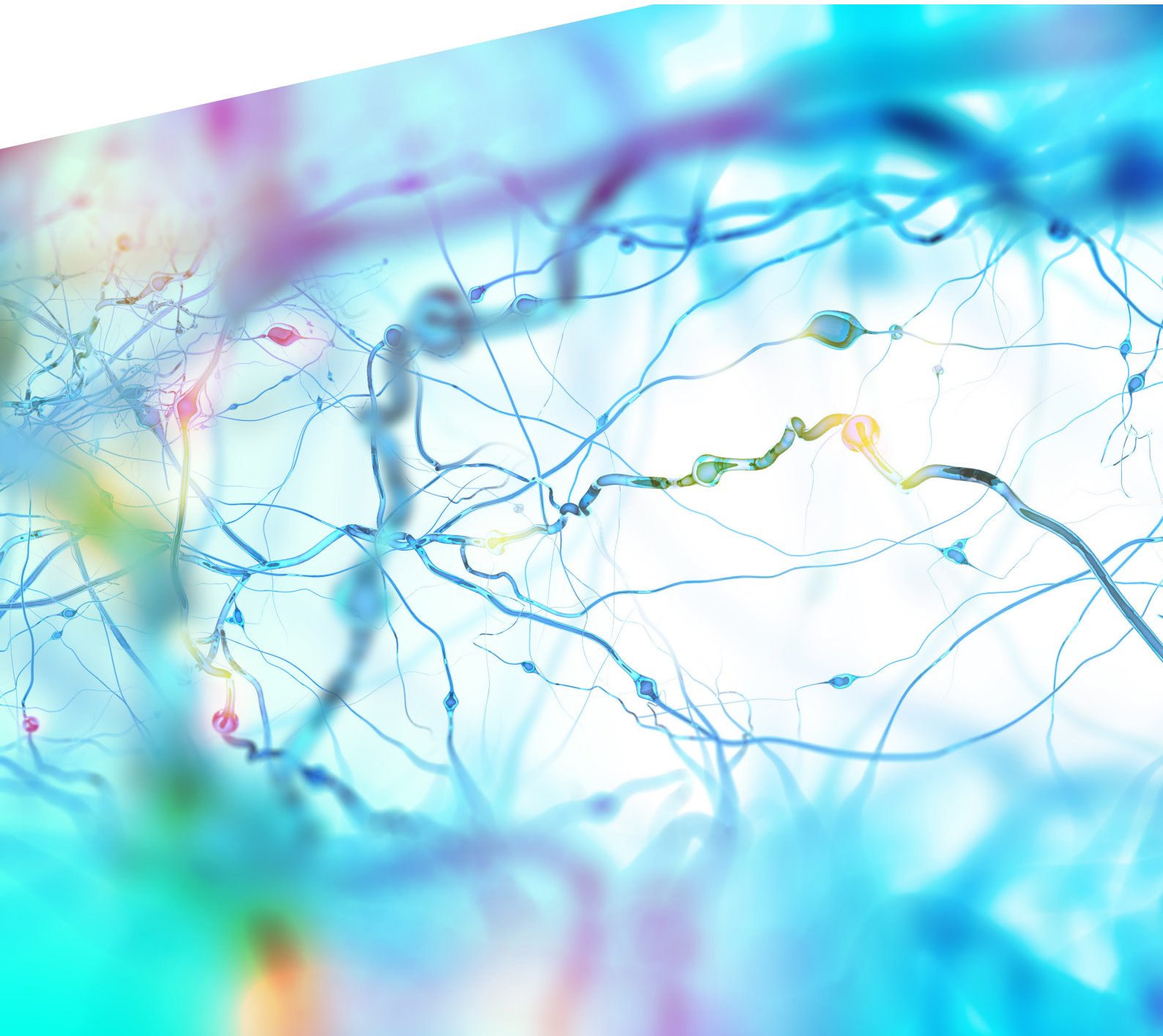
eISSN 2234-2591

The Ewha Medical Journal

Vol. 47, No. 2, 2024



eISSN 2234-2591

Aims & Scope

The Ewha Medical Journal (Ewha Med J, http://www.e-emj.org), the official publication of Ewha Womans University College of Medicine and Ewha Medical Research Institute, is published quarterly a year, last day of January, April, July, and October. The first volume was published in March, 1978. It covers all fields of medical science including clinical research and basic medical science. The Journal aims to communicate new medical information between medical personnel and to help development of medicine and propagation of medical knowledges. All manuscripts should be creative, informative and helpful for diagnosis and treatment of the medical diseases and for communication of valuable information about all fields of medicine. Subscripted manuscripts should be written out according to the instructions for the Journal. Topics include original article, case report, images and solution, letter to the editor, invited review article and special issue in the respective field of medicine. The Ewha Medical Journal is indexed/tracked/covered by KoreaMed, KoMCI, KoreaMed Synapse, WPRIM, DOI/CrossRef, Emerging Source Citation Index/Web of Science Core Collection, EMBASE, and Google Scholar.

Copyright & Permissions

The Ewha Medical Journal
Vol. 47 No. 2, April 2024

Publisher Eunhee Ha
Editor-in-Chief Sun Huh

Published by Ewha Womans University College of Medicine and Ewha Medical Research Institute
25, Magokdong-ro 2-gil, Gangseo-gu, Seoul 07804, Korea
Tel: 82-2-6986-6013, E-mail: mediewha@ewha.ac.kr, Homepage: http://www.ewhamed.ac.kr

Editorial office Ewha Medical Research Institute
25, Magokdong-ro 2-gil, Gangseo-gu, Seoul 07804, Korea
Tel: 82-2-6986-6092, E-mail: E600091@ewha.ac.kr, Homepage: http://www.e-emj.org

Printing correspondence by Guhmok Publishing Co./Guhmok Info Inc.
#609, Jungang-Deco Plaza, 148, Eulgiro, Jung-gu, Seoul 04549, Korea
Tel: 82-2-2277-3324, Fax: 82-2-2277-3390, E-mail: guhmok@guhmok.com, Homepage: http://www.guhmok.com

· The subscription is free. For inquiry, lease contact the editorial office (Tel. 82-2-6986-6092, e-mail E600091@ewha.ac.kr). All the contents are also available at the EMJ website (http://www.e-emj.org).

Editorial Board

eISSN 2234-2591
Vol. 47 No. 2, April 2024

eISSN 2234-2591
Vol. 47 No. 2, April 2024

Review

Original Articles

Health Statistics

eISSN 2234-2591
Vol. 47 No. 2, April 2024

Case Reports

Guidelines

Editorial

Ewha Med J 2024;47(2):e12
https://doi.org/10.12771/emj.2024.e12
eISSN 2234-2591

Gender equity in medicine, artificial intelligence, and other articles in this issue

Sun Huh

Institute of Medical Education, Hallym University College of Medicine, Chuncheon, Korea

Received Apr 29, 2024
Accepted Apr 29, 2024

Corresponding author
Sun Huh
Institute of Medical Education, Hallym University College of Medicine, 1 Hallymdaehak-gil, Chuncheon 24252, Korea
E-mail: shuh@hallym.ac.kr

Sex differences in medicine

In a previous editorial [1], I emphasized the *Ewha Medical Journal*'s gender equity policy. In this issue, Dr. Na-Young Kim, a gastroenterologist at Seoul National University, curated a special section on the topic of sex differences in medicine. She assembled one original and five review articles. Among these, "Health of Korean sexual and gender minorities: a narrative review of quantitative studies" by So et al. [2] is an outstanding article that provides a comprehensive overview of the health of LGBTQIA+ individuals, covering a range of identities, including lesbian, gay, bisexual, transgender, questioning, intersex, asexual, and other diverse groups. This article is unprecedented in the Korean medical literature. Individuals who belong to sexual and gender minorities in Korea face significant mental health challenges. The review highlights their high rates of depression and anxiety, as well as elevated prevalence rates of suicidal thoughts, planning, and attempts. Furthermore, these individuals report a lower perceived health-related quality of life than the general population. Sexual minorities who experience discrimination or are pressured to change their sexual orientation or gender identity are at an even higher risk of mental health issues.

The study by Choi et al. suggests that sex differences exist in the impact of obesity on the development of gastric cancer, with a positive association between excess body weight and an increased risk of gastric cancer in Koreans, particularly in highly obese men [3].

This issue of the journal contains reviews on various topics, including sex differences in metabolic dysfunction-associated fatty liver disease, sex differences in coronary atherogenesis, research on sex differences in neuroscience, and sex bias in autism spectrum disorder using preclinical rodent models. All these topics concerning sex differences in medicine are expected to engage both Korean and international readers.

Regarding diversity, equity, and inclusiveness in the target population of medical research, sexual and gender minorities often receive primary attention. Other groups in Korea, such as immigrants, disabled persons, and prisoners, also merit consideration. I plan to continue addressing issues concerning these groups in *Ewha Medical Journal*.

Deep learning and generative artificial intelligence platform

The article by Choi et al. [4] on the accurate prediction of pediatric bone age using deep learning demonstrated that "the deep learning-based Korean model exhibited higher bone age

prediction accuracy than conventional methods, a crucial advancement for accurate growth assessment and clinical decision-making." These results can be applied to deep learning algorithms for estimating bone age.

Baik and Lee discussed the role of artificial intelligence (AI) in general surgery [5]. They analyzed published research to clarify the potential applications of AI in this field. Their findings indicate that the implementation of AI in the preoperative stage is becoming feasible, although its use in the operating room requires further investigation. They recommend developing AI tools specifically for general surgery, which can be achieved by promoting interdisciplinary collaboration and leveraging insights from successful AI applications in other fields.

Since AI has become an indispensable tool in medical care and education, submissions focusing on the use or development of AI are welcomed.

Public health

Lee et al. [6] published a high-quality systematic review on the relationship between exposure to air pollution and precocious puberty. They noted that "most studies suggest that exposure to air pollutants accelerates pubertal development; however, the results from the available studies are inconsistent." Dr. Eunhee Ha, the corresponding author of this systematic review, currently serves as the dean of Ewha Womans University College of Medicine. She is recognized globally as an outstanding researcher in environmental medicine. I am pleased to publish her excellent research findings.

Mr. Seokmin Lee, an officer at the Statistics Research Institute, Statistics Korea, published an article with up-to-date comprehensive data on drug-induced deaths in Korea from 2011 to 2021 [7]. In 2021, Korea exhibited a staggering 172.7% increase in drug-induced deaths compared to 2011, with the number rising from 205 to 559 cases. The rate of drug-induced deaths per 100,000 population also increased dramatically by 153.6%, from 0.4 in 2011 to 1.1 in 2021. Although the rate of drug-induced deaths in Korea (1.1 per 100,000) remains relatively low compared to that of the United States (29.2 per 100,000), this cause of mortality has been on an alarming upward trend in recent years. Notably, drug-induced deaths disproportionately impact younger demographics, and a significant proportion involves intentional self-harm. This data will be an essential source for research in this field. I anticipate further submissions from the officers of the Statistics Research Institute in the future.

One of the most striking articles in the issue may be the study by Lee et al. [8] at the Infectious Disease Research Center, Seoul Metropolitan Government, entitled "Using an influenza epidemic threshold different from those in the United States and Europe caused longer epidemic periods in Korea during the 2018–2019, 2019–2020, and 2022–2023 seasons: a comparative study." The authors pointed out that "a low influenza epidemic threshold may have contributed to this long influenza epidemic period declared in 2022 and has continued until late 2023 in Korea" based on their comparison of the seasonal influenza epidemic thresholds in Korea, the United States, and Europe. These findings could potentially influence the Korean government's epidemic threshold, which is crucial for predicting and preventing influenza epidemics.

It has been seven months since I accepted the position of editor-in-chief in response to an offer from the Dean, Dr. Eunhee Ha, in September 2023. The primary challenge in publishing this institutional journal has been the shortage of submissions. Therefore, I have made significant efforts to attract submissions from renowned medical researchers in Korea. The April issue marks the fourth issue I have edited for this journal. The number of articles has reached a satisfactory

level and meets the eligibility criteria for evaluation by major literature databases. Regarding the quality of the articles, I have endeavored to maintain or surpass the minimum quality standards required by these databases.

I appreciate the authors who contributed to this issue for submitting high-quality articles and case reports. I believe these papers will serve as exciting resources for medical researchers, as well as for graduate and undergraduate medical students.

ORCID
Sun Huh: https://orcid.org/0000-0002-8559-8640

Authors' contributions
All work was done by Sun Huh.

Conflict of interest
Sun Huh has been the editor of the *Ewha Medical Journal* since September 2023. However, he was not involved in the peer review process or decision-making. Otherwise, no potential conflict of interest relevant to this article was reported.

Funding
Not applicable.

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Not applicable.

References

1. Huh S. Gender equity in medical journals in Korea and this issue. *Ewha Med J* 2024;47(1):e1. https://doi.org/10.12771/emj.2024.e1
2. So H, Kim S, Lee SY. Health of Korean sexual and gender minorities health: a narrative review of quantitative studies. *Ewha Med J* 2024;47(2):e14. https://doi.org/10.12771/emj.2024.e14
3. Choi Y, Jang J, Kim N. Effect of body mass index on gastric cancer risks according to sex in Korea: a nationwide cohort study and literature review. *Ewha Med J* 2024;47(2):e19. https://doi.org/10.12771/emj.2024.e19
4. Choi DH, Ahn SH, Lee R. An accurate pediatric bone age prediction model using deep learning and contrast conversion. *Ewha Med J* 2024;47(2):e23. https://doi.org/10.12771/emj.2024.e23
5. Baik SM, Lee RA. What is the role of artificial intelligence in general surgery? *Ewha Med J* 2024;47(2):e22. https://doi.org/10.12771/emj.2024.e22
6. Lee R, Oh J, Mun E, Choi JE, Kim KH, Lee JH, et al. Exposure to air pollution and precocious puberty: a systematic review. *Ewha Med J* 2024;47(2):e20. https://doi.org/10.12771/emj.2024.e20
7. Lee S. Drug-induced death statistics in Korea between 2011 and 2021. *Ewha Med J* 2024;47(2):e27. https://doi.org/10.12771/emj.2024.e27
8. Lee J, Huh S, Seo H. Using an influenza epidemic threshold different from those in the United States and Europe causes longer epidemic periods in Korea during the 2018–2019, 2019–2020, and 2022–2023 seasons: a comparative study. *Ewha Med J* 2024;47(2):e24. https://doi.org/10.12771/emj.2024.e24

Opinion

Ewha Med J 2024;47(2):e13
https://doi.org/10.12771/emj.2024.e13
eISSN 2234-2591

The new placement of 2,000 entrants at Korean medical schools in 2025: is the government's policy evidence-based?

Sun Huh

Institute of Medical Education, Hallym University College of Medicine, Chuncheon, Korea

Received Apr 29, 2024
Accepted Apr 29, 2024

Corresponding author
Sun Huh
Institute of Medical Education, Hallym University College of Medicine, 1 Hallymdaehak-gil, Chuncheon 24252, Korea
E-mail: shuh@hallym.ac.kr

Two thousand new entrants at Korean medical schools

The news section of the *BMJ* has covered the "increased number of places at Korean medical schools in 2025 by 2,000 entrants and the resignation of 10,000 residents and interns" [1]. In an opinion piece published in the *Journal of Korean Medical Sciences* [2], the author condemns the government's authoritarian attitude of treating physicians as felons, stating, "the government will annul residents' medical license, prohibit hospitals from accepting residents' resignations, or pronounce maximum criminal sentence." The author further states that the Korean Academy of Medical Sciences "looks toward the government and the medical community hoping each will take a step back and discuss these policies together to prevent public disaster." *The Lancet* has also reported on this topic in the news section:

"The Korean government aims to address a projected shortage of 15,000 doctors by 2035 through a new recruitment cap. Junior doctors, however, believe that merely increasing medical school slots won't effectively tackle physician shortages in underserved medical areas. Instead, they assert that newly qualified physicians will still gravitate toward high-paying fields like cosmetic surgery and dermatology in the greater Seoul region." [3].

It is disheartening to hear about the resignation of interns and residents, especially against the backdrop of the Korean government's threatening stance towards these young physicians during the beautiful days of spring. Each March and April, general hospitals buzz with activity as new interns and residents begin treating patients and improving their skills under the guidance of experienced supervisors. This period also fosters interaction among various health professionals, including nurses, pharmacists, dietitians, physical therapists, health information managers, and radiological technologists. These inter-professional activities mark the initial steps in alleviating the suffering of patients afflicted by illness. Furthermore, the development of rapport among health professionals, patients, and their families introduces a fresh sense of joy and fulfillment in the practice of medicine. Without these vibrant interactions, March and April become the most cruel months for young physicians.

In this opinion piece, I present the perspectives of young physicians (interns and residents) alongside the official statement from the Korean government regarding the government's threat to young physicians — namely, "*a medical license suspension at least 3 months from March 2024, and related investigations and prosecutions.*" Additionally, I will share insights from experts on critical

issues and offer suggestions from the viewpoints of educators and editors, drawing on sources such as journal articles, newspapers, government briefings, and personal communications.

Ethics statement

This is an opinion on a policy judgment; therefore, neither approval by the institutional review board nor the obtainment of informed consent was required.

Korea Interns and Residents Association Emergency Measures Committee statement 240220

The official Emergency Measures Committee statement 240220 was released by the Korea Interns and Residents Association on February 20, 2024 (Supplement 1) [4]. The Association demanded the following steps:

- ☐ Completely withdraw the essential healthcare policy package and the plan to increase the number of places at medical schools by 2,000 a year.
- ☐ Establish a body to conduct a scientific forecast of doctor supply and demand and discuss increases and decreases in the supply of physicians.
- ☐ Expand the hiring of specialists at training hospitals.
- ☐ Implement concrete measures to mitigate the legal burden on doctors due to unavoidable medical incidents.
- ☐ Improve the harsh training environment for residents, who work up to 80 hours per week.
- ☐ Withdraw all unfair orders that intimidate residents and formally apologize to them.
- ☐ Fully repeal Article 59 of the Medical Service Act, which infringes upon the fundamental rights of citizens, and comply with the Republic of Korea's Constitution and the International Labour Organization's prohibition of forced labor.

The present number of admissions to medical school is 3,058 at 40 institutions. According to a law from 2016, the maximum working hours for interns and residents is 80 hours, with an additional 8 hours allocated for educational purposes [5]. After the announcement of the statement, residents started to depart from their training hospitals individually, protesting against the government's briefing.

Why would gifted young physicians, who represent the hope of Korea, issue such a statement? This appears to be a reaction to the "Emergency Briefing on the Physician Workforce Expansion Plan" announced by the Ministry of Health and Welfare on February 6, 2024 (Supplement 2) [6]. This emergency briefing can be summarized as follows:

Four essential healthcare policy packages will be implemented to allow physicians to focus on regional and essential healthcare services: 1) expanding the healthcare workforce, 2) strengthening regional healthcare, 3) establishing a safety net for medical accidents, and 4) enhancing fairness in the compensation system. The medical school admissions quota will be increased by 2,000, from 3,058 to 5,058. Starting in 2025, an additional 2,000 students will be admitted annually.

A more specific proposal is presented in the Essential Medical Policy Package (Supplement 3).

Ongoing resignation of residents and the government's threat of judicial proceedings

The Minister of Health and Welfare stated on February 27, 2024 that "starting in March, it will be necessary to suspend the licenses of those who have not complied and to initiate related judicial proceedings" [7].

Following the statement issued by the Korea Interns and Residents Association on February 20, 2024, residents at training hospitals continued to submit their resignations. By February 28, a total of 9,997 residents, representing 80.2% of all residents, had submitted their resignations, although not all were accepted by the hospitals in compliance with a directive from the Korean government. Consequently, 9,076 residents, or 72.8% of the total, successfully left their positions [8]. The Korean government ordered 13 residents to commence work on March 1, 2024 [9] under Article 59 (2) of the Medical Service Act [10].

Article 59 (Guidance and Order) of the Medical Service Act is as follows [10]:

(2) The Minister of Health and Welfare, a relevant Mayor/Do Governor, or the head of a relevant Si/Gun/Gu may order medical personnel or founders of medical institutions to resume medical service if there is a reasonable ground to believe that suspension of medical service by the medical personnel without any justifiable ground, or temporary shutdown or closure of medical institutions by a group of the founders causes or is likely to cause significant difficulties in giving medical treatment to patients. <Amended by Act No. 8852, Feb 29, 2008; Act No. 9932, Jan 18, 2010>

Therefore, under the current circumstances, most residents who have ceased working after submitting their resignations may face criminal penalties. Moreover, doctors risk criminal punishment, which can include imprisonment or more severe consequences, potentially leading to the revocation of their medical licenses. In Korea, the professions subject to legal mandates to resume work are medical doctors and pharmacists (including medicine manufacturers and pharmacy owners) and cargo drivers (involved in freight transportation businesses and operations).

On April 1, 2024, a statement was issued by the Korean presidency addressing the nation [11]. Regarding the scale of the increase in medical school admissions, the government repeatedly stated, "the government has decided to increase medical school admissions by 2,000 students based on clear grounds and sufficient discussion," and emphasized, "if the medical community argues that the increase should be reduced from 2,000, instead of collective action, they should properly present a unified proposal with clear scientific grounds to the government." This statement also noted that "the average number of physicians of OECD countries is 3.7 per 1,000 population, but it is 2.1 in Korea."

The President continued, "if they bring a more reasonable and rational solution, we can discuss it anytime," and said, "if better opinions and rational grounds are presented, government policies can change for the better."

At the same time, he emphasized, "However, we can never accept attempts to force their will through power without proper logic and grounds. They must immediately stop illegal collective actions and bring rational proposals and grounds."

The President said, "I will create a proper medical system through medical reforms," adding, "we will make massive financial investments to make the competitiveness of our country's medical and healthcare industries the best in the world."

Mitigation of threats to young physicians by the Korean government after a meeting between the President and the leader of residents and interns

After the presidential statement, a meeting took place between the President and the President of the Korea Interns and Residents Association on April 4, 2024, in the Presidential office. It remains uncertain whether this meeting will serve as a catalyst for resolving the ongoing conflict regarding the increase in medical school admissions by 2,000 in Korea. No official announcement has been made about the content of their discussion. Following this meeting, the Korean government ceased its threats toward the residents. Nevertheless, the government continued to reject the resignations of medical residents in accordance with its directive.

After that, the Prime Minister said on April 19, 2024, "regarding the 32 universities whose medical school quota has been expanded this year, if desired, we will allow them to recruit new students autonomously only in the 2025 school year within the range of 50% to 100% of the increased number" [12].

The Minister of Health and Welfare said on April 22, 2024 that "the Special Committee on Medical Reform will be launched this week for social discussions on medical reform tasks" and "the government will do its best to present each other's opinions on major issues of medical reform, such as the direction of investment in essential medical care, through the committee, and to prepare reasonable alternatives through open discussions." He asked "the Korean Medical Association and the Korea Interns and Residents Association to participate in the special committee on medical reform so that developmental and constructive discussions can take place, not just turn a blind eye in connection with the quota of medical schools" [13].

As of late April, there were no further changes in the actions of the residents, who continued to resign from training hospitals. Despite this, the government persisted in its efforts to increase medical school quotas, although the presidents of some medical schools reduced certain quotas. The Korean Medical Association and the Korea Interns and Residents Association still refrained from participating in the government's proposed special committee.

What are the fundamental issues in the present situation in Korea: new placement of 2,000 entrants at Korean medical schools and medical residents' mass resignation

First, is the increase of medical school students by 2,000 based on scientific evidence or a policy judgment?

The Korean government has consistently stated that the increase is "based on clear grounds and sufficient discussion" and urged physicians to present a unified proposal with well-founded scientific justification [11]. Three reports are the basis of the government's argument [14–16].

However, the authors of these three reports, which the government cited as scientific evidence, have denied that they constitute a basis for the government's increase of the medical school quota by 2,000 (https://www.medicaltimes.com/Main/News/NewsView.html?ID=1157769). Professor Yun-Chul Hong at Seoul National University stated that his research [14] did not support an increase of 2,000 students. Instead, the report presented various scenarios, with the most reasonable one suggesting an increase of 500 to 1,000 students. He highlighted that although Korea will experience a physician shortage from 2045 to

2050, an oversupply is anticipated thereafter. Therefore, he recommended that medical school quotas be adjusted to reflect these projections.

Dr. Junghyun Kwon from the Korea Development Institute (KDI) has also pointed out that the government's policy inaccurately interpreted her research [15]. Her proposed scenarios include increasing admissions by 1,000 students annually starting in 2024 for a total of 4,000 additional students, maintaining a 5% annual increase until 2030 to reach 4,500 students, and 7% and 10% annual increases. Notably, her scenarios do not include one where admissions increase by 2,000 students each year for five years to add 10,000 students.

Dr. Youngseok Shin, an Honorary Fellow at the Korea Institute for Health and Social Affairs, expressed his disagreement with the government's plan to increase medical school admissions by 2,000 students, arguing that the proposed pace is too rapid [16]. He suggested that even if a total increase of 10,000 students is deemed necessary by the government, it would be more prudent to distribute this increment over 10 years rather than 5, taking into account the medical market conditions at the time the new doctors graduate.

However, the Second Vice Minister of the Ministry of Health and Welfare said, "Those three reports are policy suggestions, and when the administration makes policy decisions, of course, those suggestions are considered and referenced. We make policy decisions by taking into account all the other surrounding conditions and factors, as well as the demands of other organizations. Therefore, it is up to the government to make policy decisions" [17].

Considerations for future estimates of the number of physicians

The 18th president of the Korean Society of Epidemiology highlighted three key factors to consider when estimating future physician demand: healthy aging, the integration of artificial intelligence (AI) into medicine, and the regulation of outpatient visits. In an aging population, the prevalence of diseases may not necessarily rise; rather, the number of individuals enjoying a healthy old age is increasing. With advancements in AI that are currently difficult to foresee, we are nearing an era where AI can undertake tasks traditionally performed by medical professionals [18]. If the work performed by AI programs is recognized for medical billing, healthcare facilities might not need as many specialists. At present, AI interpretations are not accepted for billing purposes. However, AI's diagnostic capabilities are anticipated to exceed those of human specialists in various fields. When this occurs, AI-generated diagnostic results could be eligible for billing, and only the more complex cases might require the attention of a specialist.

Korea's physician-to-population ratio, at 2.6 per 1,000 people in 2021 (2.1 when excluding Oriental medicine doctors), is lower than the OECD average of 3.7 per 1,000 people. However, the disparity in physician numbers per 1,000 population between Korea and countries such as Japan (2.6), the United States (2.7), and Canada (2.8) in 2021 is minimal. Despite this, Korea records the highest number of outpatient visits per capita among OECD countries, with an annual average of 15.7 visits, significantly higher than the OECD average of 5.9 visits in 2021. The avoidable mortality rate, which quantifies the number of patients who died due to not receiving timely medical treatment, stands at 142.0 per 100,000 in 2020. This figure is less than half of the OECD average of 293.1 in 2020 [19].

The current supply of healthcare services in Korea ranks among the highest in OECD countries, and the Korean people benefit from top-level services while incurring low routine medical costs. However, it is essential to evaluate whether the high frequency of medical visits indicates an excessive demand for healthcare. It is crucial to assess whether to maintain the low

contribution rate to health insurance and unrestricted access to medical facilities, or if controls are necessary.

Dr. Jung said the following about the politics of healthcare reform in Korea [20]: "First and foremost, policymakers need to escape the recurring trap of scapegoating and blame avoidance. The government should establish a new governance framework to foster a sustainable national-physician relationship. This framework should provide a space to consider a new healthcare system that ensures health equity and accommodates demographic and technological changes. Simply increasing staff numbers will not suffice to improve the healthcare system. While it is essential to increase the number of physicians, this must be part of a broader set of policy measures. However, if the essential trust-building between these two parties continues to deteriorate in the quest to normalize interest group politics, achieving effective governance will become increasingly difficult."

Second, is threatening residents with an order of forced labor reasonable in Korea, a liberal democratic society?

The government's threat was noted above [9]. Additionally, the Second Vice-Minister of Health and Welfare provided a more detailed explanation of this matter [21]:

"Submitting the resignation letters of medical residents collectively does not genuinely reflect their intent, allowing for the possibility of legislative invalidation. I mentioned this earlier, didn't I? Consequently, this action will likely lead to a legal dispute under public law. Under the Medical Service Act, which is a part of public law, the collective resignation request was not accepted. However, if the residents fail to report to the hospital and withhold medical treatment, they violate the mandatory work commencement order. Such violations are punishable under the Medical Service Act, with penalties including up to three years of imprisonment. Non-compliance with the work commencement order will prompt actions from the Minister of Health and Welfare. Concurrently, the Korean government will initiate legal proceedings by filing a complaint and accusation. This will trigger the judicial process, starting with an investigation. The findings of the investigation will lead to an indictment, followed by a trial. If the trial results in a prison sentence, or even after the initial verdict, the government will implement administrative measures. These measures could include the revocation of the residents' medical licenses once the verdict is announced."

Although the judicial processes were halted following a meeting between the Korean President and a medical resident leader, this could still be considered a demand for forced labor. The Korean government retains the authority to resume these judicial actions at any time. Meanwhile, hospitals are unable to accept the residents' resignations due to a government directive, preventing the residents from seeking employment elsewhere. As a result, they have been without income for two months.

An opinion presented by a lawyer suggests that the order to commence work, which the Korean government is threatening to impose on resident doctors, may potentially be unconstitutional [22]. This opinion is as follows:

"The concerns raised about the directive for medical residents to commence work under Article 59 of the Medical Service Act stem from the perception that it has been crafted as an overly forceful approach, exceeding what is necessary to fulfill a specific administrative goal. In practice, as evidenced by actions taken against residents, this mandate could potentially serve as a convenient legal instrument for the government to manipulate medical professionals and

the broader medical community at its discretion. The extent to which the government might attempt to exert control over the medical sector is profoundly troubling. After all, doctors are not military physicians who are subject to punishment for insubordination or for failing to comply with legitimate orders from their superiors, are they?"

Young physicians must be liberated from despair, fear, and depression

The crux of this issue regarding the government's threats toward resident physicians revolves around the question: who is responsible for nurturing our doctors? The idea that physicians trained entirely in the private sector should be treated similarly to military doctors under Article 59 of the Medical Service Act is baffling to young doctors. Those of us born in the post-generation (1955-1963), including myself, served selflessly and without complaint, adhering to the demands of our senior physicians in the pursuit of national revival, regardless of the hardships faced. During my internship at Seoul National Hospital from May 1985 to February 1986, I recall working over 140 hours a week. When the Resident Law, which limits on-duty hours to 88 per week, including 8 hours for educational purposes, was enacted in 2015, I was thrilled. I saw this 80-hour limitation as a crucial first step toward improving both patient safety and the well-being of residents [23].

The post-war generation has successfully achieved the national revival of Korea. Today's young physicians represent a new generation with values distinct from those of the post-war era; they have a strong sense of self and are not inclined to engage in forced labor simply because they are instructed to do so. Failing to adapt to this shift will hinder our ability to understand, communicate with, and solve problems alongside the younger generation. It should be clear that coercive measures such as threatening these young physicians—who will be responsible for our future—with license suspensions and denial of resignations will not address the underlying issues. As of April 2024, the government's practice of vilifying resident physicians in public advertisements can also be seen as a form of this pressure tactic (Supplement 4). I remember the French artist, Bernard Buffet (1928–1999)'s message when he was criticized by the public: "*La haine dont je suis entouré est, pour moi, le plus merveilleux cadeau que l'on m'ait fait*" ("The hate that surrounds me is, for me, the most marvelous gift that anyone has given me").

These various measures drive our future healthcare system toward self-destruction; therefore, attempts to resolve the issue through further threats must cease. When an individual undertakes work, there must be a purpose, the ability to ascribe value to that purpose, and commensurate compensation. Moreover, tasks not chosen by oneself no longer hold meaning for this generation. We must remember that these young physicians, nurtured by our nation's people and society, are the rising stars responsible for safeguarding our health in the future.

Clinical faculty members are also under burnout due to long working hours and frequent duty at night

My junior doctors at university hospitals express feelings of helplessness, anger, and depression due to the current wave of residents' resignations. They are among the finest physicians and surgeons globally, yet they are already experiencing burnout. The primary source of this burnout has been identified as "excessive regulation by the government or university" [24]. This situation could lead them to leave university hospitals, driven by a fear of death from overwork. They require psychological support in addition to the provision of assistant personnel.

Role of the Korea Institute of Medical Education and Evaluation

In March 2024, the Korea Institute of Medical Education and Evaluation (KIMEE) released the newly revised accreditation criteria, "ASK 2026 (Accreditation Standards of KIMEE 2026, https://kimee.or.kr/board/data/)." This update replaces the previous ASK2019 standards [25]. If there are substantive changes, including an increase in the admission quota, a mandatory evaluation of the substantive change plan must be undertaken according to the KIMEE's accreditation process [26]. Failure to secure accreditation means that graduates may lose their eligibility to sit for the Korean medical licensing examination. As stipulated by the Medical and Higher Education Act [27], only graduates from KIMEE-accredited institutions are eligible to take the medical licensing exam.

The Director of KIMEE stated the following (https://www.docdocdoc.co.kr/news/articleView.html?idxno=3016783): "As the head of the accreditation agency, I am not at liberty to discuss the outcomes of forthcoming evaluations in detail. It is essential to understand that developing a single expert takes more time and effort than one might anticipate. Additionally, it is important for society to recognize that this investment is crucial for professionals to fulfill their roles effectively. Medical education should extend beyond merely passing exams; it must focus on quality and foster an appropriate educational environment. KIMEE is dedicated to improving the quality of medical education and making a significant contribution to public health, which is aligned with these objectives."

We must consider whether the accreditation system established by KIMEE for quality management in medical education is capable of effectively responding to the sudden increase in medical school enrollments by 2,000 students annually.

Editor's perspective on this policy judgement

From an editorial perspective, it is unfortunate that the three reports cited by the government as evidence have not yet been published in scholarly journals. Scientists typically submit their research findings in the form of reports, which are then published in peer-reviewed scholarly journals. Although the three studies in question are noteworthy, their scientific credibility would have been increased if they had been published in academic journals, as is customary for research reports. Looking ahead, submitting and publishing such policy-related research in international peer-reviewed journals would promote a wider understanding among researchers and readers globally, thereby enriching future policy discussions.

Since I am not a specialist in health policy but a retired basic scientist and teacher, I am unable to intervene in the current turmoil. It appears that both the Korean government and the medical residents are hesitant to abandon their positions quickly, likely due to a lack of mutual trust. Consequently, the resignations among the current residents are expected to persist for an extended period. In this situation, the two groups most at risk are the patients requiring care in university hospitals and the clinical faculty members working there. Patients concerned about delays in their medical care should consider transferring to another general hospital where residents have not resigned, especially if their wait for surgery or therapy becomes excessively long.

My greatest concern is the burnout and exhaustion of clinical faculty members who are required to care for patients during night shifts, a role typically assigned to residents. I am acutely aware of their dedication to patient care, often at the expense of their own physical

well-being, as I have personally experienced this during my four years as a clinician, both as a public health physician and an intern. Therefore, it is crucial for hospital managers to prioritize the reduction of their workload. One potential strategy could be the introduction of physician assistants, although this is not a comprehensive solution. Continued experiences of depression and hopelessness among these faculty members could lead to a breakdown in the delivery of high-quality healthcare. The primary reason I value living in Korea is the exceptional level of medical care provided by Korean physicians, healthcare personnel, and medical institutes. The thought of such a collapse is deeply troubling.

At present, the situation is dire and demands immediate action. I conclude this opinion by citing a recommendation from my senior doctor, a specialist in health policy [28]: "A country's healthcare policy must first establish a comprehensive dialogue on healthcare, and then assess the supply and demand for healthcare human resources based on clearly defined objectives and specific detailed plans. This process also necessitates the involvement and consensus of the professionals and hierarchical organizations tasked with healthcare provision. Healthcare policy is a complex and challenging issue that demands extensive collaboration. The time is now for the Korean government to formulate a cohesive short- and long-term plan, incorporating expert input, to guarantee the delivery of sustainable, top-tier healthcare services."

ORCID
Sun Huh: https://orcid.org/0000-0002-8559-8640

Authors' contributions
All work was done by Sun Huh.

Conflict of interest
Sun Huh has been the editor of the *Ewha Medical Journal* since September 2023. However, he was not involved in the peer review process or decision-making. Sun Huh is also a member of the Korean Association of Medicine (KMA); therefore, his ideas and opinions may be biased toward those of KMA and clinical faculty members, although he was a basic scientist and has already retired from his university. He may be unable to present a neutral position on the health policy issue. Otherwise, no potential conflict of interest relevant to this article was reported.

Funding
Not applicable.

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Supplementary materials are available from: https://doi.org/10.12771/emj.2024.e13.

