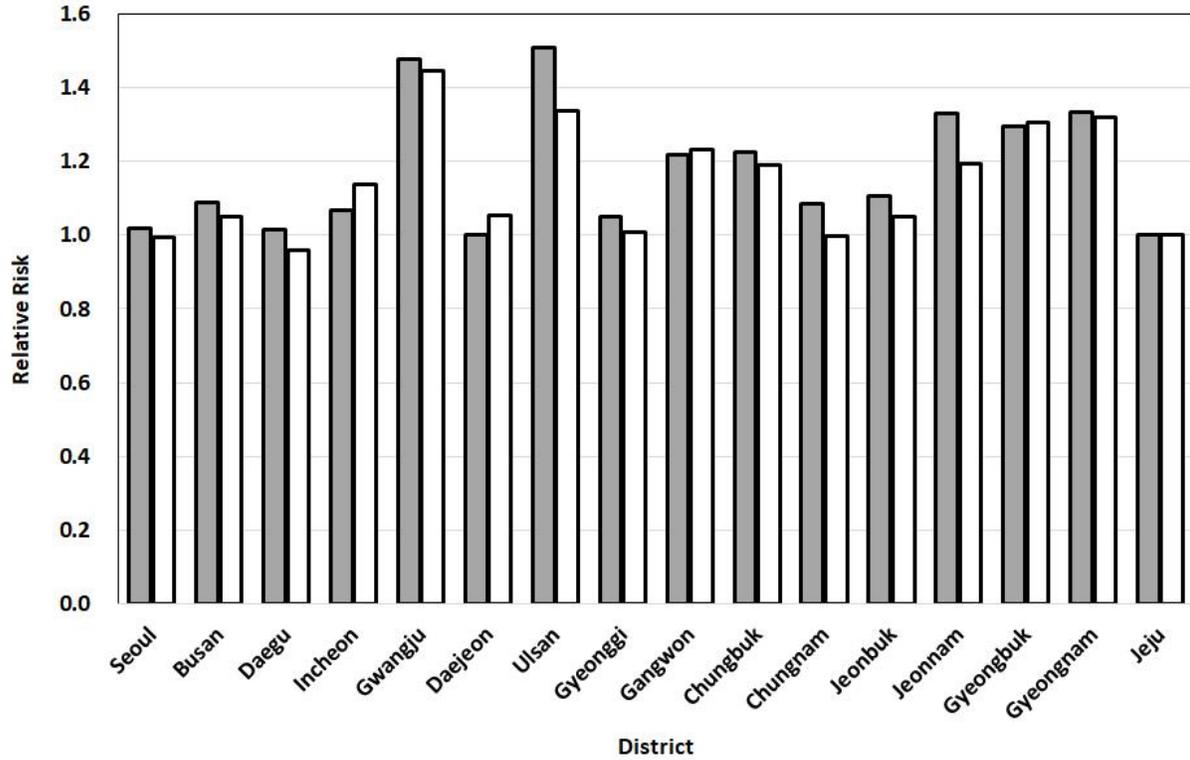
시기 효과

# 간질성 폐질환 (J84)

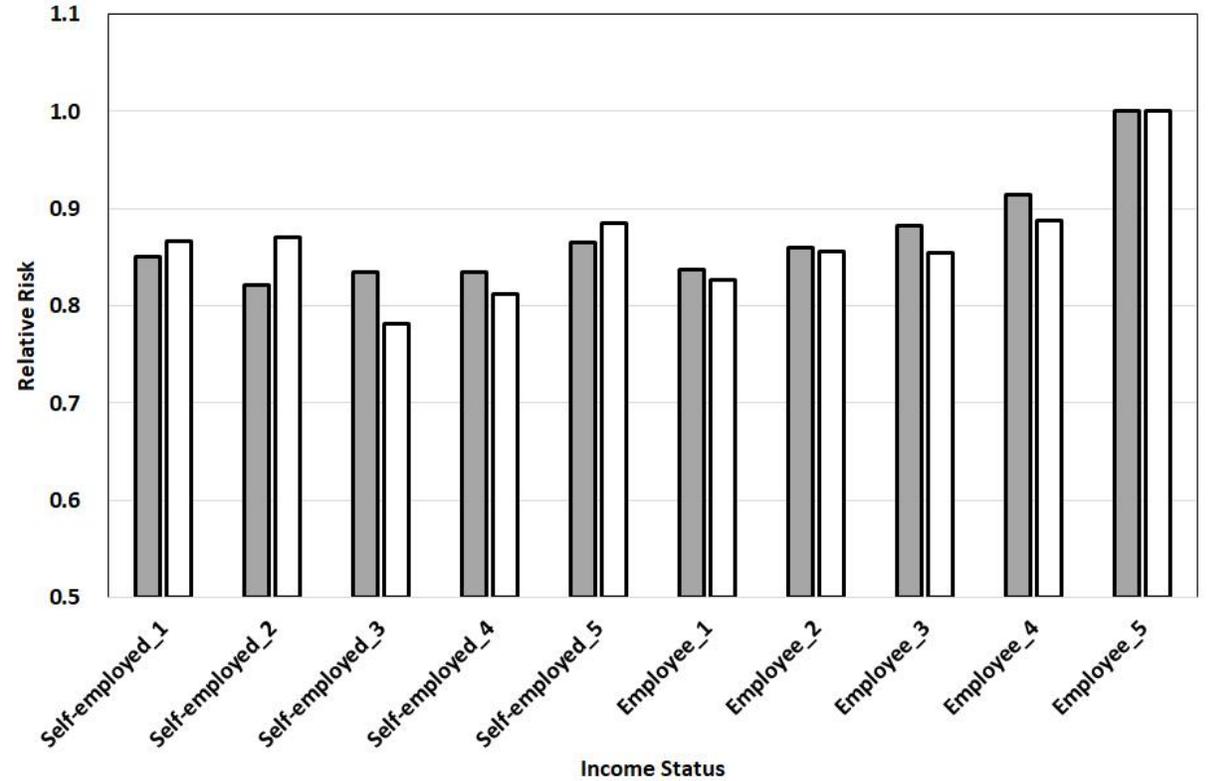


코호트 효과

# 간질성 폐질환 (J84)

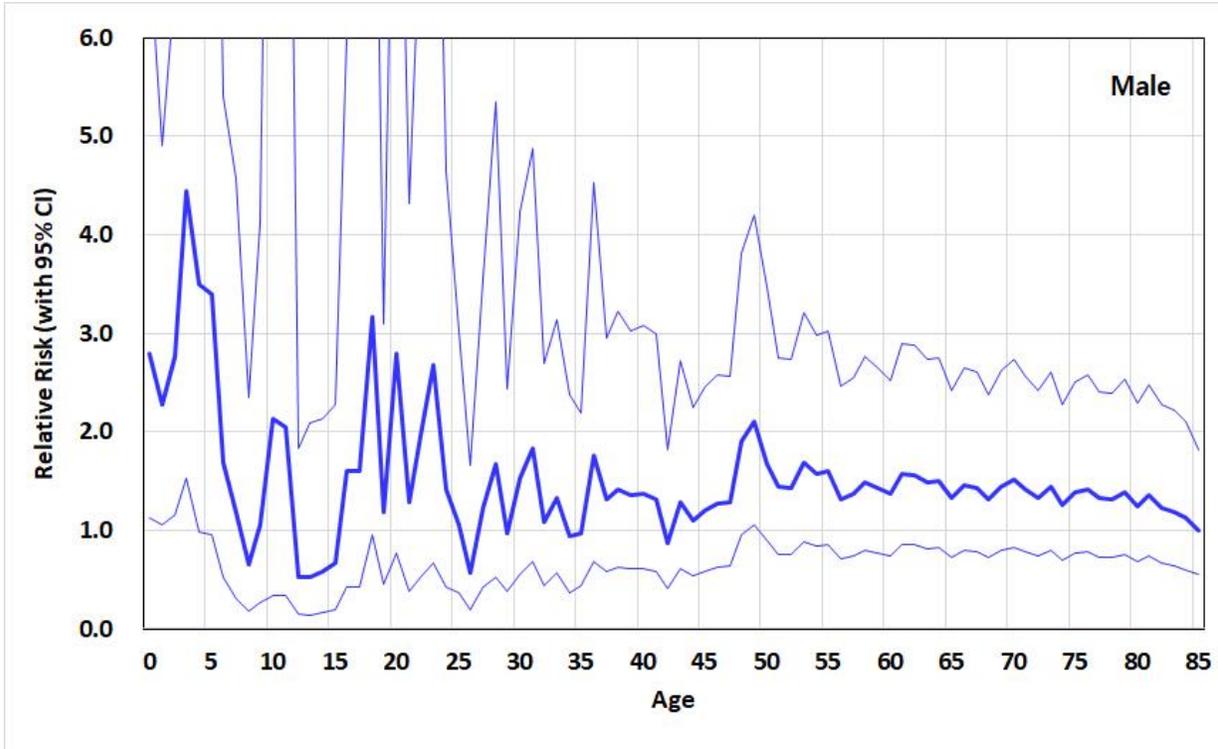


지역 효과

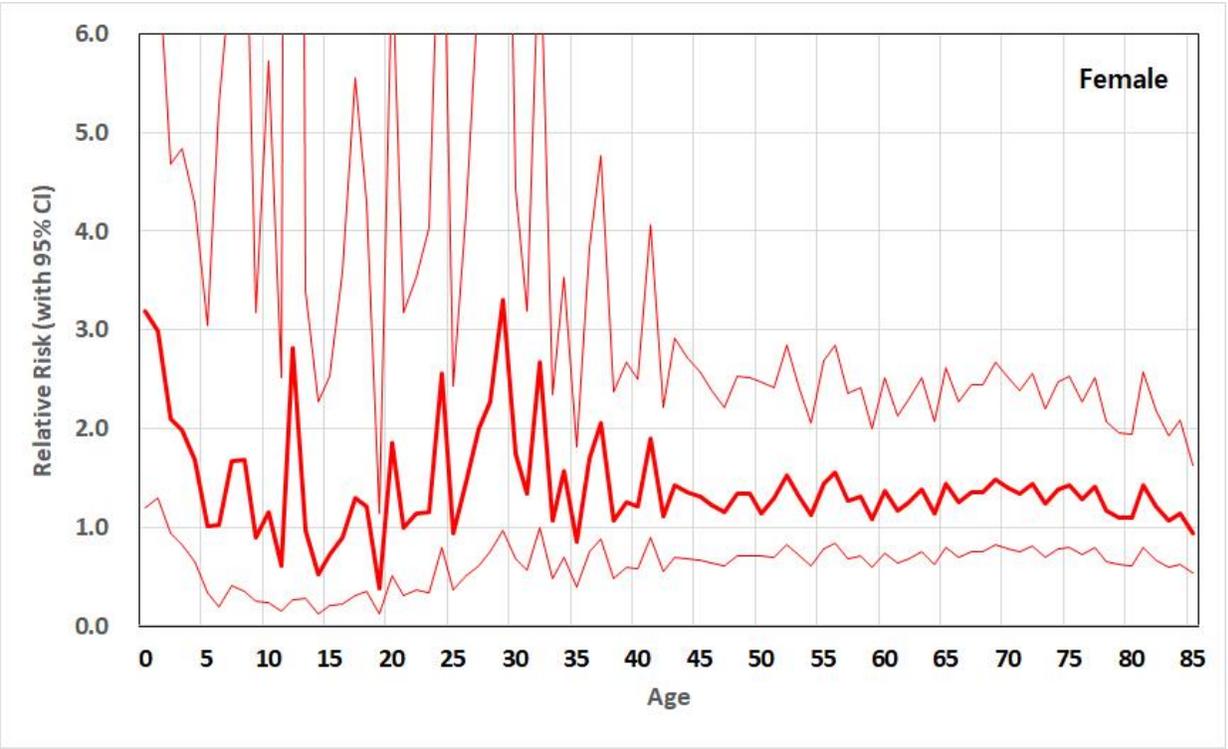


소득수준 효과

# 간질성 폐질환 (J84)



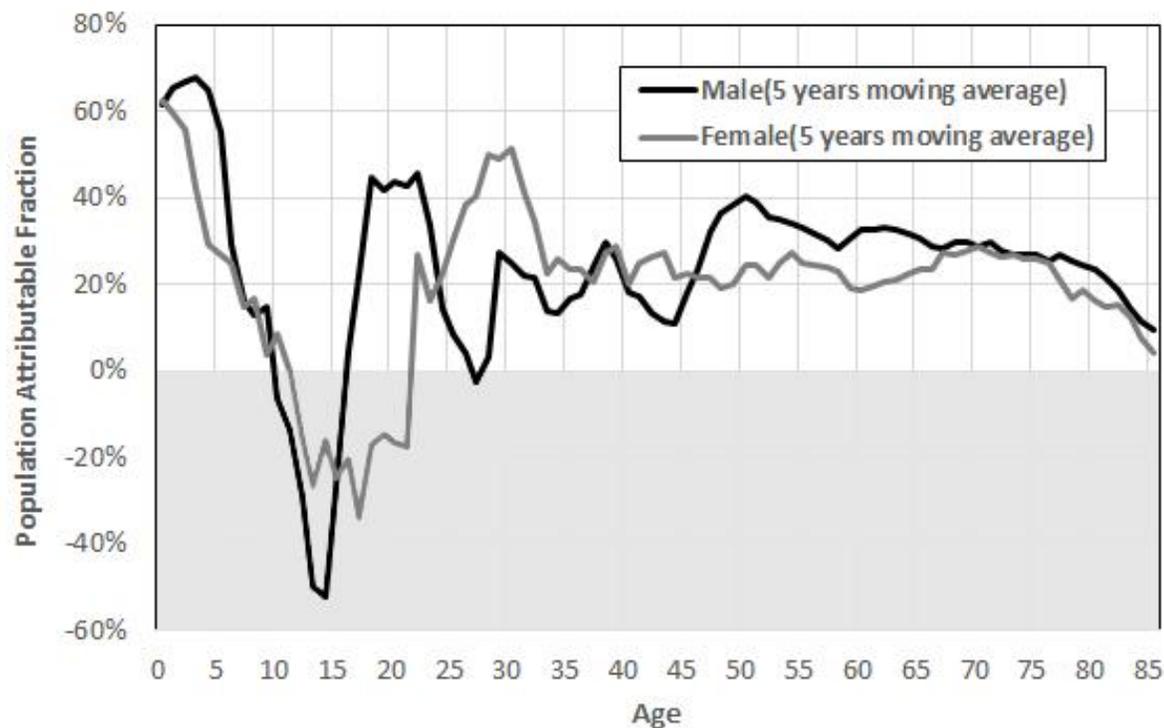
