

공개자료원을 활용한 논문 작성과 평가의 고려사항

신애선

2023

서울대학교 의과대학 예방의학교실

암연구소

협동과정 종양생물학

융합전공 혁신의과학

서울대학교병원 의학연구협력센터 임상역학실

<https://snucm.elsevierpure.com/en/persons/y-shin-9>

Editorial Experience

- Associate editor
 - J Cancer Prev (2013~)
 - J Epidemiol (2017-2023)
 - BMC Cancer (2019~)
 - Epi Health (2021~)
 - JMA Journal (2022~)
- Reviewer
 - >30 journals

목차

- 연구에 활용할 수 있는 공개자료원
 - 건강보험공단 청구자료
 - 암등록통계
 - 건강서베이
 - 질병부담연구
 - 유전체 데이터베이스
- 공개자료원 활용 연구의 장점과 주의사항

공개자료원이란?

- **Publicly available data:** data which is accessible in the public domain **without restriction**
 - 누구나 접근 가능
 - 자료제공자가 자료를 통해 생성된 결과물(논문, 보고서)에 대한 타당성을 검증하는 기능 없음 → 동료검토가 중요!

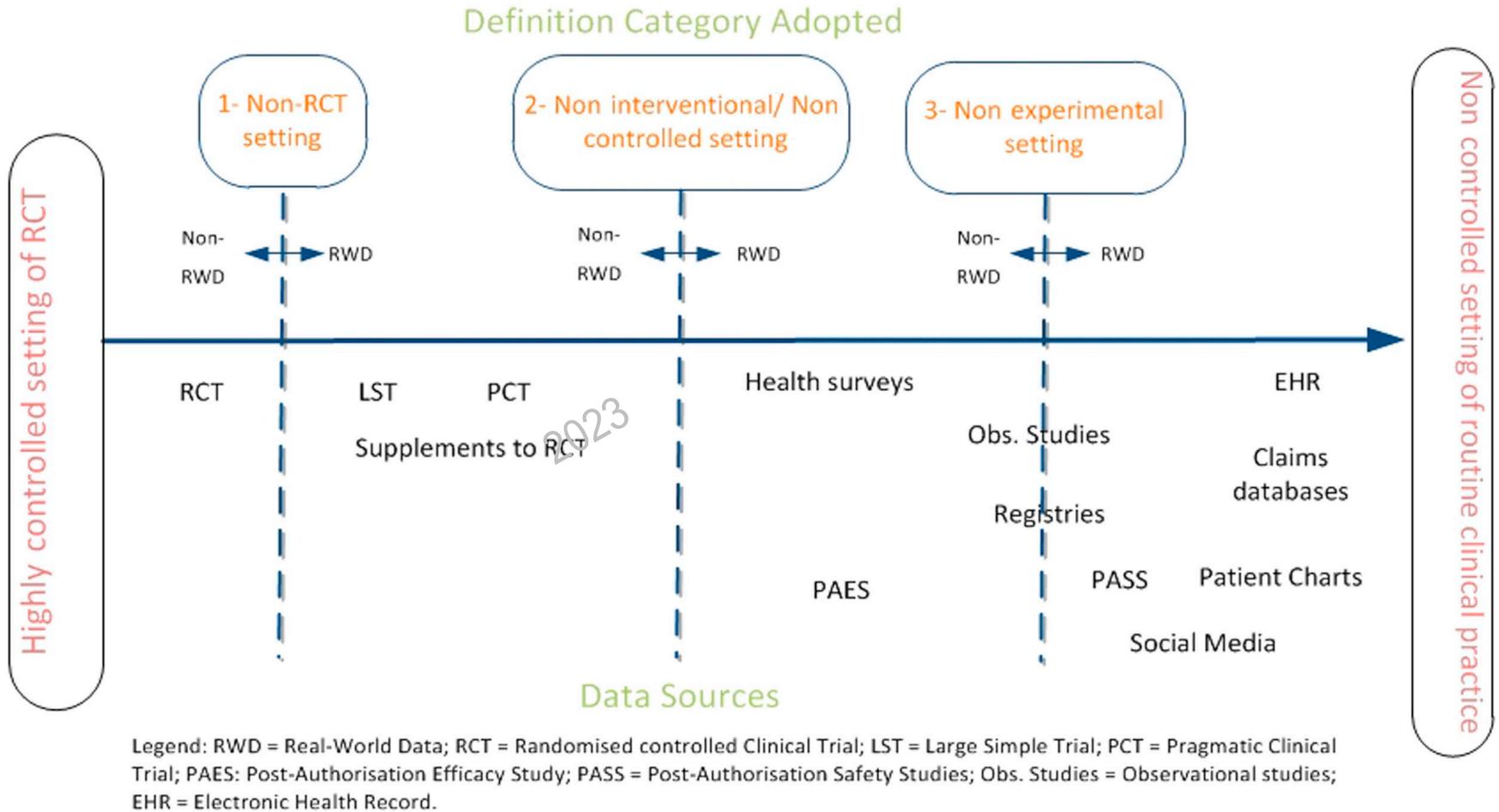
Big data

- Gartner's definition (2001): data that contains greater **variety** arriving in increasing **volumes** and with ever-higher **velocity**. (**3Vs**) (**+Veracity**)
- IBM: Big data analytics is the use of advanced analytic techniques against very large, diverse data sets that include **structured, semi-structured and unstructured data**, from **different sources**, and in different sizes from terabytes to zettabytes. Big data is a term applied to data sets whose size or type is **beyond the ability of traditional relational databases** to capture, manage and process the data with low latency.
- McKinsey Global Report (2011): Big data is data whose scale, distribution, diversity, and/or timeliness require the use of new technical architectures and analytics to enable insights that unlock new sources of business value.

Real-World Data(RWD) & Real-World Evidence (RWE)

- **Real-World Data (RWD)** are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.
- **Real-World Evidence (RWE)** is the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD.

Data spectrum in relation to RWD definition categories



건강보험 청구자료

2023

청구자료란?

- 요양기관(병원)이 환자에게 제공한 진료서비스를
보험자(건강보험공단)에게 청구하면서 발생하게 되는 데이터
- 내용: 진료내역(처치, 시술, 검사 등), 진단명, 보험자 지급 비용, 환자 부담금, 환자 인구학적 특성, 요양기관 정보
- 청구자료 이용 의료연구 분야
 - Outcome research, Comparative effectiveness studies, Quality of Care, Policy Evaluation, Prediction Modeling

((National health insurance service) OR (Health Insurance Review and Assessment)) AND (korea)

NIH National Library of Medicine National Center for Biotechnology Information

Log in

PubMed®

((National health insurance service) OR (Health Insurance Review and Assessment)) AND (korea) X Search

Advanced Create alert Create RSS User Guide

Year	Number of Results
2002	5
2003	5
2004	5
2005	5
2006	5
2007	5
2008	5
2009	5
2010	5
2011	5
2012	5
2013	5
2014	5
2015	5
2016	5
2017	5
2018	5
2019	5
2020	5
2021	5
2022	651
2023	1,161

2002: 5

2023

2023

2023

2002

2023

TEXT AVAILABILITY

At

Free full text

Full text

ARTICLE ATTRIBUTE

Association between serum lipid levels and lower-extremity functions in older adults with and without Alzheimer's dementia in South Korea: A cross-sectional analysis.

Cite Kim SJ, Kim HD.

Share Arch Gerontol Geriatr. 2023 Jul 4;115:105116. doi: 10.1016/j.archger.2023.105116. Online ahead of print.

공공데이터

의료빅데이터

의료통계정보

고객지원

시스템소개



공개는 **넌리!** 제공은 **빨리!** 이용은 **편리!**

건강보험심사평가원에서
보유하고 있는 다양한 의료데이터를
국민에게 개방합니다.