Supplement 1. The official Emergency Management Committee statement 240220 of the Korea Interns and Residents Association, released on February 20, 2024
Supplement 2. "Emergency Briefing on the Physician Workforce Expansion Plan" announced by the Ministry of Health and Welfare on February 6, 2024
Supplement 3. Essential Medical Policy Package announced by the Ministry of Health and Welfare on February 6, 2024
Supplement 4. Video file publicizing the Korean government's medical policy in public places in Korea

References

1. Dyer O. South Korea: striking junior doctors are threatened with arrest and suspension. *BMJ* 2024;384:q495. https://doi.org/10.1136/bmj.q495
2. Park HW. Encouraging message from the Korean academy of medical sciences to junior doctors in struggle. *J Korean Med Sci* 2024;39(9):e108.

https://doi.org/10.3346/jkms.2024.39.e108
3. McCurry J. South Korean doctors threaten mass resignation. *Lancet* 2024;403(10432):1124. https://doi.org/10.1016/S0140-6736(24)00578-6
4. Korea Intern Resident Association. Emergency Measures Committee statement 240220 [Internet]. Seoul (KR): Korea Intern Resident Association; c2024 [cited 2024 Apr 4]. Available from: https://youngmd.org/154
5. Huh S. Will 2016 augur well for better patient safety and health of residents in Korea according to the enactment of the Act for improving the resident training environment and enhancing resident's status? *J Educ Eval Health Prof* 2016;13.2. https://doi.org/10.3352/jeehp.2016.13.2
6. Ministry of Health and Welfare, Government of Korea. Emergency Briefing on the Physician Workforce Expansion Plan [Internet]. Sejong (KR): Ministry of Health and Welfare; c2024 [cited 2024 Apr 4]. Available from: https://www.mohw.go.kr/board.es?mid=a10503000000&bid=0027&list_no=1480186&act=view
7. Ministry of Culture, Sports, and Tourism. Korea's Policy News: the government said, "It is inevitable to proceed with judicial proceedings for medical personnel who have not returned from March" [Internet]. Sejong (KR): Ministry of Culture, Sports, and Tourism; c2024 [cited 2024 Apr 4]. Available from: https://www.korea.kr/news/policyNewsView.do?newsId=148926374
8. Ministry of Culture, Sports, and Tourism. Korea's Policy Briefing: Briefing of the Central Disaster and Safety Countermeasures Headquarters for the collective action of doctors on 2024 Feb 29 [Internet]. Sejong (KR): Ministry of Culture, Sports, and Tourism; c2024 [cited 2024 Apr 4]. Available from: https://www.korea.kr/briefing/policyBriefingView.do?newsId=156617954
9. Ministry of Health and Welfare, Government of Korea. Disclosure service of business commencement order under Article 59 (2) of the Medical Law [Internet]. Sejong (KR): Ministry of Health and Welfare; c2024 [cited 2024 Apr 4]. Available from: https://www.mohw.go.kr/board.es?mid=a10501010100&bid=0003
10. Korea Legislation Research Institute. Medical Service Act [Internet]. Sejong (KR): Korea Legislation Research Institute; c2024 [cited 2024 Apr 4]. Available from: https://elaw.klri.re.kr/kor_service/lawView.do?hseq=53532&lang=ENG
11. Ministry of Culture, Sports, and Tourism. Korea's Policy News: President Yoon said, "Medical reform must be completed... If you come up with a reasonable plan, we can discuss it as much as we want." [Internet]. Sejong (KR): Ministry of Culture, Sports, and Tourism; c2024 [cited 2024 Apr 4]. Available from: https://www.korea.kr/news/policyNewsView.do?newsId=148927692
12. Ministry of Culture, Sports, and Tourism. Korea's Policy News: The government will allow voluntary recruitment within 50 to 100% of the increase in medical schools next year [Internet]. Sejong (KR): Ministry of Culture, Sports, and Tourism; c2024 [cited 2024 Apr 26]. Available from: https://www.korea.kr/news/policyNewsView.do?newsId=148928353&pWise=mostViewNewsSub&pWiseSub=B2
13. Ministry of Culture, Sports, and Tourism. Korea's Policy News: This week's 'Special Committee on Medical Reform' is launched..."Request for participation, such as the Medical Association" [Internet]. Sejong (KR): Ministry of Culture, Sports, and Tourism; c2024 [cited 2024 Apr 26]. Available from: https://www.korea.kr/news/policyNewsView.do?newsId=148928382&pWise=mostViewNewsSub&pWiseSub=B8
14. Hong YC. A study on the appropriate numer of physicians for prepartaion of the future society. *Health Care Policy* 2020;18(3):19-23.
15. Lee C, Kwon J, Kim T. Prospects for the ripple effect of population change reflecting future population estimates in 2021 in the labor, education, and medical sectors (Feb. 2023) [Internet]. Seoul (KR): Presidential Committee on Aging Society and Population Policy; c2023 [cited 2024 Apr 26]. Available from: http://www.jnuinmun.org/board/B0068.cs?act=read&bwrId=11458&searchKeyword=&searchCondition=&searchEndDt=&m=134&searchStartDt=&pageIndex=1&pageUnit=15
16. Shin Y, Park E, Choi J, Yoon G, Shin J, Lee N, et al. A study on the comprehensive plan and mid- to long-term supply and demand estimation for health and medical menpower [Internet]. Sejong (KR): Korea Institute of Health and Social Affairs; c2020 [cited 2024 Apr 26]. Available from: https://www.mohw.go.kr/board.es?mid=a10411010100&bid=0019&act=view&list_no=1480369&tag=&nPage=1
17. Ministry of Culture, Sports, and Tourism. Korea's Policy Briefing: Briefing of the Central Disaster and Safety Countermeasures Headquarters for the collective action of doctors on 2024 Feb 20 [Internet]. Sejong (KR): Ministry of Culture, Sports, and Tourism; c2024 [cited 2024 Apr 26]. Available from: https://www.korea.kr/briefing/policyBriefingView.do?newsId=156616163&pageIndex=10&srchType=&startDate=2023-04-26&endDate=2024-04-26&period=&srchWord=
18. Natesan D, Eisenstein EL, Thomas SM, Eclov NCW, Dalal NH, Stephens SJ, et al. Health care cost reductions with machine learning-directed evaluations during radiation therapy: an economic analysis of a randomized controlled study. NEJM AI 2024 Mar 15 [Epub]. https://doi.org/10.1056/Aloa2300118
19. Organisation for Economic Co-operation and Development [OECD]. OECD health statistics [Internet]. Paris (FR): OECD; c2024 [cited 2024 Apr 26]. Available from: https://data-explorer.oecd.org/
20. Jung U. Where's the beef?: the politics of health care reform in South Korea. *J Korean Med Assoc* 2024;67(5):366-370. http://dx.doi.org/10.5124/jkma.2024.67.5.366
21. Ministry of Culture, Sports, and Tourism. Korea's Policy Briefing: Briefing of the Central Disaster and Safety Countermeasures Headquarters for the collective action of doctors on 2024 Feb 16 [Internet]. Sejong (KR): Ministry of Culture, Sports, and Tourism; c2024 [cited 2024 Apr 28]. Available from: https://www.korea.kr/briefing/policyBriefingView.do?newsId=156615747&pageIndex=1&srchType=&startDate=2023-03-01&endDate=2024-03-01&period=year&srchWord=%EC%9D%98%EC%82%AC%20%EC%A7%91%EB%8B%A8%ED%96%89%EB%8F%99%20%EC%A4%91%EC%95%99%EC%82%AC%EA%B3%A0%EC%88%98%EC%8A%B5%EB%B3%B8%EB%B6%80
22. Kim JH. A complaint about the possibility of unconstitutionality of a business commencement order. *Health Care Policy* 2020;18(4):72-74.
23. Huh S. Will the year 2016 augur well for better patient safety and health of residents in Korea according to the enactment of the Act for improving the resident training environment and enhancing resident's status? *J Educ Eval Health Prof* 2016;13:2.

https://doi.org/10.3352/jeehp.2016.13.2
24. Seo JH, Bae H, Kim BJ, Huh S, Ahn YJ, Jung SS, et al. Burnout of faculty members of medical schools in Korea. *J Korean Med Sci* 2022;37(9):e74.
https://doi.org/10.3346/jkms.2022.37.e74
25. Lee Y, Lee MJ, Ahn J, Ha C, Kang YJ, Jung CW, et al. Challenges and potential improvements in the Accreditation Standards of the Korean Institute of Medical Education and Evaluation 2019 (ASK2019) derived through meta-evaluation: a cross-sectional study. *J Educ Eval Health Prof* 2024;21:8.
https://doi.org/10.3352/jeehp.2024.21.8
26. Meng K. History of the medical education accreditation system in Korea: implementation and activities in the early stages. *J Educ Eval Health Prof* 2020;17:29.
https://doi.org/10.3352/jeehp.2020.17.29
27. Korean Law Information Center. Higher education act, article 11-2 (evaluation) [Internet]. Sejong (KR): Korean Law Information Center; c2023 [cited 2024 Apr 28]. Available from: https://www.law.go.kr/LSW/eng/engLsSc.do?menuId=2&section=lawNm&query=the+Higher+Education+Act&x=21&y=32#liBgcolor1
28. Ahn DS. Healthcare development plan: balancing accessibility and human resources in Korea. *Ewha Med J* 2023;46(4):e10.
https://doi.org/10.12771/emj.2023.e10

Review

Ewha Med J 2024;47(2):e14
https://doi.org/10.12771/emj.2024.e14
eISSN 2234-2591

Health of Korean sexual and gender minorities: a narrative review of quantitative studies

Heesung So[1], **Ssirai Kim**[2], **Sun Young Lee**[2,3,4]

[1]Department of Psychiatry, Seoul National University Hospital, Seoul, Korea
[2]Korean Initiative for Transgender Health, Seoul, Korea
[3]Public Healthcare Center, Seoul National University Hospital, Seoul, Korea
[4]Department of Human Systems Medicine, Seoul National University College of Medicine, Seoul, Korea

Received Feb 29, 2024
Revised Apr 22, 2024
Accepted Apr 23, 2024

Corresponding author
Sun Young Lee
Public Healthcare Center, Seoul National University Hospital, 101, Daehak-ro, Jongno-gu, Seoul 03080, Korea
E-mail: sy2376@snu.ac.kr

Keywords
Sexual and gender minorities; Homosexuality; Transgender persons; Gender-affirming care; Health inequities

This study reviewed quantitative research on the health of sexual and gender minorities (SGMs) in Korea and aimed to propose a role for healthcare professionals in improving their health and access to medical care. We searched PubMed through February 29, 2024 for articles published since 2000, using terms related to SGMs and the keyword "Korea." This process yielded 33 quantitative studies on Korean SGMs. Of these, 17 focused on sexual minorities and 16 on gender minorities. The findings indicate that Korean SGMs experience many symptoms of depression and anxiety, as well as high rates of suicidal ideation, planning, and attempts. They also report diminished health-related quality of life. SGM individuals who have faced discrimination or pressure to change their sexual or gender identity face an elevated risk of mental health issues. To improve the health of Korean SGMs and improve their access to healthcare, we recommend several approaches. First, more research on the health of Korean SGMs is necessary. Second, education and training programs for health professionals are essential to promote their understanding of SGM health issues and their advocacy for SGM health. Third, strategies are required to develop and implement program interventions that improve SGM health, such as increasing the availability of gender-affirming care, which is known to benefit the health of transgender and gender-diverse individuals. Finally, healthcare professionals should actively advocate for SGM health and call for shifts in public perception and institutional change, grounded in a broad understanding of SGMs and their health needs.

Introduction

Background

Definition and prevalence of sexual and gender minorities

Sexual and gender minorities (SGMs) are individuals or groups whose sexual orientation or gender identity diverges from that of the societal majority. Common international terms for SGMs include LGBT, LGBTQ, and LGBTQIA+, which encompass lesbian, gay, bisexual, transgender, questioning, intersex, asexual, and other diverse identities (as indicated by the +symbol) [1]. The American Psychiatric Association (APA) defines sexual orientation refers to an enduring pattern of emotional, romantic, and/or sexual attractions to men, women, or both sexes [2]. Heterosexual individuals are attracted to the opposite gender, homosexual people to the same gender, and

bisexual individuals to both genders. The term "lesbian" describes women who are romantically or sexually attracted to other women, while "gay" typically refers to men attracted to other men. Pansexual individuals perceive attraction regardless of gender, and asexual people experience little or no sexual attraction or interest in sexual behavior [2]. Additionally, in the medical context, the term "men who have sex with men" (MSM) describes men who engage in sexual activities with other men. Gender identity refers to a person's internal sense of being male, female or something else [3]. Transgender and gender diverse (TGD) individuals are broadly defined as those whose gender identity does not align with their sex assigned at birth, which can result in gender dysphoria [4]. This group includes transmen, who identify as men; transwomen, who identify as women; and nonbinary transgender individuals, whose identities do not conform to the male/female binary [4,5]. Intersex individual is person with variations in physical sex characteristics, including anatomy, hormones, chromosomes, or other traits, that differ from expectations generally associated with male or female body (Table 1).

Studies conducted outside of Korea estimate that SGMs comprise 4%–5% of the total population [6,7]. In the United States, an annual survey of adults aged 18 years and older found that 7.6% self-identified as SGM in 2023, an increase from 3.5% in 2012 and 5.6% in 2020 [8]. The proportion was notably higher among younger generations, with 22.3% of Generation Z (born 1997–2012) and 9.8% of millennials (born 1981–1996) identifying as SGMs. This contrasts with 4.5% of Generation X (born 1965–1980) and 2.3% of Baby Boomers (born 1946–1964). The generational difference may reflect a greater visibility and openness about their identities among younger people. Unlike countries that estimate the sizes of their SGM populations through national statistics and questionnaires, Korea lacks such data, as questions about sexual orientation and gender identity are not collected in national surveys. Applying international estimates to Korea's population of 50 million suggests a population of approximately 2–2.5 million SGM individuals.

Sexual and gender minorities and health

In the past, the understanding of SGMs was limited, and these identities were classified as mental disorders. However, with advances in scientific and social understanding, diagnostic classifications evolved, and SGM identities were depathologized. Regarding sexual orientation, homosexuality was listed in the APA's Diagnostic and Statistical Manual of Mental Disorders (DSM)-I in 1952. It was removed from the DSM-II as a mental disorder in 1973, following social

Table 1. Definitions of sexual and gender minorities

Term	Definition
Lesbian	A woman who experiences emotional, romantic, or sexual attraction to women
Gay	A man who experiences emotional, romantic, or sexual attraction to men
Bisexual	A person who is attracted to both people of their own and other genders
Asexual	A person who does not experience sexual attraction toward individuals of any gender
Transgender	An individual whose current gender identity differs from the sex assigned at birth
Nonbinary	An individual who does not identify as male or female regarding gender
Questioning	For some, the process of exploring and discovering one's own sexual orientation, gender identity, or gender expression
Intersex	A person with variations in physical sex characteristics, including anatomy, hormones, chromosomes, or other traits, that differ from expectations generally associated with male or female bodies

Data from Centers for Disease Control and Prevention [69].

https://doi.org/10.12771/emj.2024.e14

movements such as the Stonewall riots in 1969 as well as a growing understanding of SGMs. In its place, the category "sexual orientation disturbance" was introduced. By the publication of the DSM-V in 2013, homosexuality was totally deleted, with the recognition that variations in sexual orientation are not indicative of a disorder [9]. In 1998, the APA officially stated that attempts to change an individual's sexual orientation, commonly referred to as conversion therapy, are contraindicated as they can worsen mental health [9]. In the context of gender identity, "transsexualism" was first included in the DSM-III in 1980. In the 2013 release of the DSM-V, this term was replaced with the diagnosis of gender dysphoria, with the manual clarifying that gender identity is not a disorder. Gender dysphoria is now used to justify the need for gender-affirming care (GAC) rather than to label gender identity as a disorder [4].

While SGM identities are no longer classified as mental disorders, these groups continue to face health disparities in various areas. Gay, lesbian, and bisexual individuals often experience comparatively poor mental health outcomes, such as higher rates of depression, anxiety, and suicidal ideation, as well as an increased prevalence of chronic conditions like insomnia [10,11]. TGD individuals also experience elevated levels of depression, anxiety, suicidal ideation, and suicide attempts, along with higher mortality rates [12,13]. The minority stress model, which posits that stress stemming from a minority identity and status imposes additional burdens on top of general stressors, is commonly employed to explain the poor mental health observed among SGM populations [14,15]. A 2024 review published in *The Lancet* indicated that structural stigma—including societal conditions, cultural norms, and institutional policies—intensifies health disparities for SGM individuals [16].

Another critical issue for SGM health is GAC for TGD individuals. GAC includes medical interventions that enable TGD people to live in a manner consistent with their gender identity, thereby alleviating gender dysphoria [4]. This care may involve gender-affirming hormone therapy (GAHT) and gender-affirming surgery. As part of this process, many TGD individuals are diagnosed with gender identity disorder (GID, ICD-10 code F64) [4]. Although not all TGD individuals seek medical intervention, many require varying levels of GAC to relieve their gender dysphoria. Previous research has shown that GAC can enhance physical and mental health by reducing gender dysphoria, improving overall quality of life, and decreasing suicidal ideation among TGD people [17,18]. Consequently, many countries, including the United States and Germany, have implemented policies to expand access to GAC through public insurance coverage [19–21].

Health of sexual and gender minorities in Korea

Scientifically, little is known about the health of SGMs in Korea. Despite the depathologization of SGM identities, their presence has remained largely invisible within academic research for many years. Apart from acknowledging MSM as a high-risk group for HIV/acquired immunodeficiency syndrome (AIDS), research on SGM health in Korea has been limited. A systematic review of Korean SGM health, conducted in 2014, analyzed 128 papers published up to that year. This review identified 101 clinical studies and 27 social health studies [22]. Among the clinical investigations, 50 case reports pertained to intersex conditions, while 21 studies focused on surgical interventions for intersex and transgender individuals. Of the social health studies, 13 examined mental health. The review highlighted a notable shortfall in research on SGM health in Korea compared to other countries and pointed out the lack of studies on healthcare accessibility, a crucial social determinant of health [22].

As of 2024, TGD individuals in Korea are required to undergo GAC not only to alleviate gender

dysphoria but also to obtain legal gender recognition [23]. Per the Supreme Court's Family Relations Registration Guidelines No. 550, titled "Guidelines for Processing Applications for Legal Gender Recognition of Transgender Individuals," applicants are typically required to present a diagnosis of GID and evidence of their inability to reproduce [23]. Furthermore, TGD individuals assigned male at birth must receive GAC, including a GID diagnosis, to qualify for exemption from mandatory military service as stipulated by Korea's Military Service Act. In some instances, TGD individuals may undergo GAC solely for the purposes of legal gender recognition or military exemption, rather than out of personal necessity. Despite these mandates, GAC in Korea is not covered by the National Health Insurance (NHI) service, and very few healthcare facilities offer these services. Research is scarce regarding the current availability and impact of GAC for TGD individuals in Korea.

Objectives

This study aims to review existing research on the health of Korean SGMs and to propose strategies for healthcare professionals to improve their health outcomes and access to medical care.

Methods

Ethics statement

As this is a literature review study, it does not require approval from an institutional review board or individual consent.

Study design

The present study is a narrative review of peer-reviewed quantitative studies obtained through a web-based database search.

Literature search

We reviewed quantitative studies on SGM health that were published in peer-reviewed international journals after 2000s, the year marked the depathologization of SGM identities and their increased visibility within Korean society.

Inclusion and exclusion criteria

As a 2014 systematic review focused on theses and Korean domestic journals [22], our study was limited to quantitative research published in peer-reviewed international journals. We excluded PhD theses, review papers, qualitative studies, validation studies for assessment tools, and studies of surgical techniques. Additionally, research that considered SGMs within the context of HIV/AIDS risk groups was not included in our analysis.

Information source and search strategy

The following search terms were used in PubMed, combining terms related to SGMs with "Korea" [16]:

(Korea) AND {(bisexual [tiab]) OR (bisexuality [MESH terms]) OR (gay [tiab]) OR (homosexuality, female [MESH terms]) OR (homosexuality, male [MESH terms]) OR (lesbian [tiab]) OR (LGB [tiab]) OR (LGBT [tiab]) OR (sexual and gender minorities [MESH terms]) OR (sexual behavior [MESH

terms]) OR (sexual orientation [tiab]) OR (sexuality [MESH terms]) OR (transgender persons [MESH terms])}

As of February 29, 2024, the titles and abstracts of all identified papers were reviewed. Additional studies by the same authors and cited references were also examined. This led to the analysis of 33 studies that focused on Korean SGMs and included health-related variables as independent or outcome measures.

Results

The selected papers were analyzed, which involved categorizing them based on their focus on either sexual minority or gender minority health. Key aspects such as the publication year, study methods, and topics were examined, and the critical findings from each paper were reviewed.

Studies of the health of sexual minorities

Within the category of sexual minority health, a total of 17 studies were identified (Table 2). Seven studies (41%) utilized data from online surveys conducted in 2016 among gay, lesbian, and bisexual individuals. Four articles (24%) employed data from the Youth Risk Behavior Survey, which included questions regarding the sex of sexual partners. Two studies used data from online surveys of lesbian and bisexual women conducted in 2017, and two studies used data from online surveys of gay and bisexual men (that is, MSM) conducted in 2022.

The topic most frequently addressed was mental health, with six studies (35%) investigating depression, anxiety symptoms, and suicidal ideation, plans, and attempts among sexual minorities. Four studies (24%) examined health-related behaviors, including alcohol consumption and smoking. Additionally, two studies explored health-related quality of life. One study each investigated disordered eating behaviors, cervical cancer screening and HPV vaccination, HIV testing, avoidance or delay in seeking healthcare, sleep health, coronavirus disease (COVID-19) vaccination, and physical distancing from people living with HIV (PLWH).

The research findings indicated that homosexual adolescents, compared to their heterosexual counterparts, faced a higher risk of engaging in health-risk behaviors such as alcohol consumption and smoking. They also experienced higher rates of suicidal ideation, plans, and attempts [24,25]. Bisexual adolescents were more likely than homosexual adolescents to engage in disordered weight control behaviors and exhibited higher frequencies of alcohol consumption [26,27]. Among adults, gay, lesbians, and bisexuals reported poorer self-rated health, more musculoskeletal pain, higher levels of depression, and elevated risks of suicidal ideation and attempts compared to the general population [11,28]. Studies that focused exclusively on gays, lesbians, and bisexuals revealed that those with higher levels of internalized homophobia, experiences of discrimination due to sexual orientation, or experiences of bullying related to sexual orientation during adolescence were more likely to report depression [29–31]. Higher levels of internalized homophobia, experiences of sexual orientation change efforts, and experiences of bullying related to sexual orientation during adolescence were also associated with higher rates of suicidal ideation and attempts [29,31,32]. Individuals who perceived a risk of rejection because of their sexual orientation were more likely to avoid or delay seeking healthcare, and those who experienced discrimination based on sexual orientation reported poorer sleep health [33,34].

Studies focusing on lesbian and bisexual women have shown that although their physical

Table 2. List of 17 articles on the health of sexual minorities

No	Year	Authors	Primary exposure or measure	Outcome assessment	Method	Sample characteristics	Main result and significant effect association indicating adverse health effects
1	2016	Lee et al. [24]	Sexual orientation (homosexual)	Health-risk behaviors and health cognition	Youth online survey*	N=129,900, adolescent	Health risk behavior, poor health cognition.
2	2016	Cho et al. [28]	Sexual orientation (homosexual)	Stress, depression, suicidal ideation, and attempts	Online survey	N=873, MSM	Perceived stress, depression.
3	2017	Kwak et al. [25]	Sexual orientation (homosexual)	Lifestyle and suicide-related behaviors.	Youth online survey	N=3,603, adolescent	Suicidal ideation, plans, attempts, medically serious attempts.
4	2017	Yi et al. [11]	Sexual orientation (homosexual and bisexual)	Physical/mental health symptom, health-risk behaviors, suicide-related behavior	RCP1 online survey†	N=2,335, LGB	Low self-rated health, More musculoskeletal pain, depressive symptom, suicidal ideation, suicidal attempt.
5	2018	Yu et al. [26]	Sex of sexual partners (bisexual)	Disordered weight control behaviors	Youth online survey	N=67,266, adolescent	Disordered weight control behavior.
6	2019	Lee et al. [29]	Internalized Homophobia	Depressive symptoms, suicidality	RCP1 online survey	N=2,178, LGB	Depressive symptom, suicidal ideation.
7	2019	Lee et al. [30]	Discrimination	Depressive symptoms	RCP1 online survey	N=2,162, LGB	Depressive symptom.
8	2020	Kim et al. [36]	Sex of sexual partners	Cervical cancer screening and HPV	Online survey	N=671, LB (women)	Homosexual less screening and less completion of HPV vaccination than bisexual.
9	2020	Kim et al. [27]	Sexual orientation	Alcohol use behaviors	Youth online survey	N=9,014, adolescent	Alcohol use: bisexual>homosexual. Binge drinking: homosexual>bisexual, heterosexual.
10	2021	Lee et al. [32]	Sexual orientation change efforts	Depressive symptoms, suicidality	RCP1 online survey	N=2,168, LGB	Suicidal ideation and attempts.
11	2021	Kim et al. [35]	Sexual orientation	HRQoL was measured using SF-36v2	Online survey	N=736, LB (women)	Overall low mental HRQoL. Bisexual lower than homosexual.
12	2022	Lee et al. [70]	Internalized homophobia	Past 12-month HIV testing	Online survey	N=907, GB (men) with HIV (-)	Low HIV testing.
13	2022	Choo et al. [33]	Expectation of rejection	Healthcare avoidance and delay	RCP1 online survey	N=2,175, LGB	Healthcare avoidance and delay.
14	2022	Park et al. [31]	Adolescent bullying victimization due to SOGE	Adulthood suicidality and depressive symptoms	RCP1 online survey	N=2,152, LGB	Depressive symptoms, suicidal ideation, and suicide attempts.
15	2022	Choo et al. [34]	Discrimination	Poor sleep health outcomes	RCP1 online survey	N=2,192, LGB	Overall poor sleep health, discrimination experience - higher prevalence of poor sleep quality, unrestful sleep, and long sleep latency.
16	2024	Jung [37]	HIV infection (+)	COVID-19 vaccination and infection	Online survey	N=942, MSM	More COVID19 infection.
17	2024	Jung [38]	Homosexual attributes of MSM.	extent of physical distancing perceived by MSMs without HIV toward PLWH	Online survey	N=878, MSM	HIV positive acquaintances around them – less physically distance from PLWH.

*Youth online survey: Korea Youth Risk Behavior Web-based Survey.

†RCP1 online survey: rainbow connection project 1.

MSM, men who have sex with men; LGB, lesbian, gay, and bisexual; LB, lesbian and bisexual; HRQoL, health-related quality of life; SF-36v2, 36-item Short Form Health Survey version 2.0; GB, gay and bisexual; HIV, human immunodeficiency virus; SOGE, sexual orientation and gender identity; COVID-19, coronavirus disease-19; PLWH, people living with HIV.

https://doi.org/10.12771/emj.2024.e14

health-related quality of life is comparable to that of the general population, their mental health-related quality of life is lower [35]. Furthermore, rates of cervical cancer screening and HPV vaccination differ based on the sex of sexual partners, with lesbian and bisexual women who have sex exclusively with female partners tending to receive fewer screenings and vaccinations [36].

Research involving gay and bisexual men has shown that PLWH experienced higher rates of COVID-19 infection. Additionally, individuals who had PLWH as friends or acquaintances were less likely to maintain physical distancing from them, compared to those who did not had PLWH as friends or acquaintances [37,38].

Studies of the health of gender minorities

Regarding the health of gender minorities, a total of 16 studies were identified (Table 3). The earliest study, published in 2006, utilized survey data from the Military Manpower Administration. Seven studies (44%) used data from online surveys of TGD individuals conducted in 2017, while four studies (25%) used data from online surveys of this population conducted in 2020, with 1 year of follow-up. Three studies utilized hospital medical records, and one study used administrative data from the Health Insurance Review and Assessment Service. The majority of the studies (12 of 16, or 75%) were based on surveys of TGD individuals.

The most frequently researched topic was mental health, with nine studies (56%) exploring issues such as depression, anxiety symptoms, and suicidal ideation, plans, and attempts among TGD individuals. Three clinical studies focused on TGD patients who underwent GAC at a single institution; these studies reported on the detail of GAC, the physical effects of GAHT, and the time elapsed between recognizing gender dysphoria and initiating GAHT. Two studies investigated the tendency of TGD individuals to avoid or delay seeking necessary healthcare, while another study explored the barriers they faced in accessing GAC. Additionally, two studies—one using online survey data and the other analyzing administrative data—examined the demographic characteristics of the TGD population in Korea.

The research findings indicated that transgender individuals experienced higher levels of psychological distress and stress, along with lower self-esteem, compared to their non-transgender counterparts. Additionally, they faced increased risks of depression and suicidal ideation relative to the general population [39,40]. An examination of medical records revealed that 20% of TGD individuals who underwent GAC had a mental health diagnosis other than gender dysphoria [41]. Results from an online survey showed that over half (53.7%) of TGD individuals had faced discrimination due to their gender identity. Those who encountered such discrimination were more likely to avoid or postpone seeking GAC or other healthcare services [42,43]. Among TGD individuals, higher levels of internalized transphobia, stress associated with using public bathrooms, experiences of gender identity change efforts, and discrimination based on gender identity were associated with greater depression [44–47]. Higher levels of internalized transphobia and experiences of gender identity change efforts were linked to increased suicidal ideation and attempts [44,45]. TGD individuals who faced discrimination due to their gender identity reported worse sleep health. Additionally, those who avoided public bathrooms, job-seeking activities, or healthcare services because of their gender identity experienced higher levels of anxiety symptoms [48,49]. TGD individuals identified several significant barriers to accessing GAC, including costs, negative experiences in healthcare settings, a scarcity of healthcare providers and facilities with expertise in GAC, and societal stigma [50].

A study of 337 TGD individuals in Korea who underwent GAC revealed that the average age

Table 3. List of 16 articles on the health of gender minorities

No	Year	Authors	Primary exposure or measure	Outcome assessment	Sample design	Sample characteristics	Significant main effect association indicating adverse health effects of LGB
1	2006	Kim et al. [39]	Gender identity (TGD)	Psychological burdens (BDI, SADS, SES, FACES-III)	Offline survey	TGD, N=43, 49 matched non-transsexual	High depression burden, social avoidance and distress, low self-esteem scale, family adaptability and cohesion.
2	2018	Lee et al. [50]	Experiences of and barriers to transition-related healthcare	Gender affirming care	RCP2* online survey	TGD, N=278	Barrier: cost, negative experiences in healthcare settings, lack of specialized healthcare professionals and facilities, and social stigma against TGD.
3	2019	Lim et al. [41]	Demographics	Gender identity-related characteristics	Hospital medical record review	TGD, N=54	20% had mental disorder other than gender dysphoria.
4	2020	Lee et al. [44]	Internalized transphobia	Depressive symptoms, suicidal ideation, and suicide attempts.	RCP2 online survey	TGD, N=207	Depressive symptom, suicidal ideation and attempts.
5	2020	Lee et al. [40]	Gender identity (TGD)	Physical health, mental health	RCP2 online survey	TGD, N=255	Depressive symptom, suicidal ideation.
6	2021	Yun et al. [52]	Cross-sex hormone	Body composition, bone mineral density, muscle strength	Hospital medical record review	TGD women, N=11	Increase fat mass, decrease in overall lean body mass and handgrip strength.
7	2021	Lee et al. [46]	Public Bathroom-Related Stressors	Depressive Symptoms	RCP3† online survey	TGD, N=557	Depressive symptoms.
8	2022	Lee et al. [43]	Discrimination due to TGD identity	Healthcare avoidance and delay	RCP2 online survey	TGD, N=244	Healthcare avoidance and delay.
9	2022	Eom et al. [49]	Discrimination due to TGD identity	Sleep problems	RCP3 online survey	TGD, N=583	Sleep problems.
10	2022	Choo et al. [34]	Discrimination due to TGD identity	Dymptoms of depression and anxiety	RCP2 online survey	TGD, N=269	Depressive and anxiety symptoms.
11	2023	Lee et al. [45]	Gender identity change effort	Depression, PTSD, suicide attempts	RCP3 online survey	TGD, N=566	11.5% Gender identity change effort experience, more depression, panic disorder, suicide attempt.
12	2023	Lee et al. [71]	Transgender-specific COVID-19-related stressors	Past-week depressive symptoms	RCP3 online survey	TGD, N=564	30% TGD-specific COVID-19 related stressor experience and more depressive symptoms. Barrier to gender affirming care: economic hardship, limited access to hospital.
13	2023	Kim et al. [42]	Discrimination due to TGD identity	Healthcare avoidance and delay	RCP2 online survey	TGD, N=190	53.7% experienced anti-transgender discrimination at initial and one year follow up survey, and they experienced more non-transition-related healthcare avoidance and delay.
14	2023	Kim et al. [53]	Demographics	Gender identity disorder, intersex	Administrative data (HIRA)	TGD, N=8,602	For 15 years (2007-2021), 8,602 people who received the F64 codes (gender identity disorder), 45 people diagnosed intersex.
15	2024	Eom et al. [48]	Situational avoidance	Mental health (1week depression, 2week anxiety)	RCP2 online survey	TGD, N=268	50.4% experienced avoided daily activities (public bathroom use, job application, and hospital visit) and more anxiety symptom.
16	2024	Oh et al. [51]	Demographics	Onset of gender incongruence	Hospital medical record review	TGD, N=337	Mean age of onset of GI was 10.6 years (29% before age 6, 61% before age 12, and 87% before age 15), TGD lived with GI for almost 14 years before gender affirming hormone therapy.

*RCP2 online survey: rainbow connection project 2.

†RCP3 online survey: rainbow connection project 3.

TGD, transgender and gender diverse; BDI, Beck's Depression Inventory; SADS, Social Avoidance and Distress Scale; SES, Self-Esteem Scale; FACES, Family Adaptability and Cohesion Evaluation Scale; PTSD, post-traumatic stress disorder; COVID-19, coronavirus disease-19; HIRA, Health Insurance Review and Assessment Service.

https://doi.org/10.12771/emj.2024.e14

at which participants first recognized their gender dysphoria was 10.6 years (SD, 5.1 years), with 29% recognizing it before the age of 6 and 61% before age 12. Based on the median age of initiating GAHT, which is 23 years, these individuals lived with gender dysphoria for an average of approximately 14 years before beginning GAHT [51]. A study of transgender women undergoing GAHT observed physical changes that included an increase in fat mass, a decrease in hand grip strength, and a shift toward a more "feminized" body fat distribution when compared to the pre-treatment period [52].

A study utilizing Health Insurance Review and Assessment Service data found that, between 2007 and 2021, 8,602 individuals received a diagnosis of GID (ICD-10 code, F64), with an annual rate of approximately 500–600 diagnoses and an increasing trend over the years [53].

Discussion

Interpretation

Between 2000 and 2024, only 33 quantitative studies focusing on the health of Korean SGMs were published in peer-reviewed international journals. Although the number of studies has gradually increased, with one study published between 2000–2010, 13 studies from 2011–2020, and 19 studies from 2021–2024, research on this topic is still relatively limited. Starting in 2017, surveys have been conducted of Korean gay, lesbian, bisexual, and TGD individuals, contributing to the growth in published research on the health of Korean SGMs. Gay, lesbian, and bisexual Koreans have reported higher rates of depression, anxiety symptoms, suicidal ideation, and suicide attempts, as well as a lower health-related quality of life. Those with internalized homophobia, who have experienced coercive attempts to change their sexual orientation, or who have faced discrimination due to their sexual orientation are more likely to experience poor mental health and sleep issues and to avoid or delay seeking medical care. Similarly, TGD individuals report relatively high rates of depression, anxiety symptoms, suicidal thoughts, and suicide attempts. Those with greater internalized transphobia, who have encountered discrimination based on their gender identity, or who have avoided daily activities due to their gender identity are relatively likely to experience worse mental health and to avoid or delay using healthcare services.

Suggestions for improving the health of sexual and gender minorities

Based on these research findings, the following changes are needed to improve the health of the vulnerable SGM population in Korea and to increase their access to medical care:

First, more research is required on the health of Korean SGMs. While international studies encompass a wide range of topics, including chronic disease management, cancer incidence, and the long-term effects of GAC for TGD individuals, research on the health of Korean SGMs remains limited in scope. To improve the health of SGMs and to develop strategies for suitable healthcare access, it is imperative to expand the breadth of studies within the Korean healthcare system [54–56].

Based on the comparatively poor health outcomes demonstrated, many countries are actively supporting research to enhance the health of SGMs. For instance, in 2015, the US National Institutes of Health established the Sexual & Gender Minority Research Office, which developed a strategic plan to advance research on the health and well-being of SGMs [57]. Research on SGM health must be promoted, both to identify the population-level factors that contribute to their vulnerability and to scientifically explore ways to improve their health outcomes.

Additionally, national-level statistics are essential for gaining a more accurate understanding of the lives and health of SGMs. Following the examples of the United States and the United Kingdom, incorporating questions about sexual orientation and gender identity into national surveys, such as the National Health and Nutrition Examination Survey and the Community Health Survey, would represent a key first step. This would provide baseline statistics on the size and status of the SGM population, which could inform our understanding of their health in Korea [58,59].

Second, LGBTQ-friendly healthcare providers are essential for improving healthcare access for SGMs. Due to their identities, many SGM individuals face discrimination in their daily lives, which often leads to delays in seeking necessary medical care. Additionally, TGD individuals report a lack of competent healthcare providers and facilities as a significant barrier to GAC [50]. To develop LGBTQ-friendly healthcare providers, education and training on SGM health must be integrated into medical school curricula and residency training programs. In 2015, the American College of Physicians emphasized that training healthcare providers in knowledge, experience, cultural competency, and sensitivity to human rights regarding SGMs is vital for promoting SGM health and reducing health disparities [60]. While international medical schools such as Harvard University and the University of Washington offer educational programs on SGM health, and the American Medical Association provides an SGM health fellowship, educational programs on SGM health in Korea are still in their infancy [61,62]. In 2021, Seoul National University College of Medicine introduced an elective course on SGM health for second-year medical students, and in 2022, mandatory course began for all students. The need exists for a standardized curriculum on SGM health, which should be disseminated to all medical schools and training institutions nationwide. This would serve as a critical first step in training healthcare professionals who are knowledgeable about SGM health vulnerabilities and who actively participate in addressing them.

Third, research and policy efforts are needed to implement programs that have been demonstrated to enhance the health of SGMs. Social support and the legalization of same-sex marriage have been shown to improve the mental health of homosexual and bisexual individuals [63,64]. Increasing access to GAC for TGD individuals is another notable example, with many studies demonstrating its positive effects. Although systematic reviews have highlighted the beneficial impacts of GAC, and the 8th edition of the international Standards of Care for the Health of Transgender and Gender Diverse People was published in 2022, knowledge is still limited regarding the health outcomes of TGD individuals receiving GAC in Korea.

Within Korea, TGD individuals encounter barriers to accessing GAC, including prohibitive costs, a lack of competent providers and facilities, and social stigma [50]. In 2014, Soonchunhyang University Seoul Hospital opened a gender clinic within its Department of Obstetrics and Gynecology, becoming the first tertiary hospital in Korea to specialize in GAHT. Subsequently, in 2021, the LGBTQ+ Clinic at Kangdong Sacred Heart Hospital—through collaboration among the Department of Plastic Surgery, the Department of Psychiatry, and other multidisciplinary services—and the Gender Clinic of Korea University Anam Hospital's Department of Plastic Surgery opened their doors to provide gender-affirming surgery. Additionally, several primary care clinics (such as the Salim Clinic, which has administered GAHT to over 3,000 TGD individuals since 2012) have begun offering GAC, with a growing number of providers. Despite these advancements, GAC is still not accessible as a universal healthcare service. More LGBTQ-friendly healthcare facilities, capable of providing GAC rooted in current knowledge and cultural competency, are needed. Moreover, it is crucial to assess the present state of GAC in Korea

and to evaluate its impact, forming the foundation for new strategies that enhance access to these services. In 2023, eight LGBTQ-friendly healthcare facilities and researchers specializing in SGM health initiated the KITE: Korean Initiative for Transgender hEalth project, a cohort study focused on the health of Korean TGD individuals. Gathering scientific evidence on the health of the Korean TGD population and the effects of GAC is imperative. These findings can then inform societal discussions about the inclusion of GAC in the coverage provided by the NHI system.

Fourth, healthcare professionals should advocate for changes in social perceptions and institutional policies related to SGM health. Studies in Korea have shown that internalized stigma, harassment, discrimination, and exclusion due to SGM identities are associated with worsened mental health and increases in suicidal ideation and attempts [29,30,33,34,44]. To promote SGM health, the American College of Physicians recommends that healthcare professionals advocate for the rights to same-sex marriage, institutional guarantees of GAC for TGD individuals, and the inclusion of sexual orientation and gender identity as protected categories against discrimination [60].

Healthcare professionals should serve as authorities in correcting misconceptions about SGMs and spearheading institutional reforms that are closely linked to SGM health. For example, the APA has published an official statement denouncing discrimination against transgender individuals. Similarly, GLMA: Health Professionals Advancing LGBTQ+ Equality, a coalition of healthcare providers advocating for SGM equality, has consistently issued statements in favor of same-sex marriage and partnership laws, the protection of transgender individuals' healthcare rights against discrimination, and the provision of support for SGM youth in educational settings [65,66]. In a 2019 international survey, only 44% of Koreans agreed that homosexuality should be accepted by society, a figure markedly lower than those reported in Sweden (94%), the Netherlands (92%), and the United States (72%). The 2023 Social Integration Survey further indicated that Koreans exhibit a high tendency to exclude sexual minorities from societal acceptance (52.3%), ranking just below their inclination to exclude ex-convicts (72.1%) [67,68]. In Korea, same-sex marriage remains illegal, GAC for TGD individuals is not covered by the NHI, and no anti-discrimination laws yet exist that encompass SGM identities. Given this context, there is much work for healthcare professionals to do to promote SGM health.