남성



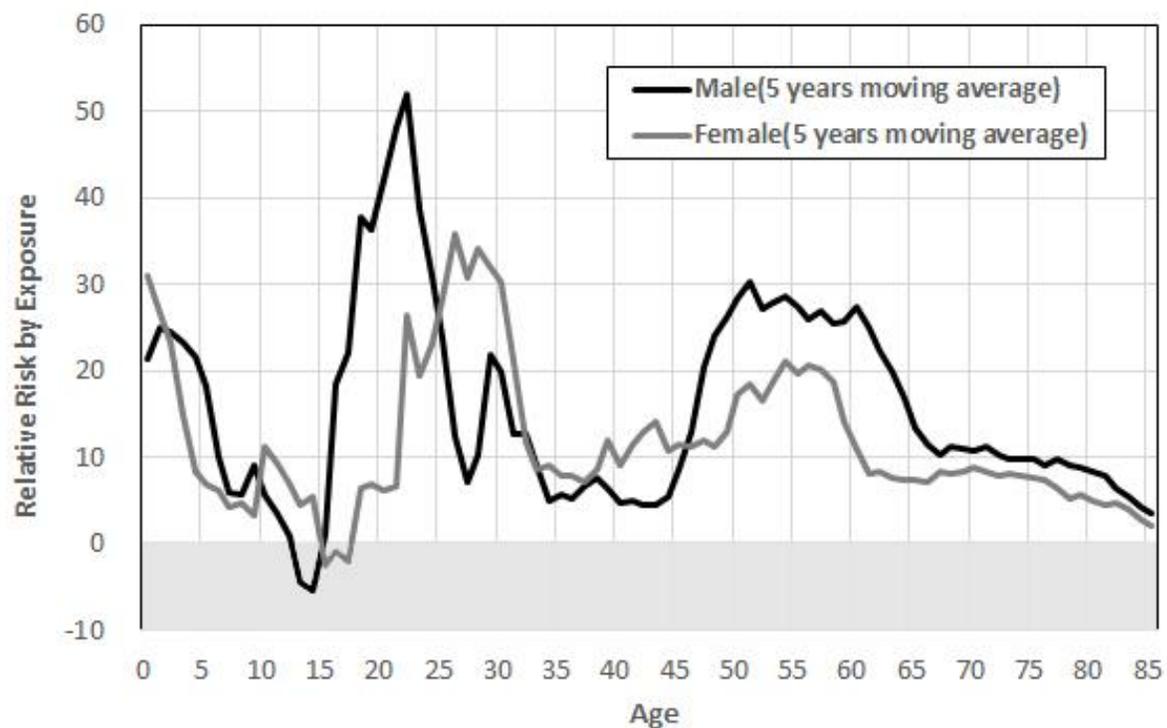
여성

판매 전후 x 연령의 교차효과

# 간질성 폐질환 (J84)

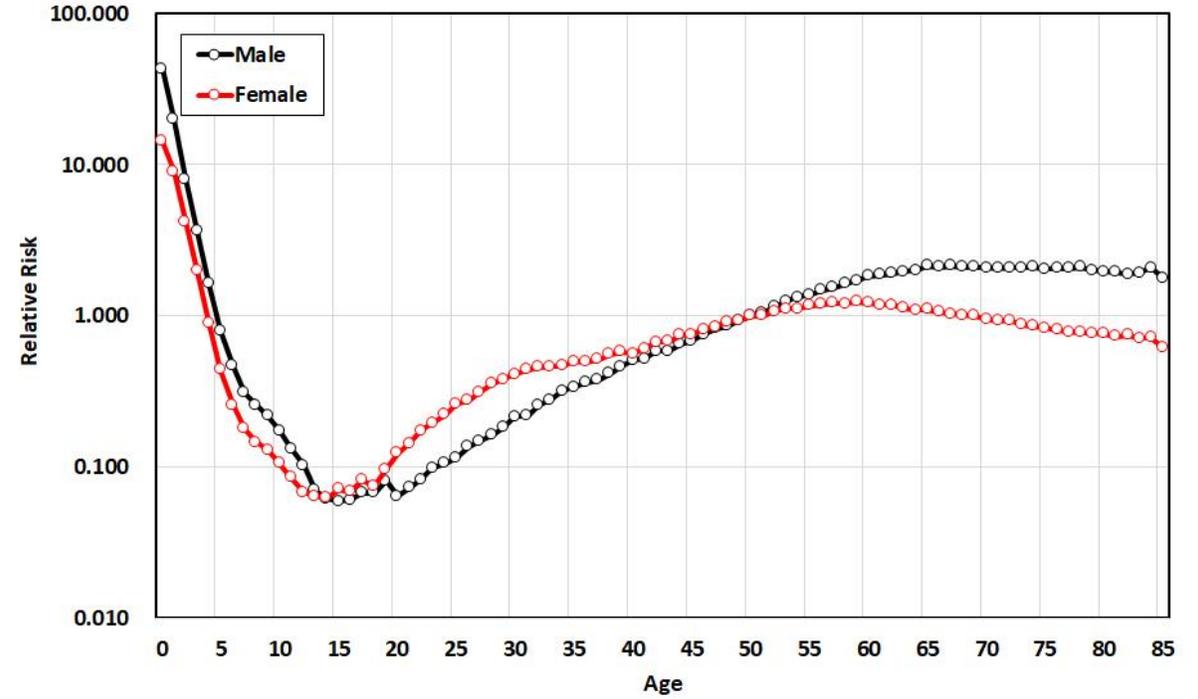
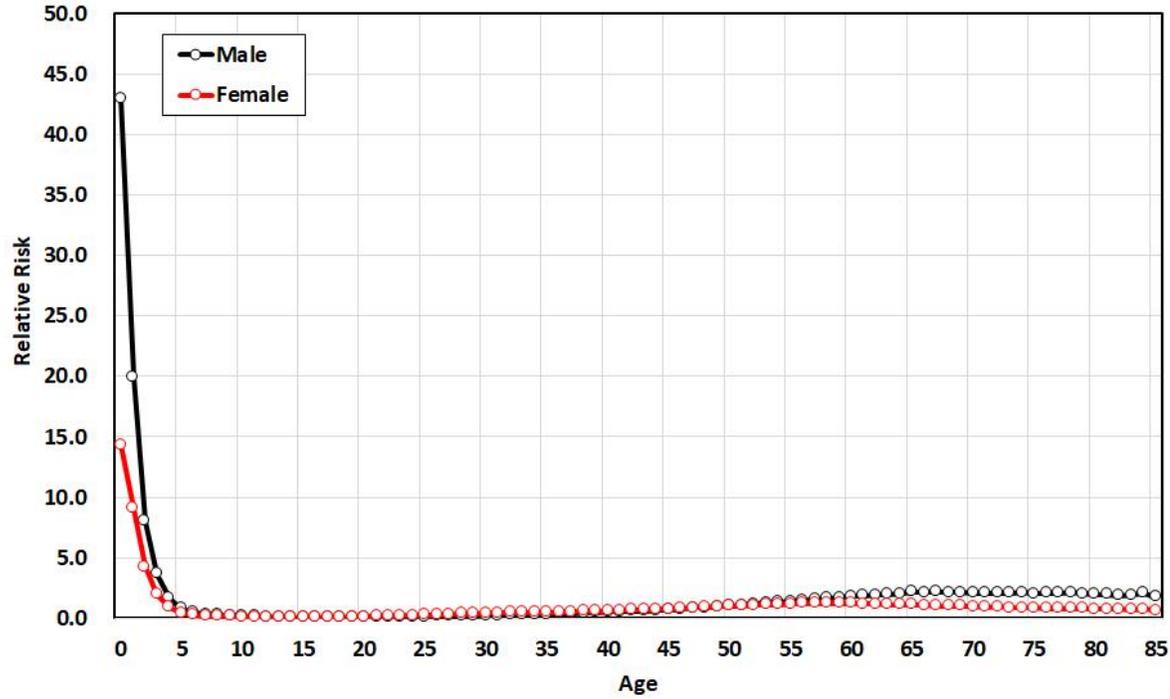


연령별 모집단기여분율



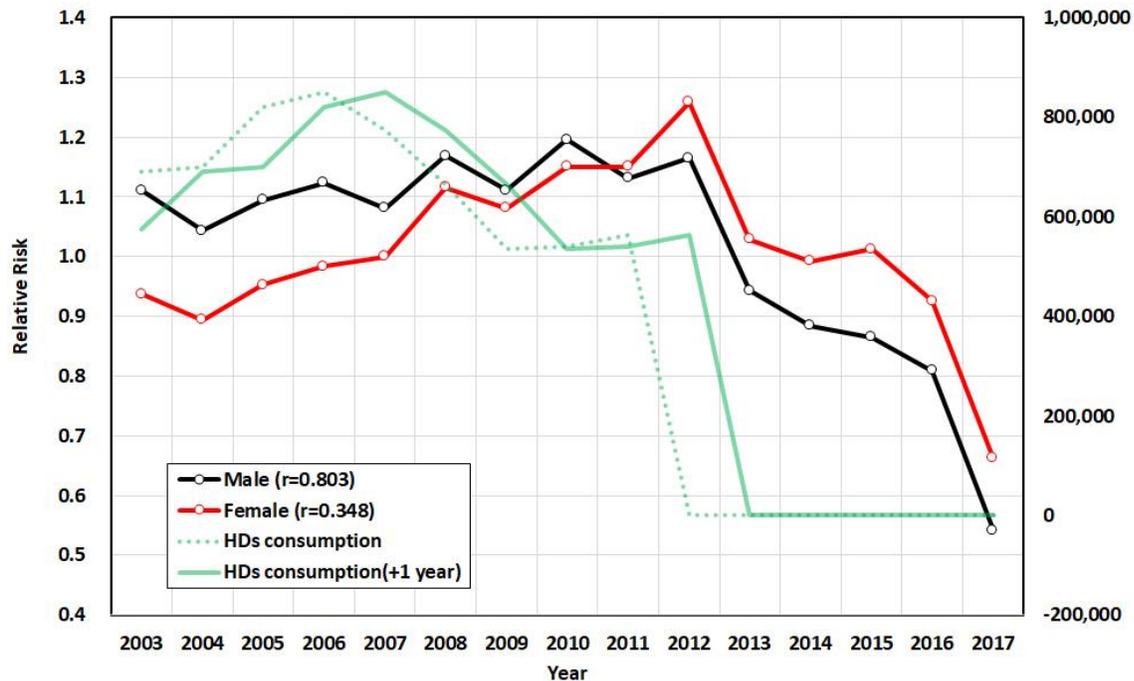
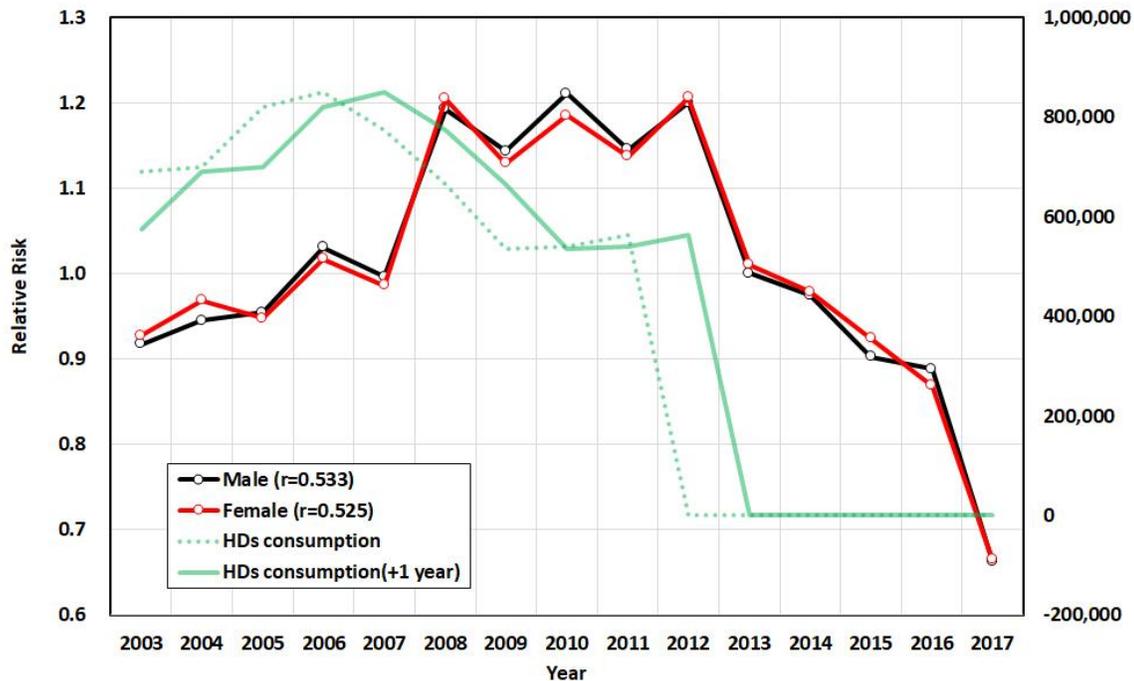
연령별 노출로 인한 상대위험도