데이터 서비스 현황

- 공공데이터 | 104 종 | **다빈도이용 Top10**
- Open API | 20 종
- 의료통계정보 | SHEET 127 | CHART 40 | Map 0
- 원격분석시스템 | 총 270 계정(사용 207 계정)

의료통계
정보



국민관심질환



국민관심
진료행위



다빈도질병



질병(소분류)



진료행위

의료빅데이터

과제목록

▶ 더보기

- 삼출성 나이관련 활반변성의 발생에 중심장액성 맥락...
- 빠진이가 있는 전이성 암환자에서 골흡수 억제제 투약...
- 조기 유방암 환자의 항암 치료 후에 수반되는 혈액암 ...
- 국내 헌팅병의 현황과 예후 예측인자

전문적인 빅데이터 분석

- 이용안내 >
- 이용신청 >
- MY 분석과제 >
- 원격분석시스템 > 바로가기
- 빅데이터분석연습 >

공공데이터 신청 안내



공공데이터

바로가기



Open API

바로가기



환자데이터셋

바로가기

공지사항

▶ 더보기

- [필독] 맞춤형 연구분석 상병정보 ... 2021-03-16
- 원격분석시스템 접속 환경 설정 2018-05-11
- Internet Explorer 사용자 필독(사용 ... 2015-06-24

HIRA 빅데이터 브리프

▶ 더보기

- [제4권4호]HIRA 빅데이터 브리프 ... 2020-12-30
- [제4권4호]II. 안전한 가명처리 방법 2020-12-30
- [제4권4호]II. 공공기관의 가명정보 ... 2020-12-30



Q&A



FAQ



용어사전



데이터
레이아웃

https://nhiss.nhis.or.kr/

h-well 국민건강보험 **NHISS** National Health Insurance Sharing Service

신애선 님 로그아웃 MY서비스 사이트맵 ENGLISH

서비스이용안내 데이터신청 성과공유 통계 의료이용지표 공공데이터 고객센터

건강보험자료 공유서비스

근거중심의 보건의료 정책 및 학술연구 지원을 선도합니다.

연구DB신청 바로가기 데이터결합신청 바로가기

2023

 연구DB →

 성과공유 →

 통계 →

 의료이용지표 →

 의료지도서비스 →

 한국인비만지수 →

 결합시스템 →

 FAQ →

심의위원회 일정

회차	마감일	심의일	심의건수
제159차	01/28	02/17	58건
제160차	02/12	02/26	36건
제161차	02/25	03/15	46건
제162차	03/11	03/23	1건

문의사항

부서별 번호확인



Data Resource Profile: The National Health Information Database of the National Health Insurance Service in South Korea

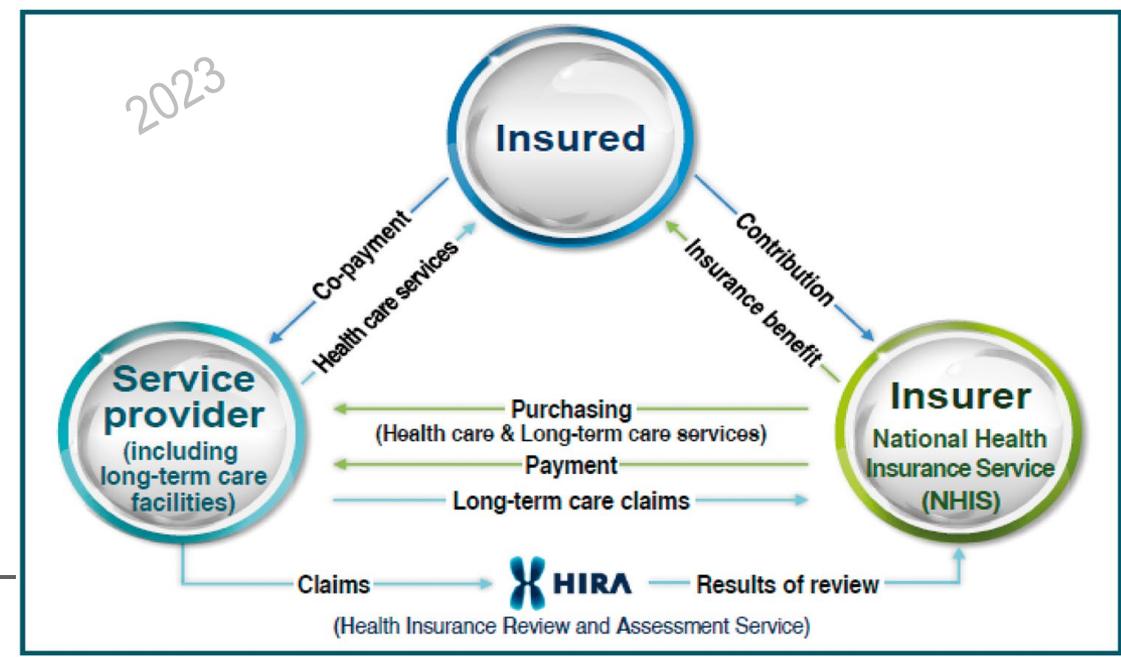
International Journal of Epidemiology, 2017, 799–800
 doi: 10.1093/ije/dyw253
 Advance Access Publication Date: 27 October 2016
 Pocket Profile



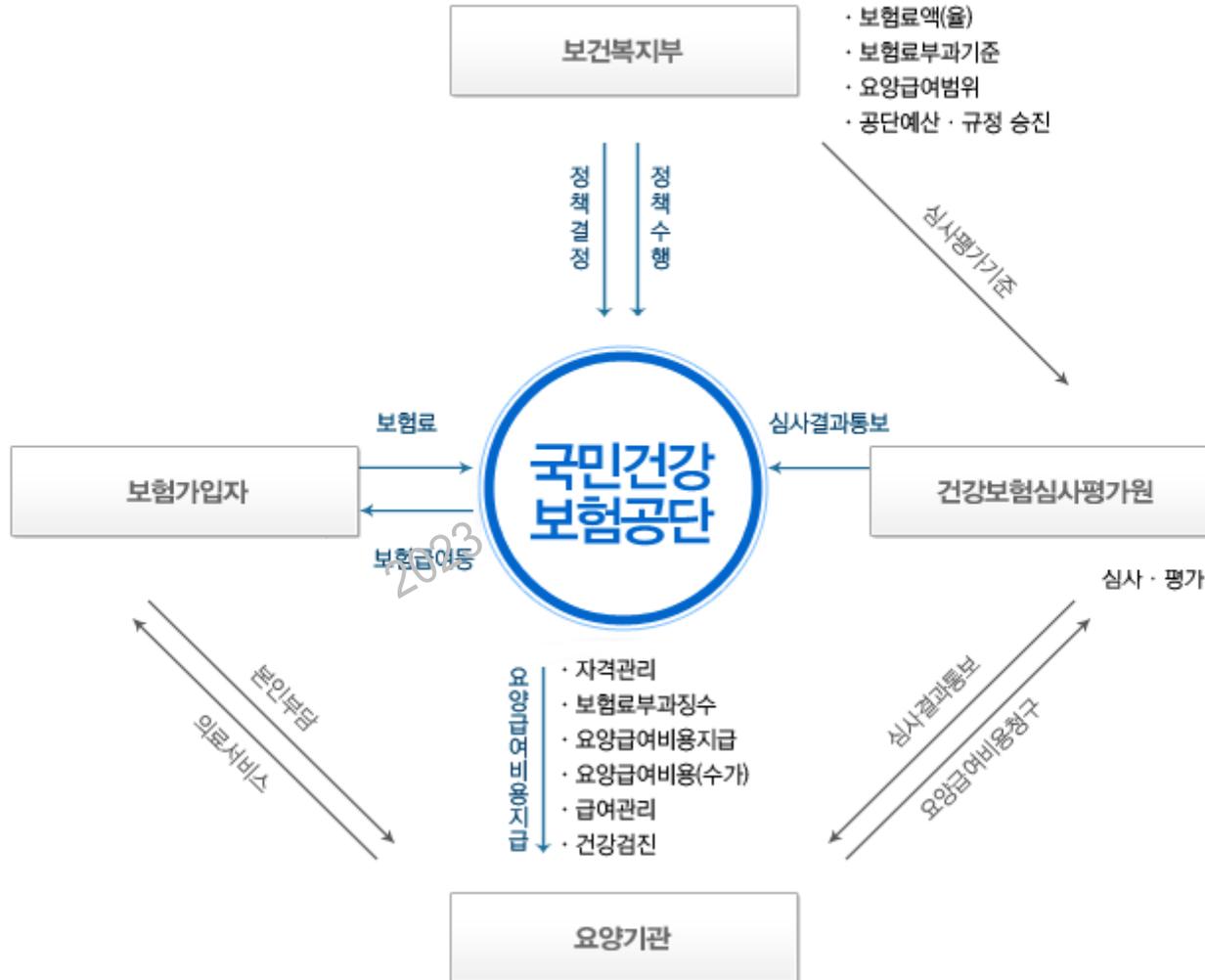
Sang Cheol Seong,¹ Yeon-Yong Kim,² Young-Ho Khang,³ Jong Heon Park,² Hee-Jin Kang,² Heeyoung Lee,⁴ Cheol-Ho Do,² Jong-Sun Song,² Ji Hyon Bang,² Seongjun Ha,² Eun-Joo Lee² and Soon Ae Shin²