Conclusion

This study aimed to propose a role for healthcare professionals in improving the health and healthcare access of Korean SGMs through a narrative review of quantitative studies on SGM health. Korean SGM individuals have been found to experience higher rates of depression, anxiety, suicidal ideation, and suicide attempts, with the risks being even greater for those subjected to discrimination or coercive efforts to alter their SGM identities.

While GAC has been shown to enhance the health and quality of life of TGD individuals, data are limited regarding the availability and accessibility of GAC within the Korean TGD population. Despite the acknowledgment that diversity in sexual orientation and gender identity is not indicative of a disorder, SGMs in Korea continue to experience poor health outcomes and a diminished health-related quality of life. Furthermore, scant research has been conducted on their health status.

In Korean society, where discrimination and hatred against SGMs are rampant, the discrimination and stigma experienced by SGMs not only worsen their mental health but also reduce their access to healthcare, further exacerbating their health vulnerabilities. Healthcare professionals should become active advocates for SGM health, grounded in a comprehensive understanding of SGMs and their health-related needs.

ORCID
Heesung So: https://orcid.org/0009-0003-6790-0679
Ssirai Kim: https://orcid.org/0000-0002-3403-6699
Sun Young Lee: https://orcid.org/0000-0002-1626-2721

Authors' contributions
Project administration: Lee SY
Conceptualization: So H, Kim S, Lee SY
Methodology & data curation: So H, Kim S, Lee SY
Funding acquisition: not applicable
Writing – original draft: So H, Lee SY
Writing – review & editing: So H, Kim S, Lee SY

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Funding
Not applicable.

Data availability
Not applicable.

Acknowledgments
We extend our gratitude to the members of the Korean Association for LGBTQ Medicine for their assistance in conducting this research. We also thank Hyein Chu from the Salim Health Welfare Social Cooperative (the Salim Clinic) for reviewing the draft of this paper and providing valuable advice.

Supplementary materials
Supplementary materials are available from: https://doi.org/10.12771/emj.2024.e14.

Supplement 1. Korean version of the present review article

References

1. Mayer KH, Bradford JB, Makadon HJ, Stall R, Goldhammer H, Landers S. Sexual and gender minority health: what we know and what needs to be done. *Am J Public Health* 2008;98(6):989-995. https://doi.org/10.2105/AJPH.2007.127811
2. American Psychological Association. Understanding sexual orientation and homosexuality [Internet]. Washington (DC): American Psychological Association; c2008 [cited 2024 Jan 24]. Available from: https://www.apa.org/topics/lgbtq/orientation
3. American Psychological Association. Understanding transgender people, gender identity and gender expression [Internet]. Washington (DC): American Psychological Association; c2023 [cited 2024 Jan 24]. Available from: https://www.apa.org/topics/lgbtq/transgender-people-gender-identity-gender-expression
4. Coleman E, Radix AE, Bouman WP, Brown GR, de Vries ALC, Deutsch MB, et al. Standards of care for the health of transgender and gender diverse people, version 8. *Int J Transgend Health* 2022;23(sup1):S1-S259. https://doi.org/10.1080/26895269.2022.2100644
5. American Psychological Association. Guidelines for psychological practice with transgender and gender nonconforming people. *Am Psychol* 2015;70(9):832-864. https://doi.org/10.1037/a0039906
6. Conron KJ, Goldberg SK. Adult LGBT population in the United States. Los Angeles: The Williams Institute; 2019.
7. Government of Netherlands. LGBTI equality in the Netherlands. The Hague: Government of Netherlands; 2018.
8. Jones JM. LGBTQ+ identification in U.S. now at 7.6% [Internet]. Washington (DC): GALLUP®; c2024 [cited 2024 Jan 24]. Available from: https://news.gallup.com/poll/611864/lgbtq-identification.aspx
9. APA Commission on Psychotherapy by Psychiatrists. Position statement on therapies focused on attempts to change sexual orientation (reparative or conversion therapies). *Am J Psychiatry* 2000;157(10):1719-1721.
10. Zeeman L, Sherriff N, Browne K, McGlynn N, Mirandola M, Gios L, et al. A review of lesbian, gay, bisexual, trans and intersex (LGBTI) health and healthcare inequalities. *Eur J Public Health* 2019;29(5):974-980. https://doi.org/10.1093/eurpub/cky226
11. Yi H, Lee H, Park J, Choi B, Kim SS. Health disparities between lesbian, gay, and bisexual adults and the general population in South Korea: Rainbow Connection Project I. *Epidemiol Health* 2017;39:e2017046. https://doi.org/10.4178/epih.e2017046
12. Reisner SL, Poteat T, Keatley J, Cabral M, Mothopeng T, Dunham E, et al. Global health burden and needs of transgender populations: a review. *Lancet* 2016;388(10042):412-436. https://doi.org/10.1016/S0140-6736(16)00684-X

13. Winter S, Diamond M, Green J, Karasic D, Reed T, Whittle S, et al. Transgender people: health at the margins of society. *Lancet* 2016;388(10042):390-400. https://doi.org/10.1016/S0140-6736(16)00683-8
14. Meyer IH. Minority stress and mental health in gay men. *J Health Soc Behav* 1995;36(1)38-56. https://doi.org/10.2307/2137286
15. Hendricks ML, Testa RJ. A conceptual framework for clinical work with transgender and gender nonconforming clients: an adaptation of the minority stress model. *Prof Psychol Res Pract* 2012;43(5):460-467. https://doi.org/10.1037/a0029597
16. Hatzenbuehler ML, Lattanner MR, McKetta S, Pachankis JE. Structural stigma and LGBTQ+ health: a narrative review of quantitative studies. *Lancet Public Health* 2024;9(2):E109-E127. https://doi.org/10.1016/S2468-2667(23)00312-2
17. White Hughto JM, Reisner SL. A systematic review of the effects of hormone therapy on psychological functioning and quality of life in transgender individuals. *Transgend Health* 2016;1(1):21-31. https://doi.org/10.1089/trgh.2015.0008
18. Mahfouda S, Moore JK, Siafarikas A, Hewitt T, Ganti U, Lin A, et al. Gender-affirming hormones and surgery in transgender children and adolescents. *Lancet Diabetes Endocrinol* 2019;7(6):484-498. https://doi.org/10.1016/S2213-8587(18)30305-X
19. Lee HL, Lee HM, Yoon JW, Park JY, Kim SS. Transgender people's access to health care in Korea. *Health Soc Welf Rev* 2015;35(4):64-94. https://doi.org/10.15709/hswr.2015.35.4.64
20. Padula WV, Baker K. Coverage for gender-affirming care: making health insurance work for transgender Americans. *LGBT Health* 2017;4(4):244-247. https://doi.org/10.1089/lgbt.2016.0099
21. Park H. A study on National Health Insurance coverage of transition-related care for transgender people. *Pub Int Hum Rights* 2018;18:191-235.
22. Lee HM, Park JY, Kim SS. LGBTQI health research in South Korea: a systematic review. *Health Soc Sci* 2014;(36):43-79.
23. Park H. Legislative suggestion on the legal gender recognition for transgender people. *Human Rights Justice* 2021;(498):41-60.
24. Lee DY, Kim SH, Woo SY, Yoon BK, Choi DS. Associations of health-risk behaviors and health cognition with sexual orientation among adolescents in school: analysis of pooled data from Korean nationwide survey from 2008 to 2012. *Medicine* 2016;95(21):e3746. https://doi.org/10.1097/MD.0000000000003746
25. Kwak Y, Kim JS. Associations between Korean adolescents' sexual orientation and suicidal ideation, plans, attempts, and medically serious attempts. *Iran J Public Health* 2017;46(4):475-484.
26. Yu KY, Kim Y, Calzo JP, Levinson JA, Bryn Austin S. Sex of sexual partners and disordered weight control behaviors in a nationally representative sample of South Korean adolescents. *Psychiatry Res* 2018;262:1-5. https://doi.org/10.1016/j.psychres.2018.01.028
27. Kim JS, Seo Y. Victimization as a mediator in the relationship between sexual orientation and adolescent alcohol use. *Arch Psychiatr Nurs* 2020;34(1):27-34. https://doi.org/10.1016/j.apnu.2019.12.005
28. Cho B, Sohn A. How do sexual identity, and coming out affect stress, depression, and suicidal ideation and attempts among men who have sex with men in South Korea? *Osong Public Health Res Perspect* 2016;7(5):281-288. https://doi.org/10.1016/j.phrp.2016.09.001
29. Lee H, Operario D, Yi H, Choo S, Kim SS. Internalized homophobia, depressive symptoms, and suicidal ideation among lesbian, gay, and bisexual adults in South Korea: an age-stratified analysis. *LGBT Health* 2019;6(8):393-399. https://doi.org/10.1089/lgbt.2019.0108
30. Lee H, Park J, Choi B, Yi H, Kim SS. Association between discrimination and depressive symptoms among 2,162 lesbian, gay, and bisexual adults in South Korea: does community connectedness modify the association? *J Homosex* 2019;68:70-87. https://doi.org/10.1080/00918369.2019.1624456
31. Park J, Lee H, Choi B, Kim JH, Yoon J, Yi H, et al. Adolescent bullying victimization at secondary school and adult suicidality and depressive symptoms among 2152 lesbian, gay, and bisexual adults in South Korea. *Asia Pacific J Public Health* 2022;34(4):338-345. https://doi.org/10.1177/10105395211073283
32. Lee H, Streed CG, Yi H, Choo S, Kim SS. Sexual orientation change efforts, depressive symptoms, and suicidality among lesbian, gay, and bisexual adults: a cross-sectional study in South Korea. *LGBT Health* 2021;8(6):427-432. https://doi.org/10.1089/lgbt.2020.0501
33. Choo S, Lee H, Yi H, Kim SS. Expectation of rejection and its association with health care avoidance and delay among 2175 Korean lesbian, gay, and bisexual adults: a nationwide cross-sectional survey. *LGBT Health* 2022;9(4):282-286. https://doi.org/10.1089/lgbt.2021.0269
34. Choo S, Kim R, Lee H, Yi H, Kim R, Kim SS. Association between discrimination and poor sleep health outcomes among 2192 South Korean gay, lesbian, and bisexual adults: a nationwide cross-sectional survey. *Sleep Health* 2022;8(6):587-592. https://doi.org/10.1016/j.sleh.2022.09.006
35. Kim S, Choi-Kwon S. Physical and mental health related quality of life and their influencing factors on sexual minority women in Korea. *Int J Environ Res Public Health* 2021;18(4):2115. https://doi.org/10.3390/ijerph18042115

36. Kim S, Lee SY, Choi-Kwon S. Cervical cancer screening and human papillomavirus vaccination among Korean sexual minority women by sex of their sexual partners. *Int J Environ Res Public Health* 2020;17(23):8924. https://doi.org/10.3390/ijerph17238924
37. Jung M. Behavioral predictors associated with COVID-19 vaccination and infection among men who have sex with men in Korea. *J Prev Med Public Health* 2024;57(1):28-36. https://doi.org/10.3961/jpmph.23.381
38. Jung M. Physical distancing for gay men from people living with HIV during the COVID-19 pandemic. J Homosex 2024 Feb 6 [Epub]. https://doi.org/10.1080/00918369.2024.2314031
39. Kim TS, Cheon YH, Pae CU, Kim JJ, Lee CU, Lee SJ, et al. Psychological burdens are associated with young male transsexuals in Korea. *Psychiatry Clin Neurosci* 2006;60(4):417-421. https://doi.org/10.1111/j.1440-1819.2006.01525.x
40. Lee H, Operario D, van den Berg JJ, Yi H, Choo S, Kim SS. Health disparities among transgender adults in South Korea. *Asia Pacific J Public Health* 2020;32(2-3):103-110. https://doi.org/10.1177/1010539520912594
41. Lim HH, Jang YH, Choi GY, Lee JJ, Lee ES. Gender affirmative care of transgender people: a single center's experience in Korea. *Obstet Gynecol Sci* 2019;62(1):46-55. https://doi.org/10.5468/ogs.2019.62.1.46
42. Kim R, Choo S, Lee H, Eom YJ, Yi H, Kim R, et al. Does discrimination prevent transgender and gender diverse people from seeking healthcare?: a nationwide cohort study in South Korea. *Int J Transgend Health* 2023;25(2):283-294. https://doi.org/10.1080/26895269.2023.2215750
43. Lee H, Operario D, Yi H, Choo S, Kim JH, Kim SS. Does discrimination affect whether transgender people avoid or delay healthcare?: a nationwide cross-sectional survey in South Korea. *J Immigr Minor Health* 2022;24(1):170-177. https://doi.org/10.1007/s10903-021-01193-9
44. Lee H, Tomita KK, Habarth JM, Operario D, Yi H, Choo S, et al. Internalized transphobia and mental health among transgender adults: a nationwide cross-sectional survey in South Korea. *Int J Transgend Health* 2020;21(2):182-193. https://doi.org/10.1080/26895269.2020.1745113
45. Lee H, Operario D, Restar AJ, Choo S, Kim R, Eom YJ, et al. Gender identity change efforts are associated with depression, panic disorder, and suicide attempts in South Korean transgender adults. *Transgend Health* 2023;8(3):273-281. https://doi.org/10.1089/trgh.2021.0171
46. Lee H, Yi H, Nic Rider G, Operario D, Choo S, Kim R, et al. Transgender adults' public bathroom-related stressors and their association with depressive symptoms: a nationwide cross-sectional study in South Korea. *LGBT Health* 2021;8(7):486-493. https://doi.org/10.1089/lgbt.2021.0007
47. Choo S, Kim R, Lee H, Eom YJ, Yi H, Kim R, et al. Associations between discrimination experiences and symptoms of depression and anxiety among transgender adults: a nationwide cohort study of 269 transgender adults in South Korea. Soc Psychiatry Psychiatr Epidemiol 2023 Aug 26 [Epub]. https://doi.org/10.1007/s00127-023-02554-5
48. Eom YJ, Lee H, Choo S, Kim R, Yi H, Kim R, et al. Situational avoidance and its association with mental health among transgender adults in South Korea: a nationwide cohort study. *LGBT Health* 2024;11(2):122-130. https://doi.org/10.1089/lgbt.2023.0060
49. Eom YJ, Lee H, Kim R, Choo S, Yi H, Kim SS. Discrimination keeps transgender people awake at night: a nationwide cross-sectional survey of 583 transgender adults in South Korea. *Sleep Health* 2022;8(6):580-586. https://doi.org/10.1016/j.sleh.2022.06.012
50. Lee H, Park J, Choi B, Yi H, Kim SS. Experiences of and barriers to transition-related healthcare among Korean transgender adults: focus on gender identity disorder diagnosis, hormone therapy, and sex reassignment surgery. *Epidemiol Health* 2018;40:e2018005. https://doi.org/10.4178/epih.e2018005
51. Oh JW, Park S, Lim S, Lee ES. Age of first experience of gender incongruence among transgender and non-binary individuals. *Obstet Gynecol Sci* 2024;67(1):132-141. https://doi.org/10.5468/ogs.23229
52. Yun Y, Kim D, Lee ES. Effect of cross-sex hormones on body composition, bone mineral density, and muscle strength in trans women. *J Bone Metab* 2021;28(1):59-66. https://doi.org/10.11005/jbm.2021.28.1.59
53. Kim DJ, Hwang NH, Lee JY, Park SH, Lee BI, Yoon ES. An analysis of the demographics and clinical characteristics of transgender and intersex populations in Korea: a retrospective study using HIRA database. *J Korean Med Sci* 2023;38(50):e385. https://doi.org/10.3346/jkms.2023.38.e385
54. Crowley F, Mihalopoulos M, Gaglani S, Tewari AK, Tsao CK, Djordjevic M, et al. Prostate cancer in transgender women: considerations for screening, diagnosis and management. *Br J Cancer* 2023;128(2):177-189. https://doi.org/10.1038/s41416-022-01989-y
55. Shadid S, Abosi-Appeadu K, De Maertelaere AS, Defreyne J, Veldeman L, Holst JJ, et al. Effects of gender-affirming hormone therapy on insulin sensitivity and incretin responses in transgender people. *Diabetes Care* 2020;43(2):411-417. https://doi.org/10.2337/dc19-1061
56. Brown JP, Kathleen Tracy J. Lesbians and cancer: an overlooked health disparity. *Cancer Causes Control* 2008;19(10):1009-1020. https://doi.org/10.1007/s10552-008-9176-z

57. National Institutes of Health. Strategic plan to advance research on the health and well-being of sexual and gender minorities. Bethesda: The National Institutes of Health; 2016.
58. Semlyen J, King M, Varney J, Hagger-Johnson G. Sexual orientation and symptoms of common mental disorder or low wellbeing: combined meta-analysis of 12 UK population health surveys. *BMC Psychiatry* 2016;16:1-9. https://doi.org/10.1186/s12888-016-0767-z
59. Ward BW, Dahlhamer JM, Galinsky AM, Joestl SS. Sexual orientation and health among US adults: National Health Interview Survey, 2013. Washington: United States Department of Health and Human Services; 2014.
60. Daniel H, Butkus R. Lesbian, gay, bisexual, and transgender health disparities: executive summary of a policy position paper from the American College of Physicians. *Ann Intern Med* 2015;163(2):135-137. https://doi.org/10.7326/M14-2482
61. Gibson AW, Gobillot TA, Wang K, Conley E, Coard W, Matsumoto K, et al. A novel curriculum for medical student training in LGBTQ healthcare: a regional pathway experience. *J Med Educ Curric Dev* 2020;7:2382120520965254. https://doi.org/10.1177/2382120520965254
62. AMA Foundation. The AMA foundation national LGBTQ+ fellowship program [Internet]. Chicago (IL): AMA Foundation; c2024 [cited 2024 Jan 24]. Available from: https://amafoundation.org/programs/lgbtq-fellowship/
63. Wight RG, LeBlanc AJ, Lee Badgett MV. Same-sex legal marriage and psychological well-being: findings from the California Health Interview Survey. *Am J Public Health* 2013;103(2):339-346. https://doi.org/10.2105/AJPH.2012.301113
64. Sattler FA, Wagner U, Christiansen H. Effects of minority stress, group-level coping, and social support on mental health of German gay men. *PLOS ONE* 2016;11(3):e0150562. https://doi.org/10.1371/journal.pone.0150562
65. American Psychiatric Association. APA expresses great disappointment in the Supreme Court Decision to lift the injunctions on the administration's ban on transgender service members [Internet]. Washington (DC): American Psychiatric Association; c2019 [cited 2024 Jan 23]. Available from: https://www.psychiatry.org/news-room/news-releases/apa-expresses-great-disappointment-in-the-supreme
66. GLMA: Health Professionals Advancing LGBTQ Equality. GLMA's policies and position statements [Internet]. Washington (DC): GLMA; c2024 [cited 2024 Jan 23]. Available from: https://www.memberleap.com/news_archive_headlines.php?org_id=GLMA&snc=969947#969947
67. Poushter J, Kent NO. The global divide on homosexuality persists: but increasing acceptance in many countries over past two decades [Internet]. Washington (DC): Pew Research Center; c2020 [cited 2024 Jan 23]. Available from: https://www.pewresearch.org/global/wp-content/uploads/sites/2/2020/06/PG_2020.06.25_Global-Views-Homosexuality_FINAL.pdf
68. Korea Institute of Public Administration. 2023 Korea social integration survey [Internet]. Seoul (KR): Korea Institute of Public Administration; c2024 [cited 2024 Jan 24]. Available from: https://www.kipa.re.kr/synap/skin/doc.html?fn=FILE_000000000196070&rs=/convert/result/201512/
69. Centers for Disease Control and Prevention. Terminology [Internet]. Atlanta (GA): Centers for Disease Control and Prevention; c2022 [cited 2024 Jan 12]. Available from: https://www.cdc.gov/healthyyouth/terminology/sexual-and-gender-identity-terms.htm
70 Lee H, Operario D, Agénor M, Yi H, Choo S, Kim SS. Internalized homophobia and HIV testing among Korean gay and bisexual men: a study in a high-income country with pervasive HIV/AIDS stigma. *AIDS Care* 2023;35(5):672-677. https://doi.org/10.1080/09540121.2022.2083056
71. Lee H, Restar AJ, Operario D, Choo S, Streed CG Jr, Yi H, et al. Transgender-specific COVID-19-related stressors and their association with depressive symptoms among transgender adults: a nationwide cross-sectional survey in South Korea. *Int J Transgend Health* 2023;24(3):334-345. https://doi.org/10.1080/26895269.2021.1989357

Review

Ewha Med J 2024;47(2):e15
https://doi.org/10.12771/emj.2024.e15
eISSN 2234-2591

Sex differences in coronary atherogenesis: a narrative review

Hack-Lyoung Kim

Division of Cardiology, Department of Internal Medicine, Boramae Medical Center, Seoul National University College of Medicine, Seoul, Korea

Received Feb 26, 2024
Revised Apr 15, 2024
Accepted Apr 18, 2024

Corresponding author
Hack-Lyoung Kim
Division of Cardiology, Department of Internal Medicine, Boramae Medical Center, Seoul National University College of Medicine, 5 Boramae-ro, Dongjak-gu, Seoul 07061, Korea
E-mail: khl2876@snu.ac.kr

Keywords
Atherogenesis; Coronary artery disease; Sex difference

Coronary artery disease (CAD) remains the leading cause of mortality worldwide, driven primarily by atherogenesis. Recent efforts to understand sex differences in CAD have revealed distinct patterns in disease burden, risk factors, and clinical presentations. This review examines these sex differences in CAD, underscoring the importance of customized diagnostic and management strategies. Although men typically have higher rates of CAD prevalence and incidence, women face unique challenges, such as delayed diagnosis, atypical symptoms, and lower rates of medication prescription. Hormonal, genetic, and lifestyle factors all play a role in these disparities, with estrogen notably reducing CAD risk in women. Nontraditional risk factors, including chronic inflammation, psychological stress, socioeconomic status, and reproductive history, also contribute to CAD development and are often neglected in clinical settings. Addressing these differences requires increased awareness, more accurate diagnosis, and equitable healthcare access for both sexes. Furthermore, greater inclusion of women in CAD research is essential to better understand sex-specific mechanisms and optimize treatment outcomes. Personalizing CAD management based on sex-specific knowledge has the potential to improve prognosis and decrease disease incidence for both men and women.

Introduction

Background

Coronary artery disease (CAD) is the leading cause of death worldwide [1]. The formation of atherosclerotic plaques, a process known as atherogenesis, is the primary factor in the development of CAD [2]. Recent years have seen a growing interest and active research into the sex differences observed in various health conditions, including CAD. These differences are striking and have become a focal point of scientific inquiry, highlighting the need for disease prevention and treatment strategies that are tailored to sex-specific characteristics [3]. To fully understand CAD, it is essential to examine the influence of sex on its pathogenesis, especially regarding atherogenesis. Studies have identified hormonal, genetic, and lifestyle factors as contributors to the distinct patterns of disease progression observed between men and women (Fig. 1).

Objectives

This review was conducted to examine the intricacies of sex differences in coronary

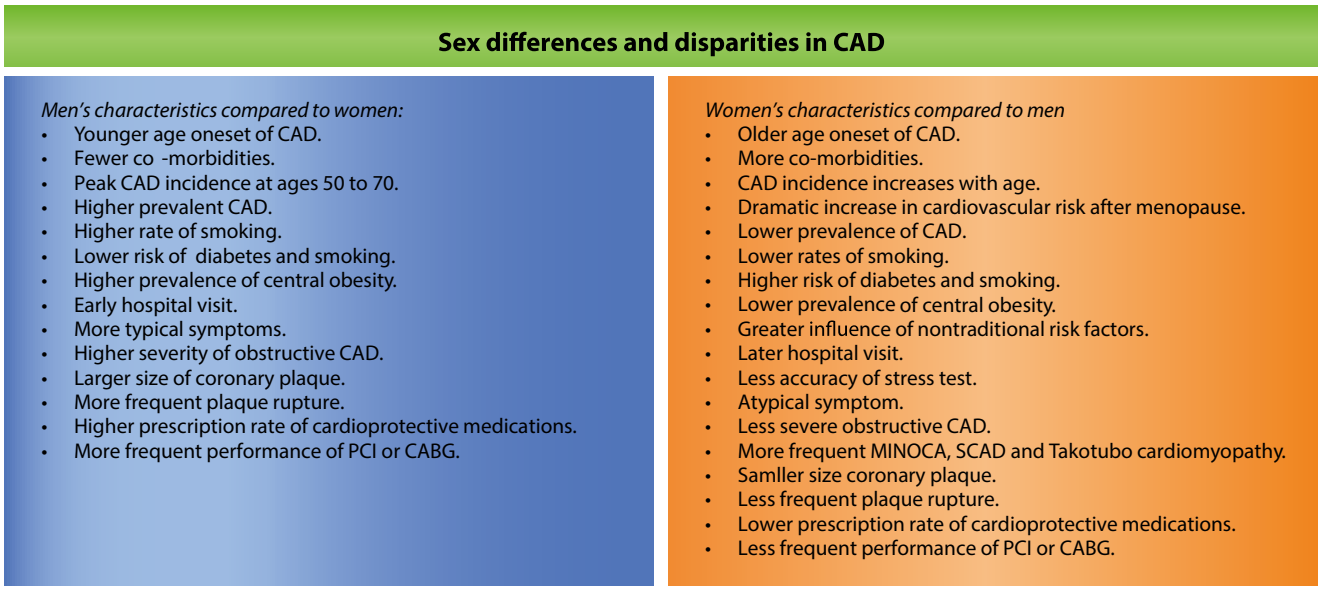


Fig. 1. Overview of sex differences and disparities in CAD. CAD, coronary artery disease; PCI, percutaneous coronary intervention; CABG, coronary artery bypass surgery; MINOCA, myocardial infarction with non-obstructive coronary arteries; SCAD, spontaneous coronary artery dissection.

atherogenesis, providing a foundation for more effective and personalized strategies in managing and preventing CAD. By highlighting the specific needs and risks associated with each sex, healthcare providers can better address the unique challenges presented by CAD, ultimately leading to improved outcomes and a reduction in disease incidence [4].

Ethics statement

It is a literature database-based review; therefore, neither approval by the institutional review board nor obtainment of informed consent was required.

Sex differences in coronary artery disease burden

Traditionally, men have exhibited a higher prevalence and incidence of CAD compared to women. Data from the United States indicate that among all adults aged 20 years and older, the prevalence of cardiovascular disease (CVD)—which includes CAD, stroke, and heart failure—was 10.9% for men and 9.2% for women between 2017 and 2020 [1]. The age-adjusted CAD prevalence in the United States in 2018 was 7.4% for men and 4.1% for women, according to the Centers for Disease Control and Prevention. In a study of 141,459 Chinese individuals, only 31.4% of those who underwent coronary angiography (CAG) for suspected CAD were women, and among those who received percutaneous coronary intervention (PCI), just 22.2% were women [5]. A Korean study that analyzed 633,907 patients hospitalized for acute myocardial infarction (AMI) between 2002 and 2018 according to the Korean National Health Insurance Claims Database found that women accounted for 40% of the cases [6]. Several studies examining CAG findings have reported that obstructive lesions are more common and more severe in men than in women [7]. While women typically experience their first cardiovascular events later in life, the sex difference in CAD prevalence and incidence diminishes with age [1]. A study of

American adults from 2008 to 2017 found that the average age of CAD onset was 57.4 years for men and 59.3 years for women [8]. Similarly, analyses of Korean patients who underwent CAG for suspected CAD or PCI showed that the average age of female patients was older than that of male patients [7,9]. According to 2021 data from the Korean Statistical Information Service, men represented the majority of overall AMI cases. However, the sex gap narrowed with age, and among individuals over 80 years old, the number of AMI cases in women was greater than that in men. In American cohort analysis data, the incidence of CAD, which was significantly higher in younger men, narrowed or even became similar to that in women aged 75 or older [10].

Role of estrogen in coronary atherogenesis

The risk of developing CAD is high for men in their 40s to 70s. In contrast, for women, the risk increases gradually with age and rises steeply from their 50s and 60s [11]. Based on analysis of the Framingham cohort, in women, 40.3% of the impact of age on CVD was found to stem from associated risk factors, greatly exceeding the 11.9% observed in men [12]. According to data from Korea's 2013 National Health Insurance Service, men represented the majority of patients with CVD until their 50s. However, this trend reversed in the 60–69-year range, and among those in their 80s or older, 72.2% of cases were found in women and 27.8% in men.

The increased risk of CAD in older women is closely related to the decrease in systemic estrogen levels. Research suggests that the number of AMIs is negatively associated with lifetime exposure to endogenous estrogen [13], while the later onset of menopause is linked to increased life expectancy [14]. Although the exact mechanisms are not yet fully understood, estrogen exerts several beneficial effects on the cardiovascular system. Notably, this hormone promotes the relaxation of blood vessel walls, leading to vasodilation [15]. This dilation helps to lower blood pressure and improve blood flow, thereby reducing the cardiac workload. Additionally, estrogen has anti-inflammatory properties that can reduce inflammation in the arteries and decrease the risk of atherosclerosis [16]. Its antioxidant characteristics enable it to neutralize harmful free radicals, thus reducing oxidative stress and preventing damage to the walls of blood vessels and other cardiovascular tissues [17]. Furthermore, estrogen influences the distribution of body fat. After menopause, declining estrogen levels can lead to an increased risk of abdominal obesity, which in turn promotes insulin resistance and elevates cardiovascular risk [18]. Estrogen also improves cholesterol profile by increasing HDL cholesterol levels and decreasing LDL levels [19]. This action helps minimize the buildup of plaque in the arteries [20]. Additionally, estrogen supports the health and function of the endothelium, the inner lining of blood vessels, which is crucial for regulating blood vessel tone, preventing clot formation, and maintaining vascular health [21]. Overall, estrogen exerts cardioprotective effects by mitigating the risk of atherosclerosis, hypertension, and thrombosis, thereby reducing the likelihood of CVDs such as heart attack and stroke [22,23].

Sex differences in risk factors for coronary atherogenesis

Traditional risk factors

As previously mentioned, age is a primary determinant of coronary atherogenesis. This factor exerts a more pronounced influence in women than in men, particularly among older individuals [12].

In individuals under 65 years old, hypertension is more prevalent in men, whereas among those over 65, the prevalence is higher in women. Women experience a marked increase in systolic

blood pressure after menopause, which can be attributed to factors such as loss of estrogen, atherosclerosis, increased salt sensitivity, decreased nitric oxide levels, and a rise in angiotensin II receptors [24]. This leads to a higher prevalence of isolated systolic hypertension in women, which is a substantial risk factor for cardiovascular complications [24]. Overall, the influence of hypertension on the development of CAD and stroke appears similar between sexes [25].

Studies indicate that diabetes mellitus presents a greater cardiovascular risk for women than for men. Research involving both individuals without diabetes and those with type 2 diabetes revealed a more pronounced increase in the risk of coronary heart disease among women compared to men over a 13-year period [26]. Furthermore, a meta-analysis revealed that relative to men with type 2 diabetes, women with the condition face a 46% higher risk of mortality from coronary heart disease [27]. Another study reported that diabetes doubles the risk of occlusive vascular mortality in men and triples it in women [28]. Consequently, intensified management of cardiovascular risk is crucial for women with diabetes.

Women generally display higher levels of HDL cholesterol, whereas men are more likely to have elevated levels of LDL cholesterol. However, after menopause, women frequently see a rise in LDL cholesterol and a reduction in HDL cholesterol, increasing their risk of CVD [29]. Although sex differences in the effects of dyslipidemia on CVD are anticipated, the available data on this subject are scarce.

Like diabetes, smoking has a greater impact on the incidence of CAD in women than men. A prospective cohort study found that the risk of AMI was 1.43-fold higher in men who smoked, whereas in women, this risk was elevated by 2.24-fold [30]. Another study indicated that female smokers experienced their first AMI earlier than male smokers [31]. To mitigate the risk of CVD, increased attention should be directed toward women who smoke, and smoking cessation education programs should be implemented.

The impact of obesity on the risk of developing CAD is slightly higher in men than in women [32]. This may be attributed to the higher prevalence of abdominal obesity in men, which further elevates the risk of CVD [33]. However, following menopause, the prevalence of abdominal obesity increases in women due to estrogen depletion, with an associated increase in cardiovascular risk [18].

Nontraditional risk factors

Nontraditional risk factors, including chronic inflammation, psychological stress, socioeconomic factors, and reproductive history, are known to influence sex differences in CAD [34].

Chronic inflammation impairs endothelial cell function, amplifies oxidative stress, and promotes vascular damage, which can lead to atherosclerosis [35]. Specifically, rheumatic diseases and autoimmune diseases, which are more prevalent in women, can trigger CAD through chronic inflammation [36].

Chronic stress elevates the secretion of stress hormones such as cortisol and adrenaline, which in turn increases blood pressure, heart rate, and inflammation. These physiological changes contribute to vascular damage and thrombus formation [37]. Although stress can induce CAD in both men and women, its effects are more pronounced in women [38].

Women tend to have lower levels of education and income compared to men, and these socioeconomic factors can contribute to the incidence of CAD. Lower socioeconomic status in women can result in restricted access to healthcare resources and preventive care for cardiovascular health [39].

Pregnancy-related factors, including gestational hypertension, preeclampsia, gestational

diabetes, miscarriage, stillbirth, and low birth weight, are also associated with an increased risk of CAD [25,40]. Consequently, women with these medical histories require more proactive management that extends beyond childbirth [41].

Overall, these nontraditional risk factors are frequently overlooked and are not sufficiently addressed in patient care relative to traditional factors. To improve cardiovascular prognoses among women, efforts must be made to ensure that these nontraditional risk factors are acknowledged and proactively managed in the prevention and treatment of CAD.

Sex differences in coronary plaque

Atherosclerosis, the primary pathology underlying CAD, is an inflammatory process driven by lipids that initiates the development of plaques within arterial walls [35]. Endothelial dysfunction permits the infiltration of LDL particles into the intima, which triggers an inflammatory cascade. Adhesion molecules and cytokines facilitate the recruitment of inflammatory cells such as monocytes, neutrophils, and T cells. Monocytes differentiate into macrophages and form foam cells, which contribute to plaque formation. Vascular smooth muscle cells migrate to the intima and establish a fibrous cap, which is crucial for plaque stability. Phenotypic changes in these smooth muscle cells further stabilize plaques through extracellular matrix production [35]. Endogenous sex hormones substantially influence this process [42]. In women, estrogen reduces the expression of adhesion molecules, the infiltration of neutrophils, and the secretion of pro-inflammatory cytokines, thereby slowing the progression of atherosclerosis. After menopause, the decline in estrogen levels results in the loss of these protective effects. In comparison, men are more likely to experience plaque rupture, leading to thrombus formation due to the rupture of the fibrous cap and the exposure of thrombotic components. Plaques with thin fibrous caps and large lipid cores are particularly vulnerable to rupture. Testosterone appears to increase inflammatory cell infiltration and cytokine secretion, promoting the development of atherosclerotic lesions [42]. Autopsy data from patients after sudden coronary death indicate that men are more susceptible to the formation of blood clots and have a higher incidence of ruptures. In contrast, women are less likely to develop thrombi, and when they do, those thrombi are more likely to be associated with erosions [43]. Another study that analyzed the culprit plaques in patients with myocardial infarction found that sex exerted a greater influence on plaque characteristics than any other clinical feature [44].

Sex differences in clinical characteristics of patients with coronary artery disease

Women with CAD typically present at an older age and with a greater number of comorbidities than men. Notably, women often experience atypical angina symptoms, which can complicate the diagnostic process [45]. Functional ischemic assessments and cardiac enzyme tests are conducted less frequently in women, and these tests tend to be less accurate in detecting CAD than in men. Furthermore, conditions such as myocardial infarction with non-obstructive coronary arteries, spontaneous coronary artery dissection, and Takotsubo cardiomyopathy are more prevalent among women [46]. In addition, women who undergo PCI face a higher risk of bleeding complications than men [9].

Sex disparities in coronary artery disease

In comparison to men, women frequently delay seeking medical attention, often present with atypical symptoms, and experience lower diagnostic accuracy. These factors lead to comparatively late or missed diagnoses and subsequent delays in initiating treatment [47]. Consequently, women are less likely to undergo invasive procedures such as PCI or coronary artery bypass graft surgery compared to men [48]. Furthermore, a sex-based disparity is evident in the prescription rates of cardioprotective drugs. Relative to men, women are less frequently prescribed essential medications such as antiplatelets, renin-angiotensin system blockers, and statins, which are crucial for managing CVD [49]. Moreover, women are significantly underrepresented in clinical research related to CAD [50]. These sex disparities ultimately lead to poorer cardiovascular prognoses among women.

Future directions

As mentioned above, sex differences are evident in the pathophysiology, risk factors, clinical manifestations, and treatment responses of CAD, as well as in diagnostic and therapeutic approaches. However, most clinical studies have not collected data on female-specific risk factors, such as pregnancy history, age at menopause, and polycystic ovary syndrome, and have not included this information in their analyses. Historically, clinical research has been male-dominated, with women frequently excluded due to factors related to fertility, breastfeeding, or menopause [51]. Awareness must be raised of these sex differences and disparities, not only among healthcare professionals but also within the general population. A survey of middle-aged and elderly women in Korea revealed a very low level of awareness regarding heart disease in women [52]. The "Go Red for Women" campaign, initiated by the American Heart Association in 2004, has raised awareness of CVD in women and effectively reduced cardiovascular mortality within this population, highlighting the impact of such awareness campaigns [53,54]. Furthermore, knowledge of sex-based differences must be integrated into clinical guidelines [34]. To conduct sex-specific analyses with sufficient statistical power, the anticipated number of patients must be at least doubled. However, creating separate guidelines for each sex presents several challenges and can be both complex and impractical. A more feasible approach in clinical practice is to address sex differences within a unified set of guidelines. In many developed Western countries, CAD recommendations consistently underscore the importance of considering sex differences. For instance, guidelines from these countries often highlight the need for cardiovascular care for women after childbirth [55]. Clinicians and researchers should be cognizant of these differences and routinely assess them in their clinical work and research endeavors. By establishing guidelines grounded in robust evidence, and by increasing the inclusion of women in randomized trials and conducting comprehensive analyses of sex differences, high-quality evidence can be progressively amassed and reflected in future recommendations [56].

Conclusion

Stark differences exist between men and women in terms of CAD burden, risk factors, plaque characteristics, and clinical features. These distinctions highlight the need for tailored diagnostic and therapeutic approaches for CAD in both sexes. It is also essential to recognize the longer diagnostic timelines, delays in procedures/surgery, and lower rates of medication prescriptions

in women. Addressing these issues requires an increased awareness of sex differences and discrimination, as well as greater inclusion of women in clinical research to gather more complete data on women's cardiovascular health.