# 천식 (J45-J46)



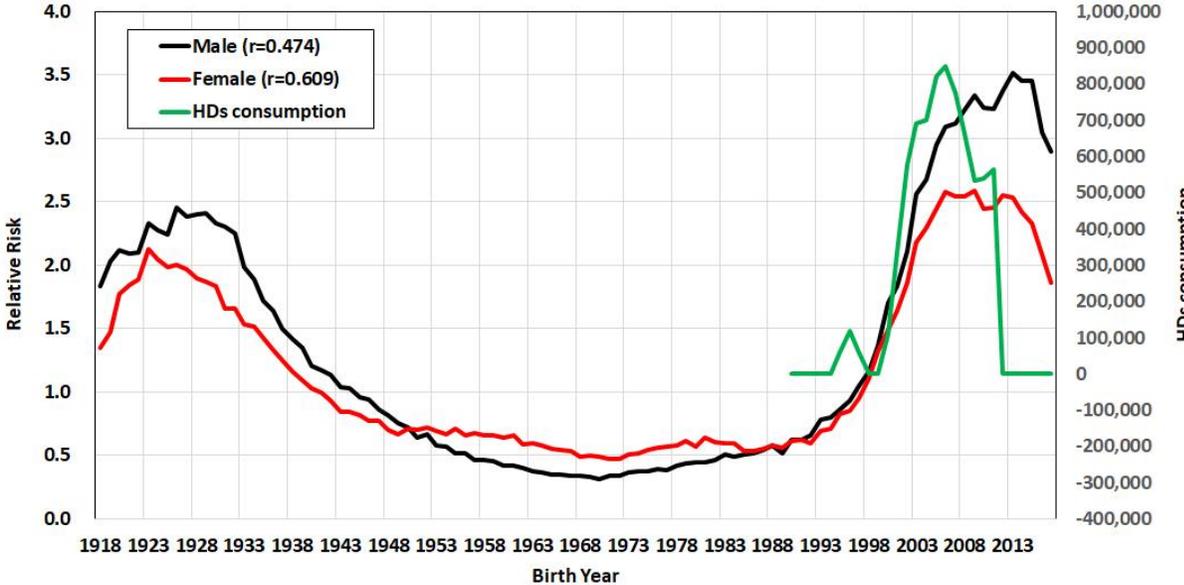
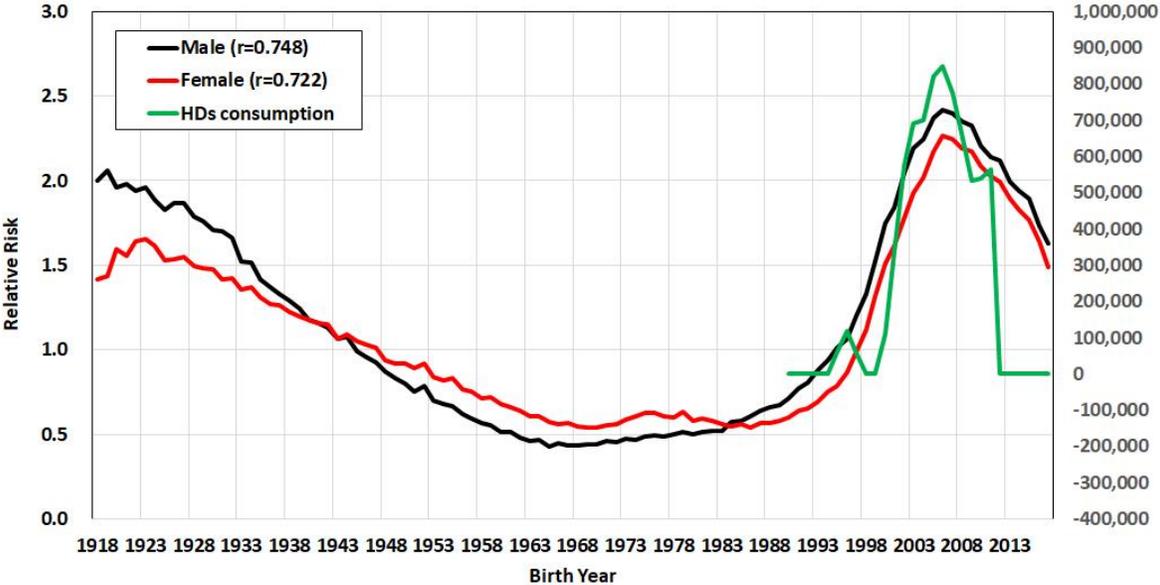
연령 효과