Corresponding author. Soon Ae Shin, E-mail: sashin513@gmail.com

Cite this as: The full version of this profile is available at *IJE* online and should be used when citing this profile



국민건강보험 관리운영 체계





International Journal of Epidemiology, 2015, 827–836

doi: 10.1093/ije/dyv098

Advance Access Publication Date: 6 June 2015

Data Resource Profile



Data Resource Profile

Data Resource Profile: Clinical Practice Research Datalink (CPRD)

Emily Herrett,^{1*} Arlene M Gallagher,^{2,3} Krishnan Bhaskaran,¹
Harriet Forbes,¹ Rohini Mathur,¹ Tjeerd van Staa^{1,3,4} and Liam Smeeth¹

¹London School of Hygiene & Tropical Medicine, London, UK, ²Clinical Practice Research Datalink, Medicines and Healthcare Products Regulatory Agency, London, UK, ³Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands and ⁴Health eResearch Centre, University of Manchester, Manchester, UK

*Corresponding author. Department of Non Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT. E-mail: Emily.Herrett@lshtm.ac.uk

Accepted 26 March 2015

Data resource profile: the National Health Insurance Research Database (NHIRD)

Liang-Yu Lin^{1,2}, Charlotte Warren-Gash¹, Liam Smeeth¹, Pau-Chung Chen^{2,3,4,5,6}

¹Department of Non-communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK; ²Institute of Occupational Medicine and Industrial Hygiene, National Taiwan University College of Public Health, Taipei, Taiwan; ³Department of Public Health, National Taiwan University College of Public Health, Taipei, Taiwan; ⁴Innovation and Policy Centre for Population Health and Sustainable Environment, National Taiwan University College of Public Health, Taipei, Taiwan; ⁵Department of Environmental and Occupational Medicine, National Taiwan University College of Medicine and Hospital, Taipei, Taiwan; ⁶Office of Occupational Safety and Health, National Taiwan University Hospital, Taipei, Taiwan

2023

Electronic health records (EHRs) can provide researchers with extraordinary opportunities for population-based research. The National Health Insurance system of Taiwan was established in 1995 and covers more than 99.6% of the Taiwanese population; this system's claims data are released as the National Health Insurance Research Database (NHIRD). All data from primary outpatient departments and inpatient hospital care settings after 2000 are included in this database. After a change and update in 2016, the NHIRD is maintained and regulated by the Data Science Centre of the Ministry of Health and Welfare of Taiwan. Datasets for approved research are released in three forms: sampling datasets comprising 2 million subjects, disease-specific databases, and full population datasets. These datasets are de-identified and contain basic demographic information, disease diagnoses, prescriptions, operations, and investigations. Data can be linked to government surveys or other research datasets. While only a small number of validation studies with small sample sizes have been undertaken, they have generally reported positive predictive values of over 70% for various diagnoses. Currently, patients cannot opt out of inclusion in the database, although this requirement is under review. In conclusion, the NHIRD is a large, powerful data source for biomedical research.

KEY WORDS: Database, Electronic health records, Information storage and retrieval, National Health Insurance Research Database, Taiwan

국민건강보험공단/심사평가원 자료 활용 시 고려사항

- 장점
 - 전국민 기반 의료이용 정보
 - 행위별수가제로 개인 단위 세부 의료이용 내역 보유
 - 다양한 외부 데이터 연계 가능성
 - 보건의료빅데이터 개방시스템을 통한 분석환경 구축
- 이용 시 고려 사항
 - 운영시스템을 통한 데이터 수집으로 연구를 위한 별도 데이터셋 구축 필요
 - 건강보험제도에 대한 이해를 통한 데이터 구조 및 속성 이해 필요(청구방식, 가산제도 등)
 - 민감정보의 비식별화 이슈 및 철저한 보안대책 필요

국민건강보험공단/심사평가원 자료 활용 시 고려사항

- 2001년 이전 자료 부재
- 급여가 인정된 의료이용만 포함
 - 비급여, 처방 없는 약 구입 정보
 - 의료급여환자 진료내역 없음
 - DRG 대상환자의 경우 구체적 진료 내역이 없음
 - 건강검진(국가암검진) 비용 내역 없음
- 진단명의 정확도
 - 청구자료 주상병과 의무기록의 일치도: 약 70%(2003년 연구)
 - 연구 목적에 맞는 질병의 조작적 정의 수립 필수
- 임상 정보 부재
- 연령, 성별 이외에 수진자에 대한 정보 부족
 - 거주지, 소득수준(보험료), 신장, 체중, 흡연, 음주력 → 공단자료 활용
- 데이터 연계에 대한 법적, 제도적 한계
 - 질병등록자료, 의료기관자료, PHR 등

Original Article

J Prev Med Public Health 2023;56:312-318 • <https://doi.org/10.3961/jpmph.23.033>

pISSN 1975-8375 eISSN 2233-4521



Journal of
Preventive Medicine
& Public Health

Operational Definitions of Colorectal Cancer in the Korean National Health Insurance Database

Hyeree Park^{1,2,3}, Yu Rim Kim⁴, Yerin Pyun^{5*}, Hyundeok Joo^{1**}, Aesun Shin^{1,2,3}

¹Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea; ²Cancer Research Institute, Seoul National University, Seoul, Korea; ³Interdisciplinary Program in Cancer Biology Major, Seoul National University College of Medicine, Seoul, Korea; ⁴College of Medicine, Ewha Womans University, Seoul, Korea; ⁵College of Nursing, Seoul National University, Seoul, Korea

Original Article

2023

<http://www.jcpjournal.org>
pISSN 2288-3649 · eISSN 2288-3657
<https://doi.org/10.15430/JCP.2023.28.2.47>



Operational Definition of Liver Cancer in Studies Using Data from the National Health Insurance Service: A Systematic Review

Yu Rim Kim¹, Ji Yoon Baek^{2,3,4}, Seung Hee Seo^{2,3,4}, Hyeree Park^{2,3,5}, Sooyoung Cho^{5,6}, Aesun Shin^{2,3,4,5,6}

¹College of Medicine, Ewha Womans University, ²Cancer Research Institute, Seoul National University, ³Interdisciplinary Program in Cancer Biology Major, Seoul National University College of Medicine, ⁴Integrated Major in Innovative Medical Science, Seoul National University Graduate School, ⁵Department of Preventive Medicine, Seoul National University College of Medicine, ⁶Medical Research Center, Genomic Medicine Institute, Seoul National University College of Medicine, Seoul, Korea

암통계자료

2023

국제암통계

- International Agency for Research on Cancer
- <https://gco.iarc.fr/>
- Incidence database
 - *GLOBOCAN* : the most recent *estimates* (for 2020) of the cancer **incidence, mortality** and **prevalence** for **36 cancers in 185 countries**
 - *CI5* (Cancer Incidence in Five Continents) : detailed information on the **incidence** of cancer *recorded* by **cancer registries** (regional or national) worldwide
 - *WHO* cancer mortality database: long time series of selected cancer **mortality** *recorded* in selected **countries** of the world, together with advanced prediction and trends analysis facilities
- Collaborative projects
 - *ACCIS* (Automated Childhood Cancer Information System)
 - *IICC* (The International Incidence of Childhood Cancer)
 - *ECO* (European Cancer Observatory)
 - *NORDCAN*
 - *SurvCan*



HOME

ABOUT

DATABASES

CANCER REGISTRY RESOURCES

HELP

The Global Cancer Observatory (GCO) is an interactive web-based platform presenting global cancer statistics to inform cancer control and research.