ORCID
Hack-Lyoung Kim: https://orcid.org/0000-0002-6703-1472

Authors' contributions
All work was done by Hack-Lyoung Kim.

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Funding
Not applicable.

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Not applicable.

References

1. Tsao CW, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, et al. Heart disease and stroke statistics—2023 update: a report from the American Heart Association. *Circulation* 2023;147(8):e93-e621. https://doi.org/10.1161/CIR.0000000000001123
2. Libby P, Buring JE, Badimon L, Hansson GK, Deanfield J, Bittencourt MS, et al. Atherosclerosis. *Nat Rev Dis Primers* 2019;5(1):56. https://doi.org/10.1038/s41572-019-0106-z
3. Vogel B, Acevedo M, Appelman Y, Noel Bairey Merz C, Chieffo A, Figtree GA, et al. The *Lancet* women and cardiovascular disease commission: reducing the global burden by 2030. *Lancet* 2021;397(10292):2385-2438. https://doi.org/10.1016/S0140-6736(21)00684-X
4. Mosca L, Barrett-Connor E, Wenger NK. Sex/gender differences in cardiovascular disease prevention: what a difference a decade makes. *Circulation* 2011;124(19):2145-2154. https://doi.org/10.1161/CIRCULATIONAHA.110.968792
5. Chen SQ, Liu J, Zhou Y, Huang ZD, Xie Y, Huang HZ, et al. Sex differences in characteristics, treatments, and in-hospital outcomes of patients undergoing coronary angiography or intervention. *Front Cardiovasc Med* 2022;9:878566. https://doi.org/10.3389/fcvm.2022.878566
6. Kim SR, Bae SA, Lee JY, Kim MS, Kim MN, Chung WJ, et al. Gender disparities in prevalence by diagnostic criteria, treatment and mortality of newly diagnosed acute myocardial infarction in Korean adults. *Sci Rep* 2023;13(1):4120. https://doi.org/10.1038/s41598-023-31014-y
7. Kim HL, Kim HJ, Kim M, Park SM, Yoon HJ, Byun YS, et al. Sex differences in coronary angiographic findings in patients with stable chest pain: analysis of data from the KoRean wOmen'S chest pain rEgistry (KoROSE). *Biol Sex Differ* 2022;13(1):2. https://doi.org/10.1186/s13293-021-00411-1
8. Okunrintemi V, Tibuakuu M, Virani SS, Sperling LS, Volgman AS, Gulati M, et al. Sex differences in the age of diagnosis for cardiovascular disease and its risk factors among US adults: trends from 2008 to 2017, the medical expenditure panel survey. *J Am Heart Assoc* 2020;9(24):e018764. https://doi.org/10.1161/JAHA.120.018764
9. Kim HL, Jang JS, Kim MA, Seo JB, Chung WY, Kim SH, et al. Gender differences of in-hospital outcomes in patients undergoing percutaneous coronary intervention in the drug-eluting stent era. *Medicine* 2019;98(20):e15557. https://doi.org/10.1097/MD.0000000000015557
10. Madhavan MV, Gersh BJ, Alexander KP, Granger CB, Stone GW. Coronary artery disease in patients ≥80 years of age. *J Am Coll Cardiol* 2018;71(18):2015-2040. https://doi.org/10.1016/j.jacc.2017.12.068
11. Wenger NK. Coronary heart disease: an older woman's major health risk. *BMJ* 1997;315(7115):1085-1090. https://doi.org/10.1136/bmj.315.7115.1085
12. Kannel WB, Vasan RS. Is age really a non-modifiable cardiovascular risk factor? *Am J Cardiol* 2009;104(9):1307-1310. https://doi.org/10.1016/j.amjcard.2009.06.051

13. Saltiki K, Doukas C, Kanakakis J, Anastasiou E, Mantzou E, Alevizaki M. Severity of cardiovascular disease in women: relation with exposure to endogenous estrogen. *Maturitas* 2006;55(1):51-57. https://doi.org/10.1016/j.maturitas.2005.12.008
14. Ossewaarde ME, Bots ML, Verbeek AL, Peeters PH, van der Graaf Y, Grobbee DE, et al. Age at menopause, cause-specific mortality and total life expectancy. *Epidemiology* 2005;16(4):556-562. https://doi.org/10.1097/01.ede.0000165392.35273.d4
15. Somani YB, Pawelczyk JA, De Souza MJ, Kris-Etherton PM, Proctor DN. Aging women and their endothelium: probing the relative role of estrogen on vasodilator function. *Am J Physiol Heart Circ Physiol* 2019;317(2):H395-H404. https://doi.org/10.1152/ajpheart.00430.2018
16. Straub RH. The complex role of estrogens in inflammation. *Endocr Rev* 2007;28(5):521-574. https://doi.org/10.1210/er.2007-0001
17. Niranjan MK, Koiri RK, Srivastava R. Expression of estrogen receptor alpha in response to stress and estrogen antagonist tamoxifen in the shell gland of *Gallus gallus* domesticus: involvement of anti-oxidant system and estrogen. *Stress* 2021;24(3):261-272. https://doi.org/10.1080/10253890.2019.1710127
18. Janssen I, Powell LH, Kazlauskaite R, Dugan SA. Testosterone and visceral fat in midlife women: the Study of Women's Health Across the Nation (SWAN) fat patterning study. *Obesity* 2010;18(3):604-610. https://doi.org/10.1038/oby.2009.251
19. Barton M. Cholesterol and atherosclerosis: modulation by oestrogen. *Curr Opin Lipidol* 2013;24(3):214-220. https://doi.org/10.1097/MOL.0b013e3283613a94
20. Burke AP, Farb A, Malcom G, Virmani R. Effect of menopause on plaque morphologic characteristics in coronary atherosclerosis. *Am Heart J* 2001;141(2):S58-S62. https://doi.org/10.1067/mhj.2001.109946
21. Taddei S, Virdis A, Ghiadoni L, Mattei P, Sudano I, Bernini G, et al. Menopause is associated with endothelial dysfunction in women. *Hypertension* 1996;28(4):576-582. https://doi.org/10.1161/01.HYP.28.4.576
22. Kim HL. Differences in risk factors for coronary atherosclerosis according to sex. *J Lipid Atheroscler* 2024;13:e12. https://doi.org/10.12997/jla.2024.13.e12
23. Iorga A, Cunningham CM, Moazeni S, Ruffenach G, Umar S, Eghbali M. The protective role of estrogen and estrogen receptors in cardiovascular disease and the controversial use of estrogen therapy. *Biol Sex Differ* 2017;8(1):33. https://doi.org/10.1186/s13293-017-0152-8
24. Brahmbhatt Y, Gupta M, Hamrahian S. Hypertension in premenopausal and postmenopausal women. *Curr Hypertens Rep* 2019;21(10):74. https://doi.org/10.1007/s11906-019-0979-y
25. Appelman Y, van Rijn BB, ten Haaf ME, Boersma E, Peters SAE. Sex differences in cardiovascular risk factors and disease prevention. *Atherosclerosis* 2015;241(1):211-218. https://doi.org/10.1016/j.atherosclerosis.2015.01.027
26. Juutilainen A, Kortelainen S, Lehto S, Rönnemaa T, Pyörälä K, Laakso M. Gender difference in the impact of type 2 diabetes on coronary heart disease risk. *Diabetes Care* 2004;27(12):2898-2904. https://doi.org/10.2337/diacare.27.12.2898
27. Huxley R, Barzi F, Woodward M. Excess risk of fatal coronary heart disease associated with diabetes in men and women: meta-analysis of 37 prospective cohort studies. *BMJ* 2006;332(7533):73-78. https://doi.org/10.1136/bmj.38678.389583.7C
28. Gnatiuc L, Herrington WG, Halsey J, Tuomilehto J, Fang X, Kim HC, et al. Sex-specific relevance of diabetes to occlusive vascular and other mortality: a collaborative meta-analysis of individual data from 980 793 adults from 68 prospective studies. *Lancet Diabetes Endocrinol* 2018;6(7):538-546. https://doi.org/10.1016/S2213-8587(18)30079-2
29. Abbey M, Owen A, Suzakawa M, Roach P, Nestel PJ. Effects of menopause and hormone replacement therapy on plasma lipids, lipoproteins and LDL-receptor activity. *Maturitas* 1999;33(3):259-269. https://doi.org/10.1016/S0378-5122(99)00054-7
30. Prescott E, Hippe M, Schnohr P, Hein HO, Vestbo J. Smoking and risk of myocardial infarction in women and men: longitudinal population study. *BMJ* 1998;316(7137):1043-1047. https://doi.org/10.1136/bmj.316.7137.1043
31. Grundtvig M, Hagen TP, German M, Reikvam A. Sex-based differences in premature first myocardial infarction caused by smoking: twice as many years lost by women as by men. *Eur J Cardiovasc Prev Rehabil* 2009;16(2):174-179. https://doi.org/10.1097/HJR.0b013e328325d7f0
32. Whitlock G, Lewington S, Sherliker P, Clarke R, Emberson J, Halsey J, et al. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet* 2009;373(9669):1083-1096. https://doi.org/10.1016/S0140-6736(09)60318-4
33. Després JP. Body fat distribution and risk of cardiovascular disease: an update. *Circulation* 2012;126(10):1301-1313. https://doi.org/10.1161/CIRCULATIONAHA.111.067264
34. Kim HL, Kim MA. Sex differences in coronary artery disease: insights from the KoRean wOmen'S chest pain rEgistry (KoROSE). *Korean Circ J* 2023;53(10):655-676. https://doi.org/10.4070/kcj.2023.0205
35. Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med* 2005;352(16):1685-1695.

https://doi.org/10.1056/NEJMra043430
36. Mason JC, Libby P. Cardiovascular disease in patients with chronic inflammation: mechanisms underlying premature cardiovascular events in rheumatologic conditions. *Eur Heart J* 2015;36(8):482-489. https://doi.org/10.1093/eurheartj/ehu403
37. Steptoe A, Kivimäki M. Stress and cardiovascular disease. *Nat Rev Cardiol* 2012;9(6):360-370. https://doi.org/10.1038/nrcardio.2012.45
38. Vaccarino V, Sullivan S, Hammadah M, Wilmot K, Al Mheid I, Ramadan R, et al. Mental stress–induced-myocardial ischemia in young patients with recent myocardial infarction: sex differences and mechanisms. *Circulation* 2018;137(8):794-805. https://doi.org/10.1161/CIRCULATIONAHA.117.030849
39. Backholer K, Peters SAE, Bots SH, Peeters A, Huxley RR, Woodward M. Sex differences in the relationship between socioeconomic status and cardiovascular disease: a systematic review and meta-analysis. *J Epidemiol Community Health* 2017;71(6):550-557. https://doi.org/10.1136/jech-2016-207890
40. O'Kelly AC, Michos ED, Shufelt CL, Vermunt JV, Minissian MB, Quesada O, et al. Pregnancy and reproductive risk factors for cardiovascular disease in women. *Circ Res* 2022;130(4):652-672. https://doi.org/10.1161/CIRCRESAHA.121.319895
41. Søndergaard MM, Hlatky MA, Stefanick ML, Vittinghoff E, Nah G, Allison M, et al. Association of adverse pregnancy outcomes with risk of atherosclerotic cardiovascular disease in postmenopausal women. *JAMA Cardiol* 2020;5(12):1390-1398. https://doi.org/10.1001/jamacardio.2020.4097
42. Yerly A, van der Vorst EPC, Baumgartner I, Bernhard SM, Schindewolf M, Döring Y. Sex-specific and hormone-related differences in vascular remodelling in atherosclerosis. *Eur J Clin Invest* 2023;53(1):e13885. https://doi.org/10.1111/eci.13885
43. Yahagi K, Davis HR, Arbustini E, Virmani R. Sex differences in coronary artery disease: pathological observations. *Atherosclerosis* 2015;239(1):260-267. https://doi.org/10.1016/j.atherosclerosis.2015.01.017
44. Arbustini E, Dal Bello B, Morbini P, Burke AP, Bocciarelli M, Specchia G, et al. Plaque erosion is a major substrate for coronary thrombosis in acute myocardial infarction. *Heart* 1999;82(3):269-272. https://doi.org/10.1136/hrt.82.3.269
45. Cho DH, Choi J, Kim MN, Kim HL, Kim YH, Na JO, et al. Gender differences in the presentation of chest pain in obstructive coronary artery disease: results from the Korean Women's Chest Pain Registry. *Korean J Intern Med* 2020;35(3):582-592. https://doi.org/10.3904/kjim.2018.320
46. La S, Beltrame J, Tavella R. Sex-specific and ethnicity-specific differences in MINOCA. *Nat Rev Cardiol* 2024;21(3):192-202. https://doi.org/10.1038/s41569-023-00927-6
47. Kaul P, Armstrong PW, Sookram S, Leung BK, Brass N, Welsh RC. Temporal trends in patient and treatment delay among men and women presenting with ST-elevation myocardial infarction. *Am Heart J* 2011;161(1):91-97. https://doi.org/10.1016/j.ahj.2010.09.016
48. Daly C, Clemens F, Lopez Sendon JL, Tavazzi L, Boersma E, Danchin N, et al. Gender differences in the management and clinical outcome of stable angina. *Circulation* 2006;113(4):490-498. https://doi.org/10.1161/CIRCULATIONAHA.105.561647
49. Zhao M, Woodward M, Vaartjes I, Millett ERC, Klipstein-Grobusch K, Hyun K, et al. Sex differences in cardiovascular medication prescription in primary care: a systematic review and meta-analysis. *J Am Heart Assoc* 2020;9(11):e014742. https://doi.org/10.1161/JAHA.119.014742
50. Agarwala A, Goldberg A. Special considerations for lipid-lowering therapy in women reflecting recent randomized trials. *Curr Atheroscler Rep* 2021;23(8):42. https://doi.org/10.1007/s11883-021-00942-3
51. Whitelaw S, Sullivan K, Eliya Y, Alruwayeh M, Thabane L, Yancy CW, et al. Trial characteristics associated with under-enrolment of females in randomized controlled trials of heart failure with reduced ejection fraction: a systematic review. *Eur J Heart Fail* 2021;23(1):15-24. https://doi.org/10.1002/ejhf.2034
52. Kim HJ, Kim HY, Kim HL, Park SM, Cho DH, Kim M, et al. Awareness of cardiovascular disease among Korean women: results from a nationwide survey. *Prev Med Rep* 2022;26:101698. https://doi.org/10.1016/j.pmedr.2022.101698
53. Mosca L, Hammond G, Mochari-Greenberger H, Towfighi A, Albert MA. Fifteen-year trends in awareness of heart disease in women: results of a 2012 American Heart Association national survey. *Circulation* 2013;127(11):1254-1263. https://doi.org/10.1161/CIR.0b013e318287cf2f
54. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, et al. Heart disease and stroke statistics—2018 update: a report from the American Heart Association. *Circulation* 2018;137(12):e67-e492. https://doi.org/10.1161/CIR.0000000000000558
55. Smith GN, Louis JM, Saade GR. Pregnancy and the postpartum period as an opportunity for cardiovascular risk identification and management. *Obstet Gynecol* 2019;134(4):851-862. https://doi.org/10.1097/AOG.0000000000003363
56. Tannenbaum C, Norris CM, Sean McMurtry M. Sex-specific considerations in guidelines generation and application. *Can J Cardiol* 2019;35(5):598-605. https://doi.org/10.1016/j.cjca.2018.11.011

Review

Ewha Med J 2024;47(2):e16
https://doi.org/10.12771/emj.2024.e16
eISSN 2234-2591

Current status and significance of research on sex differences in neuroscience: a narrative review and bibliometric analysis

Heajin Kim[1], **Heisook Lee**[2,3]

[1]Policy Research Team, Korea Center for Gendered Innovations for Science and Technology Research (GISTeR), Seoul, Korea
[2]Korea Center for Gendered Innovations for Science and Technology Research (GISTeR), Seoul, Korea
[3]Ewha Womans University, Seoul, Korea

Received Feb 29, 2024
Revised Apr 22, 2024
Accepted Apr 24, 2024

Corresponding author
Heajin Kim
Policy Research Team, Korea Center for Gendered Innovations for Science and Technology Research (GISTeR), 22 Teheran-ro 7-gil, Gangnam-gu, Seoul 06130, Korea
E-mail: khj826@gister.re.kr

Keywords
Neuroscience research, Bibliometric analysis; Mental health; Psychiatry; Sex characteristics

This review aims to highlight the importance of research on structural, functional, molecular-biological, and disease-specific sex differences in the brain, and to examine current bibliometric indicators related to research on sex differences. The Web of Science Core Collection was searched for related articles from 2010 to 2023. Structural and functional brain differences according to sex, including variations in communication patterns between hemispheres, may play a role in mental disorders. Sex differences in neurotransmitters such as serotonin, dopamine, and γ-aminobutyric acid contribute to disparities in mental health, addiction, and neurodevelopmental conditions. Neurodevelopmental disorders such as autism spectrum disorder and schizophrenia exhibit sex-based differences in prevalence, symptoms, brain changes, and neurotransmitter disruptions under hormonal influence. There is a growing body of research on depression, adolescence, the hippocampus, the amygdala, and cognition, highlighting the importance of considering sex/gender factors. Recent studies on sex differences in brain diseases have identified variations in brain structure, function, and neurophysiological substances, as well as in hormones and genes between the sexes. The incidence of psychiatric disorders such as autism spectrum disorder, depression, anxiety, and Alzheimer's disease is increasingly being linked to sex differences, and the need for research into the mechanisms underlying these differences is gaining recognition. However, there remains a significant gap in sex-specific neuroscience research related to the diagnosis, treatment, prevention, and management of these conditions. Advancing inclusive research will require comprehensive training, a consensus on methodology, diverse perspectives through collaborative frameworks, governmental/institutional support, and dedicated funding to create suitable research environments and implementation strategies.

Introduction

Background

The latest Global Burden of Disease (GBD) report [1] indicates that the worldwide need for diagnosing and treating mental disorders has surged significantly in recent years. This category includes depressive disorders, anxiety disorders, bipolar disorder, schizophrenia, autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), and neurodegenerative brain disorders. Notably, substantial sex differences have been observed in the prevalence of

depression, anxiety disorders, ASD, and neurodegenerative brain disorders (Fig. 1) [2–6].

Depressive disorders, characterized by changes in mood, interest, energy, sleep, and appetite, occur more than twice as often in women as in men. ASD, a significant brain development disorder marked by difficulties in social communication and interaction, along with repetitive behavior patterns or interests, is approximately four times more prevalent in men than in women [5,7,8]. Alzheimer's disease (AD), a common condition of aging where the brain gradually loses function and disrupts daily living, occurs twice as frequently in women as in men [9–11]. Parkinson's disease (PD), a movement disorder that impairs the ability to control movement, has been reported to be more than twice as common in men as in women [12–14]. Sex differences have been observed in several psychiatric disorders, but the causes of these differences remain largely unknown [4,15,16]. Several factors are thought to be involved, including genetic, biological, and environmental factors, but more research is needed to elucidate these influences.

Sex differences in the incidence of mental disorders serve as both direct and indirect indicators that biological or social factors may predispose different sexes to various brain diseases. This recognition has spurred a growing interest in researching the causes behind these sex differences in mental disorders [1,4,17,18]. For mental disorders with evident sex differences, it is advisable to design and conduct studies specifically aimed at identifying the underlying causes of these disparities. By analyzing the sex/gender-specific characteristics of brain diseases and uncovering the mechanisms behind sex differences, researchers can develop more effective and safer diagnostic and treatment strategies, as well as preventive and rehabilitative measures to enhance mental health.

To date, several biological factors—including variations in brain structure and function, influences of neurotransmitters and hormones, and genetics—as well as sociocultural factors, such as individual experiences and learning, have been suggested as potential explanations for sex differences in mental disorders [19,20]. In light of these factors, research into sex differences in the brain seeks to uncover the physiological and structural distinctions between male and

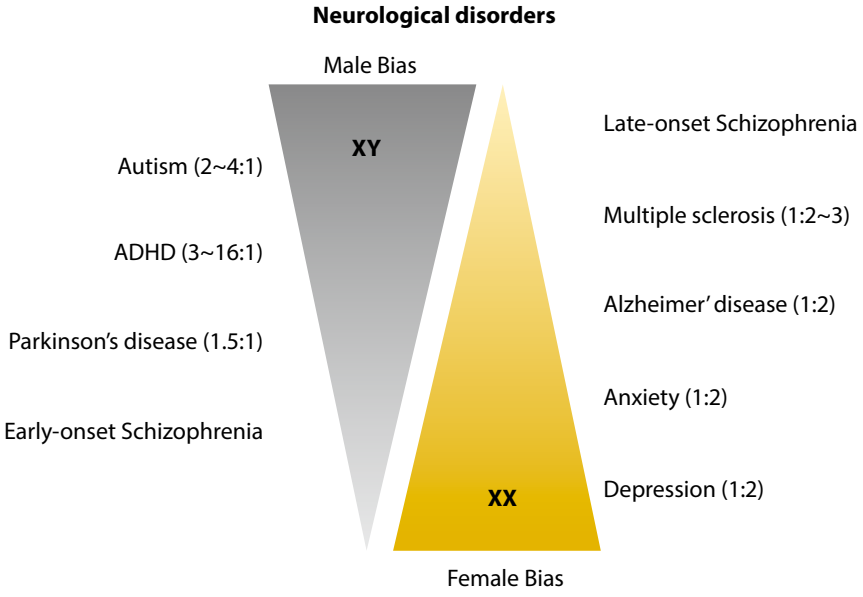


Fig. 1. Several neurological disorders that exhibit sex differences. Male bias (grey), female bias (yellow). ADHD, attention-deficit/hyperactivity disorder.

https://doi.org/10.12771/emj.2024.e16

female brains [21]. The field of research on sex differences in the brain is growing, driven by advances in various technologies such as brain imaging, genetic analysis, neural network studies, big data, and artificial intelligence.

Objectives

This review highlights the importance of research into sex differences in neuroscience. It specifically updates the following areas: structural and functional sex differences in the brain, sex differences in neurotransmitters, sex differences in mental disorders, and bibliometric findings related to sex differences in neuroscience research.

Methods

Ethics statement

This study does not involve human subjects; therefore, neither institutional review board approval nor informed consent was required.

Study design

This study was a narrative review and bibliometric study based on a literature database search.

Literature search/information source and search strategy

The Web of Science Core Collection (Clarivate) was searched for the bibliometric analysis. The authors reviewed the presence of sex/gender-specific keywords in the titles and abstracts of articles and reviews within the field of biological sciences, published from 2010 to 2023. The search utilized the keywords ["sex factor*" OR "sex characteristic*" OR "sex difference*" OR "gender factor*" OR "gender characteristic*" OR "gender difference*"] NOT ["sex* partner*" OR "sex* selection*" OR "sex* behavior*" OR "sex* behavior*"]. More specific search terms are included in Supplement 1. Additionally, to identify the main keywords and major research areas related to sex differences in neuroscience and psychiatry, we analyzed the keyword network using Vos Viewer (https://www.vosviewer.com/).

The associations between article titles and keywords, identified using sex/gender-specific search terms through Vos Viewer, revealed that research primarily focused on three main topics: brain structure and function (fMRI, amygdala, hippocampus, etc.), mental disorders (depression, anxiety, schizophrenia, etc.), and neurotransmitters (dopamine, etc.). The existing literature on these topics, with an emphasis on sex/gender differences, was thoroughly identified and reviewed.

Results

A total of 5,491 articles were identified that discussed structural and functional sex differences in the brain; 4,227 articles addressed sex differences in neurotransmitters, and 14,401 articles explored sex differences in mental disorders.

Structural and functional sex differences in the brain

Recent meta-analyses [18,22] have demonstrated structural sex differences in various brain regions. These differences include the amygdala, hippocampus, temporal lobe, and insular

regions, with men generally having a larger overall brain volume than women. Specific areas such as the left frontal gyrus, left occipital gyrus, left insula, right frontal orbital gyrus, and left occipital sulcus also showed differences. Additionally, variations in white matter regions were observed in the following sequence: midbrain, corpus callosum, right anterior cingulate gyrus, right superior colliculus, and left medial anterior cingulate gyrus [22]. Furthermore, a study on the diffusion-based structural connectome of the brain [23] confirmed that men's brains exhibit more intra-hemispheric communication, whereas women's brains show more inter-hemispheric communication. This study also indicated that brain development in men and women diverges from an early age, leading to structural and functional brain differences in adolescence and adulthood. These structural and functional changes may influence the development of neuropsychiatric disorders, particularly during adolescence—a period when physiological and behavioral differences between the sexes become more pronounced, and the risk of developing neuropsychiatric disorders increases [24]. Notably, activity in the left amygdala significantly increases in women during adolescence, which is partly associated with heightened anxiety.

Although numerous studies have analyzed the structure and function of the brain, significant gaps remain in our understanding of the behavioral and physiological differences between men and women, as well as the specific variations in brain structure and function related to mental disorders. Further research into brain structure, function, and sex-specific symptoms is essential to enhance the diagnosis and treatment of the growing prevalence of mental disorders worldwide.

Sex differences in neurotransmitters

Sex differences have also been reported in neurotransmitters that play a crucial role in regulating brain function, such as serotonin, dopamine, and γ-aminobutyric acid (GABA). Serotonin, a neurotransmitter involved in mood regulation, including depression and anxiety disorders [3,25–28], exhibits sex differences in its expression, role, and receptor [25]. Women have higher levels of serotonin in their blood compared to men [18], and these elevated serotonin levels have been linked to higher levels of estrogen, a sex hormone associated with female reproductive organs [27]. Additionally, women show significantly higher expression of 5-HT1A receptors in various cortical and subcortical brain regions than men [25]. However, the rate of serotonin synthesis is 52% faster in men than in women [26]. This sex difference in serotonin has been proposed as a potential cause for the varying incidence of depression, anxiety disorders, and bipolar disorder between sexes, suggesting that sex-specific treatments and prevention methods warrant further investigation [28].

Sex differences also exist in dopamine, which plays a crucial role in regulating the motor and reward systems [29,30]. Notably, female hormones such as β-estradiol have been shown to enhance the activity of dopamine cells, resulting in increased dopamine release [15]. Additionally, several physiological sex differences have been identified, including variations in the neuroanatomical distribution of dopamine neurons, basal dopamine levels, and the influence of ovarian hormones [30]. Sex differences in dopamine have been suggested to contribute to sex differences in addiction [29,31] and PD [12,18,29], but more detailed research on disease-, brain region-, and behavior-specific mechanisms is needed.

GABA, an inhibitory neurotransmitter that acts as an excitatory neurotransmitter in the developing brain, is regulated by sex hormones, particularly estrogen, during the perinatal period of sensitivity. Researchers have observed sex differences in the volume of certain nuclei and in the frequency and type of synapses in areas such as the hypothalamus, hippocampus,

and preoptic area [32]. Additionally, estrogen has been shown to modulate GABA receptors, controlling their synaptic inhibitory efficacy and leading to differences in signal transduction between the sexes [33]. These findings have led to the hypothesis that sex differences may contribute to conditions such as ASD, ADHD, and epilepsy that develop during this critical period [32,34,35]. However, further research is necessary to fully understand the underlying mechanisms and potential treatments.

Sex differences in mental disorders

Brain developmental disorders

ASD, the most common brain developmental disorder, affects approximately 1 in 36 children; boys are four times more likely to be affected than girls, and their symptoms tend to differ [7]. In a study involving over 2,400 individuals with ASD aged 4–18 years [5], it was found that females with ASD generally exhibited greater impairments in social communication skills, overall IQ, and adaptive functioning. In contrast, males with ASD displayed more prevalent restricted and repetitive behaviors (RRBs). Additionally, males with ASD had larger than normal volumes of grey matter, white matter, and the hippocampus, whereas females with ASD showed smaller volumes in the right hippocampus compared to typical levels [8]. Sex hormones, particularly estrogen, have been demonstrated to influence brain development by affecting the synthesis and receptor expression of GABA, an inhibitory neurotransmitter [7]. Estrogen has also been shown to increase levels of glutamate, a brain-active neurotransmitter, which affects receptor signaling and enhances NMDA receptor expression [34]. Moreover, progesterone can inhibit glutamatergic responses and exhibits sexual dimorphism in certain brain regions. Lower plasma glutamate levels have been observed in individuals with ASD [36].

Schizophrenia is classified as a brain developmental disorder. Research has shown that men are more likely to exhibit positive symptoms such as delusions, hallucinations, and aggression, whereas women tend to develop negative symptoms like depression, anxiety, and social isolation [37]. The typical age of onset varies by sex, occurring between 15 to 25 years in men and 25 to 35 years in women, with another peak occurring after menopause. Changes in brain structure and function are also evident; in men, the prefrontal lobe decreases in size and becomes more asymmetrical than normal, whereas in women, the prefrontal lobe increases in size and shows enhanced white matter connectivity. Furthermore, neurotransmitter activity differs between the sexes, with an overactivity of dopamine in males and glutamate in females [29,37].

Mood disorders

Depressive disorders, including post-traumatic stress disorder, generalized anxiety disorder, and major depressive disorder, are more common in women than in men [1,3,28,38]. Brain imaging analyses of male and female patients with depressive disorders [38] have revealed changes in the size of the hippocampus, amygdala, habenula, anterior cingulate cortex, and corpus callosum. These analyses also showed altered function in the frontal and temporal gyri, caudate nucleus, and prefrontal cortex, as well as microstructural changes in the corpus callosum and its prefrontal projections. Additionally, sex differences in brain circuitry and related systems have been identified [3]. When examining the circuits activated by a stimulus that leads to a behavioral or physiological response, comparisons between the sexes may reveal that the same circuits are involved, but the response may be more intense or prolonged in one sex compared to the other. For instance, corticotropin-releasing factor activates the arousal system

more significantly in women than in men. Alternatively, it may be activated only in one sex; for example, flight stress activates anterior limbic projections to the dorsal thalamus, mediating stress in men but not in women, and sometimes results in completely different behaviors in men and women. For instance, oxytocin activation of oxytocin receptor-containing interneurons in the medial prefrontal cortex induces anxiety in men and altruistic behavior in women. In some cases, the physiological and/or behavioral effects may be the same in both sexes, but there are sex differences in the circuits and mechanisms that produce these effects. For example, when recalling emotional content, the right amygdala is activated in men, whereas the left amygdala is activated in women. While sex differences have been observed in various brain regions and systems associated with depressive disorders, we have yet to identify sex-specific etiologies and mechanisms for treatment, management, and prevention. More focused research on sex differences is essential to unravel the complexities of brain circuitry, hormonal influences, and physiological and behavioral differences.

Neurodegenerative disorders

Neurodegenerative disorders are characterized by the loss of neurons, typically associated with aging. AD, a major neurodegenerative disorder, disproportionately affects women, with incidence rates more than twice those in men. Conversely, PD predominantly affects men, with rates more than twice as high as those in women.

AD, one of the most common neurodegenerative disorders, is a progressive condition that begins with mild memory loss and can progress to a complete inability to interact with others or respond to the environment. This progression occurs as the brain regions responsible for thinking, memory, and language become impaired. Women with AD experience a faster cognitive decline than men, and studies have shown that brain atrophy also occurs more rapidly in women [9,39]. Furthermore, depression, sleep disorders, and stress are risk factors for developing AD. Notably, depression, which is more commonly diagnosed in women, increases the risk of AD. Sleep disorders, which tend to worsen during menopause, contribute to the accumulation of amyloid beta, a protein implicated in AD. Estrogen also plays a significant role in AD. It regulates synaptic plasticity and enhances neural survival. However, the rapid fluctuations in estrogen levels after menopause are linked to an increased risk of brain damage [11].

PD is the second most common neurodegenerative disease after AD and is characterized by severe movement disorders such as bradykinesia, rigidity, tremor, and gait disturbances, which are caused by the loss of midbrain dopaminergic neurons [40]. The incidence of PD is higher in men than in women; however, women experience higher mortality rates and faster disease progression. Additionally, motor and rapid eye movement sleep behavior disorder symptoms are more prevalent in men, and studies have shown that the cortex is thinner in men than in women in the central and pre-central regions. Moreover, men exhibit greater total cortical and subcortical atrophy, and the volumes of the thalamus, caudate, insula, globular bodies, hippocampus, and brainstem are smaller in men [14]. A recent study revealed that cortical thickness varies between male and female patients with PD, influenced by age and disease duration [13]. Specifically, in men with PD, cortical thinning in six frontal lobes (bilateral caudal middle frontal gyrus, bilateral superior frontal gyrus, left frontal gyrus, and right orbitofrontal gyrus), three parietal lobes (bilateral inferior parietal gyrus and left superior parietal gyrus), and one limbic system region (right posterior cingulate gyrus) was associated with longer disease duration and older age. In contrast, in women with PD, only limbic regions showed an association with disease duration.

Bibliometric analysis of sex differences in neuroscience research papers and related keywords

A total of 57,628 articles are included in this bibliometric analysis. As research continues to uncover sex differences in brain structure, function, and neurotransmitters in various brain disorders, the importance of sex/gender-specific research is becoming increasingly recognized. Specifically, understanding the sex-specific mechanisms involved in anxiety, depression, ASD, AD, and PD is crucial for the prevention and treatment of mental disorders and for protecting sex/gender-specific mental health.

Major scientific journals such as *Nature* [41,42], *The Lancet* [43,44] and *Cell* [45] have recognized the importance of integrating sex and gender considerations into research and are actively promoting and disseminating this approach. Concurrently, there has been an increase in the publication of sex/gender-specific research articles. Our search indicated a general rise in these publications (Fig. 2), with neuroscience and psychiatry experiencing a significant uptick since 2018. In contrast, cell biology and biology have shown minimal changes. This trend underscores the widespread sex differences observed in research findings within neuroscience and psychiatry, highlighting the critical need for sex/gender-specific research in these disciplines.

It was found that the keywords related to sex differences were most commonly associated

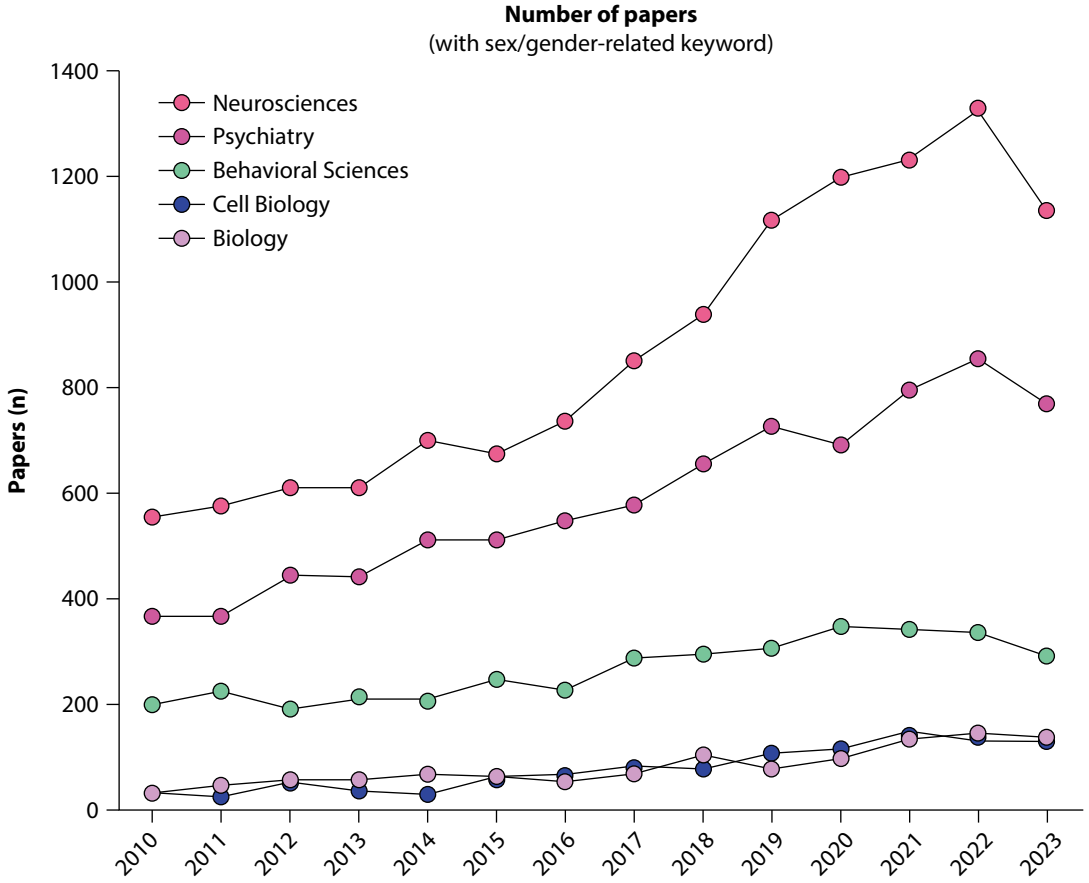


Fig. 2. Trends in sex/gender-specific papers (number of publications) in neuroscience (red), psychiatry (pink), behavioral sciences (green), cell biology (blue), and biology (purple).

https://doi.org/10.12771/emj.2024.e16

with depression, anxiety, adolescence, and stress, and are connected through brain diseases, functions, and research methods (Fig. 3). Specifically, the brain regions of the hippocampus and amygdala were identified as having strong associations with sex/gender-related keywords. Additionally, topics such as age, stress, estrogen, and cognition were found to be closely linked with studies on sex/gender differences.

Discussion

Interpretation and suggestion

In this comprehensive review, we have substantiated the existence of both structural and functional disparities between sexes in the brain, while also elucidating the underlying mechanisms that contribute to sex/gender-specific differences in various disease states. The complex orchestration of perceptual, learning, emotional, cognitive, and behavioral functions across all brain regions highlights the complexity of brain function. Consequently, the identification of sex differences in individual brain regions, along with multifaceted influencing factors, emphasizes the presence of sex/gender-specific differentials in overall brain function.