# 천식 (J45-J46)



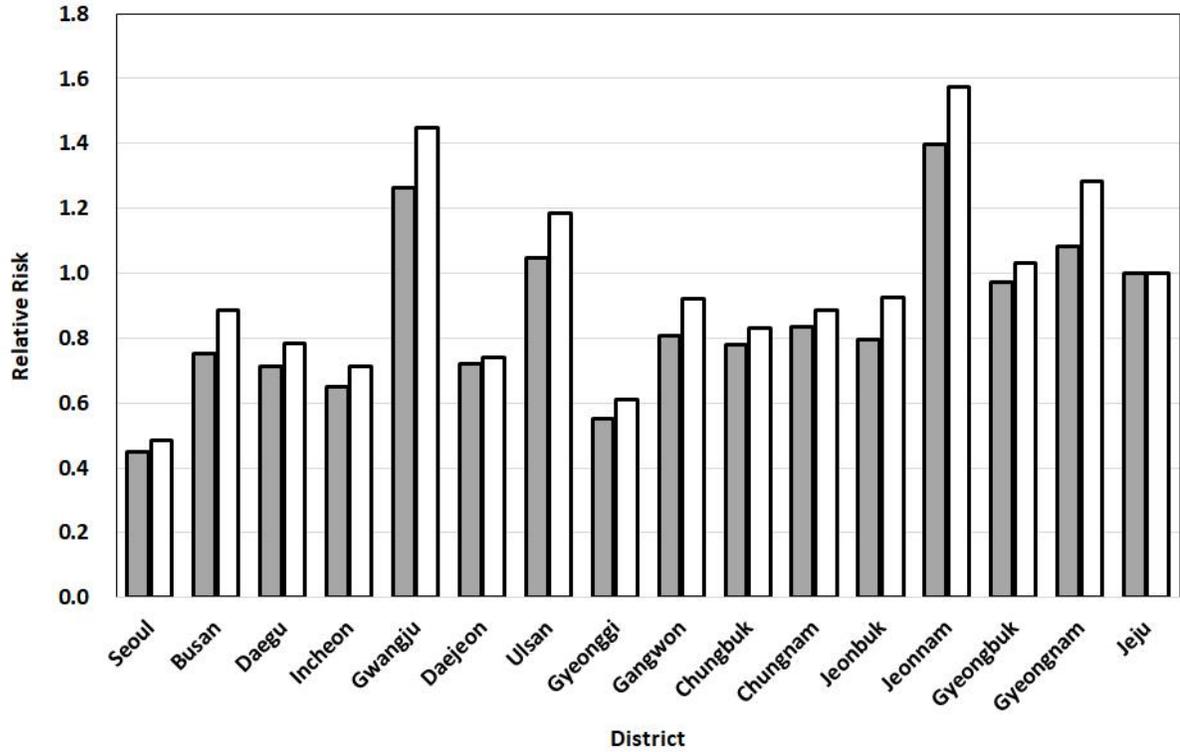
## 시기 효과

# 천식 (J45-J46)

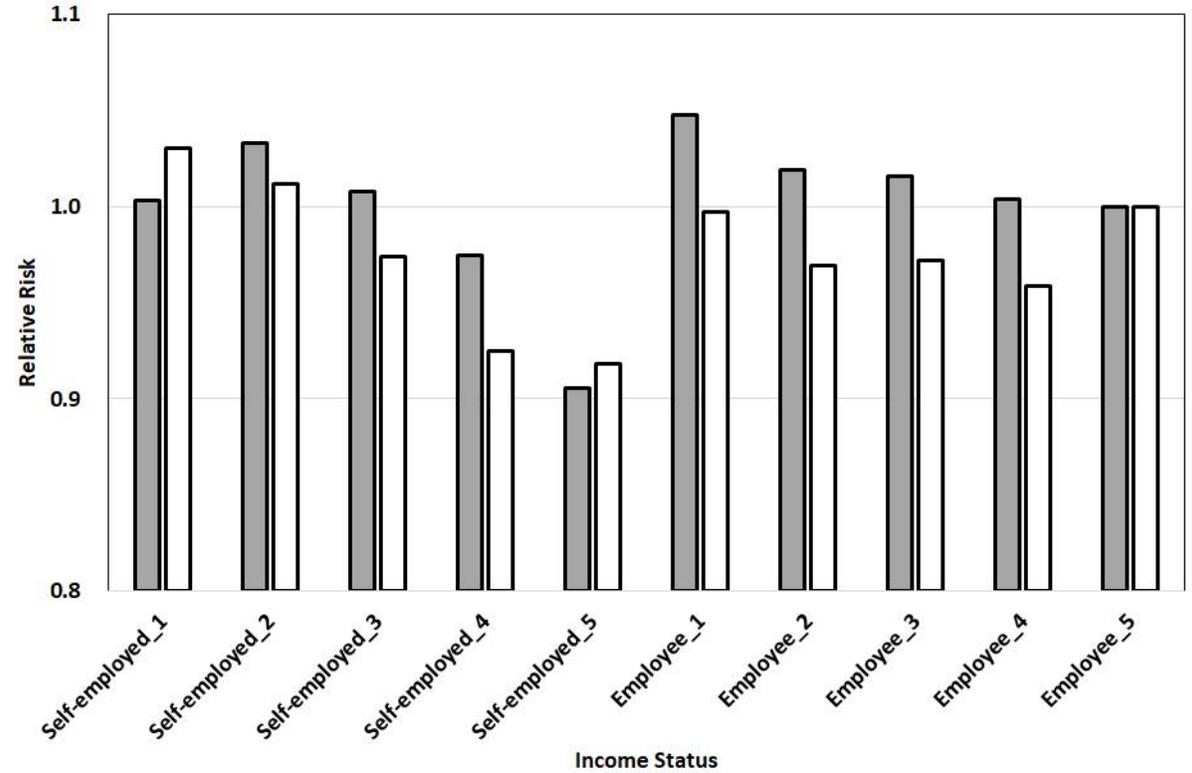


## 코호트 효과

# 천식 (J45-J46)

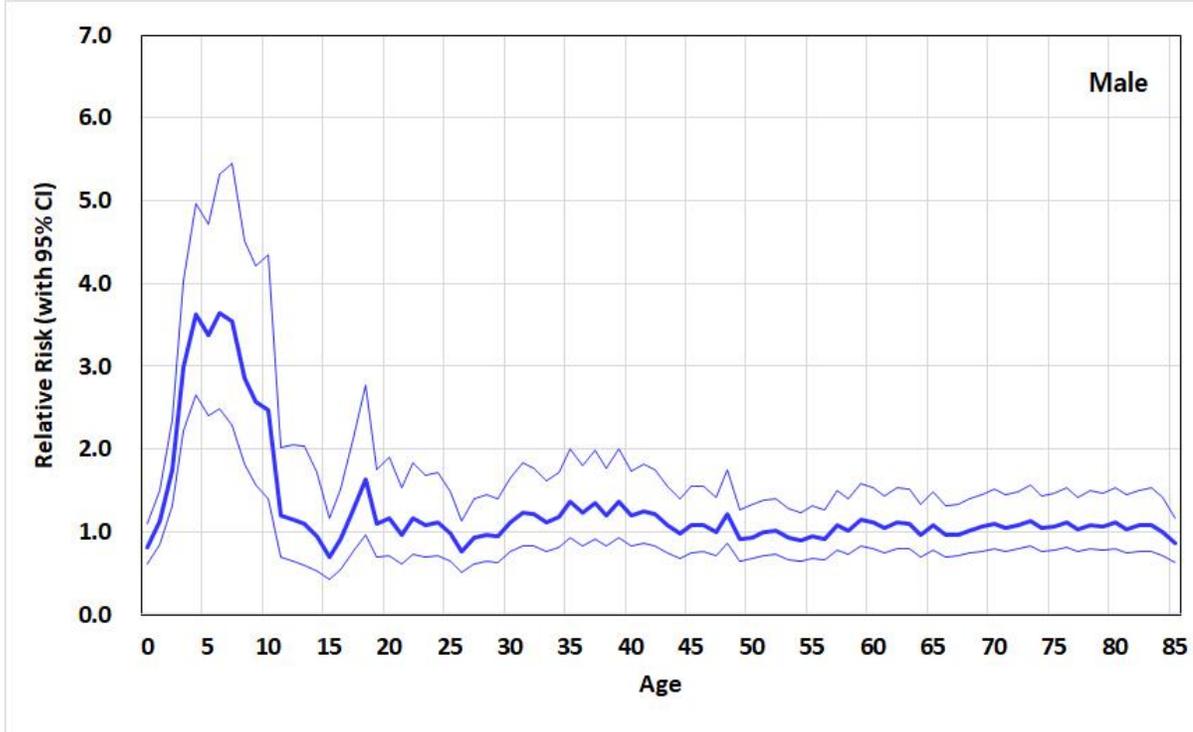


지역 효과

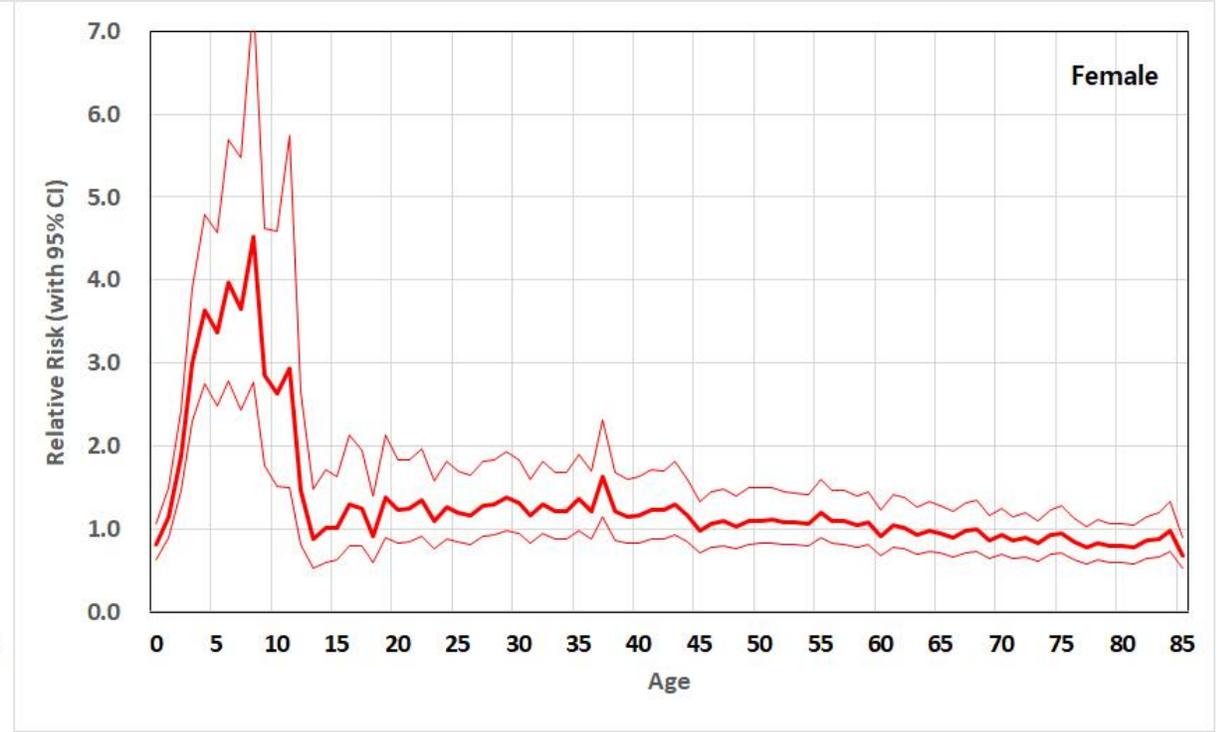


소득수준 효과

# 천식 (J45-J46)



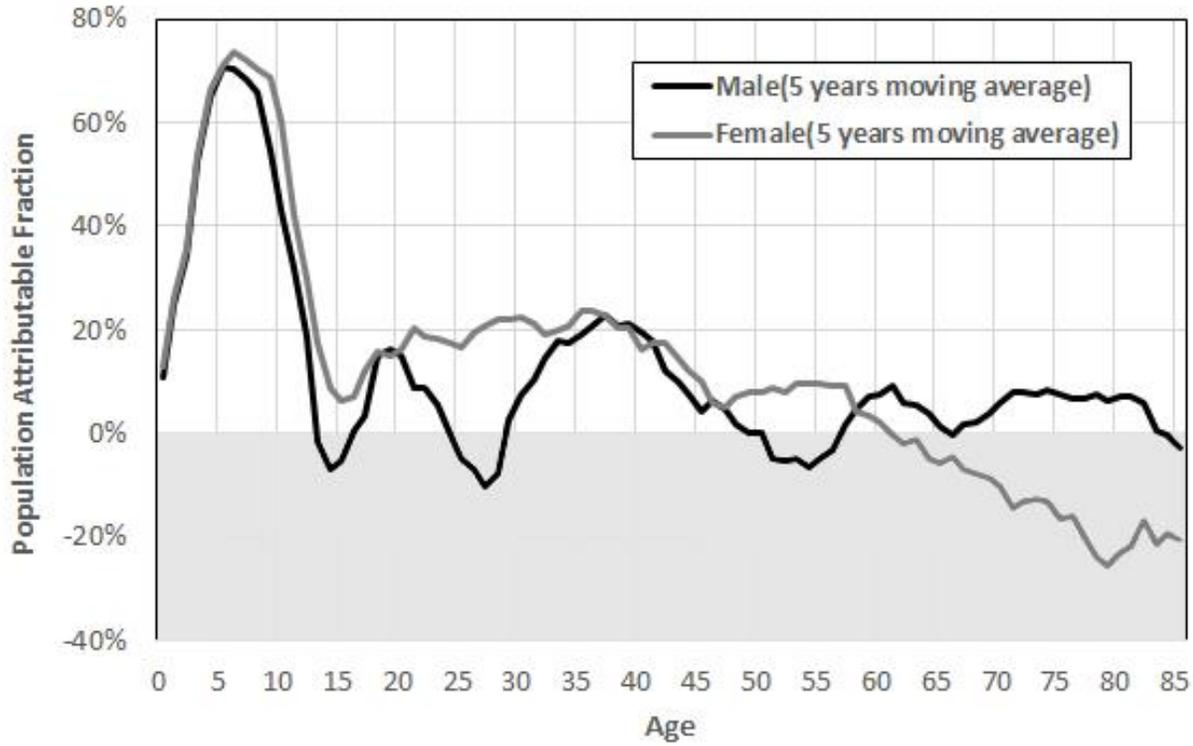
남성



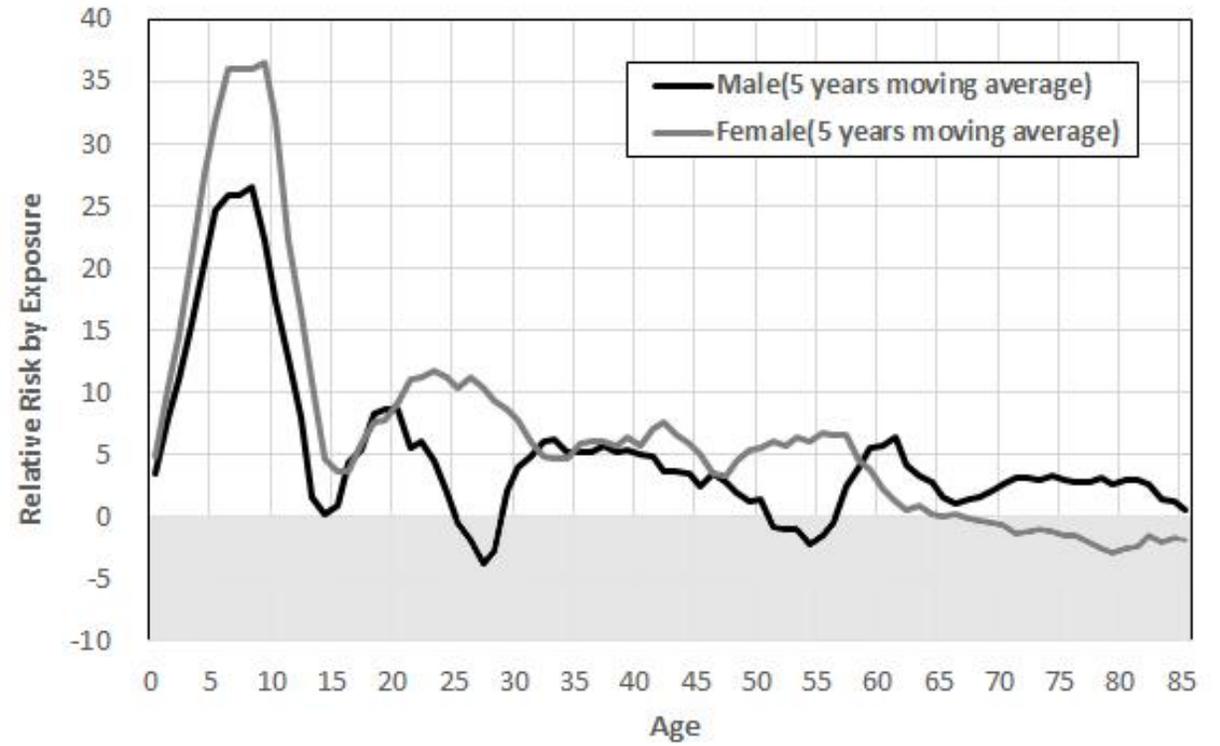
여성

판매 전후 x 연령의 교차효과

# 천식 (J45-J46)

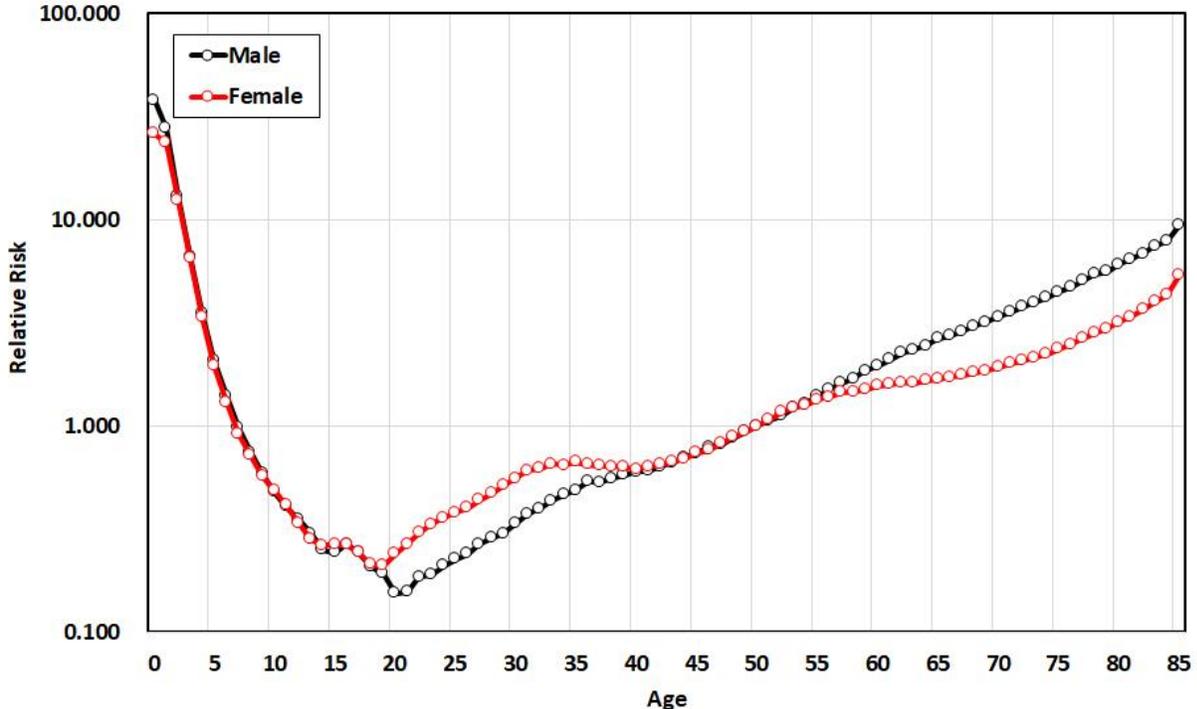
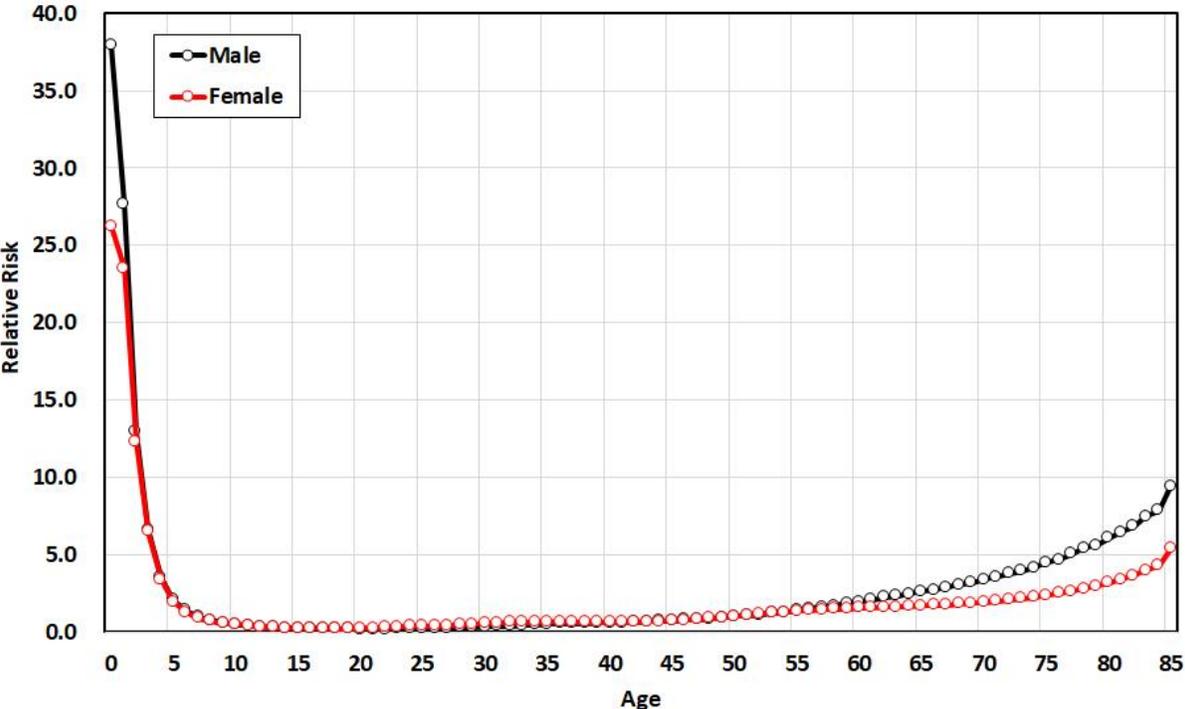