Provides data visualization tools that present current national estimates of the incidence, mortality, ...



A tool that predicts the future cancer incidence and mortality burden worldwide ...



Undergoing further development; will provide data visualization tools that document the changing ...



An expanding set of visualization tools that links the cancer burden to underlying causes, ...



Provides comprehensive survival estimates for cancers in combination with incidence and mortality ...



Cancer stories are interactive stories telling cancer facts through visualizations.

Tweets by @GLOBOCAN_GCO

Global Cancer Observatory (GCO) Retweeted

Global Cancer Observatory (GCO) @GLOBOCAN_GCO

We have a new quiz question for you!

Which cancer type is the most common?

Embed

View on Twitter

Latest publications

4 Mar 2021

The role and utility of population-based cancer registries in cervical cancer surveillance and control.

1 Mar 2021

Thyroid cancer incidence trends by histology in

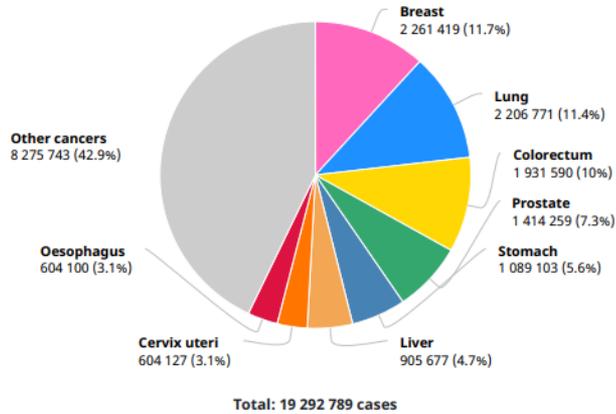
Show all

All cancers

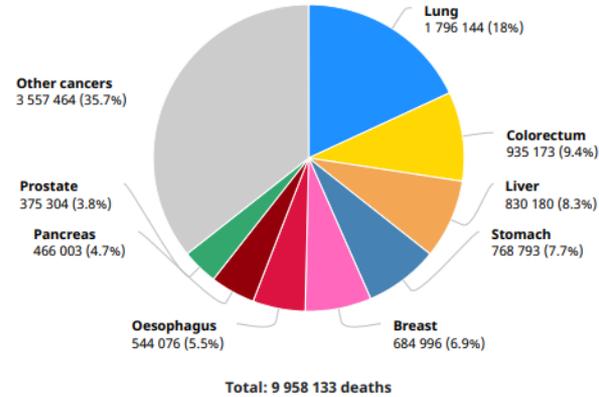
Source: Globocan 2020



Number of new cases in 2020, both sexes, all ages



Number of deaths in 2020, both sexes, all ages



2023

<https://gco.iarc.fr/>

CA CANCER J CLIN 2021;71:209-249

Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries

Hyuna Sung, PhD ¹; Jacques Ferlay, MSc, ME²; Rebecca L. Siegel, MPH ¹; Mathieu Laversanne, MSc²; Isabelle Soerjomataram, MD, MSc, PhD²; Ahmedin Jemal, DMV, PhD¹; Freddie Bray, BSc, MSc, PhD²

DATA & METHODS

Incidence

The methods used to estimate the sex- and age-specific incidence rates of cancer in a specific country fall into the following broad categories, in order of priority:

- 1 Observed national incidence rates were projected to 2020 (45 countries).
- 2 The most recently observed incidence rates (national (2a) or regional (2b)) were applied to the 2020 population (54 countries).
- 3a Rates were estimated from national mortality data by modelling, using mortality-to-incidence ratios derived from cancer registries in that country (14 countries).
- 3b Rates were estimated from national mortality estimates by modelling, using mortality-to-incidence ratios derived from cancer registries in neighbouring countries (37 countries).
- 4 Age- and sex-specific national incidence rates for all cancers combined were obtained by averaging overall rates from neighbouring countries. These rates were then partitioned to obtain the national incidence for specific sites using available cancer-specific relative frequency data (5 countries).
- 9 Rates were estimated as an average of those from selected neighbouring countries (30 countries).

Mortality

The methods used to estimate the sex- and age-specific mortality rates of cancer in a specific country fall into the following broad categories, in order of priority:

- 1 Observed national mortality rates were projected to 2020 (80 countries).
- 2 The most recently observed mortality rates (national (2a) or regional (2b)) were applied to the 2020 population (21 countries).
- 3 Rates were estimated from the corresponding national incidence estimates by modelling, using incidence-to-mortality ratios derived from cancer registries in neighbouring countries (81 countries).
- 9 Rates were estimated as an average of those from selected neighbouring countries (3 countries).

The methods used to estimate the global incidence and mortality in 2020 together with their uncertainty intervals (for all ages) can be found in [Ferlay et al. \(2018\)](#).

 [Supplemented material including the list of cancer registries that contributed to the GLOBOCAN project](#)

 [Cancer incidence and mortality data: sources and methods by country](#)

based on the GLOBOCAN cancer types by sex and

average of cancer data evaluate, compile, and I staff to improve local sed cancer registration y Development (GICR), n the coverage, quality, il efforts [here](#).

e estimates is provided

the coverage, accuracy,

y fall into the following

population (54

derived from

ratios derived from

overall rates from c sites using

into the following

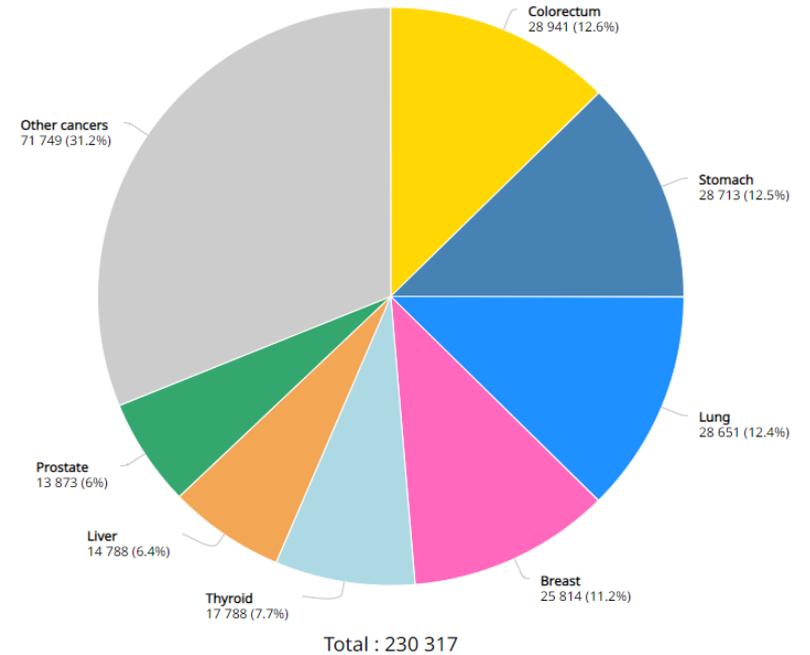
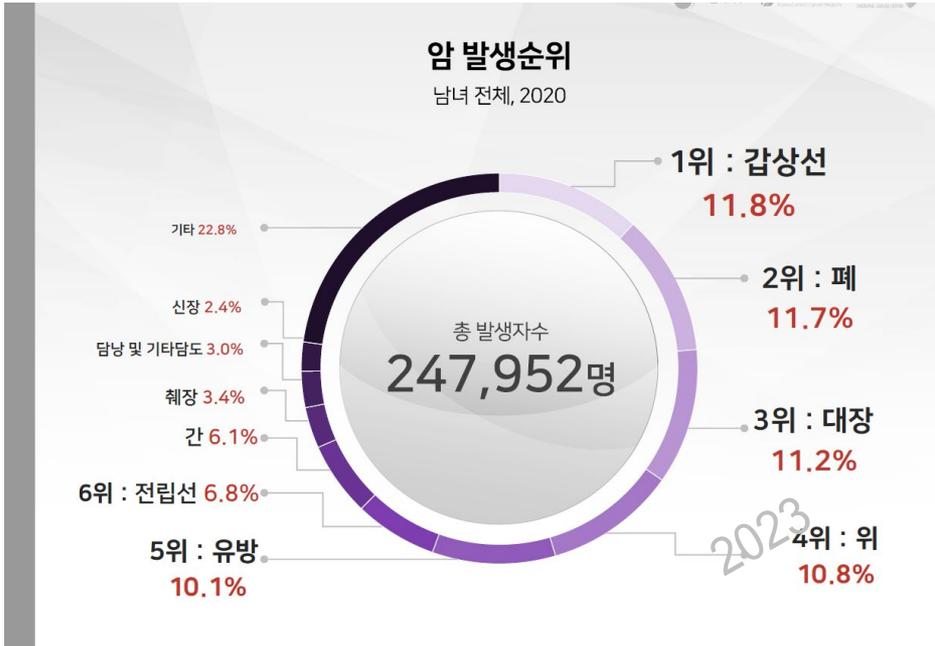
population (21