However, most existing brain research, particularly in animals, has focused on males. This

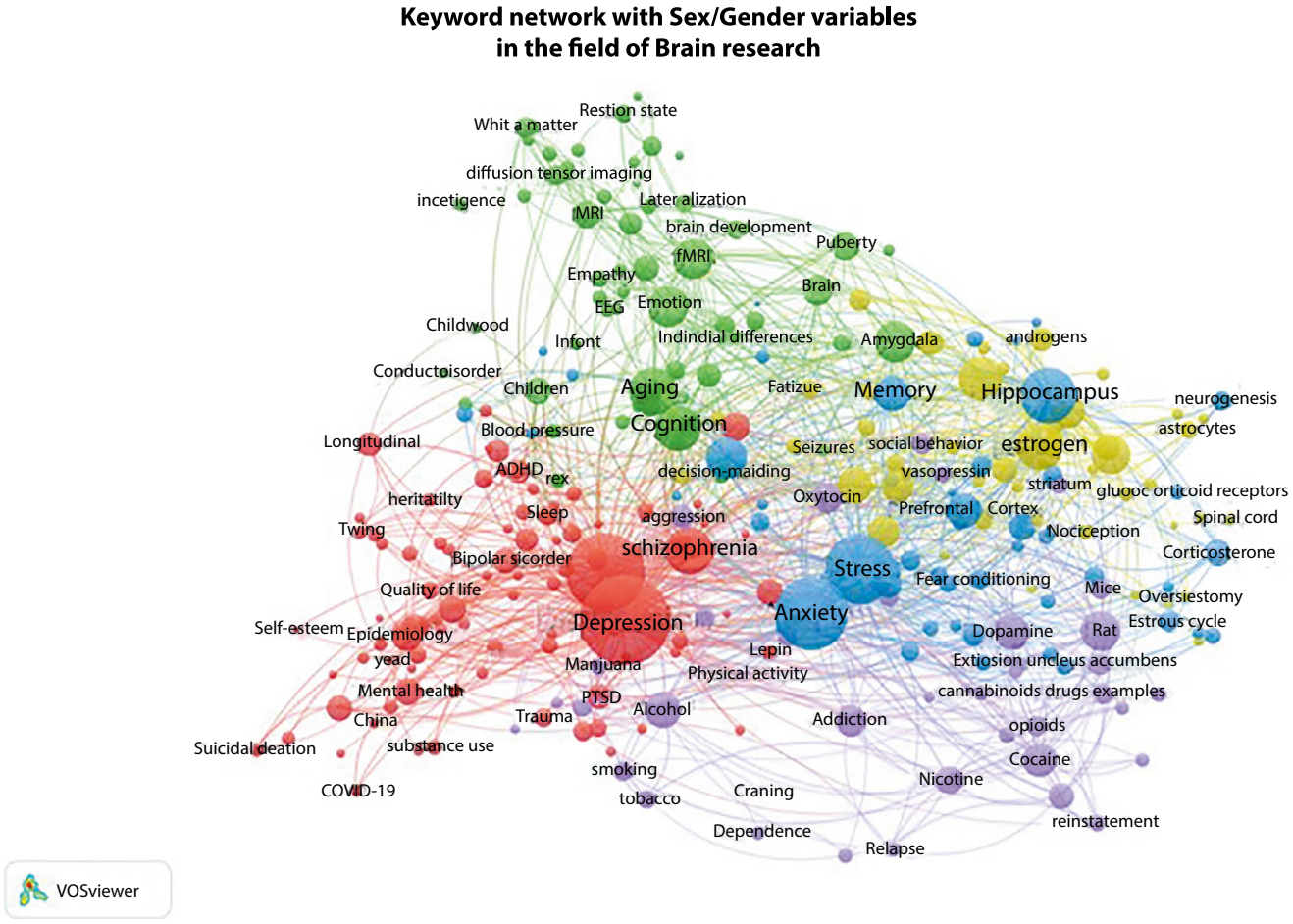


Fig. 3. Keyword co-occurrence networks in brain research. The colors of circles are used to identify the clusters resulting from analyses of the relationships provided by the VOS Viewer software (red: disorders, blue: complex, green: method, purple: addiction, cluster size: frequency).

https://doi.org/10.12771/emj.2024.e16

has resulted in a body of knowledge that is biased and lacks consideration of sex-specific differences. A deficiency in sex/gender-specific research can lead to drug side effects and inefficiencies that vary between sexes/genders. Moreover, generalizing findings can result in the absence of tailored treatments and discrimination, as well as broader health issues due to insufficient knowledge.

Recent research into the fundamental causes of sex differences in brain disorders highlights a growing consensus that sex/gender-specific vulnerabilities and resilience to diseases might originate from the brain's sexual dimorphism. This view is supported by evidence indicating that biological differences, such as hormonal and genetic variations, affect brain structure, function, and neurotransmitter systems during development.

The findings from non-clinical studies on microglia published in 2018 were particularly revealing, as they highlighted sex-dependent disparities in gene transcription, protein synthesis, and cellular function and activity [46,47]. Microglia, small cells in the central nervous system, play a crucial role in managing immune and inflammatory responses and are essential for neuron survival and the formation of neural circuits [48]. Initially observed in the early 2000s in response to estrogen, the detailed differences in microglial function between sexes were only thoroughly described in 2018. For example, it was discovered that female brains typically exhibit weaker inflammatory responses, which may provide increased resistance to certain diseases, while male brains may suffer exacerbated brain damage due to stronger inflammatory reactions [47]. Additionally, male microglia appear to be more reactive and mobile and exhibit a heightened expression of certain proteins compared to their female counterparts, which demonstrate a higher capacity for phagocytosis and gene expression related to cell repair and inflammation management [49]. These functional disparities could potentially explain the varied susceptibility to diseases between sexes, underscoring the need for further research to fully understand how microglial sexual dimorphism contributes to these differences [50].

The importance of understanding biological sex differences in the brain is paramount. These differences are essential for enhancing the effectiveness and personalization of prevention, management, and treatment strategies for sex-specific conditions. Additionally, they provide insights into the intricate interplay of biological, environmental, and developmental factors that influence human behavior and identity. This knowledge also supports ethical and social debates, fostering inclusion and understanding across various sectors of society. Consequently, global neuroscience research initiatives are increasingly concentrating on exploring these mechanisms through the lens of sex differences. Notably, the Women's Brain Project [39,51] serves as a key platform for advocating and disseminating research focused on understanding the numerous factors that affect women's brain health and the pathophysiological mechanisms of mental disorders that are commonly observed in women. This initiative is crucial in addressing the male bias in brain health research and aims to advance precision medicine by establishing a solid foundation of knowledge on sex/gender-specific brain functions.

The author suggests that research on neuroscience and psychiatric disorders should incorporate sex-specific considerations, particularly when studying conditions such as depression, anxiety, stress, adolescence, schizophrenia, and the hippocampus.

Conclusion

The predominance of single-sex studies in research has introduced biases, and a general lack of sex/gender-aware studies has hindered the expansion of informed clinical trials. These trials, which are essential for verifying the safety and efficacy of treatments, must

incorporate sex/gender considerations in their designs and outcomes. The advancement of sex/gender-inclusive research relies on comprehensive training in these concepts, agreement on methodologies, and the integration of diverse perspectives through collaborative frameworks. Additionally, governmental and institutional support, coupled with appropriate funding, is crucial for enhancing research environments and ensuring the thorough implementation of inclusive research strategies.

ORCID
Heajin Kim: https://orcid.org/0009-0004-1144-0359
Heisook Lee: https://orcid.org/0000-0002-5742-9823

Authors' contributions
Project administration: Kim H, Lee H
Conceptualization: Kim H
Methodology & data curation: Kim H, Lee H
Funding acquisition: Lee H
Writing – original draft: Kim H
Writing – review & editing: Kim H, Lee H

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Funding
Korea Center for Gendered Innovations for Science and Technology research (GISTeR), through the Center for Women in Science, Engineering and Technology (WISET) funded by the Ministry of Science and ICT (WISET202403GI01).

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Supplementary materials are available from: https://doi.org/10.12771/emj.2024.e16.

Supplement 1. Search terms used in this study

References

1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry* 2022;9(2):137-150. https://doi.org/10.1016/S2215-0366(21)00395-3
2. Babinski DE. Sex differences in ADHD: review and priorities for future research. *Curr Psychiatry Rep* 2024;26(4):151-156. https://doi.org/10.1007/s11920-024-01492-6
3. Bangasser DA, Cuarenta A. Sex differences in anxiety and depression: circuits and mechanisms. *Nat Rev Neurosci* 2021;22(11):674-684. https://doi.org/10.1038/s41583-021-00513-0
4. Bianco A, Antonacci Y, Liguori M. Sex and gender differences in neurodegenerative diseases: challenges for therapeutic opportunities. *Int J Mol Sci* 2023;24(7):6354. https://doi.org/10.3390/ijms24076354
5. Calderoni S. Sex/gender differences in children with autism spectrum disorder: a brief overview on epidemiology, symptom profile, and neuroanatomy. *J Neurosci Res* 2023;101(5):739-750. https://doi.org/10.1002/jnr.25000
6. Fernández-Artamendi S, Martínez-Loredo V, López-Núñez C. Sex differences in comorbidity between substance use and mental health in adolescents: two sides of the same coin. *Psicothema* 2021;33(1):36-43. https://doi.org/10.7334/psicothema2020.297
7. Ferri SL, Abel T, Brodkin ES. Sex differences in autism spectrum disorder: a review. *Curr Psychiatry Rep* 2018;20(2):1-17. https://doi.org/10.1007/s11920-018-0874-2
8. Napolitano A, Schiavi S, La Rosa P, Rossi-Espagnet MC, Petrillo S, Bottino F, et al. Sex differences in autism spectrum disorder: diagnostic, neurobiological, and behavioral features. *Front Psychiatry* 2022;13:889636.

https://doi.org/10.3389/fpsyt.2022.889636
9. Ferretti MT, Florencia Iulita M, Cavedo E, Andrea Chiesa P, Schumacher Dimech A, Santuccione Chadha A, et al. Sex differences in Alzheimer disease: the gateway to precision medicine. *Nat Rev Neurol* 2018;14(8):457-469. https://doi.org/10.1038/s41582-018-0032-9
10. Pinares-Garcia P, Stratikopoulos M, Zagato A, Loke H, Lee J. Sex: a significant risk factor for neurodevelopmental and neurodegenerative disorders. *Brain Sci* 2018;8(8):154. https://doi.org/10.3390/brainsci8080154
11. Zhu D, Montagne A, Zhao Z. Alzheimer's pathogenic mechanisms and underlying sex difference. *Cell Mol Life Sci* 2021;78(11):4907-4920. https://doi.org/10.1007/s00018-021-03830-w
12. Cerri S, Mus L, Blandini F. Parkinson's disease in women and men: what's the difference? *J Parkinson's Dis* 2019;9(3):501-515. https://doi.org/10.3233/JPD-191683
13. Oltra J, Segura B, Strafella AP, van Eimeren T, Ibarretxe-Bilbao N, Diez-Cirarda M, et al. A multi-site study on sex differences in cortical thickness in non-demented Parkinson's disease. *npj Parkinson's Dis* 2024;10(1):69. https://doi.org/10.1038/s41531-024-00686-2
14. Oltra J, Uribe C, Campabadal A, Inguanzo A, Monté-Rubio GC, Martí MJ, et al. Sex differences in brain and cognition in *de novo* Parkinson's disease. *Front Aging Neurosci* 2021;13:791532. https://doi.org/10.3389/fnagi.2021.791532
15. Health Canada. Health portfolio: sex- and gender-based analysis plus policy: advancing equity, diversity and inclusion [Internet]. Ottawa (ON): CIHR; c2022 [cited 2024 Jan 21]. Available from: https://www.canada.ca/en/health-canada/corporate/transparency/corporate-management-reporting/heath-portfolio-sex-gender-based-analysis-policy.html
16. Arnegard ME, Whitten LA, Hunter C, Clayton JA. Sex as a biological variable: a 5-year progress report and call to action. *J Womens Health* 2020;29(6):858-864. https://doi.org/10.1089/jwh.2019.8247
17. Clayton JA, Collins FS. Policy: NIH to balance sex in cell and animal studies. *Nature* 2014;509(7500):282-283. https://doi.org/10.1038/509282a
18. Cosgrove KP, Mazure CM, Staley JK. Evolving knowledge of sex differences in brain structure, function, and chemistry. *Biol Psychiatry* 2007;62(8):847-855. https://doi.org/10.1016/j.biopsych.2007.03.001
19. McCarthy MM, Pickett LA, VanRyzin JW, Kight KE. Surprising origins of sex differences in the brain. *Horm Behav* 2015;76:3-10. https://doi.org/10.1016/j.yhbeh.2015.04.013
20. van Eijk L, Zhu D, Couvy-Duchesne B, Strike LT, Lee AJ, Hansell NK, et al. Are sex differences in human brain structure associated with sex differences in behavior? *Psychol Sci* 2021;32(8):1183-1197. https://doi.org/10.1177/0956797621996664
21. Rechlin RK, Splinter TFL, Hodges TE, Albert AY, Galea LAM. An analysis of neuroscience and psychiatry papers published from 2009 and 2019 outlines opportunities for increasing discovery of sex differences. *Nat Commun* 2022;13(1):2137. https://doi.org/10.1038/s41467-022-29903-3
22. Ruigrok ANV, Salimi-Khorshidi G, Lai MC, Baron-Cohen S, Lombardo MV, Tait RJ, et al. A meta-analysis of sex differences in human brain structure. *Neurosci Biobehav Rev* 2014;39(100):34-50. https://doi.org/10.1016/j.neubiorev.2013.12.004
23. Ingalhalikar M, Smith A, Parker D, Satterthwaite TD, Elliott MA, Ruparel K, et al. Sex differences in the structural connectome of the human brain. *Proc Natl Acad Sci USA* 2014;111(2):823-828. https://doi.org/10.1073/pnas.1316909110
24. Kaczkurkin AN, Raznahan A, Satterthwaite TD. Sex differences in the developing brain: insights from multimodal neuroimaging. *Neuropsychopharmacology* 2019;44(1):71-85. https://doi.org/10.1038/s41386-018-0111-z
25. Jovanovic H, Lundberg J, Karlsson P, Cerin Å, Saijo T, Varrone A, et al. Sex differences in the serotonin 1A receptor and serotonin transporter binding in the human brain measured by PET. *Neuroimage* 2008;39(3):1408-1419. https://doi.org/10.1016/j.neuroimage.2007.10.016
26. Nishizawa S, Benkelfat C, Young SN, Leyton M, Mzengeza S, de Montigny C, et al. Differences between males and females in rates of serotonin synthesis in human brain. *Proc Natl Acad Sci USA* 1997;94(10):5308-5313. https://doi.org/10.1073/pnas.94.10.5308
27. Rubinow DR, Schmidt PJ, Roca CA. Estrogen–serotonin interactions: implications for affective regulation. *Biol Psychiatry* 1998;44(9):839-850. https://doi.org/10.1016/S0006-3223(98)00162-0
28. Moncrieff J, Cooper RE, Stockmann T, Amendola S, Hengartner MP, Horowitz MA. The serotonin theory of depression: a systematic umbrella review of the evidence. *Mol Psychiatry* 2023;28(8):3243-3256. https://doi.org/10.1038/s41380-022-01661-0
29. Williams OOF, Coppolino M, George SR, Perreault ML. Sex differences in dopamine receptors and relevance to neuropsychiatric disorders. *Brain Sci* 2021;11(9):1199. https://doi.org/10.3390/brainsci11091199
30. Zachry JE, Nolan SO, Brady LJ, Kelly SJ, Siciliano CA, Calipari ES. Sex differences in dopamine release regulation in the striatum. *Neuropsychopharmacology* 2021;46(3):491-499. https://doi.org/10.1038/s41386-020-00915-1

31. Gabel F, Hovhannisyan V, Berkati AK, Andry V, Goumon Y. Sex differences in neurotransmitter levels in different brain regions after acute and chronic morphine treatment in mice [Internet]. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory; c2023 [cited 2024 Jan 10]. Available from: https://www.biorxiv.org/content/10.1101/2023.01.16.524193v1.full.pdf
32. McCarthy MM, Auger AP, Perrot-Sinal TS. Getting excited about GABA and sex differences in the brain. *Trends Neurosci* 2002;25(6):307-312. https://doi.org/10.1016/S0166-2236(02)02182-3
33. Mukherjee J, Cardarelli RA, Cantaut-Belarif Y, Deeb TZ, Srivastava DP, Tyagarajan SK, et al. Estradiol modulates the efficacy of synaptic inhibition by decreasing the dwell time of GABAA receptors at inhibitory synapses. *Proc Natl Acad Sci USA* 2017;114(44):11763-11768. https://doi.org/10.1073/pnas.1705075114
34. Horder J, Petrinovic MM, Mendez MA, Bruns A, Takumi T, Spooren W, et al. Glutamate and GABA in autism spectrum disorder: a translational magnetic resonance spectroscopy study in man and rodent models. *Transl Psychiatry* 2018;8(1):106. https://doi.org/10.1038/s41398-018-0155-1
35. Zhao H, Mao X, Zhu C, Zou X, Peng F, Yang W, et al. GABAergic system dysfunction in autism spectrum disorders. *Front Cell Dev Biol* 2021;9:781327. https://doi.org/10.3389/fcell.2021.781327
36. Farkas I, Bálint F, Farkas E, Vastagh C, Fekete C, Liposits Z. Estradiol increases glutamate and GABA neurotransmission into GnRH neurons via retrograde no-signaling in proestrous mice during the positive estradiol feedback period. *eNeuro* 2018;5(4):ENEURO.0057-18.2018. https://doi.org/10.1523/ENEURO.0057-18.2018
37. Mendrek A, Mancini-Marïe A. Sex/gender differences in the brain and cognition in schizophrenia. *Neurosci Biobehav Rev* 2016;67:57-78. https://doi.org/10.1016/j.neubiorev.2015.10.013
38. Mohammadi S, Seyedmirzaei H, Amin Salehi M, Jahanshahi A, Sina Zakavi S, Dehghani Firouzabadi F, et al. Brain-based sex differences in depression: a systematic review of neuroimaging studies. *Brain Imaging Behav* 2023;17(5):541-569. https://doi.org/10.1007/s11682-023-00772-8
39. Castro-Aldrete L, Moser MV, Putignano G, Ferretti MT, Schumacher Dimech A, Santuccione Chadha A. Sex and gender considerations in Alzheimer's disease: the Women's Brain Project contribution. *Front Aging Neurosci* 2023;15:1105620. https://doi.org/10.3389/fnagi.2023.1105620
40. Kodama L, Gan L. Do microglial sex differences contribute to sex differences in neurodegenerative diseases? *Trends Mol Med* 2019;25(9):741-749. https://doi.org/10.1016/j.molmed.2019.05.001
41. Accounting for sex and gender makes for better science. *Nature* 2020;588(7837):196. https://doi.org/10.1038/d41586-020-03459-y
42. Nature journals raise the bar on sex and gender reporting in research. *Nature* 2022;605(7910):396. https://doi.org/10.1038/d41586-022-01218-9
43. A broader vision for women's health. *Lancet* 2023;402(10399):347. https://doi.org/10.1016/S0140-6736(23)01570-2
44. The gendered dimensions of COVID-19. *Lancet* 2020;395(10231):1168. https://doi.org/10.1016/S0140-6736(20)30823-0
45. Sweet DJ. New at cell press: the inclusion and diversity statement. *Cell* 2021;184(1):1-2. https://doi.org/10.1016/j.cell.2020.12.019
46. Guneykaya D, Ivanov A, Perez Hernandez D, Beule D, Kettenmann H, Wolf SA, et al. Transcriptional and translational differences of microglia from male and female brains. *Cell Rep* 2018;24(10):2773-2783.E6. https://doi.org/10.1016/j.celrep.2018.08.001
47. Villa A, Gelosa P, Castiglioni L, Cimino M, Rizzi N, Pepe G, et al. Sex-specific features of microglia from adult mice. *Cell Rep* 2018;23(12):3501-3511. https://doi.org/10.1016/j.celrep.2018.05.048
48. Prinz M, Jung S, Priller J. Microglia biology: one century of evolving concepts. *Cell* 2019;179(2):292-311. https://doi.org/10.1016/j.cell.2019.08.053
49. Han J, Fan Y, Zhou K, Blomgren K, Harris RA. Uncovering sex differences of rodent microglia. *J Neuroinflammation* 2021;18(1):74. https://doi.org/10.1186/s12974-021-02124-z
50. Lynch MA. Exploring sex-related differences in microglia may be a game-changer in precision medicine. *Front Aging Neurosci* 2022;14:868448. https://doi.org/10.3389/fnagi.2022.868448
51. Schumacher Dimech A, Ferretti MT, Sandset EC, Santuccione Chadha A. The role of sex and gender differences in precision medicine: the work of the Women's Brain Project. *Eur Heart J* 2021;42(34):3215-3217. https://doi.org/10.1093/eurheartj/ehab297

Review

Ewha Med J 2024;47(2):e17
https://doi.org/10.12771/emj.2024.e17
eISSN 2234-2591

Sex differences in metabolic dysfunction-associated steatotic liver disease: a narrative review

Sae Kyung Joo, Won Kim

Division of Gastroenterology and Hepatology, Department of Internal Medicine, Seoul Metropolitan Government Boramae Medical Center, Seoul National University College of Medicine , Seoul, Korea

Received Mar 20, 2024
Revised Apr 22, 2024
Accepted Apr 22, 2024

Corresponding author
Won Kim
Division of Gastroenterology and Hepatology, Department of Internal Medicine, Seoul Metropolitan Government Boramae Medical Center, Seoul National University College of Medicine, 20, Boramae-ro 5 gil, Dongjak-gu, Seoul 07061, Korea
E-mail: drwon1@snu.ac.kr

Keywords
Gastrointestinal microbiome; Hepatocellular carcinoma; Liver cirrhosis; Postmenopause; Sex characteristics

Understanding the effects of sex and sex differences on liver health and disease is crucial for individualized healthcare and informed decision-making for patients with liver disease. The impact of sex on liver disease varies according to its etiology. Women have a lower prevalence of metabolic dysfunction-associated steatotic liver disease (MASLD) than men. However, postmenopausal women face a higher risk of advanced liver fibrosis due to hormonal influences. Sex differences affect the pathogenesis of MASLD, which involves a complex process involving several factors such as hormones, obesity, and the gut microbiome. Furthermore, sex-related differences in the development of MASLD-related hepatocellular carcinoma have been observed. The sex-specific characteristics of MASLD necessitate an individualized management approach based on scientific evidence. However, research in this area has been lacking. This article reviews the current understanding of sex differences in MASLD.

Introduction

Background

There has been growing interest in sex differences in medical conditions both in Korea and around the world [1,2]. Additionally, Korea has established its first institute focusing on sex differences in medicine, mirroring a global trend towards increased awareness of these differences.

Sex-specific medicine strives to deliver optimal personalized care for both men and women, grounded in scientific evidence. Physiological differences between the sexes, including hormone levels and fat distribution, along with variations in social and cultural factors such as dietary habits and physical activity, can affect the onset and progression of liver disease. Consequently, it is crucial for systematic research to concentrate on exploring the sex-specific differences in disease development.

Metabolic dysfunction-associated steatotic liver disease (MASLD) is a significant health concern, affecting roughly one-third of the global population. Its prevalence differs by sex and reproductive status [3–5]. The nature of MASLD as a sex-differentiated disease necessitates an individualized management approach based on scientific evidence. However, research on this topic has been limited.

Objectives

This article aimed to review sex differences in the epidemiology and pathophysiology of MASLD.

Methods

Ethics statement

Neither approval by the institutional review board nor obtainment of informed consent was required, since this was a literature-based study.

Identifying the literature

A comprehensive literature search was conducted in March 2024 using the PubMed databases to identify relevant studies. The search keywords included "sex differences," "sex characteristics," "estrogen," "postmenopause," "gastrointestinal microbiome," "gut-liver axis," "hepatocellular carcinoma," "liver cirrhosis," "nonalcoholic fatty liver disease," "metabolic dysfunction-associated steatotic disease," "nonalcoholic steatohepatitis," and "metabolic dysfunction-associated steatohepatitis." These keywords were used individually or in combination. The initial search yielded published reports from 2012 to August 2023. From this extensive list of search results, studies that met the following criteria were included in this review: published after 2015, review articles, and articles on MASLD or MASLD-related epidemiology and pathophysiology. Initially, the type of study was reviewed, followed by a screening of abstracts to identify suitable studies. Ultimately, ten studies were included in this review and subjected to a comprehensive evaluation in terms of the epidemiology and pathophysiology of MASLD (Supplement 1).

Epidemiology

MASLD is caused by the excessive accumulation of fat in the liver. Differences in the distribution of adipose tissue and various associated mechanisms between sexes are linked to variations in MASLD, which subsequently influence its epidemiology, risk factors, complications, and treatment [6]. The prevalence of MASLD varies among studies [4]. Studies have shown that MASLD is more prevalent in men than in women during the premenopausal stage; however, its prevalence is higher in postmenopausal women [7–11]. A recent meta-analysis revealed a lower prevalence of MASLD among women; however, no significant sex differences were observed in metabolic dysfunction-associated steatohepatitis (MASH) [12]. Advanced liver fibrosis is more prevalent in women, particularly those in the postmenopausal stage [6,13,14]. The lower incidence of MASLD in postmenopausal women has been associated with the use of hormone replacement therapy [15]. Furthermore, a multinational study of histologically confirmed MASLD and advanced liver fibrosis reported poorer survival and a higher incidence of hepatocellular carcinoma (HCC) among older individuals and men, suggesting that estrogen may have a protective effect against MASLD progression.

Pathophysiology, adiposity, and estrogen

Obesity manifests differently in men and women. Men generally accumulate more visceral fat within the abdominal cavity, leading to upper body or apple-shaped obesity. Visceral fat is known for its high lipolysis rate and an inflammatory adipokine profile. Anatomically, this type of fat drains directly into the hepatic portal vein, which results in the liver being exposed to elevated levels of lipids and inflammatory adipokines. In contrast, premenopausal women tend

to accumulate more subcutaneous fat around the lower body and hips, resulting in lower body or pear-shaped obesity [6]. This subcutaneous fat is characterized by a lower rate of lipolysis, a higher capacity for fat storage, and increased potential for fat browning, and it is associated with the release of adiponectin, which protects against metabolic syndrome and MASLD [16]. However, the redistribution of fat following menopause heightens the risk of MASLD [7,17]. Moreover, metabolic syndrome is more common among both men and women in the postmenopausal stage than in premenopausal women [18].

Estrogen influences the interactions between the liver and adipose tissue; consequently, women generally have a higher percentage of body fat than men. However, women tend to accumulate a lower ratio of visceral fat to subcutaneous fat. Research indicates that the expandability and browning capacity of adipose tissue are more pronounced in women than in men, which helps to reduce the metabolic load on the liver. Women exhibit higher blood levels of adiponectin and leptin, along with increased expression and activation of downstream adiponectin signaling elements such as AMP-activated protein kinase, peroxisome proliferator-activated receptor-α, and peroxisome proliferator-activated receptor-γ coactivator-1 α. Additionally, women are shielded from intrahepatic fat accumulation due to enhanced mitochondrial biosynthesis and heightened fatty acid oxidation. Moreover, women have higher expression of antioxidant enzymes compared to men, leading to reduced oxidative stress and preventing the continuous activation of c-Jun N-terminal kinase in response to various stimuli, including fatty acids and pro-inflammatory cytokines. In contrast, men are more prone to sustained activation of c-Jun N-terminal kinase, which can lead to insulin resistance and liver damage through apoptotic necrosis. Although the expression of fibroblast growth factor 21 is stimulated by peroxisome proliferator-activated receptor-γ, no differences in blood levels have been noted between sexes in humans. Fibroblast growth factor 21 primarily affects adipose tissue, promoting glucose uptake, fat browning, and adiponectin expression. Levels of intrahepatic cytokines, such as retinol-binding protein 4 and certain angiopoietin-like isoforms, are elevated in men [16].

Additionally, estrogen plays a role in the development of MASLD in premenopausal women by suppressing the expression of adipogenesis-related genes in hepatocytes, inhibiting the release of inflammatory cytokines from Kupffer cells, and reducing the expression of fibrosis-related genes in hepatic stellate cells [19]. Summarizing these findings, the epidemiology and pathophysiology of MASLD are influenced by age and hormonal changes during the premenopausal and postmenopausal stages [20]. For example, early menarche may heighten the risk of MASLD in adulthood, a risk partially mediated by excessive obesity. Ovarian aging, due to estrogen deficiency, eventually leads to the progression of hepatic steatosis and liver fibrosis through metabolic dysregulation. This metabolic dysregulation also leads to type 2 diabetes, hypertriglyceridemia, and visceral obesity, which are commonly observed post-menopause. Thus, sex-based differences in adiposity and other metabolic risk factors contribute to variations in disease progression based on sex.

Microbiome and bile acids

It is well established that alterations in gut microbiota and bile acids contribute to the development of MASLD, MASH, and HCC [21,22]. In a healthy gut, the microbiome provides nutrients and energy, protects against cancer, inhibits pathogens, and supports normal gastrointestinal immune functions and bowel movements [23]. However, when the gut microbiota is disrupted, bacterial metabolites and commensal components compromise

intestinal epithelial integrity and facilitate access to the liver via the portal vein [24]. These byproducts of the microbiome contribute to inflammation, intrahepatic steatosis, liver injury, and, ultimately, MASLD and MASH [25].

The gut microbiota regulates the gut-liver axis via farnesoid X receptor signaling. This signaling pathway releases fibroblast growth factor 15 and fibroblast growth factor 19, which modulate bile acid synthesis, lipid metabolism, and glucose metabolism. Bile acids, products of cholesterol metabolism, are secreted into the intestine through the biliary tree and regulate energy homeostasis through hepatic and extrahepatic metabolism [26].

Scientific evidence indicates that various factors, including age, hormones, ethnicity, diet, antibiotics, stress, and physical activity, influence the diversity and composition of the gut microbiota [27]. The gut microbiome evolves in response to age-related changes in sex hormones. Clinical studies have shown body mass index-specific sex differences and dimorphism in the gut microbiota associated with the menopausal stage, highlighting a strong connection between the gut microbiota and sex hormones [28–30]. However, further research is needed to explore the effects of the gut microbiome and bile acids on MASLD related to sex differences.

Metabolic dysfunction-associated steatotic liver disease with hepatocellular carcinoma

HCC is more prevalent in men than in women, regardless of its etiology [31]. A large study of patients with MASH and cirrhosis revealed that the incidence of HCC in men was two to seven times higher than in women [32,33]. Women have a higher survival rate associated with HCC until the age of 55 years; after this age, the trend reverses [34]. Chronic injury and inflammation are well-known precursors to HCC. Additionally, the incidence of HCC in patients with MASH increases alongside the incidence of liver fibrosis. A cross-sectional study involving 87 patients with MASH indicated that men are more likely to develop HCC at earlier stages of liver fibrosis compared to women [35]. Therefore, this study suggests that there are sex differences in the development of MASH-related HCC.

Suggestions for future research

Sex and sex hormones play crucial roles in biological differences in MASLD. Despite evident sex differences in the mechanisms of MASLD, the development of tailored treatments is hindered by a lack of sufficient evidence. Therefore, it is vital to consider potential sex or hormonal influences in population-level analyses. Additional research, encompassing preclinical studies, epidemiological surveys, and clinical trials, is necessary to investigate how sex differences and reproductive status influence disease risk in women with MASLD.

Conclusion

The prevalence of MASLD is rapidly increasing worldwide. Furthermore, MASLD can progress to MASH and cirrhosis, with a rising incidence of MASLD-associated HCC. MASLD has clinical significance due to its association with cardiovascular disease and the development of malignant neoplasms. The complex and multifactorial mechanisms underlying MASLD involve factors such as female hormones, adipose tissue distribution, gut microbiota, and bile acids. However, the development of tailored treatments is contingent upon the availability of sufficient evidence. Consequently, studies that take into account variables such as sex, age, and hormonal status could pave the way for evidence-based and personalized clinical treatments that alleviate the burden of MASLD.

ORCID
Sae Kyung Joo: https://orcid.org/0000-0002-4615-7607
Won Kim: https://orcid.org/0000-0002-2926-1007

Authors' contributions
Project administration: Kim W
Conceptualization: Kim W
Methodology & data curation: Joo SK
Funding acquisition: not applicable
Writing – original draft: Joo SK
Writing – review & editing: Joo SK, Kim W

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Funding
Not applicable.

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Supplementary materials are available from: https://doi.org/10.12771/emj.2024.e17.

Supplement 1. List of studies finally included in this review for a comprehensive evaluation in terms of the epidemiology and pathophysiology of metabolic dysfunction-associated steatotic liver disease (MASLD).

References

1. Kim N. Sex/gender-specific medicine in the gastrointestinal diseases. Singapore: Springer; 2022.
2. Kim N. Sex/gender-specific medicine in clinical areas. Singapore: Springer Nature; 2024.
3. Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol* 2018;15(1):11-20. https://doi.org/10.1038/nrgastro.2017.109
4. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease: meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016;64(1):73-84. https://doi.org/10.1002/hep.28431
5. Rinella M, Charlton M. The globalization of nonalcoholic fatty liver disease: prevalence and impact on world health. *Hepatology* 2016;64(1):19-22. https://doi.org/10.1002/hep.28524
6. Lonardo A, Nascimbeni F, Ballestri S, Fairweather D, Win S, Than TA, et al. Sex differences in nonalcoholic fatty liver disease: state of the art and identification of research gaps. *Hepatology* 2019;70(4):1457-1469. https://doi.org/10.1002/hep.30626
7. Park SH, Jeon WK, Kim SH, Kim HJ, Park DI, Cho YK, et al. Prevalence and risk factors of non-alcoholic fatty liver disease among Korean adults. *J Gastroenterol Hepatol* 2006;21(1):138-143. https://doi.org/10.1111/j.1440-1746.2005.04086.x
8. Wong VWS, Chu WCW, Wong GLH, Chan RSM, Chim AML, Ong A, et al. Prevalence of non-alcoholic fatty liver disease and advanced fibrosis in Hong Kong Chinese: a population study using proton-magnetic resonance spectroscopy and transient elastography. *Gut* 2012;61(3):409-415. https://doi.org/10.1136/gutjnl-2011-300342
9. Eguchi Y, Hyogo H, Ono M, Mizuta T, Ono N, Fujimoto K, et al. Prevalence and associated metabolic factors of nonalcoholic fatty liver disease in the general population from 2009 to 2010 in Japan: a multicenter large retrospective study. *J Gastroenterol* 2012;47(5):586-595. https://doi.org/10.1007/s00535-012-0533-z
10. Wang Z, Xu M, Hu Z, Hultström M, Lai E. Sex-specific prevalence of fatty liver disease and associated metabolic factors in Wuhan, south central China. *Eur J Gastroenterol Hepatol* 2014;26(9):1015-1021. https://doi.org/10.1097/MEG.0000000000000151
11. Long MT, Pedley A, Massaro JM, Hoffmann U, Ma J, Loomba R, et al. A simple clinical model predicts incident hepatic steatosis in a community-based cohort: the Framingham Heart Study. *Liver Int* 2018;38(8):1495-1503. https://doi.org/10.1111/liv.13709

12. Balakrishnan M, Patel P, Dunn-Valadez S, Dao C, Khan V, Ali H, et al. Women have a lower risk of nonalcoholic fatty liver disease but a higher risk of progression vs men: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2021;19(1):61-71.E15.
https://doi.org/10.1016/j.cgh.2020.04.067
13. Kim W. Epidemiologic landscape of nonalcoholic fatty liver disease is changed during lifetime by menstrual and reproductive status and sex hormonal factors. *Clin Gastroenterol Hepatol* 2021;19(6):1114-1116.
https://doi.org/10.1016/j.cgh.2020.10.054
14. Yuan L, Kardashian A, Sarkar M. NAFLD in women: unique pathways, biomarkers, and therapeutic opportunities. *Curr Hepatol Rep* 2019;18(4):425-432.
https://doi.org/10.1007/s11901-019-00495-9
15. Clark JM, Brancati FL, Diehl AM. Nonalcoholic fatty liver disease. *Gastroenterology* 2002;122(6):1649-1657.
https://doi.org/10.1053/gast.2002.33573
16. Morán-Costoya A, Proenza AM, Gianotti M, Lladó I, Valle A. Sex differences in nonalcoholic fatty liver disease: estrogen influence on the liver-adipose tissue crosstalk. *Antioxid Redox Signal* 2021;35(9):753-774.
https://doi.org/10.1089/ars.2021.0044
17. Lovejoy JC, Champagne CM, de Jonge L, Xie H, Smith SR. Increased visceral fat and decreased energy expenditure during the menopausal transition. *Int J Obes* 2008;32(6):949-958.
https://doi.org/10.1038/ijo.2008.25
18. Link JC, Reue K. Genetic basis for sex differences in obesity and lipid metabolism. *Annu Rev Nutr* 2017;37:225-245.
https://doi.org/10.1146/annurev-nutr-071816-064827
19. Buzzetti E, Parikh PM, Gerussi A, Tsochatzis E. Gender differences in liver disease and the drug-dose gender gap. *Pharmacol Res* 2017;120:97-108.
https://doi.org/10.1016/j.phrs.2017.03.014
20. Ballestri S, Nascimbeni F, Baldelli E, Marrazzo A, Romagnoli D, Lonardo A. NAFLD as a sexual dimorphic disease: role of gender and reproductive status in the development and progression of nonalcoholic fatty liver disease and inherent cardiovascular risk. *Adv Ther* 2017;34(6):1291-1326.
https://doi.org/10.1007/s12325-017-0556-1
21. Sharpton SR, Ajmera V, Loomba R. Emerging role of the gut microbiome in nonalcoholic fatty liver disease: from composition to function. *Clin Gastroenterol Hepatol* 2019;17(2):296-306.
https://doi.org/10.1016/j.cgh.2018.08.065
22. Chávez-Talavera O, Tailleux A, Lefebvre P, Staels B. Bile acid control of metabolism and inflammation in obesity, type 2 diabetes, dyslipidemia, and nonalcoholic fatty liver disease. *Gastroenterology* 2017;152(7):1679-1694.E3.
https://doi.org/10.1053/j.gastro.2017.01.055
23. Flint HJ, Scott KP, Louis P, Duncan SH. The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* 2012;9(10):577-589.
https://doi.org/10.1038/nrgastro.2012.156
24. Balmer ML, Slack E, de Gottardi A, Lawson MAE, Hapfelmeier S, Miele L, et al. The liver may act as a firewall mediating mutualism between the host and its gut commensal microbiota. *Sci Transl Med* 2014;6(237):237ra66.
https://doi.org/10.1126/scitranslmed.3008618
25. Lang S, Schnabl B. Microbiota and fatty liver disease: the known, the unknown, and the future. *Cell Host Microbe* 2020;28(2):233-244.
https://doi.org/10.1016/j.chom.2020.07.007
26. Arab JP, Karpen SJ, Dawson PA, Arrese M, Trauner M. Bile acids and nonalcoholic fatty liver disease: molecular insights and therapeutic perspectives. *Hepatology* 2017;65(1):350-362.
https://doi.org/10.1002/hep.28709
27. Osadchiy V, Martin CR, Mayer EA. The gut–brain axis and the microbiome: mechanisms and clinical implications. *Clin Gastroenterol Hepatol* 2019;17(2):322-332.
https://doi.org/10.1016/j.cgh.2018.10.002
28. Santos-Marcos JA, Rangel-Zuñiga OA, Jimenez-Lucena R, Quintana-Navarro GM, Garcia-Carpintero S, Malagon MM, et al. Influence of gender and menopausal status on gut microbiota. *Maturitas* 2018;116:43-53.
https://doi.org/10.1016/j.maturitas.2018.07.008
29. Haro C, Rangel-Zúñiga OA, Alcalá-Díaz JF, Gómez-Delgado F, Pérez-Martínez P, Delgado-Lista J, et al. Intestinal microbiota is influenced by gender and body mass index. *PLoS ONE* 2016;11(5):e0154090.
https://doi.org/10.1371/journal.pone.0154090
30. Jašarević E, Morrison KE, Bale TL. Sex differences in the gut microbiome–brain axis across the lifespan. *Philos Trans R Soc Lond B Biol Sci* 2016;371(1688):20150122.
https://doi.org/10.1098/rstb.2015.0122
31. Villa E. Role of estrogen in liver cancer. *Womens Health* 2008;4(1):41-50.
https://doi.org/10.2217/17455057.4.1.41
32. Ascha MS, Hanouneh IA, Lopez R, Tamimi TAR, Feldstein AF, Zein NN. The incidence and risk factors of hepatocellular carcinoma in patients with nonalcoholic steatohepatitis. *Hepatology* 2010;51(6):1972-1978.
https://doi.org/10.1002/hep.23527
33. Vilar-Gomez E, Calzadilla-Bertot L, Wai-Sun Wong V, Castellanos M, Aller-de la Fuente R, Metwally M, et al. Fibrosis severity as a determinant of cause-specific mortality in patients with advanced nonalcoholic fatty liver disease: a multi-national cohort

study. *Gastroenterology* 2018;155(2):443-457.E17.
https://doi.org/10.1053/j.gastro.2018.04.034
34. Yang D, Hanna DL, Usher J, LoCoco J, Chaudhari P, Lenz HJ, et al. Impact of sex on the survival of patients with hepatocellular carcinoma: a surveillance, epidemiology, and end results analysis. *Cancer* 2014;120(23):3707-3716.
https://doi.org/10.1002/cncr.28912
35. Yasui K, Hashimoto E, Komorizono Y, Koike K, Arii S, Imai Y, et al. Characteristics of patients with nonalcoholic steatohepatitis who develop hepatocellular carcinoma. *Clin Gastroenterol Hepatol* 2011;9(5):428-433.
https://doi.org/10.1016/j.cgh.2011.01.023

Review

Ewha Med J 2024;47(2):e18
https://doi.org/10.12771/emj.2024.e18
eISSN 2234-2591

Etiologies underlying sex bias in autism spectrum disorder: a narrative review of preclinical rodent models

Taeyoung Lee[1], **Eunha Kim**[1,2]

[1]BK21 Graduate Program, Department of Biomedical Sciences, Korea University College of Medicine, Seoul, Korea
[2]Department of Neuroscience, Korea University College of Medicine, Seoul, Korea

Received Mar 11, 2024
Revised Apr 22, 2024
Accepted Apr 24, 2024

Corresponding author
Eunha Kim
BK21 Graduate Program, Department of Biomedical Sciences, Korea University College of Medicine, Seoul 02841, Korea
E-mail: Eunha_Kim@korea.ac.kr

Keywords
Autism spectrum disorder; Genetic variation; Pregnancy; Risk factors; Sex characteristics

Neurodevelopmental disorders, which emerge early in development, include a range of neurological phenotypes and exhibit marked differences in prevalence between sexes. A male predominance is particularly pronounced in autism spectrum disorder (ASD). Although the precise cause of ASD is still unknown, certain genetic variations and environmental influences have been implicated as risk factors. Preclinical ASD models have been instrumental in shedding light on the mechanisms behind the sexual dimorphism observed in this disorder. In this review, we explore the potential processes contributing to sex bias by examining both intrinsic differences in neuronal mechanisms and the influence of external factors. We organize these mechanisms into six categories: 1) sexually dimorphic phenotypes in mice with mutations in ASD-associated genes related to synaptic dysfunction; 2) sex-specific microglial activity, which may disrupt neural circuit development by excessively pruning synapses during critical periods; 3) sex steroid hormones, such as testosterone and allopregnanolone, that differentially influence brain structure and function; 4) escape from X chromosome inactivation of the O-linked-N-acetylglucosamine transferase gene in the placenta; 5) sexually dimorphic activation of the integrated stress response pathway following maternal immune activation; and 6) immunological responses that are differentially regulated by sex. Understanding these mechanisms is essential for deciphering the underlying causes of ASD and may offer insights into other disorders with notable sex disparities.