연령별 모집단기여분율



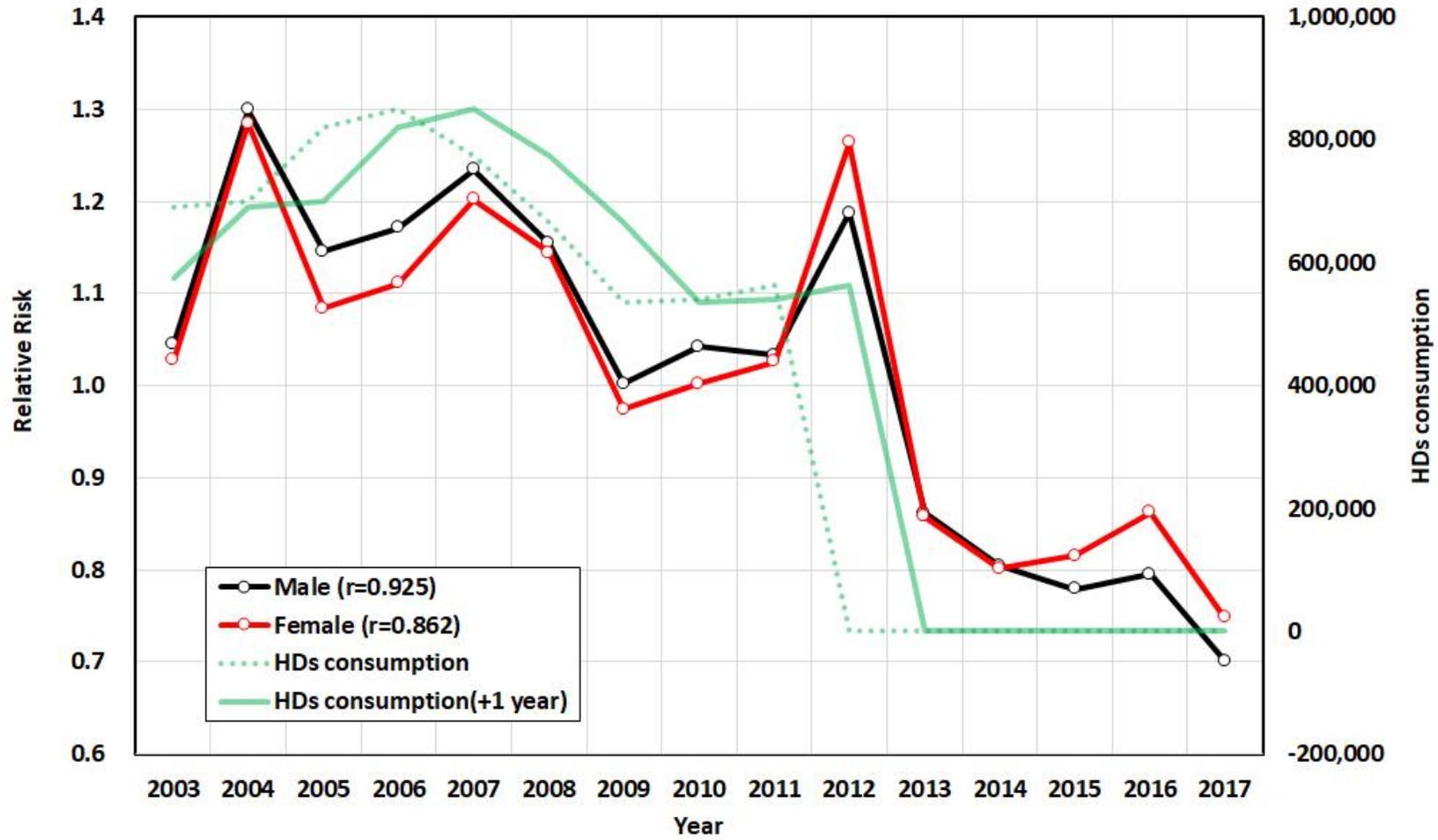
연령별 노출로 인한 상대위험도

# 폐렴 (J12-J18)



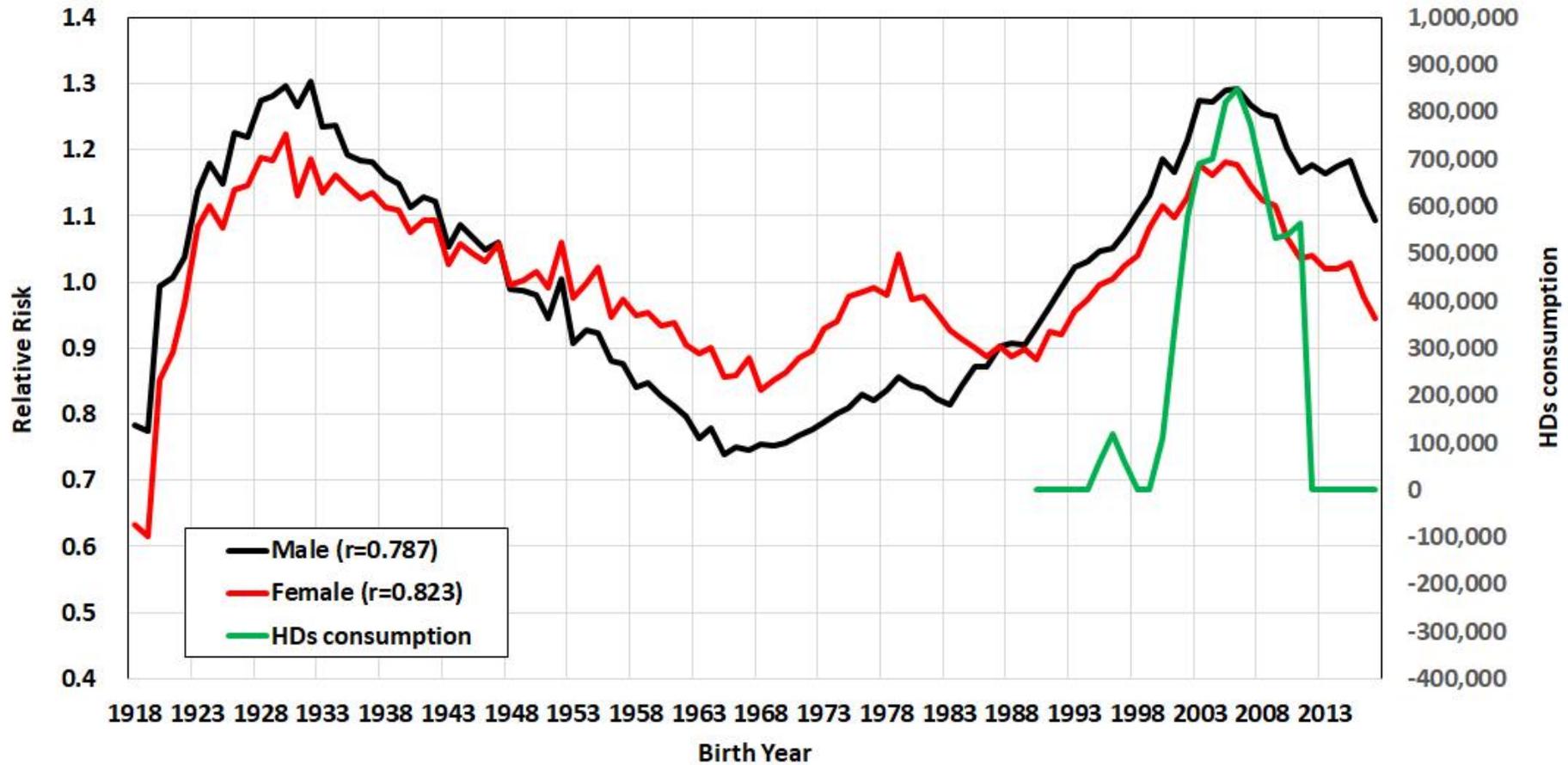
## 연령 효과

# 페렴 (J12-J18)



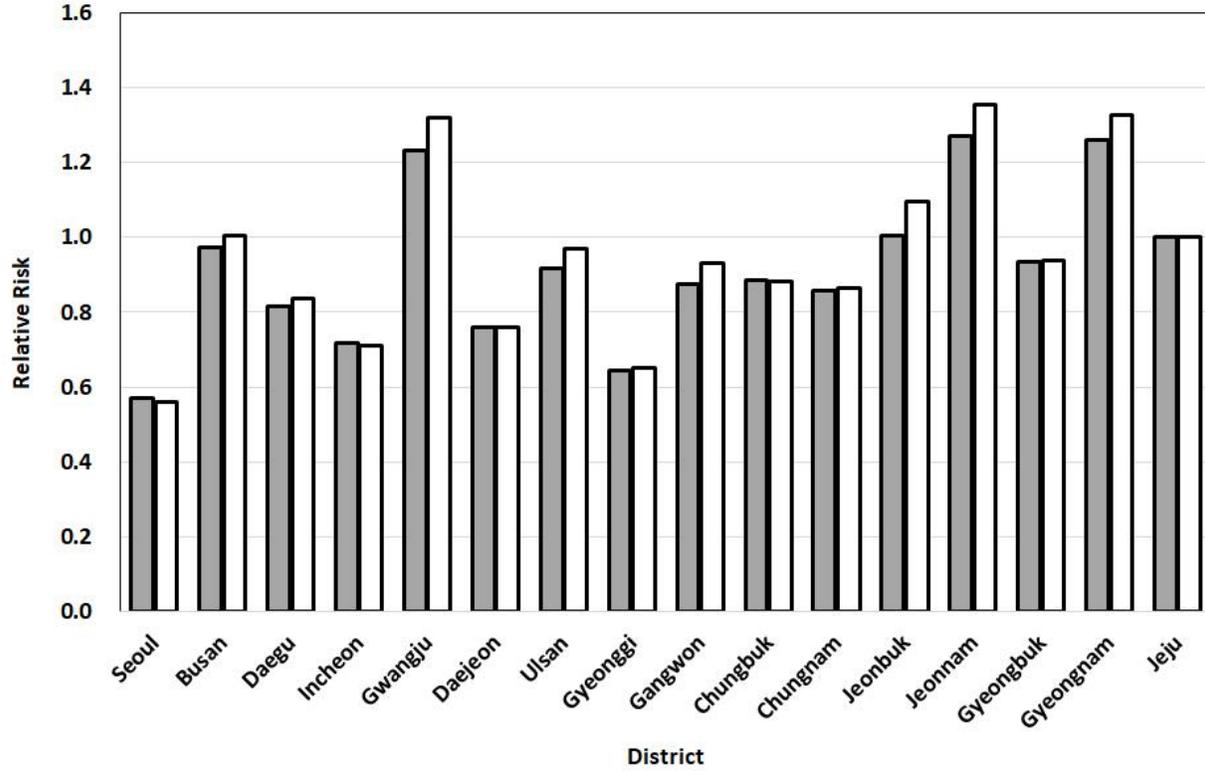
시기 효과

# 페렴 (J12-J18)

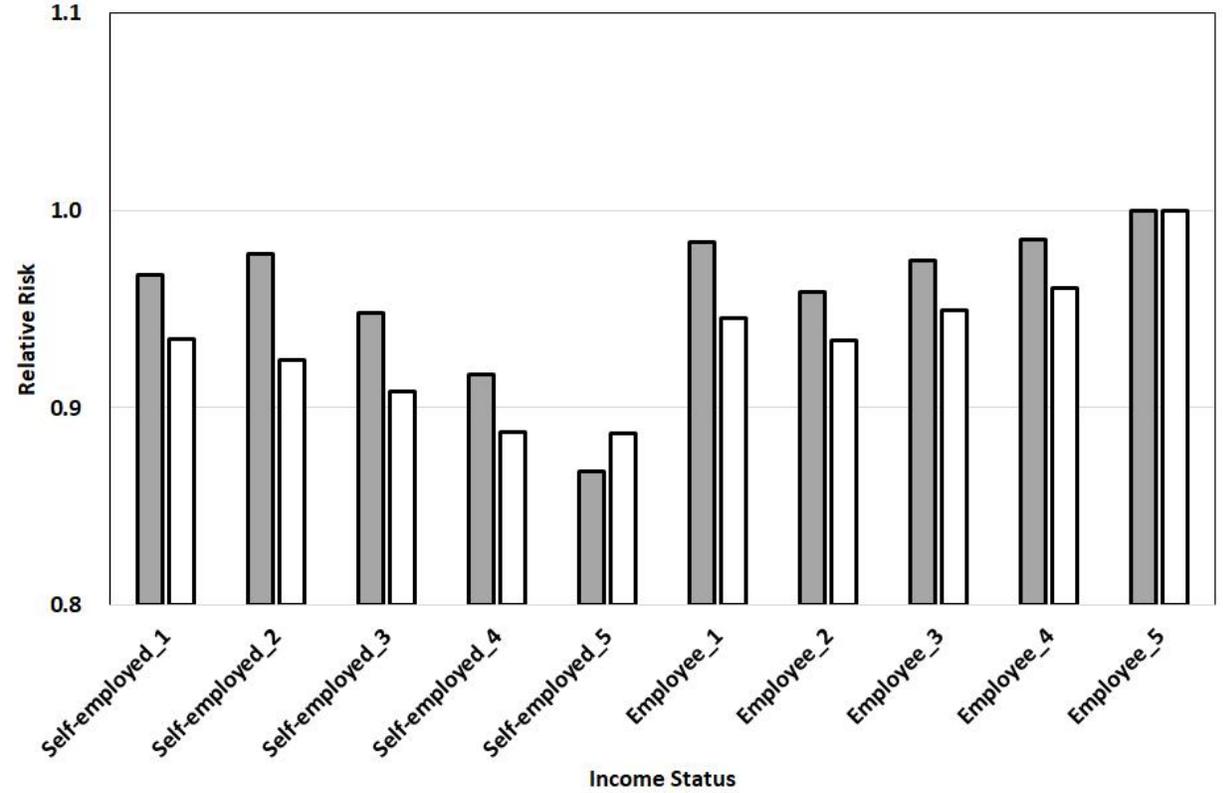


코호트 효과

# 폐렴 (J12-J18)

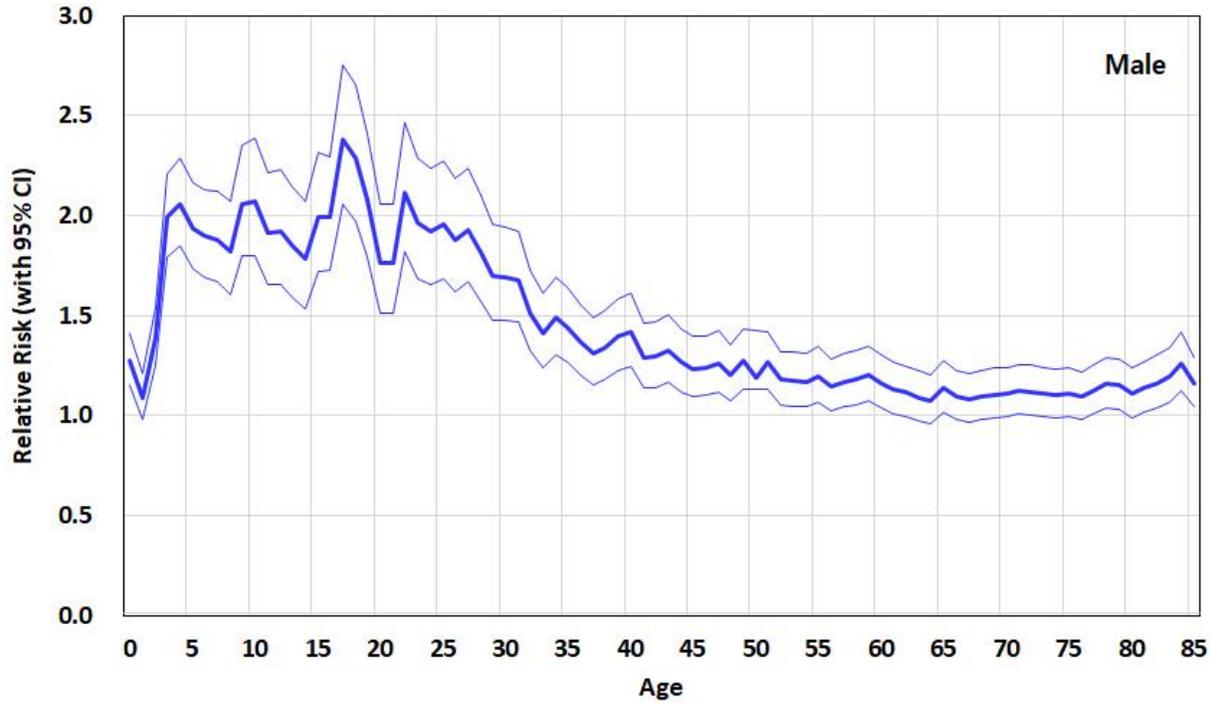


지역 효과

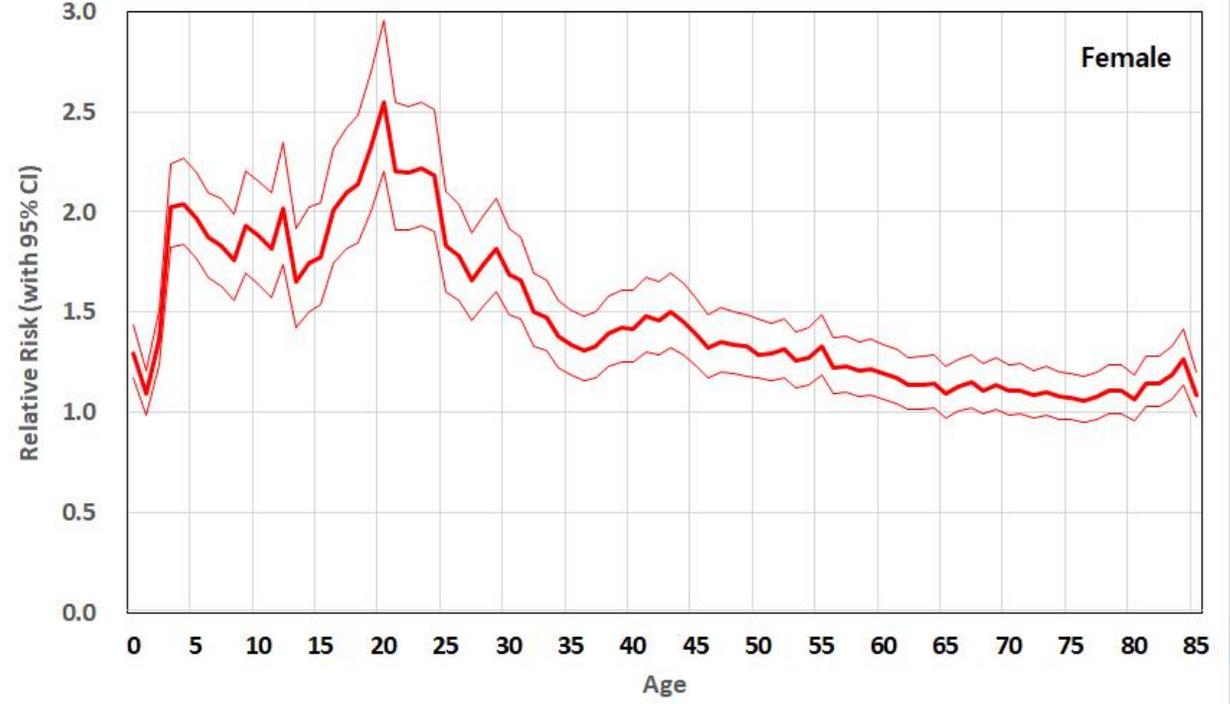


소득수준 효과

# 폐렴 (J12-J18)



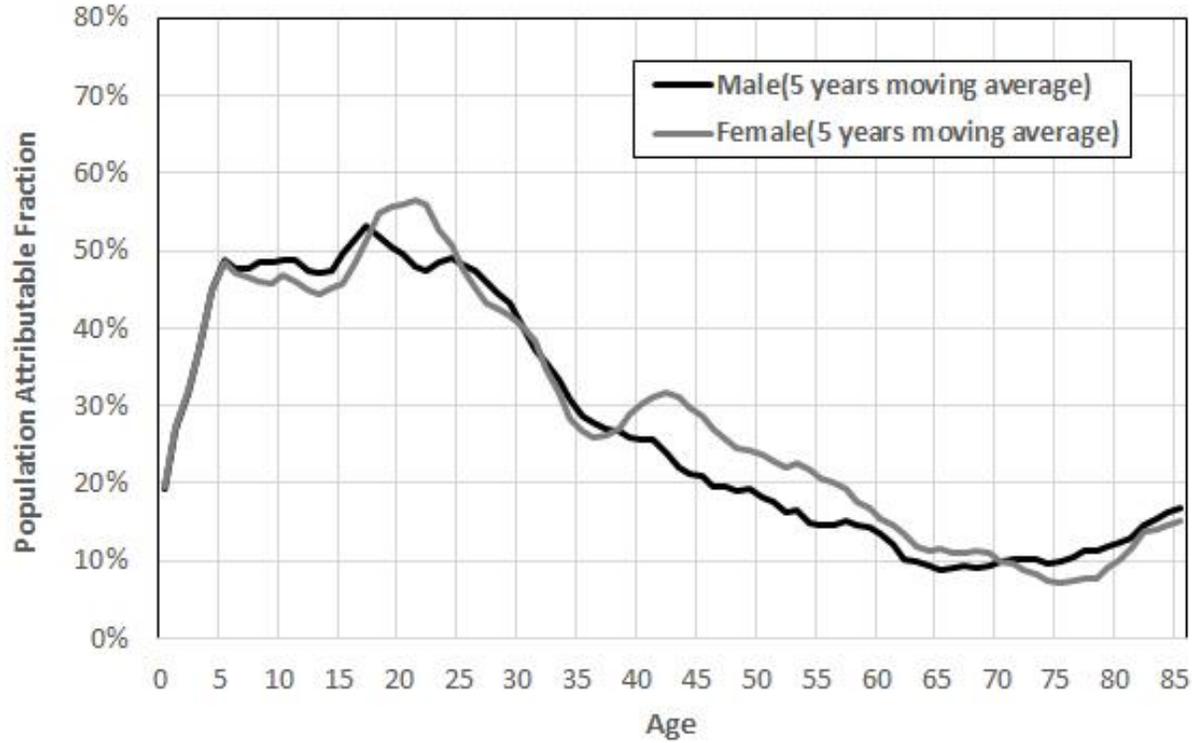
남성



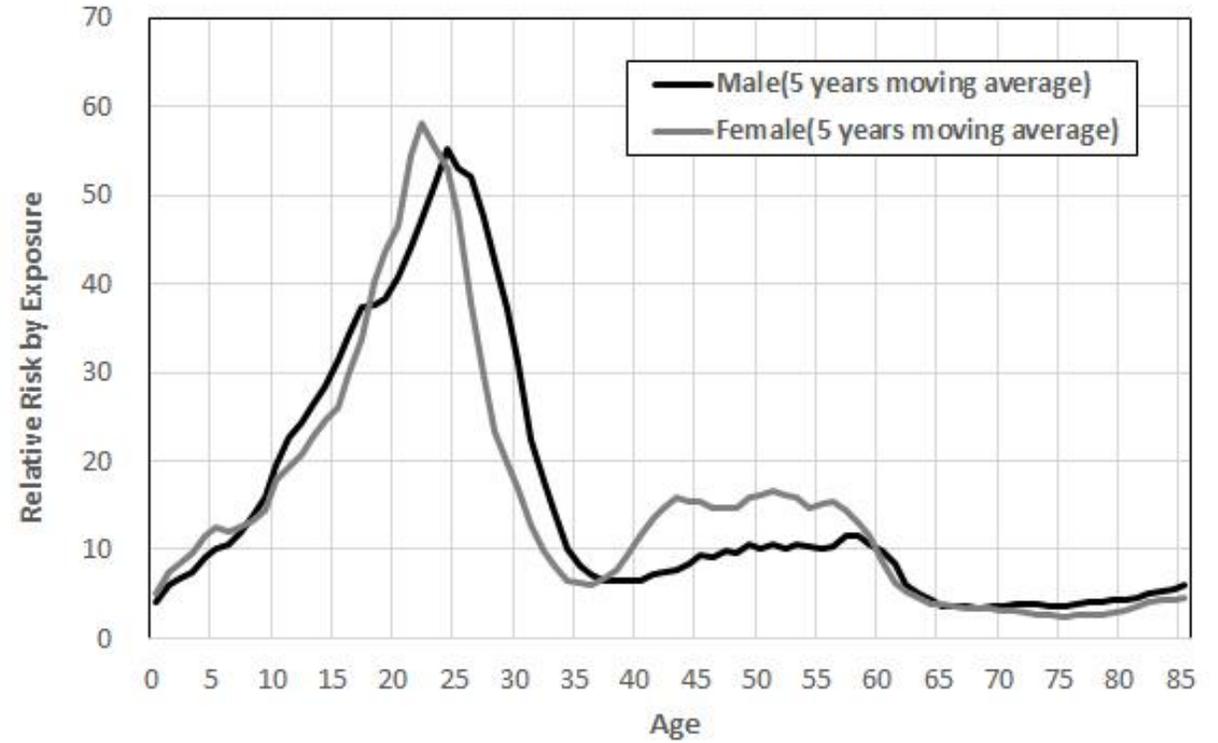
여성

판매 전후 x 연령의 교차효과

# 폐렴 (J12-J18)



연령별 모집단기여분율



연령별 노출로 인한 상대위험도

# [기자 출연] '가습기살균제' 방대한 데이터, 어떻게 조사했나?

손병산 | 기사입력 2018-08-05 20:10 | 최종수정 2018-08-05 20:52