incidence-to-mortality

중앙암등록, 2020 (2022년 12월 발표)

GLOBOCAN 2020

Estimated number of new cases in 2020, Korea, Republic of, both sexes, all ages



Letter to the Editor

Open Access

Colorectal Cancer Incidence in Korea Is Not the Highest in the World

Aesun Shin, MD, PhD^{1,2}, Kyu-Won Jung, MS³, Hyeongtaek Woo, MD¹, Seung-Yong Jeong, MD, PhD^{2,4}

¹Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, ²Cancer Research Institute, Seoul National University, Seoul, ³The Korea Central Cancer Registry, National Cancer Center, Goyang,

⁴Department of Surgery, Seoul National University College of Medicine, Seoul, Korea



pISSN 1598-2998, eISSN 2005-9256

<https://doi.org/10.4143/crt.2022.1612>

Cancer Res Treat. 2023;55(3):1058-1060

Correspondence

Right Then, Wrong Now: Early-Onset Colorectal Cancer in Korea

Aesun Shin^{1,2,3,4}, Kyu-Won Jung^{5,6}, Seung-Yong Jeong^{2,7}

¹Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, ²Cancer Research Institute, Seoul National University, Seoul, ³Interdisciplinary Program in Cancer Biology Major, Seoul National University College of Medicine, Seoul, ⁴Integrated Major in Innovative Medical Science, Seoul National University Graduate School, Seoul, ⁵Korea Central Cancer Registry, National Cancer Center, Goyang, ⁶Division of Cancer Registration and Surveillance, National Cancer Control Institute, National Cancer Center, Goyang, ⁷Department of Surgery, Seoul National University College of Medicine, Seoul, Korea

미국 SEER <https://seer.cancer.gov/>

seer.cancer.gov

NIH NATIONAL CANCER INSTITUTE
Surveillance, Epidemiology, and End Results Program

Search SEER

Home Cancer Statistics SEER Data & Software Registry Operations News About

SEER IS AN AUTHORITATIVE SOURCE FOR CANCER STATISTICS IN THE UNITED STATES.

The Surveillance, Epidemiology, and End Results (SEER) Program provides information on cancer statistics in an effort to reduce the cancer burden among the U.S. population. SEER is supported by the Surveillance Research Program (SRP) in NCI's Division of Cancer Control and Population Sciences (DCCPS).



Cancer Stat Facts

Statistical summaries for a number of common cancer types:

All Cancers

[View Stat Facts](#)

SEER*Explorer

A new way of exploring cancer statistics in the U.S. [Read More...](#)

[Start Exploring](#)

Did You Know? Video Series

Did You Know? videos give a snapshot of cancer impact in the U.S.

[Watch Latest Video](#)

Latest Releases & Highlights

SEER News



암통계자료 활용 시 주의점

- 자료의 특성 정확히 파악
 - 실제 등록 자료인지(WHO mortality database, CI5, SEER) 추정한 자료인지 (GLOBOCAN) 이해하고 활용
- Knowledge gap
 - 기존 동일자료원을 사용한 연구와 비교하여 이 연구의 독창성은 무엇인가?
 - 선행연구의 충실한 리뷰가 필수적

서베이 자료

2023



주요 게시물 ▶ || 1 2 3 4 5 핵심 서비스

2021 국민건강통계 발간 및 국민건강영양조사 제8기 3차년도

건강행태, 영양, 만성질환 250여개 지표로 정리된 「2021 국민건강통계」 발간

건강정책개발, 역학연구 등을 위한 원시자료 공개

질병관리청  지역사회건강조사
Community Health Survey

🏠 ≡

사업소개 조사내용 지표 원시자료요청 수요조사 자료실 게시판

건강행태	 19.3%	 35.6%	
영양섭취	 1,859Kcal	 158.6%	
만성질환	 37.1%	 21.4%	
	원세염연율	일간폭음율	
	에너지 섭취	나트륨 목표섭취대비	
	비만	고혈압	

지역사회건강조사

함께하는 건강조사 함께하는 건강내일

"2008년부터의
우리 지역사회 건강기록"

지역건강통계 2023 한눈에 보기

건강통계집 >

지역사회 건강프로파일

건강프로파일인포그래픽 >

우리지역 건강통계

찾으시는 지역을 선택해주세요.

선택 검색

질병관리청 <https://www.cdc.go.kr/index.es?sid=a2>



8 선별진료소



- 감염병
- 건강정보
- 정책정보
- 연구개발
- 알림·자료
- 간행물·통계
- 민원·정보공개
- 기관소개
- 사업별홈페이지

코로나바이러스감염증-19

코로나19 백신 및 예방접종



2 예방접종



- 감염병
- 건강정보
- 정책정보
- 연구개발
- 알림·자료
- 간행물·통계
- 민원·정보공개
- 기관소개
- 사업별홈페이지

감염병포털	예방접종도우미	해외감염병NOW	국립검역소	KISED
결핵제로	국가건강정보포털	희귀질환헬프라인	국민건강영양조사	지역사회건강조사
청소년건강행태조사	프리온질환저해물질	국립의과학지식센터	오송PHRP	임상연구정보서비스
임상연구관리시스템	임상유전체생명정보	줄기세포주등록정보	One Health AMR 포털시스템	국립중앙인체자원은행
국가병원체자원은행	첨단재생의료포털			

인플루엔자

항생제 내성 예방 수칙

마스크 착용

미세먼지 건강수칙

한눈에 보기

전국

조회



건강행태

- 금연
- 흡연
- 간접흡연
- 전자담배
- 음주
- 고위험음주

건강결과

- 손상경험
- 협심증 또는 심근경색증
- 비만
- 고혈압
- 뇌졸중
- 당뇨병

건강상태

- 삶의 질

의료이용

- 의료이용
- 미충족의료
- 건강검진

심정지·손상

- 발생
- 목적 및 응급처치
- 생존 결과
- 전체 퇴원환자
- 특정손상 퇴원환자
- 생애주기 손상 퇴원환자

검색어를 입력해주세요

국내통계

"관심주제설정" 에서 특정 주제를 선택한 경우 해당 주제의 목록만 보여집니다. 전체 목록은 "주제 전체보기"를 클릭하세요.

통계목록

버튼형태보기

- 추계인구('21)
- 출생아수('19)
- 합계출산율('20)
- 사망자수('19)
- 기대수명('19)

주제별 통계

기관별 통계

e-지방지표(통계표)

e-지방지표(시각화)

과거·중지통계

나의통계

통계표, 게시글 등 나만의 스크랩 목록을 빠르게 찾아갈 수 있습니다.