Introduction

Background

Neurodevelopmental disorders (NDDs) are a heterogenous group of conditions that manifest during the developmental period and are characterized by impairments in various aspects of neurological functioning [1]. As defined by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, NDDs include autism spectrum disorder (ASD), intellectual disability/developmental delay, attention-deficit/hyperactivity disorder, motor and tic disorders, and specific language disorders. These conditions can lead to a range of developmental deficits, from specific challenges in learning or executive function management to more extensive impairments in social skills or intellectual abilities [2]. This review is primarily focused on ASD, a NDD marked by deficits in social interaction and engagement in repetitive and stereotyped behaviors [2].

The prevalence, age of onset, pathophysiology, and symptomatology of many NDDs vary substantially by sex. A pronounced male bias is evident in ASD, with a male-to-female prevalence of approximately four to one [3,4]. The potential underdiagnosis of females with ASD has raised concerns, suggesting the need for distinct diagnostic criteria in their assessment [5,6]. Even apart from variations in diagnostic practices, a sex bias persists in the prevalence of ASD, with a male-to-female ratio ranging from at least 2:1 to 3:1. This highlights the critical need to explore the biological basis of sexual dimorphism, which may be key to understanding the processes underlying ASD pathogenesis [7].

Here, we explore the potential mechanisms contributing to the sex-biased prevalence disparity in ASD using various preclinical models (Table 1). While preclinical rodent models of ASD cannot fully replicate the spectrum of human ASD phenotypes [8], potentially due to differences in brain structures and developmental trajectories [9], they remain invaluable for gaining mechanistic

Table 1. Potential contributing mechanisms underlying the sex-biased prevalence of ASD as demonstrated in preclinical rodent models

Potential contributing mechanisms	Preclinical models showing sexually dimorphic ASD-like phenotypes	Suggested mechanism(s)	References
Synaptic dysfunction	*Shank3* KO	Reduced levels of mGluR5 in male mice	[27]
	Chd8$^{+/N2373K}$	Sexually dimorphic changes in neuronal activity, synaptic transmission, and transcriptomic profiles	[40]
	Fmr1 KO	Sexually dimorphic upregulation of ASD risk genes (male↑: *Ctnnb1*[a] and *Grin1*[a], female↑: *Homer1*[a], *Ptgs2* [a], *Drd1*[a], *Pik3ca*[b], and *Csnk1g1*[b])	[44]
Microglial abnormalities	*Cntnap2* KO	Activated morphology and phagocytosis of synaptic structures in male microglia	[58]
	DEP/MS	Hyper-ramified phenotype in male microglia	[63]
Hormones	VPA-induced ASD mouse model	Lower levels of TH expression in the AVPV of male mice	[70]
	Placenta-specific *Akr1c14* KO	Male mouse-specific abnormalities in cerebellar white matter	[75]
Escape from X chromosome inactivation	Prenatal stress model	Placental OGT expression levels are twice as high for female fetuses as for male fetuses; this results in sexually distinct gene expression in trophoblasts through epigenetic modulation by histone methylation	[79,80]
Integrated stress response pathway	MIA (Poly[I:C])	Hyperactivation of the ISR pathway in male MIA offspring, resulting in reduced nascent protein synthesis in the brain	[85]
Immune pathways	Prenatal GBS infection	Heightened levels of pro-inflammatory cytokines and chemokines such as IL-1β and CINC-1/CXCL1 in male fetuses	[98]
	MIA (LPS)	Male MIA offspring exhibit heightened cortical hypoxia, reduced mitosis of radial glial cells, disrupted E/I balance within the brain, severe placental necrosis, elevated inflammation, and reduced placental growth	[99]
	MIA (two-hit model)	The anti-inflammatory cytokines IL-10 and TGF-β1 are decreased in male offspring but increased in female mice	[100]

↑: upregulated, [a]: high-confidence risk genes for ASD, [b]: suggestive risk genes for ASD.

ASD, autism spectrum disorder; *Shank3*, SH3 and multiple ankyrin repeat domains 3; KO, knockout; mGluR5, metabotropic glutamate receptor 5; *Chd8*, chromodomain helicase DNA-binding protein 8; *Fmr1*, fragile X mental retardation 1; *Ctnnb1*, catenin beta 1; *Grin1*, glutamate ionotropic receptor NMDA type subunit 1; *Homer1*, homer scaffold protein 1; *Ptgs2*, prostaglandin-endoperoxide synthase 2; *Drd1*, dopamine receptor D1; *Pik3ca*, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; *Csnk1g1*, casein kinase 1 gamma 1; *Cntnap*2, contactin-associated protein 2; DEP/MS, diesel exhaust particles and maternal stress; VPA, valproic acid; TH, tyrosine hydroxylase; AVPV, anteroventral periventricular nucleus; *Akr1c14*, aldo-keto reductase family 1 member C4; OGT, O-linked-N-acetylglucosamine transferase; MIA, maternal immune activation; poly(I:C), polyinosinic:polycytidylic acid; ISR, integrated stress response; GBS, Group B *Streptococcus*; IL-1β, interleukin 1 beta; CINC-1/CXCL1, cytokine-induced neutrophil chemoattractant-1; LPS, lipopolysaccharide; E/I, excitation/inhibition; IL-10, interleukin 10; TGF-β1, transforming growth factor beta 1.

https://doi.org/10.12771/emj.2024.e18

insights into the pathogenesis of ASD and for exploring potential therapeutic strategies [10,11].

Objectives

This review aims to elucidate the potential contributing mechanisms underlying the sex-biased prevalence of ASD as demonstrated in preclinical rodent models. These include synaptic dysfunction, microglial abnormalities, the influence of sex hormones, escape from X chromosome inactivation, the integrated stress response (ISR) pathway, and immune pathways.

Methods

Ethics statement

The present study is based on a review of the literature; consequently, neither approval from an institutional review board nor the acquisition of informed consent was necessary.

Study design

This study is a narrative review.

Literature search: information sources and search strategies

We searched the PubMed database for articles published from 1990 up to April 2024. Only articles published in English were included.

Results

This review encompasses a total of 104 articles. A list of articles pertaining to each topic is available in the references.

Synaptic dysfunction

Previous studies have reported that those with ASD exhibit different brain connectivity patterns compared to typically developing individuals [12]. Patterns of widespread cortical underconnectivity, local overconnectivity, or a combination of these suggest that disrupted brain connectivity may represent a potential neural signature of ASD [13]. Brain connectivity is largely determined by the characteristics of neurons and synapses, with synapses being highly specialized, asymmetric cell-to-cell junctions that constitute the fundamental units of brain communication [14].

According to the Simons Foundation Autism Research Initiative (SFARI) gene database, hundreds of genes have been identified as being associated with ASD [15–18]. Among these, genes such as those of the SH3 and multiple ankyrin repeat domains (*SHANK*) family, fragile X mental retardation 1 (*FMR1*), and chromodomain helicase DNA-binding protein 8 (*CHD8*) are linked to common cellular pathways that converge at synapses [19–21]. This convergence suggests that synaptic dysfunction may contribute to the development of ASD, potentially leading to functional and cognitive impairments [14].

SHANK, also known as ProSAP, is a family of postsynaptic proteins found at glutamatergic synapses and includes three major isoforms: SHANK1, SHANK2, and SHANK3. These proteins act as master scaffolding proteins at excitatory synapses [22,23]. They interact with over 30 synaptic proteins across multiple domains and are critical for synaptic formation, glutamate receptor trafficking, and neuronal signaling [24]. Genetic screenings have identified mutations, rare variants,

or disruptions of the *SHANK3* gene in patients with ASD [22]. Mice with a genetic disruption of *Shank3* display compulsive/repetitive behaviors and social interaction deficits, which reflect clinical features of ASD [25]. Studies using *Shank3* knockout (KO) mouse models have reported sexually dimorphic phenotypes [26,27]. Matas et al. found that male *Shank3* KO mice with a mutation in the C-terminal regions (exons 21–22) [28] exhibit more pronounced gait deficits than their female siblings [27]. Further research into cerebellar glutamate levels and postsynaptic receptors showed that metabotropic glutamate receptor 5 levels were reduced only in male *Shank3* KO mice, suggesting a potential cause for the varied behavioral outcomes [27].

CHD8, a chromatin remodeling factor, is essential for regulating the transcription of a wide variety of genes [29,30], including approximately 1,000 ASD risk genes identified in the SFARI gene database [30]. Mice with homozygous deletions of *Chd8* die early in embryonic development [31], while those with heterozygous mutations or gene knockdown display a range of ASD-like phenotypes [32–36]. These include impaired social interaction, repetitive behaviors, and cognitive impairments, resembling characteristics of individuals with *CHD8* mutations [20,30,37–39]. Jung et al. found that a heterozygous mutation in *Chd8*, specifically the substitution of asparagine with lysine at position 2373—the first mutation identified as an ASD risk factor in human *CHD8*—results in sexually dimorphic effects that range from transcriptional to behavioral changes in mice [40]. Male $Chd8^{+/N2373K}$ mice exhibited various abnormal behaviors at the pup, juvenile, and adult stages, such as increased ultrasonic vocalizations when seeking their mother, heightened attachment upon reunion with their mother, and increased self-grooming when isolated. In contrast, their female counterparts did not exhibit these behaviors. This behavioral disparity is thought to be associated with sexual dimorphism in neuronal activity, synaptic transmission, and transcriptomic profiles.

The *FMR1* gene encodes the fragile X mental retardation protein (FMRP), which acts as a messenger RNA-binding translational suppressor. It also modulates activity-dependent calcium signaling during critical developmental periods [41]. In mice, FMRP is most abundantly expressed in the hippocampus and cerebral cortex, with peak levels occurring between 2 to 4 weeks postnatally—a crucial time frame for synaptic development and maturation [42]. A deficiency in FMRP results in abnormal synaptic plasticity and structural remodeling [43]. Notably, male *Fmr1* KO mice exhibit more severe anxiety, deficiencies in social preference, and repetitive behaviors than their female counterparts [44]. Differential gene expression analysis of the hippocampus in wild-type (WT) versus *Fmr1* KO mice revealed that in male *Fmr1* KO mice, *Ctnnb1* and *Grin1*—genes considered high-confidence risk factors for ASD—are highly upregulated. In contrast, female *Fmr1* KO mice exhibited upregulation of genes such as *Homer1*, *Ptgs2*, and *Drd1*, which are strong ASD risk gene candidates, as well as *Pik3ca* and *Csnk1g1*, which provide suggestive evidence of risk for ASD. These findings suggest that the loss of FMRP leads to sexually dimorphic phenotypes, potentially due to different patterns of gene expression regulation resulting from the absence of FMR1.

The collective evidence from these reports suggests that synaptic dysfunction and disrupted connectivity could be responsible for sex-specific functional and cognitive impairments observed in ASD [45].

Microglial abnormalities

The balance between excitation and inhibition (E/I) in neural circuits is critical for maintaining brain homeostasis [46]. Disruption of this E/I balance has been implicated as a potential cause of behavioral phenotypes associated with ASD [47]. Microglia, the phagocytic cells that reside

in the brain from the developmental period, engulf the synaptic materials, thus pruning synapses and supporting synaptic maturation. [48]. When microglial function is compromised, improper synaptic pruning can disrupt the E/I balance and potentially contribute to the pathogenesis of ASD [49]. Notably, microglia exhibit sexually dimorphic transcriptional and translational profiles [50]. Furthermore, the morphology and number of microglia in the developing rat brain differ between male and female rats [51]. During the early postnatal period, male rats have significantly higher numbers of microglia compared to female rats. These sex-based differences in microglial numbers appear to be functionally related to sex-specific behaviors [52].

The contactin-associated protein 2 (*CNTNAP2*) gene encodes the CASPR2 protein, which is a neurexin-related synaptic cell adhesion molecule. A study utilizing high-density single nucleotide polymorphisms identified *CNTNAP2* as a strong candidate gene implicated in the etiology of ASD [53]. Subsequent loss-of-function studies in *Cntnap2* KO mice demonstrated that the absence of *Cntnap2* leads to a decrease in dendritic spine density [54], disruptions in synaptic function [55], imbalances in E/I signaling, and impaired neural oscillations [56]. These *Cntnap2* KO mice also display core ASD-like behavioral phenotypes, including impairments in sociability and repetitive behaviors [57]. Dawson et al. found that male *Cntnap2* KO mice exhibited pronounced social deficits, whereas their female counterparts did not. Further investigation into the anterior cingulate cortex—a region critical for social behavior regulation through its connections with other intracortical and subcortical areas—revealed a more activated morphology and increased phagocytosis of synaptic structures in male KO mice compared to WT mice, a distinction not observed in female KO versus WT mice [58].

In addition to genetic models, differences in microglial morphology and function have been observed in preclinical models that incorporate environmental risk factors. High levels of air pollution, particularly during development [59,60], and maternal stress (MS) during gestation [61,62] have been linked to an increased risk of ASD. Smith et al. investigated the combined effects of these two risk factors and found that prenatal exposure to air pollution—specifically diesel exhaust particles (DEP)—along with MS in mice led to sociability deficits exclusively in male offspring [63]. These behavioral impairments were paralleled by alterations in microglial morphology and gene expression, with DEP/MS exposure resulting in a hyper-ramified microglial phenotype in male but not female animals.

The collective evidence from these reports suggests that sexually dimorphic microglial activity could play a role in the etiology of ASD. This activity may disrupt the development of neural circuits responsible for social behavior by excessively pruning synapses during a critical period of development [49].

Hormones

Sex steroid hormones are known to contribute to sex differences in neural activity and behaviors in mammals through their interactions with specific nuclear hormone receptors [64]. Testosterone plays a crucial role during prenatal development in shaping sex differences, influencing brain structure, neurotransmitter and receptor levels, neurogenesis, immune responses, neuropeptide signaling, and cellular processes such as apoptosis, migration, and differentiation [65]. Clinical reports have correlated high levels of testosterone with autistic behavior [66,67], and this association is supported by Erdogan et al., who showed that prenatal testosterone exposure led to ASD-like behaviors in the offspring of Wistar rats [68]. Both male and female rats exposed to testosterone exhibited reduced interaction times with a stranger rat during the three-chamber sociability and social novelty test, indicating a decrease in social

interaction and a phenotype with characteristics resembling ASD. In line with these findings, studies on the valproic acid (VPA)-induced ASD mouse model, which is based on a medication known to increase the risk of ASD in humans [69], have shown that elevated plasma testosterone levels resulting from VPA treatment led to significantly lower levels of tyrosine hydroxylase (TH) expression in the anteroventral periventricular nucleus of male mice. In contrast, TH levels in female mice were unaffected [70].

Allopregnanolone (ALLO), a 3α, 5α progesterone metabolite [71], is a key GABAergic neurosteroid [72,73]. Reduced ALLO levels are correlated with a greater severity of restricted and repetitive behaviors [74]. Penn and colleagues have shown that ALLO plays a vital role as a placental hormone in shaping the fetal brain, leading to sexually dimorphic behavioral outcomes [75]. Specifically, a deficiency of placental ALLO in mice resulted in male-specific abnormalities in cerebellar white matter and core ASD symptoms, such as diminished social preference and increased repetitive behaviors. Notably, this study observed sex-linked dysregulation of myelin proteins in the cerebellar vermis of preterm infants, which aligns with human data.

These results highlight the influence of hormones in molding the early brain environment, potentially leading to sexually dimorphic behavioral outcomes.

Escape from X chromosome inactivation

In a mouse model of early prenatal stress, male offspring exposed to MS during gestation exhibited certain NDD phenotypes [76,77]. The placenta plays a critical role during pregnancy, acting as a mediator in response to disturbances within the intrauterine environment [78]. MS leads to sexually dimorphic changes in the placental expression of O-linked-N-acetylglucosamine transferase (OGT), an X-linked gene essential for the regulation of proteins involved in chromatin remodeling [79]. Notably, OGT escapes X chromosome inactivation in the placenta, resulting in placental levels that are approximately twice as high in female animals than in male animals. Crucially, this finding also translates to humans: levels of both OGT and its biochemical marker, O-GlcNAcylation, have been found to be considerably lower for male fetuses and are further reduced by prenatal stress [79]. Nugent et al. demonstrated that OGT levels establish a sex-specific gene expression pattern in trophoblasts through regulation of a canonical histone repressive mark, H3K27me3 [80]. Higher placental levels of H3K27me3 for female offspring provided a protective effect against the altered hypothalamic programming associated with prenatal stress exposure. Consequently, lower levels of OGT may predispose male offspring to a higher risk of ASD. Future studies should explore the molecular mechanisms underlying this increased male susceptibility.

Integrated stress response pathway

Maternal immune activation (MIA) during pregnancy is linked to a heightened risk of ASD in offspring [81]. This phenomenon has been extensively investigated using a rodent MIA model. In this model, pregnant mice received intraperitoneal injections of polyinosinic:polycytidylic acid (poly[I:C]), a synthetic analog of double-stranded RNA that simulates viral infection. The offspring exhibited significant neurodevelopmental impairments, including diminished social interaction and increased repetitive behaviors [82–84]. Kalish et al. found that MIA exerts a sexually dimorphic effect in utero, leading to different behavioral outcomes. Male offspring exhibited MIA-induced behavioral abnormalities, whereas female offspring did not [85]. Notably, when gene expression was examined at the single-cell level, changes in the fetal cortex were observed to be sexually dimorphic. In male fetuses, these changes were

predominantly characterized by reduced gene expression related to protein translation, followed by an overactive ISR pathway. In eukaryotic cells, the ISR signaling pathway regulates protein synthesis in response to various stresses, both physiological and pathological, to restore cellular homeostasis [86]. Dysregulation of protein synthesis has been implicated in ASD-related traits and other neurological disorders [87–89]. Male-specific activation of the ISR is dependent on the maternal induction of interleukin (IL)-17A following MIA, which has been shown to be necessary for the development of MIA-induced ASD-like behaviors in mouse offspring [83]. The genetic and pharmacological inhibition of ISR pathway hyperactivation was sufficient to protect male offspring from MIA-induced behavioral abnormalities. This study offers valuable insights into potential preventative strategies for ASD-like phenotypes that may result from prenatal immune activation.

Immune pathways

Preterm delivery is associated with a higher likelihood of ASD in children compared to those born at full term [90–93]. Chorioamnionitis, caused by Group B *Streptococcus* (GBS; *Streptococcus agalactiae*), is one of the most common maternal infections and accounts for 40% to 70% of preterm births [94–96]. This condition typically involves an inflammatory intrauterine environment, even in the absence of bacterial translocation from mother to fetus [97]. Allard et al. demonstrated that prenatal infection with live GBS in rats resulted in social impairments in male but not in female offspring [98]. A prominent inflammatory state was noted in male animals, with higher levels of the pro-inflammatory cytokine IL-1β and the cytokine-induced neutrophil chemoattractant-1 (CINC-1/CXCL1), compared to female rats. These findings suggest that sex-specific inflammatory profiles may contribute to the observed sexually dimorphic behavioral outcomes [98].

Consistent with this notion, in a model of MIA induced by lipopolysaccharide (LPS), a toll-like receptor 4 agonist, Braun et al. investigated sex-specific pro-inflammatory responses in both the placenta and fetus, as well as their effects on behavioral outcomes. Male offspring of mothers exposed to LPS exhibited behavioral abnormalities in social interaction and learning, as well as increased repetitive behavior, whereas female offspring were unaffected [99]. Male MIA offspring showed increased cortical hypoxia, decreased mitosis of radial glial cells, and disrupted E/I balance in the brain. Additionally, severe placental necrosis, heightened inflammation, and reduced placental growth were specifically observed in male mice affected by MIA, suggesting that unique sex-specific placental characteristics may make male offspring more susceptible to intrauterine disturbances.

Carlezon Jr et al. demonstrated sex-specific behavioral effects and immune responses in the brain using a combined rodent "two-hit" immune activation model. This model involved treatment with poly (I:C) to induce MIA and the administration of LPS to produce postnatal immune activation [100]. Exposure to early-life immune activation (EIA) was shown to lead to reduced social interaction and increased repetitive behaviors in male animals, while female rodents displayed no significant changes. Molecular studies indicated that EIA resulted in pronounced sex-specific alterations in the expression of inflammation-related genes in the brain. Both male and female rodents exposed to EIA exhibited elevated levels of pro-inflammatory factors in the brain, such as tumor necrosis factor alpha, inducible nitric oxide synthase, IL-6, and IL-1β. Conversely, the expression of anti-inflammatory factors like IL-10 and transforming growth factor beta 1 was reduced in male mice but elevated in female animals [100].

The collective findings of these studies suggest that sexually dimorphic inflammatory

responses could potentially contribute to the sex-specific effects of MS on the neurobehavioral outcomes of offspring.

Discussion

Implication and suggestion

In this review, we have comprehensively examined the potential mechanisms by which genetic variants and environmental factors contribute to sex differences, as demonstrated by preclinical models. Our analysis encompassed both intrinsic differences in the brain, such as synaptic connectivity and microglial activity, and the potential influence of extrinsic factors, including sex hormones and the placenta. These elements may either increase male susceptibility or bolster female resilience. Notably, beyond the intrinsic factors of the brain, the hormonal profile, epigenetic landscape, and immune pathways associated with the placenta have been implicated in contributing to sexually dimorphic outcomes in mouse models exhibiting ASD-like behaviors. This indicates that a deeper understanding of the placenta as a temporary but dynamic interface during prenatal development could provide valuable insights into the sex biases observed in NDDs.

Although preclinical mouse models of ASD have important limitations, such as their inability to engage neural circuitry comparable to that observed in humans or to recapitulate all human ASD phenotypes [8], they remain valuable tools for gaining mechanistic insights into ASD pathogenesis and for developing potential therapeutic approaches [10,11]. Further investigation is imperative to unravel the mechanistic basis of sexual dimorphism in ASD, as well as in other NDDs. For example, utilizing large-scale transcriptomics from postmortem brain studies [101], generating a single-cell atlas [102], and conducting multi-omic profiling of somatic mutations [103] in brains exhibiting ASD could help uncover sex-specific changes in genes or pathways during early neurodevelopment. Given that ASD is a developmental disorder, it is crucial to conduct further longitudinal investigations to understand how risk factors evolve across the developmental trajectory. For instance, a longitudinal cohort study of children with developmental disabilities suggested that *de novo* protein-truncating variants were correlated with clinical characteristics [104].

Conclusion

Overall, understanding sex-specific mechanisms is pivotal for comprehending the fundamental causes of ASD and may illuminate the pathologies of other diseases characterized by prominent sex biases.

ORCID
Taeyoung Lee: https://orcid.org/0009-0005-2317-3451
Eunha Kim: https://orcid.org/0000-0001-7041-1727

Authors' contributions
Project administration: not applicable
Conceptualization: Lee T, Kim E
Methodology & data curation: not applicable
Funding acquisition: Kim E
Writing – original draft: Lee T
Writing – review & editing: Lee T, Kim E

Conflict of interest
Eunha Kim serves as a consultant for Interon Laboratories.

Funding
Eunha Kim received support through a Korea University grant (K2225821).

Data availability
Not applicable.

Acknowledgments
We would like to express our gratitude to all members of the Kim laboratory for their insightful comments. We also give special thanks to Hyun Je for providing technical assistance in structuring the manuscript.

Supplementary materials
Not applicable.

References

1. Morris-Rosendahl DJ, Crocq MA. Neurodevelopmental disorders—the history and future of a diagnostic concept. *Dialogues Clin Neurosci* 2020;22(1):65-72. https://doi.org/10.31887/DCNS.2020.22.1/macrocq
2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). 5th ed. Washington: American Psychiatric Association; 2013.
3. Fombonne E. Epidemiology of pervasive developmental disorders. *Pediatr Res* 2009;65(6):591-598. https://doi.org/10.1203/PDR.0b013e31819e7203
4. Werling DM, Geschwind DH. Sex differences in autism spectrum disorders. *Curr Opin Neurol* 2013;26(2):146-153. https://doi.org/10.1097/WCO.0b013e32835ee548
5. Loomes R, Hull L, Mandy WPL. What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *J Am Acad Child Adolesc Psychiatry* 2017;56(6):466-474. https://doi.org/10.1016/j.jaac.2017.03.013
6. Hull L, Petrides KV, Mandy W. The female autism phenotype and camouflaging: a narrative review. *Rev J Autism Dev Disord* 2020;7(4):306-317. https://doi.org/10.1007/s40489-020-00197-9
7. Zwaigenbaum L, Bryson SE, Szatmari P, Brian J, Smith IM, Roberts W, et al. Sex differences in children with autism spectrum disorder identified within a high-risk infant cohort. *J Autism Dev Disord* 2012;42(12):2585-2596. https://doi.org/10.1007/s10803-012-1515-y
8. Silverman JL, Thurm A, Ethridge SB, Soller MM, Petkova SP, Abel T, et al. Reconsidering animal models used to study autism spectrum disorder: current state and optimizing future. *Genes Brain Behav* 2022;21(5):e12803. https://doi.org/10.1111/gbb.12803
9. Pensado-López A, Veiga-Rúa S, Carracedo Á, Allegue C, Sánchez L. Experimental models to study autism spectrum disorders: hiPSCs, rodents and zebrafish. *Genes* 2020;11(11):1376. https://doi.org/10.3390/genes11111376
10. Kazdoba TM, Leach PT, Yang M, Silverman JL, Solomon M, Crawley JN. Translational mouse models of autism: advancing toward pharmacological therapeutics. *Curr Top Behav Neurosci* 2016;28:1-52. https://doi.org/10.1007/7854_2015_5003
11. Jabarin R, Netser S, Wagner S. Beyond the three-chamber test: toward a multimodal and objective assessment of social behavior in rodents. *Mol Autism* 2022;13(1):41. https://doi.org/10.1186/s13229-022-00521-6
12. Mohammad-Rezazadeh I, Frohlich J, Loo SK, Jeste SS. Brain connectivity in autism spectrum disorder. *Curr Opin Neurol* 2016;29(2):137-147. https://doi.org/10.1097/WCO.0000000000000301
13. Maximo JO, Cadena EJ, Kana RK. The implications of brain connectivity in the neuropsychology of autism. *Neuropsychol Rev* 2014;24(1):16-31. https://doi.org/10.1007/s11065-014-9250-0
14. Guang S, Pang N, Deng X, Yang L, He F, Wu L, et al. Synaptopathology involved in autism spectrum disorder. *Front Cell Neurosci* 2018;12:470. https://doi.org/10.3389/fncel.2018.00470
15. Banerjee-Basu S, Packer A. SFARI Gene: an evolving database for the autism research community. *Dis Model Mech* 2010;3(3-4):133-135. https://doi.org/10.1242/dmm.005439
16. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* 2013;4(1):36. https://doi.org/10.1186/2040-2392-4-36
17. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* 2014;515(7526):216-221.

https://doi.org/10.1038/nature13908
18. Choi L, An JY. Genetic architecture of autism spectrum disorder: lessons from large-scale genomic studies. *Neurosci Biobehav Rev* 2021;128:244-257.
https://doi.org/10.1016/j.neubiorev.2021.06.028
19. Wang T, Guo H, Xiong B, Stessman HAF, Wu H, Coe BP, et al. *De novo* genic mutations among a Chinese autism spectrum disorder cohort. *Nat Commun* 2016;7:13316.
https://doi.org/10.1038/ncomms13316
20. Stessman HAF, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M, et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat Genet* 2017;49(4):515-526.
https://doi.org/10.1038/ng.3792
21. Ellingford RA, Panasiuk MJ, de Meritens ER, Shaunak R, Naybour L, Browne L, et al. Cell-type-specific synaptic imbalance and disrupted homeostatic plasticity in cortical circuits of ASD-associated *Chd8* haploinsufficient mice. *Mol Psychiatry* 2021;26(7):3614-3624.
https://doi.org/10.1038/s41380-021-01070-9
22. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, et al. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet* 2007;39(1):25-27.
https://doi.org/10.1038/ng1933
23. Betancur C, Buxbaum JD. *SHANK3* haploinsufficiency: a "common" but underdiagnosed highly penetrant monogenic cause of autism spectrum disorders. *Mol Autism* 2013;4(1):17.
https://doi.org/10.1186/2040-2392-4-17
24. Monteiro P, Feng G. SHANK proteins: roles at the synapse and in autism spectrum disorder. *Nat Rev Neurosci* 2017;18(3):147-157.
https://doi.org/10.1038/nrn.2016.183
25. Peca J, Feliciano C, Ting JT, Wang W, Wells MF, Venkatraman TN, et al. *Shank3* mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* 2011;472(7344):437-442.
https://doi.org/10.1038/nature09965
26. Maloney SE, Sarafinovska S, Weichselbaum C, McCullough KB, Swift RG, Liu Y, et al. A comprehensive assay of social motivation reveals sex-specific roles of autism-associated genes and oxytocin. *Cell Rep Methods* 2023;3(6):100504.
https://doi.org/10.1016/j.crmeth.2023.100504
27. Matas E, Maisterrena A, Thabault M, Balado E, Francheteau M, Balbous A, et al. Major motor and gait deficits with sexual dimorphism in a *Shank3* mutant mouse model. *Mol Autism* 2021;12(1):2.
https://doi.org/10.1186/s13229-020-00412-8
28. Moretto E, Murru L, Martano G, Sassone J, Passafaro M. Glutamatergic synapses in neurodevelopmental disorders. *Prog Neuropsychopharmacol Biol Psychiatry* 2018;84(Pt B):328-342.
https://doi.org/10.1016/j.pnpbp.2017.09.014
29. Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat Rev Neurosci* 2015;16(9):551-563.
https://doi.org/10.1038/nrn3992
30. Barnard RA, Pomaville MB, O'Roak BJ. Mutations and modeling of the chromatin remodeler CHD8 define an emerging autism etiology. *Front Neurosci* 2015;9:477.
https://doi.org/10.3389/fnins.2015.00477
31. Nishiyama M, Nakayama K, Tsunematsu R, Tsukiyama T, Kikuchi A, Nakayama KI. Early embryonic death in mice lacking the *β*-catenin-binding protein Duplin. *Mol Cell Biol* 2004;24(19):8386-8394.
https://doi.org/10.1128/MCB.24.19.8386-8394.2004
32. Katayama Y, Nishiyama M, Shoji H, Ohkawa Y, Kawamura A, Sato T, et al. CHD8 haploinsufficiency results in autistic-like phenotypes in mice. *Nature* 2016;537(7622):675-679.
https://doi.org/10.1038/nature19357
33. Durak O, Gao F, Kaeser-Woo YJ, Rueda R, Martorell AJ, Nott A, et al. Chd8 mediates cortical neurogenesis via transcriptional regulation of cell cycle and Wnt signaling. *Nat Neurosci* 2016;19(11):1477-1488.
https://doi.org/10.1038/nn.4400
34. Gompers AL, Su-Feher L, Ellegood J, Copping NA, Asrafuzzaman Riyadh M, Stradleigh TW, et al. Germline *Chd8* haploinsufficiency alters brain development in mouse. *Nat Neurosci* 2017;20(8):1062-1073.
https://doi.org/10.1038/nn.4592
35. Platt RJ, Zhou Y, Slaymaker IM, Shetty AS, Weisbach NR, Kim JA, et al. *Chd8* mutation leads to autistic-like behaviors and impaired striatal circuits. *Cell Rep* 2017;19(2):335-350.
https://doi.org/10.1016/j.celrep.2017.03.052
36. Suetterlin P, Hurley S, Mohan C, Riegman KLH, Pagani M, Caruso A, et al. Altered neocortical gene expression, brain overgrowth and functional over-connectivity in *Chd8* haploinsufficient mice. *Cereb Cortex* 2018;28(6):2192-2206.
https://doi.org/10.1093/cercor/bhy058
37. Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O, et al. Disruptive *CHD8* mutations define a subtype of autism early in development. *Cell* 2014;158(2):263-276.
https://doi.org/10.1016/j.cell.2014.06.017
38. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012;338(6114):1619-1622.

https://doi.org/10.1126/science.1227764
39. Merner N, Forgeot d'Arc B, Bell SC, Maussion G, Peng H, Gauthier J, et al. A *de novo* frameshift mutation in chromodomain helicase DNA-binding domain 8 (CHD8): a case report and literature review. *Am J Med Genet A* 2016;170(5):1225-1235. https://doi.org/10.1002/ajmg.a.37566
40. Jung H, Park H, Choi Y, Kang H, Lee E, Kweon H, et al. Sexually dimorphic behavior, neuronal activity, and gene expression in Chd8-mutant mice. *Nat Neurosci* 2018;21(9):1218-1228. https://doi.org/10.1038/s41593-018-0208-z
41. Davis JK, Broadie K. Multifarious functions of the fragile X mental retardation protein. *Trends Genet* 2017;33(10):703-714. https://doi.org/10.1016/j.tig.2017.07.008
42. Pacey LKK, Xuan ICY, Guan S, Sussman D, Mark Henkelman R, Chen Y, et al. Delayed myelination in a mouse model of fragile X syndrome. *Hum Mol Genet* 2013;22(19):3920-3930. https://doi.org/10.1093/hmg/ddt246
43. Qin M, Entezam A, Usdin K, Huang T, Liu ZH, Hoffman GE, et al. A mouse model of the fragile X premutation: effects on behavior, dendrite morphology, and regional rates of cerebral protein synthesis. *Neurobiol Dis* 2011;42(1):85-98. https://doi.org/10.1016/j.nbd.2011.01.008
44. Wang Z, Qiao D, Chen H, Zhang S, Zhang B, Zhang J, et al. Effects of *Fmr1* gene mutations on sex differences in autism-like behavior and dendritic spine development in mice and transcriptomic studies. *Neuroscience* 2023;534:16-28. https://doi.org/10.1016/j.neuroscience.2023.10.001
45. Luo J, Norris RH, Gordon SL, Nithianantharajah J. Neurodevelopmental synaptopathies: insights from behaviour in rodent models of synapse gene mutations. *Prog Neuropsychopharmacol Biol Psychiatry* 2018;84(Pt B):424-439. https://doi.org/10.1016/j.pnpbp.2017.12.001
46. Tatti R, Haley MS, Swanson OK, Tselha T, Maffei A. Neurophysiology and regulation of the balance between excitation and inhibition in neocortical circuits. *Biol Psychiatry* 2017;81(10):821-831. https://doi.org/10.1016/j.biopsych.2016.09.017
47. Hollestein V, Poelmans G, Forde NJ, Beckmann CF, Ecker C, Mann C, et al. Excitatory/inhibitory imbalance in autism: the role of glutamate and GABA gene-sets in symptoms and cortical brain structure. *Transl Psychiatry* 2023;13(1):18. https://doi.org/10.1038/s41398-023-02317-5
48. Paolicelli RC, Bolasco G, Pagani F, Maggi L, Scianni M, Panzanelli P, et al. Synaptic pruning by microglia is necessary for normal brain development. *Science* 2011;333(6048):1456-1458. https://doi.org/10.1126/science.1202529
49. Koyama R, Ikegaya Y. Microglia in the pathogenesis of autism spectrum disorders. *Neurosci Res* 2015;100:1-5. https://doi.org/10.1016/j.neures.2015.06.005
50. Guneykaya D, Ivanov A, Hernandez DP, Haage V, Wojtas B, Meyer N, et al. Transcriptional and translational differences of microglia from male and female brains. *Cell Rep* 2018;24(10):2773-2783.E6. https://doi.org/10.1016/j.celrep.2018.08.001
51. Schwarz JM, Sholar PW, Bilbo SD. Sex differences in microglial colonization of the developing rat brain. *J Neurochem* 2012;120(6):948-963. https://doi.org/10.1111/j.1471-4159.2011.07630.x
52. McCarthy MM, Nugent BM, Lenz KM. Neuroimmunology and neuroepigenetics in the establishment of sex differences in the brain. *Nat Rev Neurosci* 2017;18(8):471-484. https://doi.org/10.1038/nrn.2017.61
53. Alarcón M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, Bomar JM, et al. Linkage, association, and gene-expression analyses identify *CNTNAP2* as an autism-susceptibility gene. *Am J Hum Genet* 2008;82(1):150-159. https://doi.org/10.1016/j.ajhg.2007.09.005
54. Varea O, Martin-de-Saavedra MD, Kopeikina KJ, Schürmann B, Fleming HJ, Fawcett-Patel JM, et al. Synaptic abnormalities and cytoplasmic glutamate receptor aggregates in contactin associated protein-like *2/Caspr2* knockout neurons. *Proc Natl Acad Sci USA* 2015;112(19):6176-6181. https://doi.org/10.1073/pnas.1423205112
55. Anderson GR, Galfin T, Xu W, Aoto J, Malenka RC, Südhof TC. Candidate autism gene screen identifies critical role for cell-adhesion molecule CASPR2 in dendritic arborization and spine development. *Proc Natl Acad Sci USA* 2012;109(44):18120-18125. https://doi.org/10.1073/pnas.1216398109
56. Lazaro MT, Taxidis J, Shuman T, Bachmutsky I, Ikrar T, Santos R, et al. Reduced prefrontal synaptic connectivity and disturbed oscillatory population dynamics in the CNTNAP2 model of autism. *Cell Rep* 2019;27(9):2567-2578.E6. https://doi.org/10.1016/j.celrep.2019.05.006
57. Peñagarikano O, Geschwind DH. What does *CNTNAP2* reveal about autism spectrum disorder? *Trends Mol Med* 2012;18(3):156-163. https://doi.org/10.1016/j.molmed.2012.01.003
58. Dawson MS, Gordon-Fleet K, Yan L, Tardos V, He H, Mui K, et al. Sexual dimorphism in the social behaviour of *Cntnap2*-null mice correlates with disrupted synaptic connectivity and increased microglial activity in the anterior cingulate cortex. *Commun Biol* 2023;6(1):846. https://doi.org/10.1038/s42003-023-05215-0
59. Rahman MM, Shu YH, Chow T, Lurmann FW, Yu X, Martinez MP, et al. Prenatal exposure to air pollution and autism spectrum disorder: sensitive windows of exposure and sex differences. *Environ Health Perspect* 2022;130(1):017008-1-017008-9.