가습기살균제

건강보험

건강

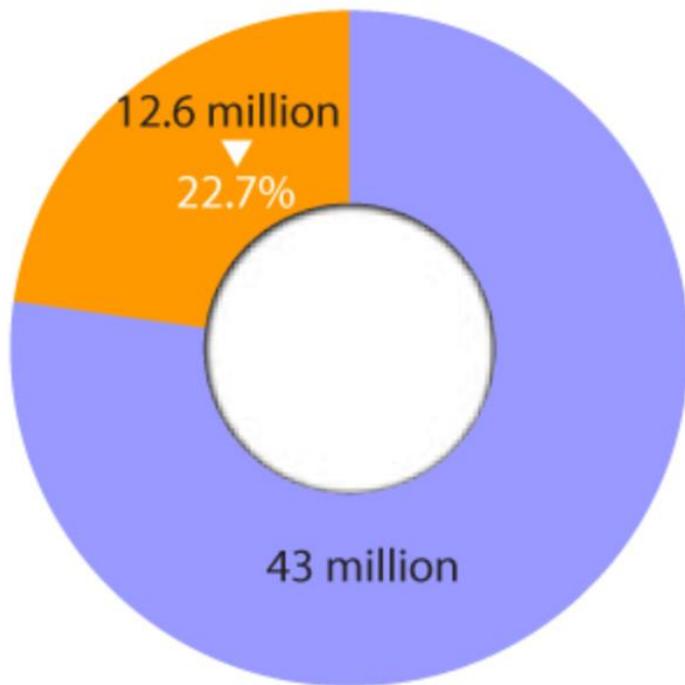
질병



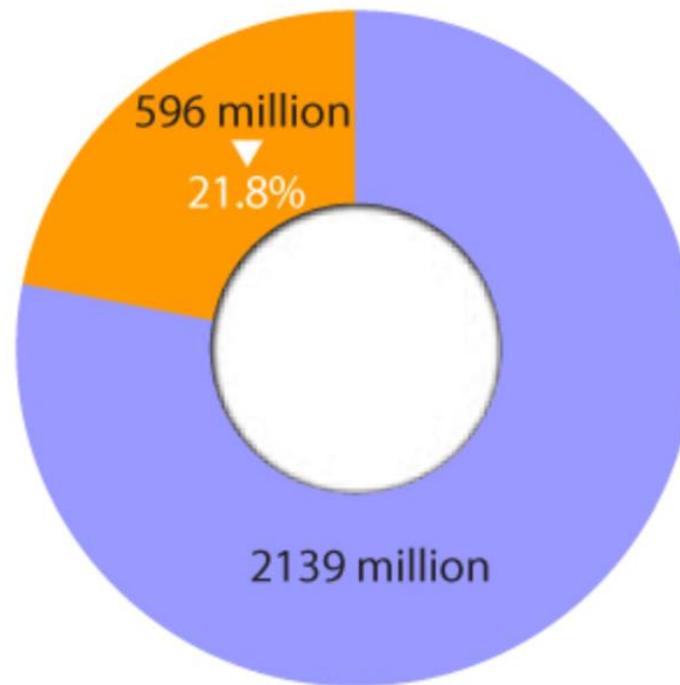
[http://imnews.imbc.com/replay/2018/nwdesk/article/4739543\\_22663.html?menuid=nwdesk&XAREA=mbmain\\_list](http://imnews.imbc.com/replay/2018/nwdesk/article/4739543_22663.html?menuid=nwdesk&XAREA=mbmain_list)



Deaths (millions)



DALYs (millions)



- Attributable to the environment
- Not attributable to the environment

Source: Preventing disease through healthy environments: a global assessment of the burden of disease from environmental risks. WHO, 2016.

# 환경성 질환관리를 위한 정보 활용?

비료공장 들어선 2001년부터 '장점마을의 비극'은 시작됐다

송윤경 기자 kyung@kyunghyang.com