복합통계표조회

여러개의 통계표를 하나의 통계표로 합쳐

주제별 통계

✓ 주제전체보기 관심주제설정 통계목록검색 🔍 검색 전체목록받기 부분목록받기

- 인구
- 사회일반
- 범죄·안전
- 노동
- 소득·소비·자산
- 보건

- ▶ 사망원인통계
- ▶ 건강보험통계
- ▶ 지역사회건강조사
- ▶ 국민건강영양조사
- ▶ 건강검진통계
- ▶ 청소년건강행태조사

미국 National Health and Nutrition Examination Survey(NHANES)

cdc.gov/nchs/nhanes/index.htm



[Español](#) | [Other Languages](#)



Search

Search NCHS

National Center for Health Statistics

CDC > NCHS

National Health and Nutrition Examination Survey

About NHANES +

What's New +

Webinar

Questionnaires, Datasets, and Related Documentation +

Survey Participants +

Biospecimen Program +

New Content and Proposal Guidelines

Publications and Products +

Tutorials +

Listserv

Contact Us



[Print](#)

National Health and Nutrition Examination Survey

Survey Participants



If you were selected, learn more about participating

Survey Data and Documentation



Access data, documentation & response rates

Publications and Products



View health and nutrition reports & CDC Growth Charts

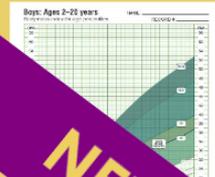
Data Analysis Tutorials



Review step-by-step guidance on using NHANES data



CDC extended BMI-for-age growth charts



Charts extend to a BMI of 60 with percentile curves above the 95th

[Click to learn more.](#)

What's New

Data Release
Laboratory Data

- [Flame Retardants – Urine](#) (P_FR 2017-March 2020)
- [Pyrethroids, Herbicides, & Organophosphorus](#)

Global Burden of Disease (GBD) project

healthdata.org/research-analysis/gbd#:~:text=The%20Global%20Burden%20of%20Disease,be%20improved%20and%20disparities%20eliminated.



Contact us

Research library

Data catalog

Give now

THE INSTITUTE FOR HEALTH METRICS AND EVALUATION



Research and analysis

Data tools and practices

News and events

About us



Home > Research and analysis

Global Burden of Disease (GBD)

The Global Burden of Disease (GBD) study provides a comprehensive picture of mortality and disability across countries, time, age, and sex. It quantifies health loss from hundreds of diseases, injuries, and risk factors, so that health systems can be improved and disparities eliminated.

281,586

Data sources were synthesized to estimate mortality, health outcomes, and risks from the GBD 2019 study.

3.5 billion+

Highly standardized and comprehensive estimates of health outcome and health system measures.

350+

Health outcomes and risk factors, providing a powerful basis for insights on global health trends and challenges.

10,000+

Individuals from over 160 countries and territories collaborate in vetting GBD data sources and estimates.

On this page:

[Overview](#)

[Research library](#)

[About GBD](#)

Key findings from GBD 2019

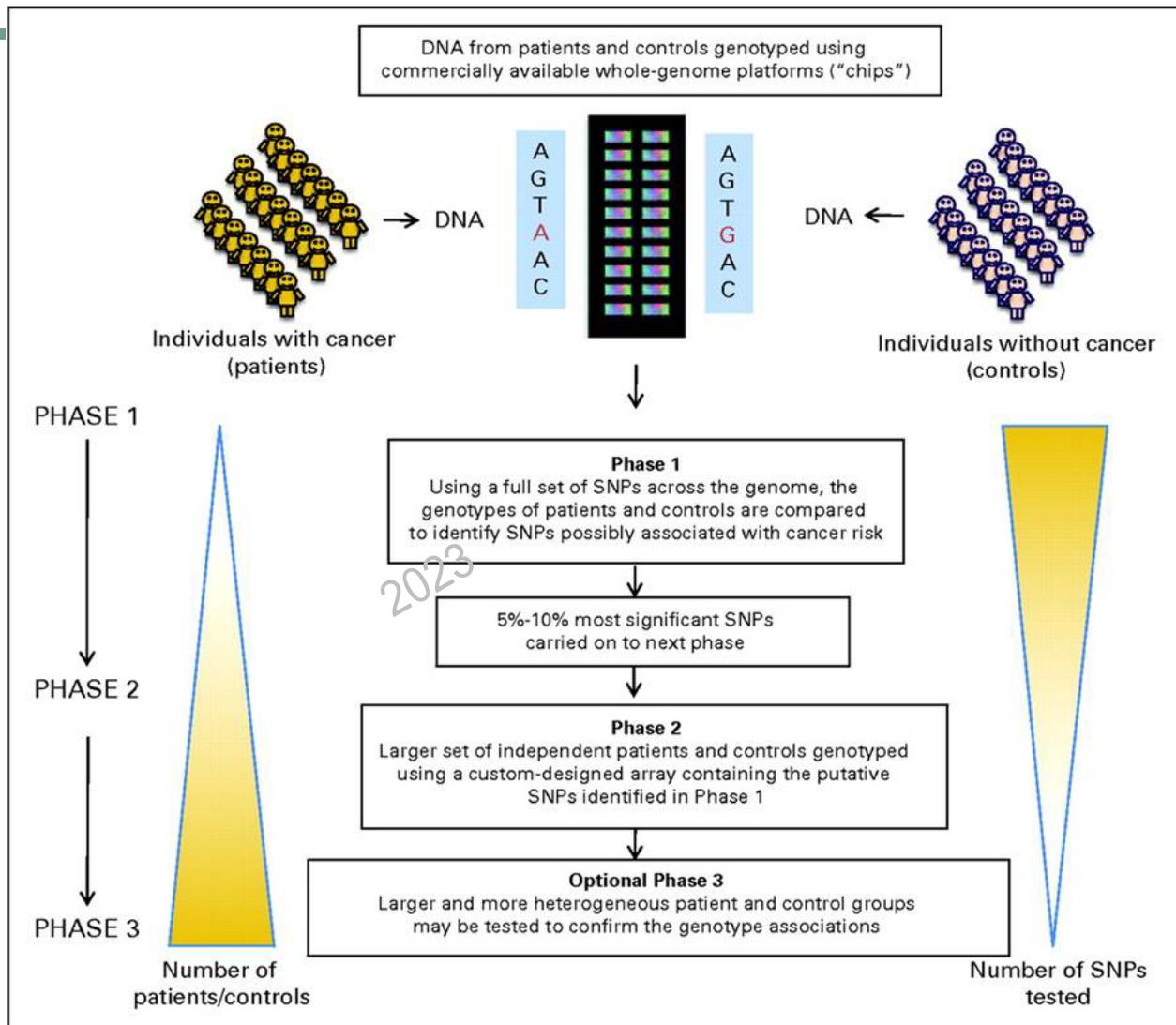
Our most recent GBD estimates were released in 2020, incorporating data through 2019. These findings provide new insights on underlying health before the COVID-19 pandemic, and quantify the challenges of protecting against further pandemic threats.

[READ KEY FINDINGS](#)