https://doi.org/10.1289/EHP9509
60. Volk HE, Lurmann F, Penfold B, Hertz-Picciotto I, McConnell R. Traffic-related air pollution, particulate matter, and autism. *JAMA Psychiatry* 2013;70(1):71-77. https://doi.org/10.1001/jamapsychiatry.2013.266
61. Roberts AL, Koenen KC, Lyall K, Ascherio A, Weisskopf MG. Women's posttraumatic stress symptoms and autism spectrum disorder in their children. *Res Autism Spectr Disord* 2014;8(6):608-616. https://doi.org/10.1016/j.rasd.2014.02.004
62. Kinney DK, Munir KM, Crowley DJ, Miller AM. Prenatal stress and risk for autism. *Neurosci Biobehav Rev* 2008;32(8):1519-1532. https://doi.org/10.1016/j.neubiorev.2008.06.004
63. Smith CJ, Rendina DN, Kingsbury MA, Malacon KE, Nguyen DM, Tran JJ, et al. Microbial modulation via cross-fostering prevents the effects of pervasive environmental stressors on microglia and social behavior, but not the dopamine system. *Mol Psychiatry* 2023;28(6):2549-2562. https://doi.org/10.1038/s41380-023-02108-w
64. Gegenhuber B, Wu MV, Bronstein R, Tollkuhn J. Gene regulation by gonadal hormone receptors underlies brain sex differences. *Nature* 2022;606(7912):153-159. https://doi.org/10.1038/s41586-022-04686-1
65. Ferri SL, Abel T, Brodkin ES. Sex differences in autism spectrum disorder: a review. *Curr Psychiatry Rep* 2018;20(2):9. https://doi.org/10.1007/s11920-018-0874-2
66. Baron-Cohen S, Auyeung B, Nørgaard-Pedersen B, Hougaard DM, Abdallah MW, Melgaard L, et al. Elevated fetal steroidogenic activity in autism. *Mol Psychiatry* 2015;20(3):369-376. https://doi.org/10.1038/mp.2014.48
67. Majewska MD, Hill M, Urbanowicz E, Rok-Bujko P, Bieńkowski P, Namyslowska I, et al. Marked elevation of adrenal steroids, especially androgens, in saliva of prepubertal autistic children. *Eur Child Adolesc Psychiatry* 2014;23(6):485-498. https://doi.org/10.1007/s00787-013-0472-0
68. Erdogan MA, Bozkurt MF, Erbas O. Effects of prenatal testosterone exposure on the development of autism-like behaviours in offspring of Wistar rats. *Int J Dev Neurosci* 2022;83(2):201-215. https://doi.org/10.1002/jdn.10248
69. Christensen J, Grønborg TK, Sørensen MJ, Schendel D, Parner ET, Pedersen LH, et al. Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *JAMA* 2013;309(16):1696-1703. https://doi.org/10.1001/jama.2013.2270
70. Grgurevic N. Testing the extreme male hypothesis in the valproate mouse model; sex-specific effects on plasma testosterone levels and tyrosine hydroxylase expression in the anteroventral periventricular nucleus, but not on parental behavior. *Front Behav Neurosci* 2023;17:1107226. https://doi.org/10.3389/fnbeh.2023.1107226
71. Diviccaro S, Cioffi L, Falvo E, Giatti S, Melcangi RC. Allopregnanolone: an overview on its synthesis and effects. *J Neuroendocrinol* 2021;34(2):e12996. https://doi.org/10.1111/jne.12996
72. Lambert JJ, Belelli D, Hill-Venning C, Peters JA. Neurosteroids and $GABA_A$ receptor function. *Trends Pharmacol Sci* 1995;16(9):295-303. https://doi.org/10.1016/S0165-6147(00)89058-6
73. Pinna G, Uzunova V, Matsumoto K, Puia G, Mienville JM, Costa E, et al. Brain allopregnanolone regulates the potency of the $GABA_A$ receptor agonist muscimol. *Neuropharmacology* 2000;39(3):440-448. https://doi.org/10.1016/S0028-3908(99)00149-5
74. Chew L, Sun KL, Sun W, Wang Z, Rajadas J, Flores RE, et al. Association of serum allopregnanolone with restricted and repetitive behaviors in adult males with autism. *Psychoneuroendocrinology* 2021;123:105039. https://doi.org/10.1016/j.psyneuen.2020.105039
75. Vacher CM, Lacaille H, O'Reilly JJ, Salzbank J, Bakalar D, Sebaoui S, et al. Placental endocrine function shapes cerebellar development and social behavior. *Nat Neurosci* 2021;24(10):1392-1401. https://doi.org/10.1038/s41593-021-00896-4
76. Mueller BR, Bale TL. Early prenatal stress impact on coping strategies and learning performance is sex dependent. *Physiol Behav* 2007;91(1):55-65. https://doi.org/10.1016/j.physbeh.2007.01.017
77. Mueller BR, Bale TL. Sex-specific programming of offspring emotionality after stress early in pregnancy. *J Neurosci* 2008;28(36):9055-9065. https://doi.org/10.1523/JNEUROSCI.1424-08.2008
78. Marsit CJ, Maccani MA, Padbury JF, Lester BM. Placental 11-beta hydroxysteroid dehydrogenase methylation is associated with newborn growth and a measure of neurobehavioral outcome. *PLoS ONE* 2012;7(3):e33794. https://doi.org/10.1371/journal.pone.0033794
79. Howerton CL, Morgan CP, Fischer DB, Bale TL. O-GlcNAc transferase (OGT) as a placental biomarker of maternal stress and reprogramming of CNS gene transcription in development. *Proc Natl Acad Sci USA* 2013;110(13):5169-5174. https://doi.org/10.1073/pnas.1300065110
80. Nugent BM, O'Donnell CM, Neill Epperson C, Bale TL. Placental H3K27me3 establishes female resilience to prenatal insults. *Nat Commun* 2018;9(1):2555.

https://doi.org/10.1038/s41467-018-04992-1
81. Estes ML, Kimberley McAllister A. Maternal immune activation: implications for neuropsychiatric disorders. *Science* 2016;353(6301):772-777. https://doi.org/10.1126/science.aag3194
82. Smith SEP, Li J, Garbett K, Mirnics K, Patterson PH. Maternal immune activation alters fetal brain development through interleukin-6. *J Neurosci* 2007;27(40):10695-10702. https://doi.org/10.1523/JNEUROSCI.2178-07.2007
83. Choi GB, Yim YS, Wong H, Kim S, Kim H, Kim SV, et al. The maternal interleukin-17a pathway in mice promotes autism-like phenotypes in offspring. *Science* 2016;351(6276):933-939. https://doi.org/10.1126/science.aad0314
84. Shin Yim Y, Park A, Berrios J, Lafourcade M, Pascual LM, Soares N, et al. Reversing behavioural abnormalities in mice exposed to maternal inflammation. *Nature* 2017;549(7673):482-487. https://doi.org/10.1038/nature23909
85. Kalish BT, Kim E, Finander B, Duffy EE, Kim H, Gilman CK, et al. Maternal immune activation in mice disrupts proteostasis in the fetal brain. *Nat Neurosci* 2021;24(2):204-213. https://doi.org/10.1038/s41593-020-00762-9
86. Pakos-Zebrucka K, Koryga I, Mnich K, Ljujic M, Samali A, Gorman AM. The integrated stress response. *EMBO Rep* 2016;17(10):1374-1395. https://doi.org/10.15252/embr.201642195
87. Kelleher RJ 3rd, Bear MF. The autistic neuron: troubled translation? *Cell* 2008;135(3):401-406. https://doi.org/10.1016/j.cell.2008.10.017
88. Torossian A, Saré RM, Loutaev I, Smith CB. Increased rates of cerebral protein synthesis in *Shank3* knockout mice: implications for a link between synaptic protein deficit and dysregulated protein synthesis in autism spectrum disorder/intellectual disability. *Neurobiol Dis* 2021;148:105213. https://doi.org/10.1016/j.nbd.2020.105213
89. Wood H. Integrated stress response mediates cognitive decline in Down syndrome. *Nat Rev Neurol* 2020;16(1):3. https://doi.org/10.1038/s41582-019-0298-6
90. Dudova I, Kasparova M, Markova D, Zemankova J, Beranova S, Urbanek T, et al. Screening for autism in preterm children with extremely low and very low birth weight. *Neuropsychiatr Dis Treat* 2014;10:277-282. https://doi.org/10.2147/NDT.S57057
91. Guy A, Seaton SE, Boyle EM, Draper ES, Field DJ, Manktelow BN, et al. Infants born late/moderately preterm are at increased risk for a positive autism screen at 2 years of age. *J Pediatr* 2015;166(2):269-275.E3. https://doi.org/10.1016/j.jpeds.2014.10.053
92. Kuzniewicz MW, Wi S, Qian Y, Walsh EM, Armstrong MA, Croen LA. Prevalence and neonatal factors associated with autism spectrum disorders in preterm infants. *J Pediatr* 2014;164(1):20-25. https://doi.org/10.1016/j.jpeds.2013.09.021
93. Limperopoulos C, Bassan H, Sullivan NR, Soul JS, Robertson RL Jr, Moore M, et al. Positive screening for autism in ex-preterm infants: prevalence and risk factors. *Pediatrics* 2008;121(4):758-765. https://doi.org/10.1542/peds.2007-2158
94. Jain VG, Willis KA, Jobe A, Ambalavanan N. Chorioamnionitis and neonatal outcomes. *Pediatr Res* 2022;91(2):289-296. https://doi.org/10.1038/s41390-021-01633-0
95. Horvath B, Lakatos F, Tóth C, Bödecs T, Bódis J. Silent chorioamnionitis and associated pregnancy outcomes: a review of clinical data gathered over a 16-year period. *J Perinat Med* 2014;42(4):441-447. https://doi.org/10.1515/jpm-2013-0186
96. Larsen JW, Sever JL. Group B *Streptococcus* and pregnancy: a review. *Am J Obstet Gynecol* 2008;198(4):440-450. https://doi.org/10.1016/j.ajog.2007.11.030
97. Nasef N, Shabaan AE, Schurr P, Iaboni D, Choudhury J, Church P, et al. Effect of clinical and histological chorioamnionitis on the outcome of preterm infants. *Am J Perinatol* 2013;30(01):059-068. https://doi.org/10.1055/s-0032-1321501
98. Allard MJ, Bergeron JD, Baharnoori M, Srivastava LK, Fortier LC, Poyart C, et al. A sexually dichotomous, autistic-like phenotype is induced by group B *Streptococcus* maternofetal immune activation. *Autism Res* 2016;10(2):233-245. https://doi.org/10.1002/aur.1647
99. Braun AE, Carpentier PA, Babineau BA, Narayan AR, Kielhold ML, Moon HM, et al. "Females are not just 'Protected' males": sex-specific vulnerabilities in placenta and brain after prenatal immune disruption. *eNeuro* 2019;6(6):ENEURO.0358-19.2019. https://doi.org/10.1523/ENEURO.0358-19.2019
100. Carlezon WA Jr, Kim W, Missig G, Finger BC, Landino SM, Alexander AJ, et al. Maternal and early postnatal immune activation produce sex-specific effects on autism-like behaviors and neuroimmune function in mice. *Sci Rep* 2019;9(1):16928. https://doi.org/10.1038/s41598-019-53294-z
101. Werling DM, Pochareddy S, Choi J, An JY, Sheppard B, Peng M, et al. Whole-genome and RNA sequencing reveal variation and transcriptomic coordination in the developing human prefrontal cortex. *Cell Rep* 2020;31(1):107489. https://doi.org/10.1016/j.celrep.2020.03.053
102. Kim S, Lee J, Koh IG, Ji J, Kim HJ, Kim E, et al. An integrative single-cell atlas to explore the cellular and temporal specificity of neurological disorder genes during human brain development [Internet]. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory; c2024 [cited 2024 Apr 11]. Available from: https://www.biorxiv.org/content/10.1101/2024.04.09.588220v1

103. Chung C, Yang X, Bae T, Vong KI, Mittal S, Donkels C, et al. Comprehensive multi-omic profiling of somatic mutations in malformations of cortical development. *Nat Genet* 2023;55(2):209-220. https://doi.org/10.1038/s41588-022-01276-9
104. Lee T, Lee H, Kim S, Park KJ, An JY, Kim HW. Brief report: risk variants could inform early neurodevelopmental outcome in children with developmental disabilities. J Autism Dev Disord 2022 Sep 7 [Epub]. https://doi.org/10.1007/s10803-022-05735-4

Original Article

Ewha Med J 2024;47(2):e19
https://doi.org/10.12771/emj.2024.e19
eISSN 2234-2591

Effect of body mass index on gastric cancer risk according to sex in Korea: a nationwide cohort study and literature review

Yonghoon Choi[1,*], **Jieun Jang**[2,*], **Nayoung Kim**[1,3]

[1]Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Korea
[2]Division of Clinical Research, Research Institute, National Cancer Center, Goyang, Korea
[3]Department of Internal Medicine and Liver Research Institute, Seoul National University College of Medicine, Seoul, Korea

*These authors contributed equally to this work.

Received Mar 14, 2024
Accepted Apr 18, 2024

Corresponding author
Nayoung Kim
Department of Internal Medicine, Seoul National University Bundang Hospital, 82 Gumi-ro 173beon-gil, Bundang-gu, Seongnam 13620, Korea
E-mail: nakim49@snu.ac.kr

Keywords
Body mass index; Obesity; Sex characteristics; Stomach neoplasm; Waist circumference

Objectives: Gastric cancer (GC) demonstrates a sex disparity that may also be associated with body mass index (BMI). This study explored whether the effect of BMI on the risk of GC varies by sex.

Methods: The study cohort included 341,999 Koreans aged 40 years or older from the National Health Insurance Service–Health Screening Cohort, with a median follow-up period of 10 years. Participants were categorized into five groups based on their BMI. The effect of BMI was evaluated using Cox proportional hazard regression. Additionally, stratification analysis was performed according to waist circumference.

Results: An increased risk of developing GC was observed across the study population among those with obesity (BMI 25.0–29.9 kg/m^2; hazard ratio [HR], 1.11; 95% CI , 1.03–1.20) and severe obesity (BMI ≥30.0 kg/m^2; HR, 1.22; 95% CI, 1.01–1.47), considering a 2-year latency period. Notably, the rise in GC risk was particularly pronounced among women with obesity and men with severe obesity. In the age-stratified analysis, severe obesity (BMI ≥30.0 kg/m^2) was associated with an increased risk of GC in men under 50 years old (HR, 1.83; 95% CI, 0.99–3.37). For individuals aged ≥50 years, obesity was linked to a heightened risk of GC in both sexes. Furthermore, normal BMI (18.5–22.9 kg/m^2) was associated with an increased GC risk in women.

Conclusion: These findings indicate a positive association between excess body weight and the risk of GC in Koreans, particularly among men with severe obesity.

Introduction

Background

Gastric cancer (GC) represents the fifth most common malignancy worldwide and exhibits the highest incidence rates in Eastern Asia, including South Korea [1,2]. Traditional risk factors for GC encompass *Helicobacter pylori* infection, dietary patterns, and exposure to risk factors such as alcohol consumption and smoking [3–5]. Furthermore, the influence of obesity on the development of non-communicable diseases, including GC, has become more pronounced [6], paralleling the global rise in obesity rates [7].

Despite considerable interest and research, the relationship between obesity and GC remains less clear than for other cancers, such as colon cancer [8,9]. This ambiguity is partly because

the pathogenesis of GC varies by anatomical location. Cardia and non-cardia GCs each have unique pathological and etiological features [10]. In cardia cancers, the risk of developing GC due to obesity is higher, and the association with obesity is more pronounced. In contrast, non-cardia cancers do not exhibit a significant link with obesity [11]. However, a recent study from Korea suggested that underweight was associated with an increased risk of developing GC [12], with a U-shaped pattern of risk increase. Therefore, the relationship between underweight or overweight and the risk of developing GC necessitates further investigation. GC is also recognized as a male-dominant disease [13], with this sex difference typically attributed to variations in exposure to risk factors and the influence of sex hormones [14,15]. The effect of weight outside the normal range on the development of GC likely differs between sexes, given that estrogen, a key sex hormone, is associated with obesity [15].

Objectives

Consequently, our goal was to analyze the impact of body mass index (BMI) on the risk of developing GC by sex, drawing on data from a large-scale retrospective cohort study and a review of the existing literature.

Methods

Ethics statement

The study protocol was approved by the institutional review board (IRB) of Seoul National University Bundang Hospital (IRB No. X-2209-780-901). The requirement for informed consent was waived by the IRB.

Study design

This retrospective cohort study was conducted in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology statement (https://www.strobe-statement.org/).

Setting

In February and March of 2024, the authors conducted a search regarding BMI and GC within the National Health Insurance Service–Health Screening Cohort (NHIS-HEALS) database, which contains records from 2002 to 2019. The selected data were subsequently analyzed by the authors.

Data sources

Korea operates the National Health Insurance Service (NHIS), a single, mandatory health insurance system that covers approximately 97% of the Korean population. The NHIS administers a biennial health checkup program for adults aged 20 years and older, known as the National Health Screening (NHS). The NHS program encompasses over 70% of the total population in Korea [16]. During these NHS health checkups, various data are collected from the examinees. These include anthropometric measurements such as height, weight, and waist circumference (WC), as well as sociodemographic factors and health behaviors, including alcohol consumption and smoking status. Additionally, family and medical histories are recorded alongside laboratory test results [16,17].

For research purposes, the NHIS constructed a sampled retrospective cohort consisting of

514,866 participants. These individuals were randomly selected from participants in the NHS programs in 2002–2003, a cohort referred to as NHIS-HEALS. This cohort has undergone annual follow-up through 2019 to gather information on healthcare utilization and mortality.

Participants

Inclusion and exclusion criteria

Since the NHS has collected WC data since 2009, we defined 2009–2010 as the baseline period. Among the participants in the NHIS-HEALS (n=514,866), those lacking information on BMI for 2009–2010 were excluded. Additionally, we excluded participants with a history of cancer, operationally defined as those with claims data containing a major diagnosis code beginning with "C" from 2002–2008. In Korea, the NHIS offers a program that enables adults over 40 years of age to undergo GC screening every 2 years. We posited that a minimum of 2 years is necessary for BMI at a given time point to influence the development of GC. Consequently, GC cases diagnosed within 2 years following BMI measurement were excluded. Ultimately, 341,999 participants, including 4,277 GC cases, were selected for the study (Fig. 1).

Measurements

Study participants were categorized into four groups based on their BMI, following the classifications used in a prior study [18]: underweight (BMI <18.5 kg/m^2), normal weight (BMI 18.5–22.9 kg/m^2), overweight (BMI 23.0–24.9 kg/m^2), obesity (BMI 25.0–29.9 kg/m^2), and severe obesity (BMI ≥30.0 kg/m^2). Additionally, WC reference values of 90 cm for male and 85 cm for female participants were adopted in line with the Korean diagnostic criteria for metabolic syndrome [18].

We operationally defined GC cases as participants who had medical claims with the corresponding diagnosis code for GC, specifically the International Classification of Diseases-10 code C16, along with a history of hospital admission. The follow-up period was defined as the time from the index date (the date of BMI measurement) to the date of GC diagnosis, death, or the end of the follow-up period (December 31, 2019), whichever occurred first.

Variables

The outcome variables included demographic findings, such as a diagnosis of GC, along with age, WC, and BMI.

Bias

Since participants were selected from the cohort database according to the inclusion criteria, selection bias was not a concern.

Study size

A sample size estimation was not performed, as all target participants were included.

Statistical methods

To compare differences in baseline characteristics across BMI levels, we conducted univariable tests, utilizing the chi-square test for categorical variables and analysis of variance for continuous variables. The characteristics considered included sociodemographic factors (such as age, sex, and income), health behaviors (including smoking status and physical activity),

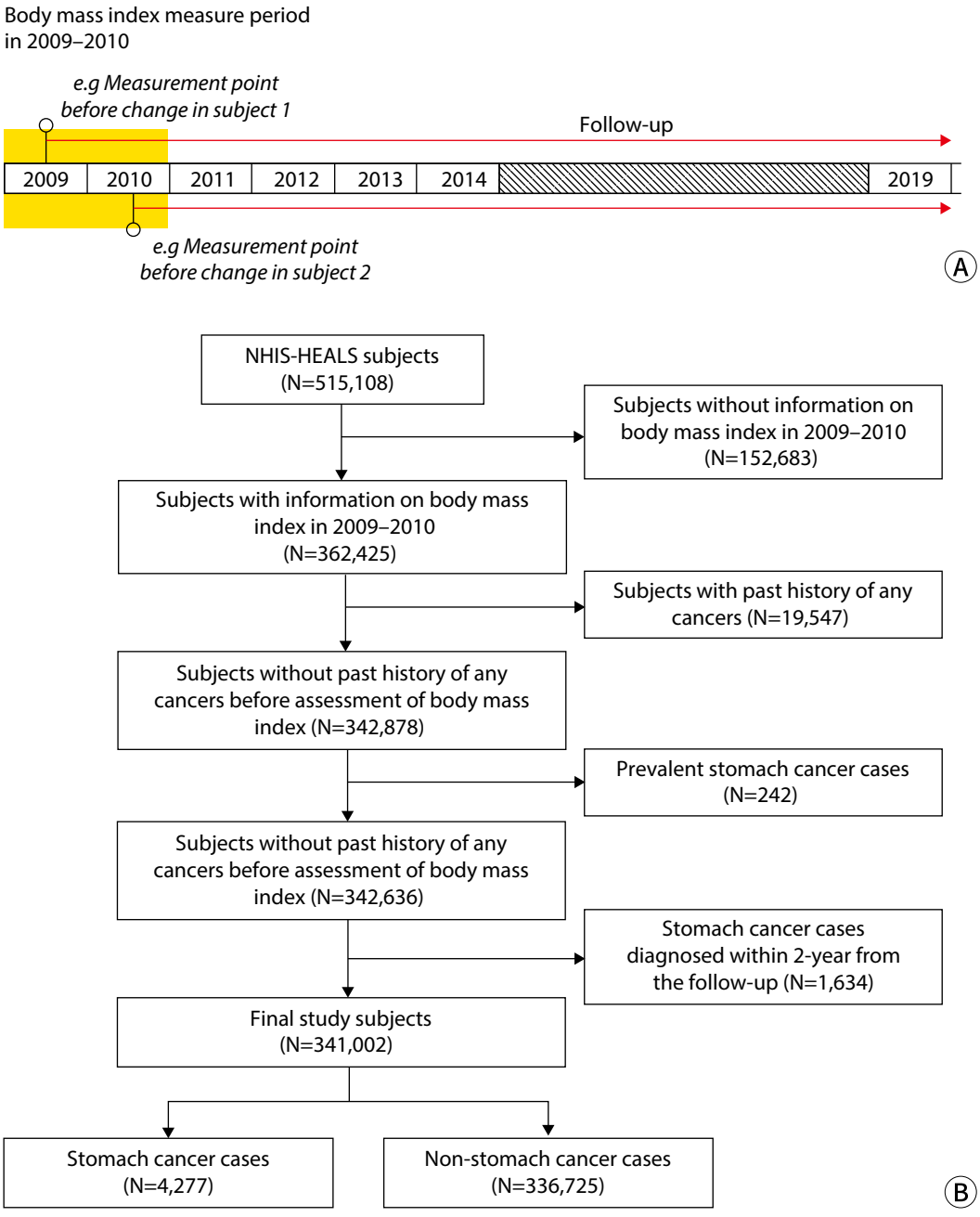


Fig. 1. Study scheme. (A) Time points for measuring body mass index (BMI) and initiating follow-up. The orange box represents the period during which BMI was measured in 2009–2010, with follow-up extending until 2019. (B) Flowchart depicting the selection of study participants. NHIS-HEALS, National Health Insurance Service-Health Screening Cohort.

fasting glucose level, total cholesterol, blood pressure at the time of BMI measurement, and comorbidities such as diabetes mellitus, hypertension, and dyslipidemia recorded before the most recent BMI measurement. We assessed GC risk in relation to BMI by using the overweight group (BMI of 23.0–24.9 kg/m^2) as the reference category in Cox proportional hazard regression analysis.

https://doi.org/10.12771/emj.2024.e19

To account for the influence of baseline characteristics on the development of GC, we constructed a Cox proportional hazards regression model. This model included age, sex, alcohol consumption, smoking status, and history of diabetes mellitus, hypertension, and dyslipidemia as exploratory variables.

Although some baseline characteristics (namely, smoking status and alcohol consumption) had missing values, the proportions of missing data were relatively low: 1.9% for smoking status and 3.1% for alcohol consumption. To address the issue of missing values, we conducted data imputation using the PROC MI procedure within SAS (SAS Institute, Cary, NC, USA).

All data management and statistical analyses were performed using SAS version 9.4. The protocol of this study was approved by the IRB of Seoul National University Bundang Hospital (IRB No. X-2209-780-901).

Results

Baseline characteristics of participants

The baseline characteristics of the participants are summarized in Table 1. Most participants fell into the normal weight, overweight, and obesity categories, with only a small percentage classified as underweight (2.1%) or having severe obesity (2.8%). Among BMI categories, the severe obesity group contained the highest proportions of women, non-smokers, and non-drinkers. Furthermore, the prevalence rates of diabetes, hypertension, and dyslipidemia increased with rising BMI levels.

Risk of gastric cancer according to body mass index

Although we hypothesized that both underweight and obesity were associated with an increased risk of GC development, only those with obesity (BMI ≥25.0 kg/m^2) exhibited an elevated GC risk compared to participants with a BMI of 23.0–24.9 kg/m^2 (BMI 25.0–29.9 kg/m^2: hazard ratio [HR], 1.11; 95% CI, 1.03–1.20; BMI ≥30.0 kg/m^2: HR, 1.22; 95% CI, 1.01–1.47). However, an elevated GC risk was not observed in the underweight population in this study. Although the highest increase in GC risk was found in those with severe obesity, this association was only noted in the male population (HR, 1.36; 95% CI, 1.12–1.74).

In a sensitivity analysis considering various latency periods, we assessed the association between BMI and GC risk by repeatedly excluding GC cases diagnosed within 2, 3, and 4 years of follow-up (Table 2). The analysis showed that the increase in GC risk was consistently higher with greater severity of obesity, regardless of the latency period considered. Furthermore, a significantly elevated GC risk in the severe obesity group was observed exclusively in the male population. Consequently, we applied a 2-year latency period to subsequent analyses.

Age-stratified risk of gastric cancer according to body mass index

A stratification analysis by age was conducted using a threshold of 50 years (Table 3). Notably, the NHIS-HEALS dataset was characterized by a disparity in participant person-years, with approximately 49,000 for individuals under 50 years and over 3 million for those aged 50 and above. This resulted in limitations regarding the statistical power for the subset of participants younger than 50 years.

In individuals under 50 years old, severe obesity (BMI ≥30.0 kg/m^2) was associated with a heightened risk of GC in men (HR, 1.83; 95% CI, 0.99–3.37). However, no significant associations between BMI and GC risk were found in the other BMI categories.

Table 1. Baseline characteristics of study participants according to body mass index

Characteristics	Body mass index (kg/m^2)					P-value
	<18.5	18.5–22.9	23.0–24.9	25.0–29.9	≥30.0	
	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)	
No.	7,177	118,958	96,244	109,192	9,628	
Age (mean±SD)	62.5±11.0	58.6±9.2	58.3±8.6	58.5±8.5	58.7±8.5	<0.01*
Stomach cancer	99 (1.4)	1,420 (1.2)	1,170 (1.2)	1,465 (1.3)	123 (1.3)	
Sex						
Male	3,714 (51.7)	60,260 (50.7)	54,792 (56.9)	62,982 (57.7)	4,110 (42.7)	<0.01*
Female	3,463 (48.3)	58,698 (49.3)	41,452 (43.1)	46,210 (42.3)	5,518 (57.3)	
Smoking status						
Non-smoker	4,445 (61.9)	77,509 (65.2)	59,666 (62.0)	67,062 (61.4)	6,803 (70.7)	<0.01*
Past smoker	799 (11.1)	17,288 (14.5)	18,612 (19.4)	22,420 (20.5)	1,484 (15.4)	
Current smoker	1,815 (25.3)	21,994 (18.5)	16,099 (16.7)	17,552 (16.1)	1,188 (12.3)	
Missing	118 (1.7)	2,167 (1.8)	1,867 (1.9)	2,158 (2.0)	153 (1.6)	
Alcohol consumption						
Non-drinker	4,877 (67.9)	73,374 (61.7)	54,922 (57.1)	61,705 (56.5)	6,310 (65.5)	<0.01*
2–3/month to 1–2/week	1,248 (17.4)	28,760 (24.2)	26,634 (27.6)	30,289 (27.8)	2,087 (21.7)	
≥3–4/week	724 (10.1)	12,685 (10.6)	11,738 (12.2)	14,097 (12.9)	1,004 (10.4)	
Missing	328 (4.6)	4,139 (3.5)	2,950 (3.1)	3,101 (2.8)	227 (2.4)	
Income						
Q1	1,161 (16.2)	16,800 (14.1)	12,688 (13.2)	14,642 (13.4)	1,451 (15.1)	<0.01*
Q2	1,095 (15.3)	16,762 (14.1)	12,457 (12.9)	13,867 (12.7)	1,291 (13.4)	
Q3	1,259 (17.5)	19,494 (16.4)	14,828 (15.4)	17,346 (15.9)	1,710 (17.8)	
Q4	1,377 (19.2)	24,423 (20.5)	20,024 (20.8)	23,256 (21.3)	2,160 (22.4)	
Q5	2,267 (31.6)	41,281 (34.7)	36,142 (37.6)	39,912 (36.6)	2,996 (31.1)	
Beneficiary	18 (0.2)	198 (0.2)	105 (0.1)	169 (0.2)	20 (0.2)	
Fasting glucose (mean±SD)	96.7±24.6	98.1±23.8	101.0±25.0	103.8±26.2	108.2±29.1	<0.01*
Total cholesterol (mean±SD)	189.3±36.5	197.4±36.6	201.3±37.5	203.1±38.3	205.1±39.7	<0.01*
SBP (mean±SD)	120.0±16.4	122.3±15.3	125.3±14.8	128.3±14.8	131.8±14.9	<0.01*
DBP (mean±SD)	74.3±10.2	75.8±9.9	77.6±9.7	79.5±9.8	81.5±10.0	<0.01*
History of diabetes						
No	5,969 (83.2)	97,668 (82.1)	75,985 (78.9)	82,342 (75.4)	6,523 (67.7)	<0.01*
Yes	1,208 (16.8)	21,290 (17.9)	20,259 (21.1)	26,850 (24.6)	3,105 (32.3)	
History of hypertension						
No	5,246 (73.1)	83,946 (70.6)	60,242 (62.6)	57,661 (52.8)	3,593 (37.3)	<0.01*
Yes	1,931 (26.9)	35,012 (29.4)	36,002 (37.4)	51,531 (47.2)	6,035 (62.7)	
History of dyslipidemia						
No	6,043 (84.2)	93,083 (78.2)	69,842 (72.6)	73,383 (67.2)	5,765 (59.9)	<0.01*
Yes	1,134 (15.8)	25,875 (21.8)	26,402 (27.4)	35,809 (32.8)	3,863 (40.1)	

With consideration of a 2-year latency period.
No., number; Q, quartile; SBP, systolic blood pressure; DBP, diastolic blood pressure.
*Indicates statistical significance.

https://doi.org/10.12771/emj.2024.e19

Table 2. Association between body mass index and gastric cancer risk according to sex

BMI (kg/m^2)	Total			Male			Female		
	Person-years	No. of cases	HR (95% CI)*	Person-years	No. of cases	HR (95% CI)†	Person-years	No. of cases	HR (95% CI)†
2-year latency									
<18.5	64,244	99	1.04 (0.85–1.28)	32,116	77	1.08 (0.85–1.36)	32,128	22	0.94 (0.60–1.45)
18.5–22.9	1,148,614	1,420	1.00 (0.93–1.08)	573,269	1,021	0.96 (0.88–1.05)	575,344	399	1.13 (0.97–1.32)
23.0–24.9	939,227	1,170	Ref	530,335	909	Ref	408,892	261	Ref
25.0–29.9	1,065,977	1,465	1.11 (1.03–1.20)‡	610,902	1,082	1.07 (0.98–1.17)	455,076	383	1.24 (1.06–1.45)‡
≥30.0	93,809	123	1.22 (1.01–1.47)‡	39,737	87	1.36 (1.12–1.74)‡	54,072	36	0.97 (0.69–1.38)
3-year latency									
<18.5	64,206	84	1.06 (0.85–1.33)	32,090	67	1.13 (0.88–1.46)	32,116	17	0.86 (0.53–1.42)
18.5–22.9	1,148,125	1,219	1.01 (0.93–1.10)	572,917	876	0.97 (0.88–1.07)	575,208	343	1.14 (0.96–1.35)
23.0–24.9	938,826	1,004	Ref	530,026	781	Ref	408,800	223	Ref
25.0–29.9	1,065,447	1,247	1.10 (1.01–1.19)‡	610,487	912	1.05 (0.95–1.15)	455,961	335	1.27 (1.07–1.51)‡
≥30.0	93,767	105	1.21 (0.99–1.48)	39,714	77	1.43 (1.13–1.80)‡	54,053	28	0.89 (0.60–1.33)
4-year latency									
<18.5	64,170	74	1.08 (0.85–1.37)	32,058	58	1.13 (0.87–1.49)	32,112	16	0.92 (0.55–1.54)
18.5–22.9	1,147,469	1,033	0.97 (0.89–1.06)	572,454	744	0.93 (0.84–1.04)	575,015	289	1.08 (0.90–1.30)
23.0–24.9	938,421	889	Ref	529,710	691	Ref	408,711	198	Ref
25.0–29.9	1,064,868	1,083	1.08 (0.98–1.18)	610,040	785	1.02 (0.92–1.13)	454,828	298	1.27 (1.06–1.52)‡
≥30.0	93,728	94	1.22 (0.98–1.51)	39,691	70	1.46 (1.14–1.87)‡	54,038	24	0.86 (0.56–1.32)

BMI, body mass index; HR, hazard ratio.
*Adjusted for age, sex, smoking status, alcohol consumption frequency per week, history of diabetes mellitus, hypertension, and dyslipidemia with consideration of a 2-year latency period.
†Adjusted for age, smoking status, alcohol consumption frequency per week, history of diabetes mellitus, hypertension, and dyslipidemia with consideration of a 2-year latency period.
‡Indicates statistical significance.

In participants over 50 years of age, we observed an increased risk of GC in both men and women with obesity, which is consistent with the results of the preceding analysis. Furthermore, an elevated risk of GC was also identified in women with normal BMI (18.5–22.9 kg/m^2).

Gastric cancer risk according to body mass index and waist circumference

Additional analysis considering WC was performed to account for abdominal obesity, as BMI only reflects the height and weight of participants (Table 4). We aimed to clarify the impact of

Table 3. Association between body mass index and gastric cancer risk in an age-stratified population

BMI (kg/m^2)	Total			Male			Female		
	Person-years	No. of cases	HR (95% CI)*	Person-years	No. of cases	HR (95% CI)†	Person-years	No. of cases	HR (95% CI)†
<50 years									
<18.5	8,291	8	1.46 (0.7–3.00)	4,045	6	1.65 (0.72–3.80)	4,246	2	0.99 (0.23–4.22)
18.5–22.9	185,647	119	0.97 (0.74–1.27)	96,466	85	1.03 (0.76–1.41)	89,181	34	0.78 (0.46–1.33)
23.0–24.9	138,637	99	Ref	91,841	76	Ref	46,796	23	Ref
25.0–29.9	148,725	97	0.86 (0.65–1.14)	109,552	84	0.92 (0.67–1.25)	39,173	13	0.65 (0.33–1.29)
≥30.0	12,266	12	1.34 (0.73–2.45)	7,732	12	1.83 (0.99–3.37)	4,534	0	N/A
≥50 years									
<18.5	55,953	91	1.02 (0.83–1.27)	28,071	71	1.06 (0.83–1.35)	27,882	20	0.93 (0.59–1.47)
18.5–22.9	962,966	1,301	1.01 (0.93–1.09)	476,803	936	0.96 (0.87–1.05)	486,163	365	1.16 (0.99–1.37)
23.0–24.9	800,590	1,071	Ref	438,494	833	Ref	362,096	238	Ref
25.0–29.9	917,252	1,368	1.13 (1.04–1.23)‡	501,349	998	1.08 (0.99–1.19)	415,903	370	1.29 (1.09–1.52)‡
≥30.0	81,543	111	1.21 (0.99-1.47)	32,005	75	1.34 (1.06–1.70)‡	49,538	36	1.05 (0.74–1.49)

BMI, body mass index; HR, hazard ratio; N/A, not applicable.
*Adjusted for age, sex, smoking status, alcohol consumption frequency per week, history of diabetes mellitus, hypertension, and dyslipidemia with consideration of a 2-year latency period.
†Adjusted for age, smoking status, alcohol consumption frequency per week, history of diabetes mellitus, hypertension, and dyslipidemia with consideration of a 2-year latency period.
‡Indicates statistical significance.

underweight on the development of GC in those with a low WC (<90 cm for men and <85 cm for women), as well as the impact of obesity on GC development in the group with a high WC (≥90 cm for men and ≥85 cm for women). Even among those with a small WC, we observed no significant increase in GC risk among underweight men (men: HR, 1.10; 95% CI, 0.87–1.40; women: HR, 1.01; 95% CI, 0.65–1.57). However, we did find that the risk of GC increased in the male population with severe obesity and a WC of ≥90 cm. Furthermore, the magnitude of increased GC risk in men with severe obesity and a high WC (HR, 1.41; 95% CI, 1.07–1.85) was greater than that observed when WC was not considered (HR, 1.36; 95% CI, 1.12–1.74). These results suggest that obesity, particularly abdominal obesity as opposed to simple weight gain, plays a role in elevating the risk of GC.