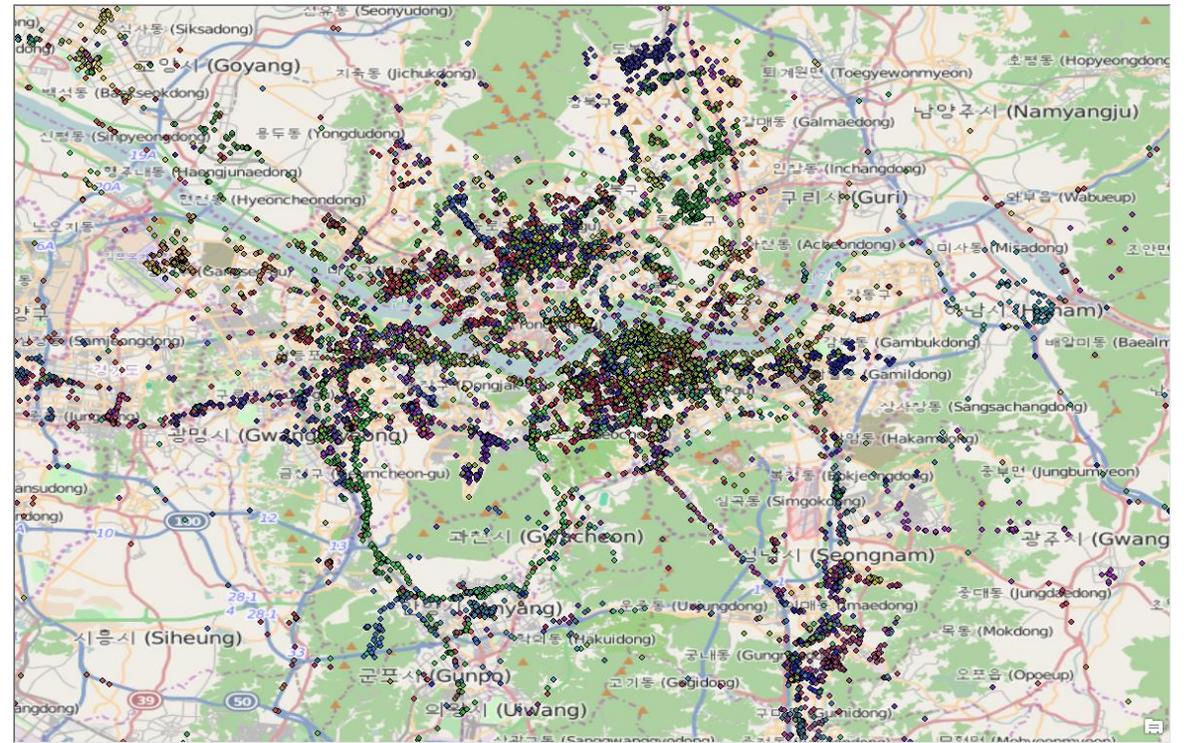
851



입력 : 2018.01



## 감염병 관리를 위한 정보 활용 예시



**Big  
Data**

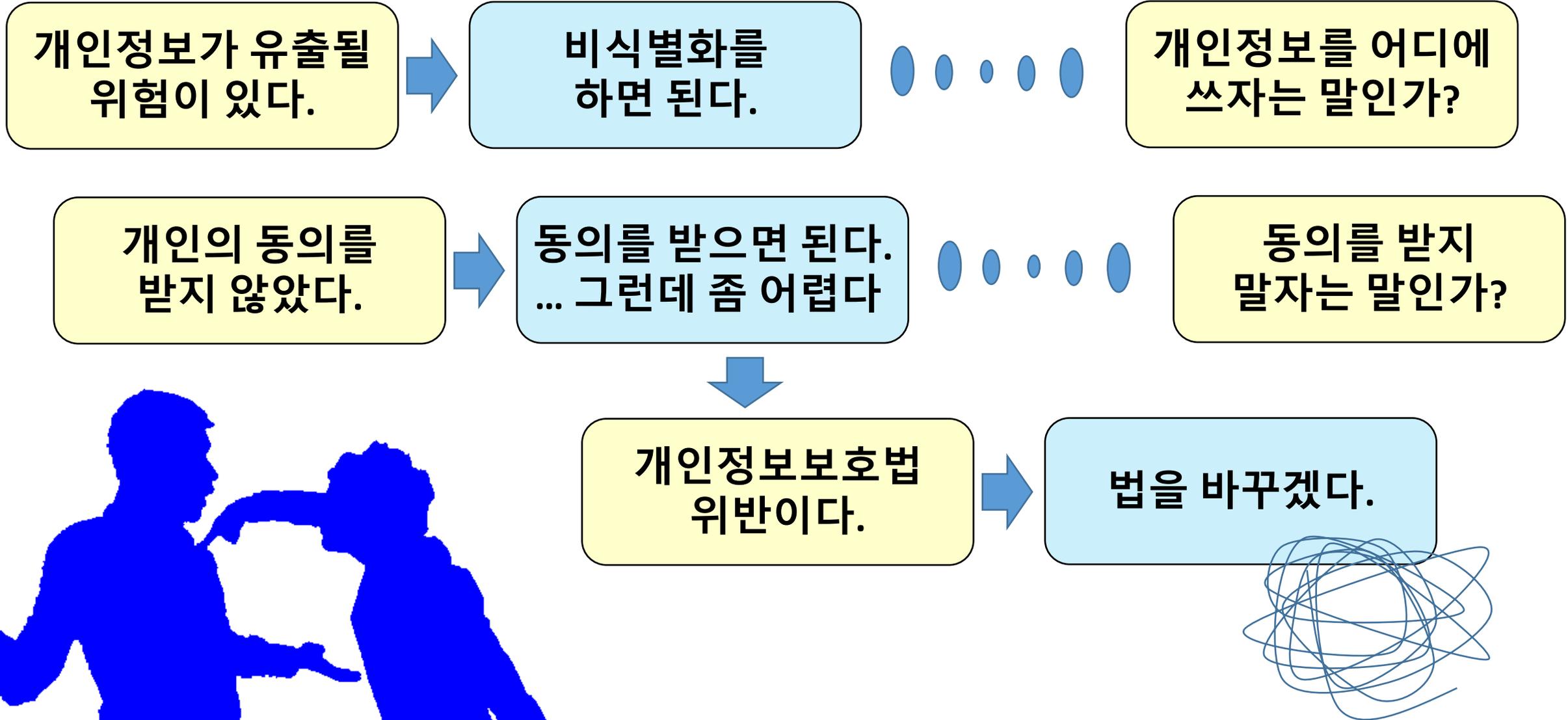
**근거기반  
공공정책**

**Evidence  
Based  
Public policy**

**산업화  
사익 추구**

**New version  
Of  
Panopticon**

# 그 동안의 논쟁 패턴

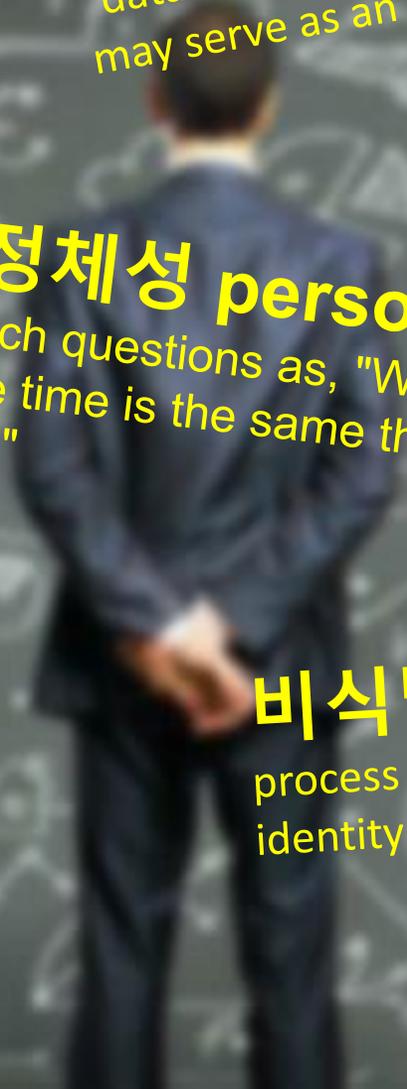


**익명화 anonymization**  
"data is scrubbed for any information that may serve as an identifier of a data subject."

**가명화 pseudonymization**  
"the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information."

**개인의 정체성 personal identity**  
deals with such questions as, "What makes it true that a person at one time is the same thing as a person at another time?"