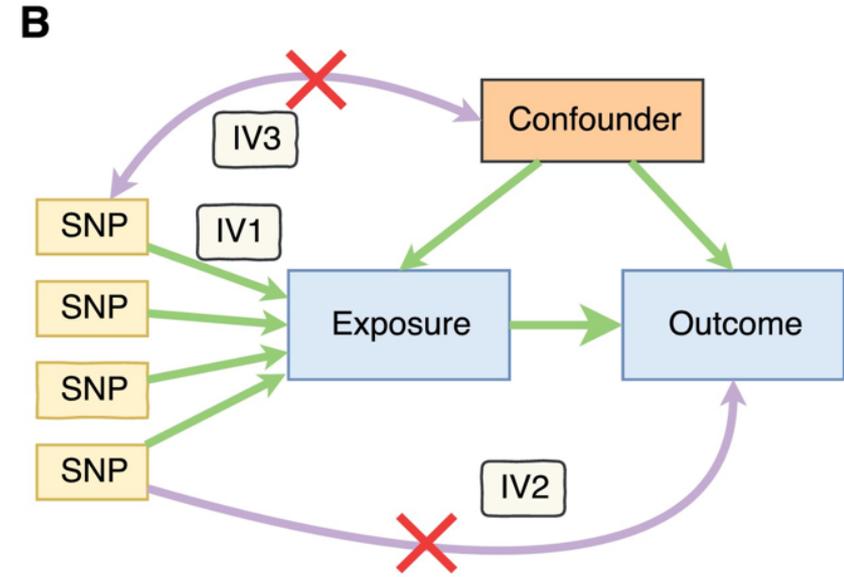
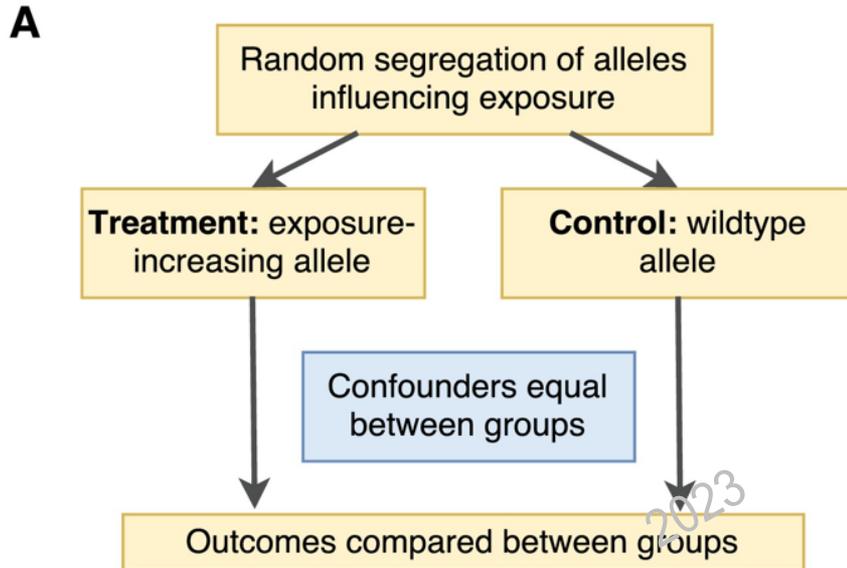
유전체 자료

2023

Genome-Wide Association Study (GWAS)



Mendelian randomization

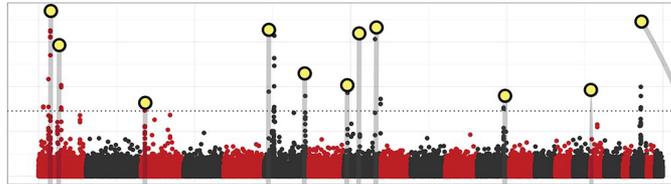


Instrumental variable (IV)

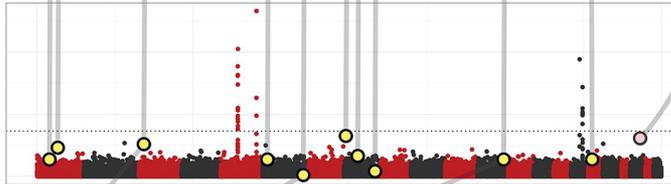
1. Be causally associated with the exposure
2. Affect the outcome *only* through the exposure
3. Not be associated with any confounders (known or unknown) of the association between the exposure and the outcome

2-sample Mendelian randomization (2SMR)

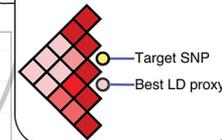
Obtain instruments from exposure GWAS



Extract SNP effects from outcome GWAS



LD Proxies
If an exposure instrument is not available in the outcome GWAS then look for LD proxies in 1000 genomes

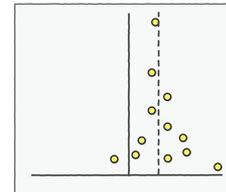
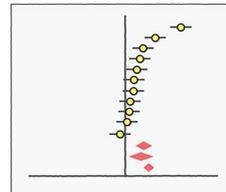
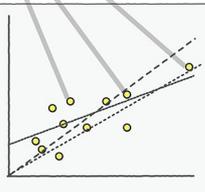


Harmonise exposure and outcome effects

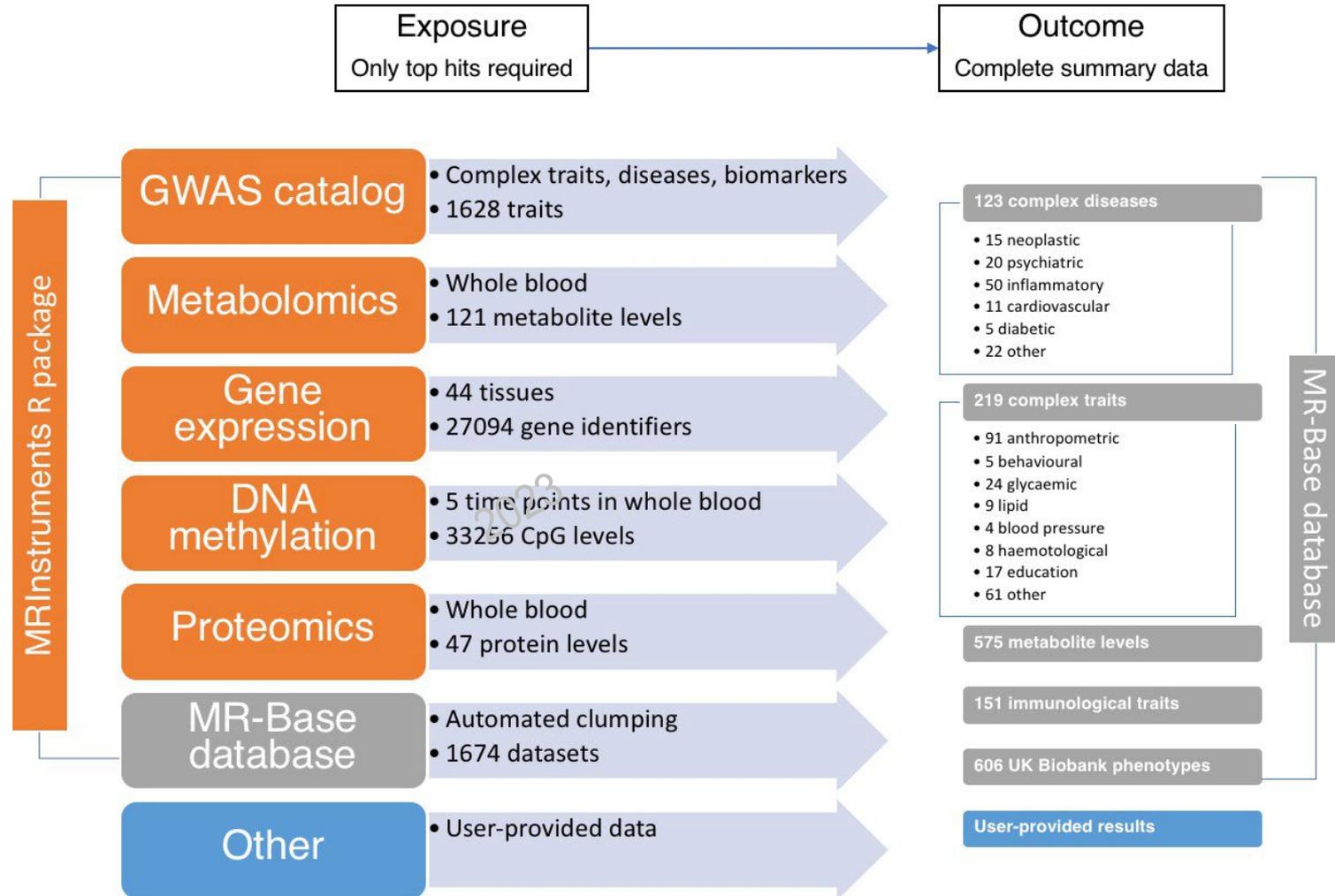
SNP	Exposure GWAS				Outcome GWAS			
	Effect	Effect allele	Other allele	Effect allele frequency	Effect	Effect allele	Other allele	Effect allele frequency
rs12345	0.132	A	G	0.28	0.022	A	G	0.26
rs23456	-0.485	G	T	0.41	0.056	T	G	0.61
rs34567	0.203	G	C	0.11	-0.046	G	C	0.88

SNP	Exposure GWAS				Outcome GWAS			
	Effect	Effect allele	Other allele	Effect allele frequency	Effect	Effect allele	Other allele	Effect allele frequency
rs12345	0.132	A	G	0.28	0.022	A	G	0.26
rs23456	-0.485	G	T	0.41	-0.056	G	T	0.39
rs34567	0.203	G	C	0.11	0.046	G	C	0.12