**비식별화 de-identification**  
process used to prevent a person's identity from being connected with information



우리는 동의 또는 익명화 패러다임 내에서 방법을 찾는 것이 점점 더 어려워질 것이라고 믿는다.

POLICY



European Journal of Human Genetics (2016) 24, 956–960  
© 2016 Macmillan Publishers Limited All rights reserved 1018-4813/16

[www.nature.com/ejhg](http://www.nature.com/ejhg)

# Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach

This paper has been amended since online publication and a corrigendum also appears in this issue.

Menno Mostert\*, Annelien L Bredenoord, Monique CIH Biesart and Johannes JM van Delden

Medical research is increasingly becoming data-intensive; sensitive data are being re-used, linked and analysed on an unprecedented scale. The current EU data protection law reform has led to an intense debate about its potential effect on this processing of data in medical research. To contribute to this evolving debate, this paper reviews how the dominant ‘consent or anonymise approach’ is challenged in a data-intensive medical research context, and discusses possible ways forwards within the EU legal framework on data protection. A large part of the debate in literature focuses on the acceptability of adapting consent or anonymisation mechanisms to overcome the challenges within these approaches. **We however believe that the search for ways forward within the consent or anonymise paradigm will become increasingly difficult.** Therefore, we underline the necessity of an appropriate research exemption from consent for the use of sensitive personal data in medical research to take account of all legitimate interests. **The appropriate conditions of such a research exemption are however subject to debate, and we expect that there will be minimal harmonisation of these conditions in the forthcoming EU Data Protection Regulation.** Further deliberation is required to determine when a shift away from consent as a legal basis is necessary and proportional in a data-intensive medical research context, and what safeguards should be put in place when such a research exemption from consent is provided.

*European Journal of Human Genetics* (2016) 24, 956–960; doi:10.1038/ejhg.2015.239; published online 11 November 2015



# [사설]개·망·신 法 울타리 갇혀선 4차 산업혁명 미래 없다

동아일보    입력 2018-08-11 00:00    수정 2018-08-11 00:00



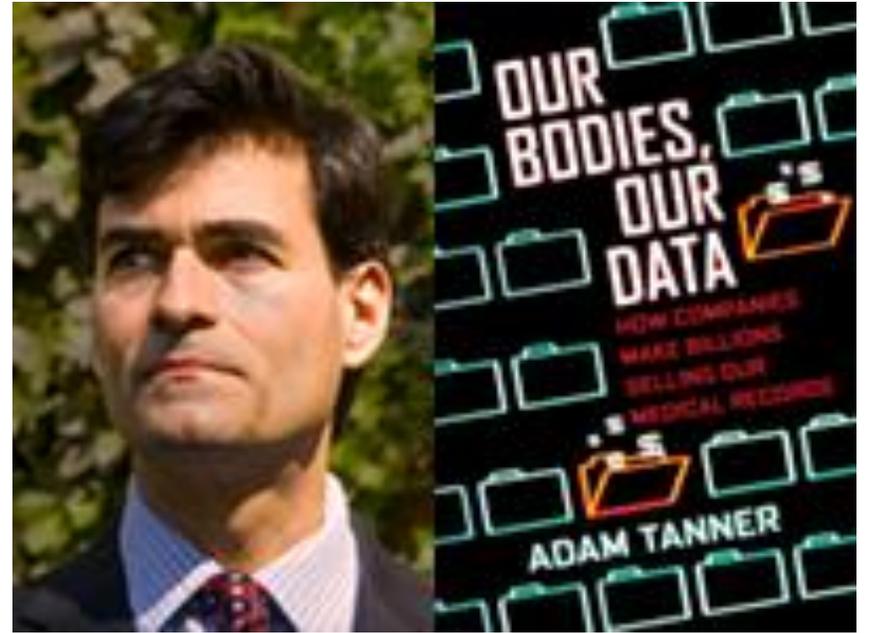
수많은 정보가 모인 빅데이터는 금융과 의료, 유통, 농업 등 전방위 분야에서 신산업의 원재료(原材料)로 활용된다. 빅데이터가 '4차 산업혁명의 석유'로 불리는 이유다. 해외에서는 개인정보를 수집, 가공해 비식별 정보로 만들어 필요한 곳에 공급하는 '데이터 브로커' 사업까지 성장일로에 있다.

한국무역협회는 2020년 전 세계 빅데이터 시장이 2100억 달러, 우리 돈 약 235조 원 규모까지 커질 것이라고 전망했다. 개인정보 보호라는 족쇄에 묶여 있는 한 우리로서는 군침만 흘리면서 쳐다볼 수밖에 없는 시장이다.

건강정보 회사인 엠디온(Emdeon) 부사장인 Kris Joshi는 .... "보건의료는 주입된 정보의 양에 비해서 정말 매우 적은 정도로 작은 양의 가치밖에 만들지 못하고 있다."

모든 과학적 약속들 (scientific promise)에도 불구하고 막상 기업들은 이 방대한 정보 수집물을 주로 세속적인 상업적 목적을 위해 사용한다고 강조한다: "예를 들어, 이것은 의약품 판매에 사용되는데 어느 의사가 무슨 약을 처방하는지 알아내는데 쓰이고 있다."

그들은 우리의 가장 은밀한 정보들을 사고 판다. 이것은 선견지명이 있는 경영진과 이상주의적인 의사, 정보과학자, 영리한 장사꾼 (salespeople), 그리고 자신의 내밀한 정보가 상업과 과학에 이전보다 훨씬 필수적인 요소가 되어 버려 사면초가에 처한 환자들로 구성된 세계다. 그리고 이 세계는 우리의 안녕이 가장 큰 잠재적 위험에 노출된 세계이기도 하다.



# 뉴질랜드 통계청 Integrated Data Infrastructure



- 연구자 심사
- 프로젝트 심사
- 보안 시설
- 데이터 심사
- 연구결과물 심사

# A new organisational structure

In 2017-18, we overhauled NHS Digital's organisational structure to allow us to meet our commitments to the health and care system.

Our programmes, services and corporate functions are now grouped together into seven directorates, which provide clearer lines of accountability, more integrated delivery and a strategic perspective on what our customers and the health and care system require.

## Corporate Services

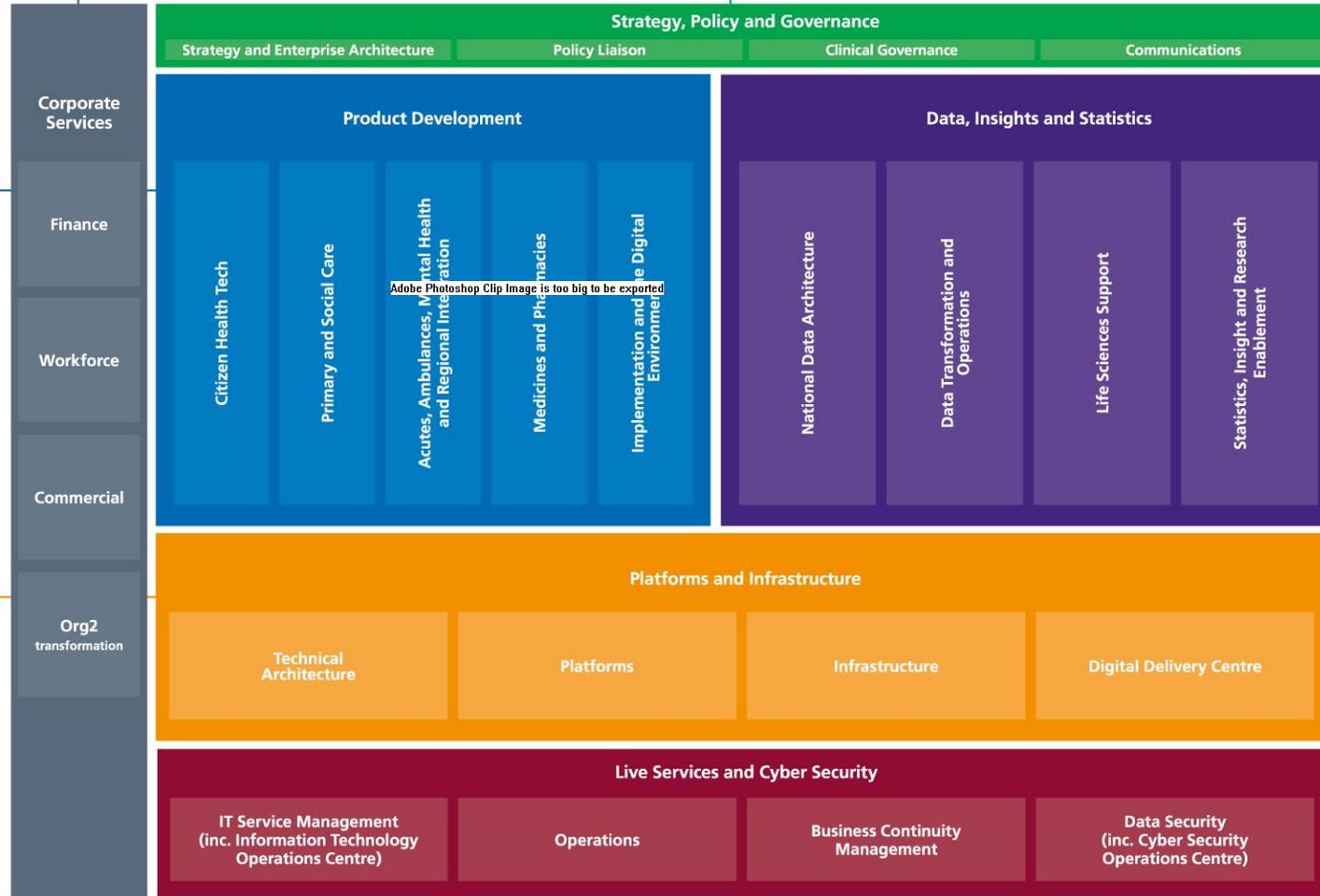
Includes our financial, commercial, people and workforce functions. It is responsible for delivering the improvement programme Org2 over the next two years.

## Product Development

Designs and builds the new applications and services that make a difference for the public and improve health and care services. The directorate works closely with the healthcare technology market and is structured to align directly with stakeholders in key areas: in citizen technology; primary and social care; acute, ambulance and mental health services; and pharmacy.

## Platforms and Infrastructure

Develops the mission-critical IT infrastructure of the NHS and social care system. Our teams are building new platforms to support NHS Digital's data management and product development and facilitating cloud usage, identity and access management, and a fully interoperable IT infrastructure that provides more flexibility for providers and developers.



## Strategy, Policy and Governance

Defines our strategic agenda based on the needs of our customers and the political, technical, government and market environment. It provides clinical and information governance guidance and oversight and works with the Department of Health and Social Care and our national and local partners to shape our policies and governance.

## Assurance and Risk Management

Ensures all our work is quality-assessed and assured from business and technical perspective. The team monitors, evaluates and reports on programme and service delivery as well as all aspects of operational risk and mitigation planning, working closely with internal and external audit functions and our board's Assurance and Risk Committee.

## Data, Insights and Statistics

Fulfills our role as the data custodian for the health and care system and our responsibility to support healthcare research and improve the information available to clinicians and system leaders.

## Live Services and Cyber Security

Responsible for the reliable, secure and effective operation of all live systems that we operate for the health and care system, including core services like e-referrals, electronic prescriptions, NHSmail and cancer screening. The directorate fulfills our responsibilities as the national lead on cyber security for the health and care system.

## Assurance and Risk Management

Information Assurance

Technical Assurance

Programme Assurance

# 우리 사회는 보건의로 빅데이터의 공익적 활용에 성공할 수 있을까?

- 사회적 편견과 낙인을 강화하는 연구?
- 개인의 자기검열을 부추기는 연구?
- 근거가 없는 사이버 과학(pseudo-science)?
- 자기결정권을 악용하는 집단의 도덕적 해이와 불공정 거래(my data)
- 무책임하고 일관성이 없는 정치 관행

# 전제 조건들

- 학계와 전문가 그룹의 공익적 책무성 (통제된-검증된 기득권)
- 공공 부문의 역량 및 전문성, 정치적 중립성 강화 (공공≠공권력)
- 보건의료 부문의 공익성 및 비영리성, 환자중심성 강화 (의사≠신)
- 개인과 프라이버시, 인권에 대한 사회적 민감성 (사람>돈)
- 사회적 통합과 개방성의 추구 → 신뢰 체계의 추구

→ 빅데이터 시대에 우리 국민 모두가 마지막에 웃으려면....

# 의문 사항들

## - Big data의 공익적 활용을 위한 big picture의 결여

- 분야별 검증되지 않은 주장들과 프로젝트 발주, 논문/보고서로만 종료되는 무책임성
- 빅데이터를 활용한 공공부문의 전략적 자기개발 계획 자체가 결여
- 아이디어 중심으로 민간 위탁에 의존 → 현장 활용 실패의 반복

## - 공익적 활용 목적을 담보할 만한 인프라의 결여

- 사회적 논의 우회 / 기관별 자기중심적 세계 구축 / 훈련된 공공 인력 결핍
- 공공조직 내 연구개발 인력들의 상황 (행정직 중심, 기존 현업 중심, 기술역전 현상)