MR estimates and sensitivity analyses



The data available through MR-Base



Sources of GWAS summary statistics

Consortium	Full consortium name	Summary statistics link
ALSKP	ALS Knowledge portal	http://alskp.org/informational/data
CARDIoGRAMplusC4D	Coronary ARtery Disease Genome wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) Genetics	http://www.cardiogramplusc4d.org/data-downloads/
CDKP/ISGC	Cerebrovascular Disease Knowledge portal/International Stroke Genetics Consortium	https://cd.hugeamp.org/downloads.html
CHARGE	Cohorts for Heart and Aging Research in Genetic Epidemiology	http://www.chargeconsortium.com/main/results
CKDGen	Chronic Kidney Disease Genetics Consortium	http://ckdgen.imbi.uni-freiburg.de
CMDKP	Common Metabolic Diseases Knowledge portal	https://hugeamp.org/downloads.html
CVDKP	Cardiovascular Disease Knowledge portal	https://cvd.hugeamp.org/downloads.html
deCODE	deCODE genetics	https://www.decode.com/summarydata/
Diagram	DIAbetes Genetics Replication And Meta-analysis	http://diagram-consortium.org/downloads.html
EAGLE	EAGLE eczema consortium	http://data.bris.ac.uk/datasets/tar/28uchsdpmub118uex26ylacqm.zip
EGG	Early Growth Genetics Consortium	http://egg-consortium.org/
GEFOS	GEnetic Factors for OSteoporosis Consortium	http://www.gefos.org
GIANT	Genetic Investigation of ANthropometric Traits	http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files
GLGC	Global Lipids Genetics Consortium	http://csg.sph.umich.edu//abecasis/public/lipids2013/
GRASP	Genome-Wide Repository of Associations Between SNPs and	https://grasp.nhlbi.nih.gov/FullResults.aspx

Just because it is possible
in no way means
that it should be done.

2023

Medical big data analysis vs. classical statistical analysis

	Medical big data analysis	Classical statistical analysis
Application	Hypothesis-generating	Hypothesis-testing
Questions of interest	Overcoming the limitation of locally or temporally stable association with continually updating the data and algorithm	Trying to prove causal relationships
Domain knowledge	More important in interpretation of the results	Important both in collection of data and interpretation of the results
Coverage of data to be analyzed	Substantial fraction of entire population	Small data samples from a specific population with some assumptions of their distribution
Nature of data	Unstructured and structured	Mainly structured
Data quality	Rarely clean	Quality controlled
Research questions of data analysis	May be different from those of data collection	Same as those of data collection
Underlying assumption of the model	Frequently absent	Based on various underlying probability distribution function
Analytic tools	Frequently automated with data mining algorithm	Manually by expert with classical statistics
Main outputs of analysis	Prediction, models, patterns identified	Statistical score contrasted against random chance
Privacy & ethics	Concerns about privacy and ethical issues	Data collection according to the pre-approved protocol; informed consent from the participants

공개자료원을 이용한 데이터사이언스 논문

- 장점
 - Hypothesis-generating
 - Analytic tools -Frequently automated with data mining algorithm
- 주의사항
 - **Domain knowledge: More important in interpretation of the results**
 - 연구설계와 자료원에 대한 깊은 이해 필요
 - 생물학적 개연성에 기반하지 않은 기계적 분석 지양
 - 투망식 연구 지양 :탐색적 연구라면 그에 맞는 분석방법 선택
 - 나만 쓸 수 있는 자료가 아니다: 기존 문헌 검색과 리뷰

Domains of health data science

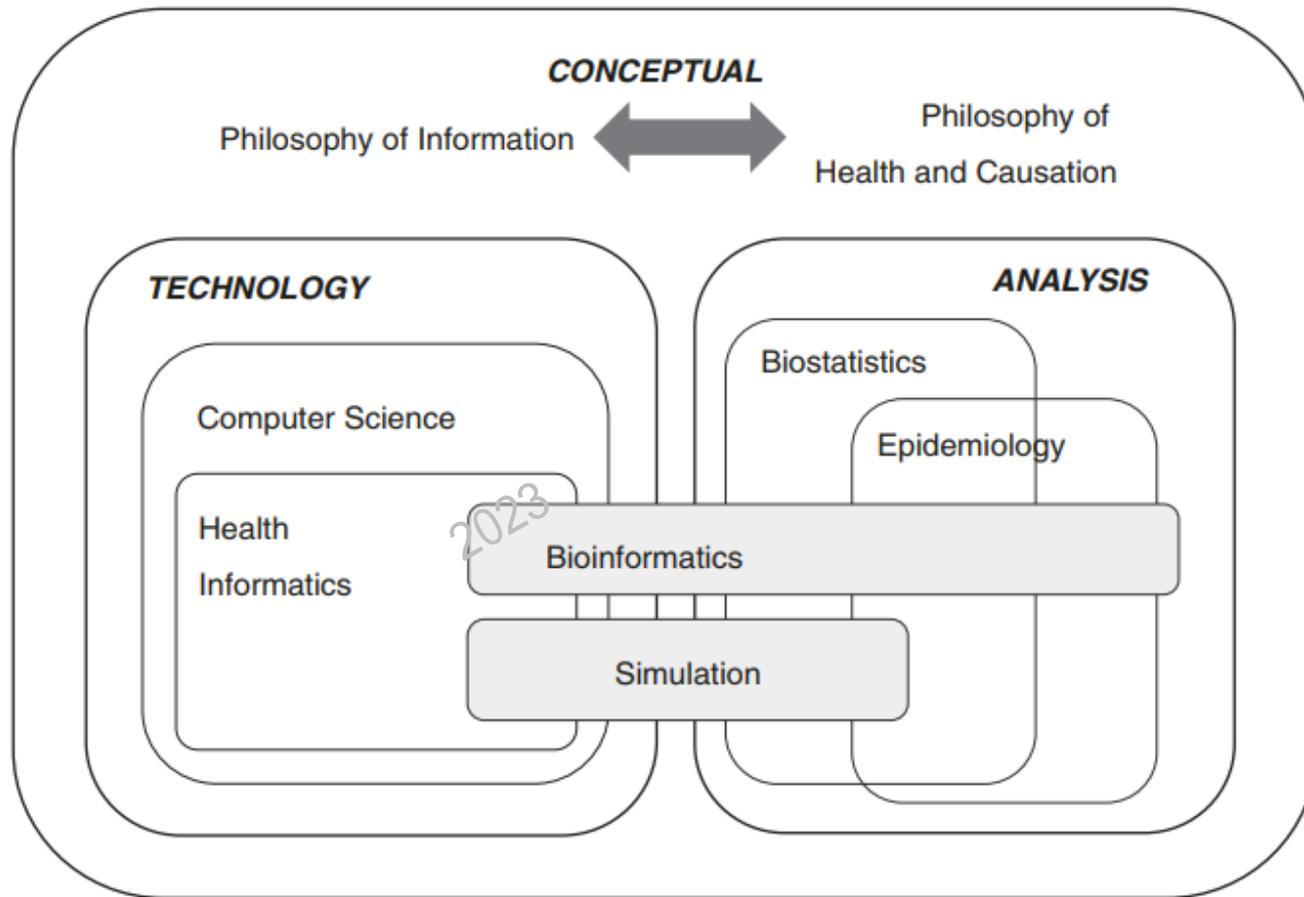


Fig. 2.1 Domains of health data science

Summary

- 공개자료원의 장점
 - 낮은 연구 수행 비용
 - 다양한 대규모 인구집단 자료로 일반화 가능성 높음
- 공개자료원 사용 연구 시 주의점
 - 투망적 연구 지양: 생물학적 개연성과 설명가능 여부에 기반한 주제
 - 기존 연구에 비하여 독창적인 요소 필요: Knowledge gap
 - 자료의 생성과정에 대한 이해 필요: 실측 자료인가 연구진의 추정치인가?
 - 자료의 제한점 이해: 건강보험 청구자료