Discussion

Key results

In this study, we aimed to analyze the influence of BMI on GC risk according to sex, based on a large-scale retrospective cohort analysis. The results indicated that GC risk was higher in

Table 4. Impact of body mass index on gastric cancer risk according to waist circumference and sex

WC (cm) by sex	BMI (kg/m^2)	Total			Male			Female		
		Person-year	No. of cases	HR (95% CI)*	Person-year	No. of cases	HR (95% CI)†	Person-year	No. of cases	HR (95% CI)†
<90 (M) <85 (F)	<18.5	63,692	99	1.08 (0.87–1.33)	31,940	77	1.10 (0.87–1.40)	31,752	22	1.01 (0.65–1.57)
	18.5–22.9	1,120,480	1,370	1.01 (0.93–1.10)	562,289	989	0.96 (0.88–1.06)	558,192	381	1.18 (1.00–1.40)‡
	23.0–24.9	809,799	963	Ref	460,118	759	Ref	349,681	204	Ref
	25.0–29.9	517,213	634	1.09 (0.98–1.20)	295,048	478	1.06 (0.94–1.19)	222,166	156	1.19 (0.97–1.47)
	≥30.0	5,979	5	0.85 (0.35–2.04)	2,275	3	0.85 (0.27–2.64)	3,704	2	0.89 (0.22–3.58)
≥90 (M) ≥85 (F)	<18.5	551	0	N/A	176	0	N/A	375	0	N/A
	18.5–22.9	28,134	50	1.08 (0.79–1.47)	10,981	32	1.17 (0.80–1.72)	17,153	18	0.92 (0.54–1.57)
	23.0–24.9	129,427	207	Ref	70,217	150	Ref	59,210	57	Ref
	25.0–29.9	548,764	831	1.07 (0.92–1.25)	315,854	604	1.05 (0.88–1.26)	232,910	227	1.12 (0.83–1.49)
	≥30.0	87,830	118	1.19 (0.94–1.49)	37,462	84	1.41 (1.07–1.85)‡	50,368	34	0.86 (0.59–1.33)

WC, waist circumference; BMI, body mass index; HR, hazard ratio; M, male; F, female.

*Adjusted for age, sex, smoking status, alcohol consumption frequency per week, history of diabetes mellitus, hypertension, and dyslipidemia with consideration of a 2-year latency period.

†Adjusted for age, smoking status, alcohol consumption frequency per week, history of diabetes mellitus, hypertension, and dyslipidemia with consideration of a 2-year latency period.

‡Indicates statistical significance.

the obesity and severe obesity groups. Furthermore, the increase in GC risk was particularly pronounced in women with obesity and men with severe obesity. Normal BMI was also associated with increased GC risk in women aged 50 years and older. When considering BMI and WC together, the risk of GC was elevated in men with severe obesity and high WC.

Interpretation and comparison with previous research

Previous studies have identified a heightened risk of GC associated with overweight or obesity, particularly concerning cardia cancer [11,19–25]. These studies suggest that both overweight and obesity are linked to an increased risk of cardia cancer. Consequently, the World Cancer Research Fund and the International Agency for Research on Cancer have recognized overweight and obesity as risk factors for cardia GC [26,27]. However, no significant difference in the risk of non-cardia GC has been observed between individuals with normal weight and those with obesity [11]. One proposed mechanism for the link between obesity and cardia cancer involves the development of gastroesophageal reflux disease (GERD) due to obesity [28,29]. Specifically, the rise in intra-abdominal pressure caused by abdominal obesity may lead to the reflux of gastric acid and the gastric contents, resulting in a higher incidence of GERD among individuals with obesity [30]. Consequently, obesity could increase the risk of cardia GC through a cascade of events, extending to GERD to Barrett esophagus and ultimately to gastroesophageal junction adenocarcinoma, given that Barrett esophagus is recognized as a

precursor to esophagogastric junction adenocarcinoma [31,32]. Additionally, other mechanisms such as hyperinsulinemia and an increase in insulin-like growth factors, as well as elevated levels of adipokines (leptin and adiponectin), tumor necrosis factor-alpha, and interleukin 6 secreted from adipose tissues, have been proposed as potential contributors to the increased risk of GC under obese and diabetic conditions [33,34].

Our data revealed a sex difference in the increased risk of GC, with the difference being more pronounced in men with obesity [35]. This may be attributed to the distinct patterns of obesity between the sexes. Typically, men exhibit a central distribution of adipose tissue, while women tend to have a peripheral distribution, particularly in the limbs and hips [36]. The greater visceral adiposity found in central obesity is linked to adverse metabolic outcomes, including increased postprandial insulin, free fatty acids, and triglyceride levels [36]. Given that factors secreted from adipose tissues have been proposed among potential mechanisms for elevating GC risk, this could provide an explanation. Another factor to consider is sex hormones. The aforementioned sex-based obesity patterns are largely due to variations in sex hormones and their receptors. For instance, estrogen is known to exert a protective effect against GC, particularly the intestinal type [37]. This protective role of estrogen has been repeatedly proposed to account for the disparity in GC risk between male and female individuals [14,38,39]. Estrogen is synthesized not only in the gonads, but also in adipose tissue. Thus, its levels may become particularly high in obese women, which could further explain the sex difference in GC risk within the obese population. However, few studies have concurrently considered BMI, sex, and the anatomical location of GC, which complicates the ability to draw definitive conclusions. Therefore, additional research in this area is warranted.

Another notable result of this study is the observed association between normal weight and an increased risk of GC in women aged 50 years and older. Recent reports have indicated a rise in GC risk among both underweight and overweight individuals, forming a so-called “U-shaped pattern” [12,40]. While the link between overweight, obesity, and GC risk has been described previously, it appears that being underweight or of normal weight also increases GC risk compared to those who are overweight. To explain this phenomenon, the authors propose two potential mechanisms. The first suggests that precursor lesions of GC, such as atrophic gastritis or metaplasia, could lead to malabsorption and consequently underweight [41]. This theory is supported by the fact that the underweight group in our study was older than the other groups. The second mechanism posits that cigarette smoking could link underweight to an increased risk of GC [12]. Additionally, it is worth considering that being underweight might cause gonadal dysfunction or a decrease in estrogen levels, which could negate the protective effects of estrogen. To date, the causal relationship or sequence of events linking underweight or normal weight with GC risk remains unclear. Moreover, one previous study found an association between underweight and non-cardia cancer [12], while another study reported a link between underweight and cardia cancer [40], indicating some inconsistencies in the findings. In conclusion, when investigating the correlation between obesity and GC risk, it is essential to consider not only sex but also age and the anatomical location of the cancer.

Limitations

This study has several limitations. First, since WC data have been collected since 2009 from the NHIS-HEALS dataset, we analyzed data with a relatively short follow-up period compared to other observational studies. Second, we examined all GC cases without differentiating between anatomical locations—namely, cardia or non-cardia sites. To address these issues, we

plan to conduct long-term follow-up studies using BMI for comparison. Third, our age-specific analysis was constrained by the small number of participants and cancer cases under the age of 50 years, which reflects the characteristics of the NHIS-HEALS dataset. Given the rising concern over GC in younger populations, future studies should include datasets with a larger representation of young adults. Lastly, we were unable to account for other GC risk factors, such as *H. pylori* infection status and dietary habits, due to the inherent limitations of observational studies. Despite these constraints, we were able to identify the risk of developing GC associated with weight gain, while noting some sex differences.

Conclusion

This study demonstrated a positive association between excess body weight and the risk of GC in Koreans, indicating that the risk of GC was elevated in individuals with obesity and severe obesity. Additionally, sex-specific differences were observed in the impact of obesity on GC development, with men who were severely obese and had a high WC facing a particularly increased risk.

ORCID
Yonghoon Choi: https://orcid.org/0000-0002-1331-969X
Jieun Jang: https://orcid.org/0000-0001-6970-9374
Nayoung Kim: https://orcid.org/0000-0002-9397-0406

Authors' contributions
Project administration: Kim N
Conceptualization: Choi Y, Jang J, Kim N
Methodology & data curation: Jang J
Funding acquisition: Kim N
Writing – original draft: Choi Y
Writing – review & editing: Choi Y, Jang J, Kim N

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Funding
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00337453).

Data availability
The data for this study has been made publicly available and the ownership of the data belongs to the National Health Insurance Service-Health Screening Cohort (NHIS-HEALS). Thus, the corresponding author has no authority to share this data. Any researchers can access to this NHIS-HEALS database after submitting and receiving approval of study proposal by NHIS review committee via the Health Insurance Data Service home page (http://nhiss.nhis.or.kr).

Acknowledgments
Not applicable.

Supplementary materials
Not applicable.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68(6):394-424. https://doi.org/10.3322/caac.21492
2. López MJ, Carbajal J, Alfaro AL, Saravia LG, Zanabria D, Araujo JM, et al. Characteristics of gastric cancer around the world. *Crit Rev Oncol Hematol* 2023;181:103841. https://doi.org/10.1016/j.critrevonc.2022.103841
3. Kumar S, Metz DC, Ellenberg S, Kaplan DE, Goldberg DS. Risk factors and incidence of gastric cancer after detection of

Helicobacter pylori infection: a large cohort study. *Gastroenterology* 2020;158(3):527-536.E7.
https://doi.org/10.1053/j.gastro.2019.10.019
4. Shin A, Park S, Shin HR, Park EH, Park SK, Oh JK, et al. Population attributable fraction of infection-related cancers in Korea. *Ann Oncol* 2011;22(6):1435-1442.
https://doi.org/10.1093/annonc/mdq592
5. Tramacere I, Negri E, Pelucchi C, Bagnardi V, Rota M, Scotti L, et al. A meta-analysis on alcohol drinking and gastric cancer risk. *Ann Oncol* 2012;23(1):28-36.
https://doi.org/10.1093/annonc/mdr135
6. Pi-Sunyer X. The medical risks of obesity. *Postgrad Med* 2009;121(6):21-33.
https://doi.org/10.3810/pgm.2009.11.2074
7. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *Lancet* 2017;390(10113):2627-2642.
https://doi.org/10.1016/S0140-6736(17)32129-3
8. Basen-Engquist K, Chang M. Obesity and cancer risk: recent review and evidence. *Curr Oncol Rep* 2011;13(1):71-76.
https://doi.org/10.1007/s11912-010-0139-7
9. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008;371(9612):569-578.
https://doi.org/10.1016/S0140-6736(08)60269-X
10. Mukaisho K, Nakayama T, Hagiwara T, Hattori T, Sugihara H. Two distinct etiologies of gastric cardia adenocarcinoma: interactions among pH, *Helicobacter pylori*, and bile acids. *Front Microbiol* 2015;6:412.
https://doi.org/10.3389/fmicb.2015.00412
11. Chen Y, Liu L, Wang X, Wang J, Yan Z, Cheng J, et al. Body mass index and risk of gastric cancer: a meta-analysis of a population with more than ten million from 24 prospective studies. *Cancer Epidemiol Biomarkers Prev* 2013;22(8):1395-1408.
https://doi.org/10.1158/1055-9965.EPI-13-0042
12. Jang J, Wang T, Cai H, Ye F, Murphy G, Shimazu T, et al. The U-shaped association between body mass index and gastric cancer risk in the *Helicobacter pylori* biomarker cohort consortium: a nested case-control study from eight East Asian cohort studies. *Int J Cancer* 2020;147(3):777-784.
https://doi.org/10.1002/ijc.32790
13. Lou L, Wang L, Zhang Y, Chen G, Lin L, Jin X, et al. Sex difference in incidence of gastric cancer: an international comparative study based on the Global Burden of Disease Study 2017. *BMJ Open* 2020;10(1):e033323.
https://doi.org/10.1136/bmjopen-2019-033323
14. Choi Y, Kim N, Kim KW, Jo HH, Park J, Yoon H, et al. Sex-based differences in histology, staging, and prognosis among 2983 gastric cancer surgery patients. *World J Gastroenterol* 2022;28(9):933-947.
https://doi.org/10.3748/wjg.v28.i9.933
15. Sanikini H, Biessy C, Rinaldi S, Navionis AS, Gicquiau A, Keski-Rahkonen P, et al. Circulating hormones and risk of gastric cancer by subsite in three cohort studies. *Gastric Cancer* 2023;26(6):969-987.
https://doi.org/10.1007/s10120-023-01414-0
16. Kyoung DS, Kim HS. Understanding and utilizing claim data from the Korean National Health Insurance Service (NHIS) and Health Insurance Review & Assessment (HIRA) Database for Research. *J Lipid Atheroscler* 2022;11(2):103-110.
https://doi.org/10.12997/jla.2022.11.2.103
17. Kim MK, Han K, Lee SH. Current trends of big data research using the Korean national health information database. *Diabetes Metab J* 2022;46(4):552-563.
https://doi.org/10.4093/dmj.2022.0193
18. Kim KK, Haam JH, Kim BT, Kim EM, Park JH, Rhee SY, et al. Evaluation and treatment of obesity and its comorbidities: 2022 update of clinical practice guidelines for obesity by the Korean Society for the Study of Obesity. *J Obes Metab Syndr* 2023;32(1):1-24.
https://doi.org/10.7570/jomes23016
19. Samanic C, Gridley G, Chow WH, Lubin J, Hoover RN, Fraumeni JF Jr. Obesity and cancer risk among white and black United States veterans. *Cancer Causes Control* 2004;15(1):35-43.
https://doi.org/10.1023/B:CACO.0000016573.79453.ba
20. Lindblad M, Rodríguez LAG, Lagergren J. Body mass, tobacco and alcohol and risk of esophageal, gastric cardia, and gastric non-cardia adenocarcinoma among men and women in a nested case-control study. *Cancer Causes Control* 2005;16(3):285-294.
https://doi.org/10.1007/s10552-004-3485-7
21. MacInnis RJ, English DR, Hopper JL, Giles GG. Body size and composition and the risk of gastric and oesophageal adenocarcinoma. *Int J Cancer* 2006;118(10):2628-2631.
https://doi.org/10.1002/ijc.21638
22. Samanic C, Chow WH, Gridley G, Jarvholm B, Fraumeni JF Jr. Relation of body mass index to cancer risk in 362,552 Swedish men. *Cancer Causes Control* 2006;17(7):901-909.
https://doi.org/10.1007/s10552-006-0023-9
23. Merry AHH, Schouten LJ, Goldbohm RA, van den Brandt PA. Body mass index, height and risk of adenocarcinoma of the oesophagus and gastric cardia: a prospective cohort study. *Gut* 2007;56(11):1503-1511.
https://doi.org/10.1136/gut.2006.116665

24. Corley DA, Kubo A, Zhao W. Abdominal obesity and the risk of esophageal and gastric cardia carcinomas. *Cancer Epidemiol Biomarkers Prev* 2008;17(2):352-358. https://doi.org/10.1158/1055-9965.EPI-07-0748
25. O'Doherty MG, Freedman ND, Hollenbeck AR, Schatzkin A, Abnet CC. A prospective cohort study of obesity and risk of oesophageal and gastric adenocarcinoma in the NIH–AARP Diet and Health Study. *Gut* 2012;61(9):1261-1268. https://doi.org/10.1136/gutjnl-2011-300551
26. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K, et al. Body fatness and cancer: viewpoint of the IARC Working Group. *N Engl J Med* 2016;375(8):794-798. https://doi.org/10.1056/NEJMsr1606602
27. World Cancer Research Fund, American Institute for Cancer Research. Diet, nutrition, physical activity and oesophageal cancer. London: World Cancer Research Fund International; 2018.
28. Derakhshan MH, Malekzadeh R, Watabe H, Yazdanbod A, Fyfe V, Kazemi A, et al. Combination of gastric atrophy, reflux symptoms and histological subtype indicates two distinct aetiologies of gastric cardia cancer. *Gut* 2008;57(3):298-305. https://doi.org/10.1136/gut.2007.137364
29. Alemán JO, Eusebi LH, Ricciardiello L, Patidar K, Sanyal AJ, Holt PR. Mechanisms of obesity-induced gastrointestinal neoplasia. *Gastroenterol Hepatol* 2014;146(2):357-373. https://doi.org/10.1053/j.gastro.2013.11.051
30. Olefson S, Moss SF. Obesity and related risk factors in gastric cardia adenocarcinoma. *Gastric Cancer* 2015;18(1):23-32. https://doi.org/10.1007/s10120-014-0425-4
31. Cameron AJ, Lomboy CT, Pera M, Carpenter HA. Adenocarcinoma of the esophagogastric junction and Barrett's esophagus. *Gastroenterology* 1995;109(5):1541-1546. https://doi.org/10.1016/0016-5085(95)90642-8
32. Fullard M, Kang JY, Neild P, Poullis A, Maxwell JD. Systematic review: does gastro-oesophageal reflux disease progress? *Aliment Pharmacol Ther* 2006;24(1):33-45. https://doi.org/10.1111/j.1365-2036.2006.02963.x
33. Cohen DH, LeRoith D. Obesity, type 2 diabetes, and cancer: the insulin and IGF connection. *Endocr Relat Cancer* 2012;19(5):F27-F45. https://doi.org/10.1530/ERC-11-0374
34. Tseng CH, Chen CJ, Landolph JR. Diabetes and cancer: epidemiological, clinical, and experimental perspectives. *Exp Diabetes Res* 2012;2012:101802. https://doi.org/10.1155/2012/101802
35. Muscogiuri G, Verde L, Vetrani C, Barrea L, Savastano S, Colao A. Obesity: a gender-view. *J Endocrinol Invest* 2024;47(2):299-306. https://doi.org/10.1007/s40618-023-02196-z
36. Guglielmi V, Sbraccia P. Obesity phenotypes: depot-diferences in adipose tissue and their clinical implications. *Eat Weight Disord* 2018;23(1):3-14.
37. Chandanos E, Lagergren J. Oestrogen and the enigmatic male predominance of gastric cancer. *Eur J Cancer* 2008;44(16):2397-2403. https://doi.org/10.1016/j.ejca.2008.07.031
38. Camargo MC, Goto Y, Zabaleta J, Morgan DR, Correa P, Rabkin CS. Sex hormones, hormonal interventions, and gastric cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2012;21(1):20-38. https://doi.org/10.1158/1055-9965.EPI-11-0834
39. Kim SM, Min BH, Lee J, An JY, Lee JH, Sohn TS, et al. Protective effects of female reproductive factors on Lauren intestinal-type gastric adenocarcinoma. *Yonsei Med J* 2018;59(1):28-34. https://doi.org/10.3349/ymj.2018.59.1.28
40. Jo HH, Kim N, Jang J, Choi Y, Park J, Park YM, et al. Impact of body mass index on survival depending on sex in 14,688 patients with gastric cancer in a tertiary hospital in South Korea. *Gut Liver* 2023;17(2):243-258. https://doi.org/10.5009/gnl220104
41. Watabe H, Mitsushima T, Derakhshan MH, Yamaji Y, Okamoto M, Kawabe T, et al. Study of association between atrophic gastritis and body mass index: a cross-sectional study in 10,197 Japanese subjects. *Dig Dis Sci* 2009;54(5):988-995. https://doi.org/10.1007/s10620-008-0468-7

Review

Ewha Med J 2024;47(2):e20
https://doi.org/10.12771/emj.2024.e20
eISSN 2234-2591

Exposure to air pollution and precocious puberty: a systematic review

Rosie Lee[1], Jongmin Oh[2,3,4], Eunji Mun[1], Jung Eun Choi[1], Kyung Hee Kim[1], Ji Hyen Lee[1,3], Hae Soon Kim[1,3], Eunhee Ha[2,3,5]

[1]Department of Pediatrics, Ewha Womans University College of Medicine, Seoul, Korea
[2]Department of Environmental Medicine, Ewha Womans University College of Medicine, Seoul, Korea
[3]Institute of Ewha-SCL for Environmental Health (IESEH), Ewha Womans University College of Medicine, Seoul, Korea
[4]Department of Human Systems Medicine, College of Medicine, Seoul National University, Seoul, Korea
[5]System Health & Engineering Major in Graduate School (BK21 Plus Program), Ewha Womans University, Seoul, Korea

Received Feb 28, 2024
Revised Apr 21, 2024
Accepted Apr 22, 2024

Corresponding author
Hae Soon Kim
Department of Pediatrics, Ewha Womans University College of Medicine, 260, Gonghang-daero, Gangseo-gu, Seoul, 07804, Korea
E-mail: hyesk@ewha.ac.kr

Eunhee Ha
Department of Environmental Medicine, Ewha Womans University College of Medicine, 25, Magokdong-ro 2-gil, Gangseo-gu, Seoul 07804, Korea
E-mail: eunheeha@ewha.ac.kr

Keywords
Air pollution; Endocrine disruptors; Menarche; Particulate matter; Puberty, precocious

The worldwide incidence of precocious puberty, which is associated with negative health outcomes, is increasing. Several studies have suggested that environmental factors contribute to the development of precocious puberty alongside genetic factors. Some epidemiological studies have provided limited evidence suggesting an association between exposure to air pollution and changes in pubertal development. This systematic review aimed to summarize existing evidence on the association between air pollution exposure and precocious puberty. Following the Preferred Reporting Items for Systematic reviews and Meta-Analyses guidelines, we searched two databases (PubMed and Web of Science) until August 2023. The included studies assessed the association between air pollutant exposure and the risk of precocious puberty, early menarche, or pubertal development. Two authors independently performed study selection and data extraction. A meta-analysis and analysis of the risk of bias were infeasible due to the limited number of studies and the heterogeneity among them. The literature search resulted in 184 studies, from which we included six studies with sample sizes ranging from 437 to 4,074 participants. The studies reported heterogeneous outcomes. Four studies found that increased exposure to air pollution was related to earlier pubertal onset. One study was inconclusive, and another suggested that air pollutant exposure may delay the onset of thelarche. Most studies suggest that exposure to air pollutants accelerates pubertal development; however, the results from the available studies are inconsistent. More extensive and well-designed longitudinal studies are required for a comprehensive understanding of the association between air pollution and precocious puberty.

Introduction

Background

The increasing incidence of precocious puberty is emerging as a significant medical and social issue worldwide [1,2]. A meta-analysis from 2020 reported a trend of breast development beginning approximately 0.24 years earlier every decade from 1977 to 2023 [3]. Additionally, there has been a notable decrease in the age of menarche from the 19th to the 20th century [4]. The onset of puberty is determined by both genetic and environmental factors [5,6]. Recent studies have highlighted that non-genetic lifestyle factors, such as adiposity, exposure to endocrine-disrupting chemicals (EDCs), and air pollution, might influence the timing of pubertal

onset [7]. Furthermore, several epidemiological studies have found that exposure to ambient air pollution is linked to an earlier onset of menarche in girls [8,9].

Ambient air pollution consists of a mixture of particulate matter (PM) and gaseous pollutants that originate from both human activities and natural sources. This pollution primarily includes sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), and ozone (O_3). PM contains various components such as heavy metals, polycyclic aromatic hydrocarbons, and EDCs, all of which can interfere with the endocrine system [10,11]. When inhaled, ambient air pollution can enter the human bloodstream and be transported to various organs [12], leading to a range of health outcomes, including endocrine disruption [10]. PM can interact with estrogen receptors, triggering the release of kisspeptin, which in turn stimulates the secretion of gonadotropin-releasing hormone, thereby initiating the onset of puberty [13]. Moreover, fine particulate matter ($PM_{2.5}$), with a diameter of less than 2.5 μm, as opposed to particulate matter (PM_{10}), which has a diameter of less than 10 μm, can penetrate deeper into the body upon inhalation, potentially causing more severe adverse effects.

Currently, epidemiological studies that investigate the association between exposure to air pollution and pubertal development are limited, and their findings are inconsistent. It is important to note that, to date, no systematic review has been conducted on the relationship between exposure to air pollution and pubertal development.

Objectives

This systematic review investigated the impact of exposure to ambient air pollution on pubertal development and the risk of precocious puberty.

Methods

Ethics statement

Since this research did not involve any direct human participants or human-derived materials, it did not require approval from an institutional review board or the obtainment of informed consent.

Study design

We conducted a systematic review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (Fig. 1). The study protocol was registered with PROSPERO on September 19, 2023 (CRD42023465050). Revisions to the protocol were necessary as assessing the risk of bias and conducting a meta-analysis proved infeasible due to the heterogeneity among the included studies.

Eligibility criteria

To investigate our systematic review, we defined the population, exposures, comparison, outcomes, and study designs (PECOS) parameters. The details of PECOS are as follows: a) population: infants, children, adolescents, or pediatric patients; b) exposure: air pollutants, such as $PM_{2.5}$, PM_{10}, SO_2, NO, NO_2, or O_3; c) comparison: exposure to lower or higher levels of each specific type of air pollutants in the same population or in a control population; with provision of a measure of risk (e.g., relative risk [RR], OR, hazard ratio [HR], or mean difference [MD]); d) outcome: precocious puberty risk, early menarche risk, or pubertal development stage; and e) study design: human epidemiological studies, including prospective and retrospective cohort

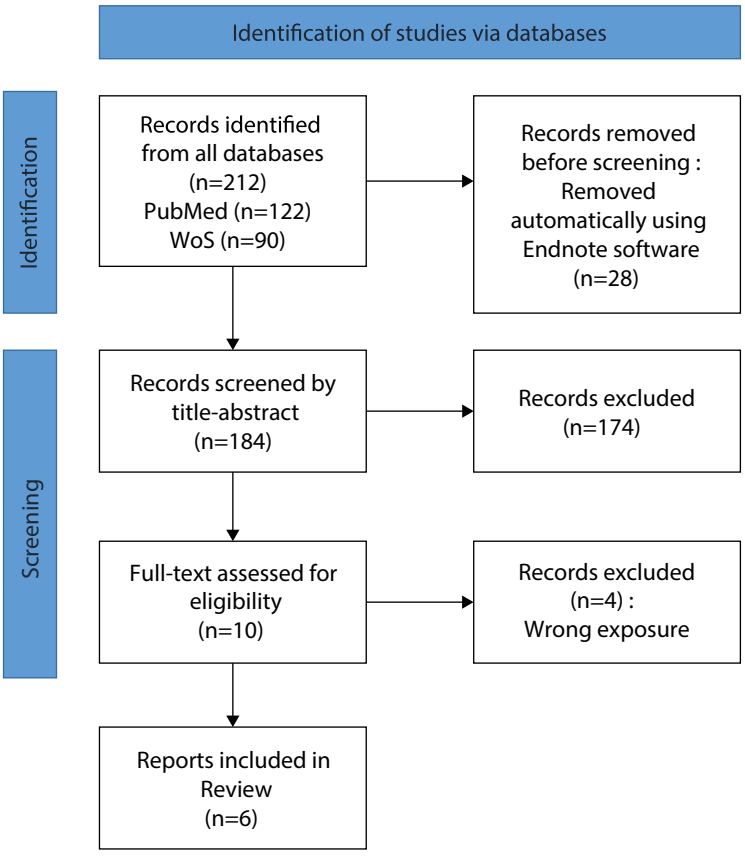


Fig. 1. Flow diagram summarizing the process of literature search and selection. WoS, Web of Science.

studies, case-control, and cross-sectional studies. We excluded abstracts, case reports, editorials, animal studies, *in vivo* studies, and commentaries.

Information sources

We performed a systematic search of the 1) PubMed and 2) Web of Science (WoS) databases from their inception to August 23, 2023. Only articles in English were included.

Search strategy

The search terms for the search strategy can be found in Supplement 1. Duplicate articles were eliminated using EndNote and the manual method.

Selection process

Two authors (RSL and EJM) independently reviewed the titles and abstracts and selected potentially eligible articles. Subsequently, the full texts of selected studies were examined by two authors (RSL and EJM) with the participation of other authors (JMO, KHK, JHL, HSK, and EHH) to address any disagreements and reach a consensus.

Data collection process

Two authors (RSL and JMO) extracted the following data of interest in a Microsoft Excel sheet: first author, year of publication, country, study design, follow-up period, sample size,

population characteristics, exposure details, details on outcome assessment, and confounders. Furthermore, we also extracted measures of effect (i.e., RR, OR, HR, MD, and time ratio [TR]) for the association between air pollution and precocious puberty.

Data synthesis

In our initial protocol for the systematic review, we planned to conduct a meta-analysis if at least three studies shared similar study designs, analysis methods, and effect sizes. However, a meta-analysis proved infeasible due to the limited number of studies (<3) and the diverse range of study designs, including case-crossover, cohort, and cross-sectional studies. Additionally, the outcomes varied among the studies, encompassing the risk of precocious puberty, the risk of early menarche, and stages of pubertal development. Given the heterogeneity among the included studies, we created a table to outline the characteristics of the studies and the relationships between the exposures and outcomes. As a result, we were unable to perform a meta-analysis and instead provided a descriptive summary.

Results

Study selection

In our systematic search, we identified 184 studies out of the initial 212 (PubMed: 122 and WoS: 90), following the removal of duplicates. Of these, 10 full-text studies were assessed for eligibility, with 6 ultimately being included [8,9,14–17]. Fig. 1 illustrates the flow diagram of the study selection process. The studies that were excluded, along with the reasons for their exclusion, are detailed in Supplement 2.

Characteristics of the included studies

Table 1 summarizes the characteristics of the six studies. Three studies were designed as cohorts [8,16,17], two were conducted as cross-sectional studies [9,15], and one was a case-crossover study [14]. These studies were conducted in China, Hong Kong, the USA, South Korea, Poland, and Germany. The sample sizes ranged from 437 to 4,074 participants, with follow-up durations spanning from 3 to 14 years. Five studies investigated exposure to ambient air pollutants such as $PM_{2.5}$, PM_{10}, SO_2, NO, NO_2, and O_3 [8,9,14–16], whereas one study analyzed traffic-related metrics [17]. Yang et al. measured air pollutant levels using inverse distance weighting; Wronka et al. assessed air quality based on data from the chief inspectorate for environmental protection; Zhao et al. utilized a land-use regression model; Jung et al. relied on air monitoring network stations; Huang et al. gathered data from a monitoring station; and McGuinn et al. used annualized traffic data from the California Department of Transportation Highway Performance Monitoring System. The outcomes of these studies varied. Two studies [8,9] focused on the risk of early menarche as their primary outcome, with the age at menarche being self-reported. One study [14] examined the incidence of precocious puberty, defined by a professional pediatrician as the activation of the hypothalamic-pituitary-gonadal axis function and the onset of secondary sexual characteristics before the age of 8 in girls and 9 in boys. The remaining studies [15–17] assessed pubertal development at a specific age using Tanner staging or measured sex hormone levels.

Synthesis of results

Relationship between air pollution exposure and pubertal development

Table 2 summarizes the findings on the impact of air pollution on pubertal development,

Table 1. Characteristics of studies assessing the association between air pollutant exposure and pubertal development

Author	Country	Study design	Study period	Sample size	Age (years)	Sex Female (%)	Measured pollutants	Outcomes
Yang et al. [14]	China	Case-crossover	2015–2021	2,201	7.47±1.24	96.6	PM_{10}, $PM_{2.5}$, SO_2, NO_2, O_3	The risk of precocious puberty
Wronka et al. [8]	Poland	Longitudinal cohort	2015–2018*	1,257	19–25	100	PM_{10}, $PM_{2.5}$, benzene, SO_2, NO	The risk of early menarche
Zhao et al. [15]	Germany	Cross-sectional	1995–2009†	1,945	10	48.4	PM_{10}, $PM_{2.5}$, NO_2, O_3	Pubertal development at age 10 years assessed with estradiol and testosterone
Jung et al. [9]	South Korea	Cross-sectional	2010–2012	639	13–17	100	PM_{10}	The risk of early menarche
Huang et al. [16]	Hong Kong	Birth cohort	1997–2008‡	4,074	9–12	47.5	PM_{10}, SO_2, NO, NO_2	Pubertal stage at age 11 years assessed with Tanner stage
McGuinn et al. [17]	USA	Longitudinal cohort	2005–2012	437	6–8	100	Traffic metrics	Pubertal stage at 6–8 years assessed with Tanner stage

PM, particulate matter; SO_2, sulfur dioxide; NO_2; nitrogen dioxide; O_3, ozone.
*Girls were born between 1993 and 1998.
†The authors assessed children born between 1995 and 1999 when they reached the age of 10, between 2005 and 2009.
‡The authors assessed children born in 1997 when they reached the age of 9–12, between 2005 and 2008.

precocious puberty, and age at menarche across various studies. Four out of six studies [8,9,14,17] indicated that exposure to air pollution accelerates pubertal development stages and promotes precocious puberty. Jung et al. found that a 1 μg/m^3 increase in PM_{10} was associated with a higher risk of early menarche (OR=1.08; 95% CI, 1.04–1.12) and accelerated age at menarche by 0.046 years (95% CI, −0.064 to −0.027) on a 1-year average. The authors reported that the results were consistent across the 2-year average (OR=1.06; 95% CI, 1.02–1.10; 0.038 years; 95% CI, −0.059 to −0.018) and 3-year model average (OR=1.05; 95% CI, 1.01–1.09; 0.031 years; 95% CI, −0.047 to −0.015). Wronka et al. found that the risk of early menarche (age below 11) was higher in the group living in areas with high PM levels. The ORs were calculated as 3.18 (95% CI, 2.29–4.69) for PM_{10} and 3.25 (95% CI, 2.34–4.8) for $PM_{2.5}$. Yang et al. used a distributed lag nonlinear model to determine the OR of the lag effect of $PM_{2.5}$ and PM_{10} on the incidence of precocious puberty. They reported that the most significant effects of $PM_{2.5}$ and PM_{10} on precocious puberty were observed in lag 27 (OR=1.72; 95% CI, 1.01–2.92) and lag 16 (OR=1.95; 95% CI, 1.33–2.85), respectively. McGuinn et al. found that girls living within 150 m of a major road or highway had a higher likelihood of experiencing early pubarche (TR=0.96; 95% CI, 0.93–0.99), but not thelarche (TR=0.99; 95% CI, 0.97–1.02). The authors used accelerated failure time models, and calculated TRs, where a TR of <1.0 indicated an earlier age at pubertal development than the reference group. In contrast, Huang et al. reported that exposure to PM_{10} during the prenatal and infantile periods could delay thelarche. Exposure to PM_{10}, SO_2, NO, and NO_2 was considered as z-scores for comparability, and the outcomes were the MD in Tanner stage per SD increment in each type of air pollutant. In girls, higher PM_{10} exposure *in utero* (MD: −0.05; 95% CI, −0.08 to −0.02) and in infancy (MD: −0.03; 95% CI, −0.06 to −1.2) was associated with a lower pubertal stage. In boys, higher SO_2 exposure *in utero* (MD: −0.03; 95% CI, −0.05, −0.01) and during childhood (MD: −0.06; 95% CI, −0.08, −0.04) were associated with lower pubertal

https://doi.org/10.12771/emj.2024.e20

Table 2. Relationship between air pollutant exposure and pubertal development

Author	Air pollutant exposure		Outcomes	Effect measure	Result with 95% CI	Adjusted variables	Average concentration of main air pollutants
Yang et al. [14]	PM_{10}		The risk of precocious puberty	OR	1.95 (1.33–2.85)†	Age, sex, SO_2, NO_2, CO, O_3	69.77±41.07 μg/m^3
	$PM_{2.5}$				1.72 (1.01–2.92)†		38.81±26.36 μg/m^3
Wronka et al. [8]	PM_{10}	Low* Medium High	The risk of early menarche	OR	3.18 (2.29–4.69)†	Urbanization, mother's education, father's education, number of siblings, financial conditions	NA
	$PM_{2.5}$				3.25 (2.34–4.80)†		
	Benzene				1.11 (0.90–1.64)		
	SO_2				1.22 (1.01–2.14)		
	NO				1.47 (0.65–1.35)		
Zhao et al. [15]	PM_{10}	10 μg/m^3 increase in pollutant concentrations	Pubertal development at age 10 years assessed with estradiol and testosterone	OR	Female: 0.896 (0.379–2.122) Male: 0.821 (0.383–1.759)	Age, sex, body mass index, secondhand smoke exposure, time spent outside and in front of a screen, physical activity level, season, and time of the blood sampling, household income, parental education, maternal age at birth, single parent status	21.95±3.26 μg/m^3
	$PM_{2.5}$				Female: 0.163 (0.022–1.166) Male: 1.089 (0.156–7.605)		14.76±2.13 μg/m^3
	NO_2				Female: 0.892 (0.581-1.369) Male: 1.152 (0.768–1.728)		22.03±3.86 μg/m^3
	O_3				Female: 0.900 (0.605–1.339) Male: 0.830 (0.573–1.203)		69.18±4.9 μg/m^3
Jung et al. [9]	PM_{10}	1 μg/m^3 increase in PM_{10}	The risk of early menarche	OR	1.08 (1.04–1.12)†	Body mass index, city size, household income level, maternal age at menarche, second-hand smoke exposure at home	NA
Huang et al. [16]	PM_{10}	SD increase in pollutants	Pubertal stage at age 11 years assessed with Tanner stage	Mean difference in Tanner stage	Female *In utero*: −0.05 (−0.08 to −0.02)† Infancy: −0.03 (−0.06 to −1.2)†	Neighborhood and household income per person, mother's migration status, highest parental educational level, age, maternal age at birth, parity, maternal smoking	NA
	SO_2				Male *In utero*: −0.03 (−0.05 to −0.01)† Childhood: −0.06 (−0.08 to −0.04)†		
	NO				Statistically insignificant		
	NO_2				Male *In utero*: −0.03 (−0.04 to −0.02)† Childhood: −0.02 (−0.04 to −0.01)†		
McGuinn et al. [17]	Traffic metrics	Distance to road (meters)	Pubertal stage at 6–8 years assessed with Tanner stage	TR	Pubarche: 0.96 (0.93–0.99)† Thelarche: 0.99 (0.97–1.02)	Race/ethnicity, household income, girl's cotinine level	NA

PM, particulate matter; SO_2, sulfur dioxide; NO_2; nitrogen dioxide; CO, carbon monoxide; O_3, ozone; TR, time ratio; NA, not applicable.

*Low: annual pollutant values and the number of days per year with exceedances were below the allowable limit; Medium: annual values below the permissible limit, but with the number of days exceeding the normal above the limit; High: included zones above the limit.

†Statistically significant results.

https://doi.org/10.12771/emj.2024.e20

stage. Furthermore, higher NO_2 exposure *in utero* (MD: −0.03; 95% CI, −0.04, −0.02) and during childhood (MD: −0.02; 95% CI, −0.04, −0.01) was associated with a lower pubertal stage. Zhao et al. found no statistically significant associations between air pollution exposure and pubertal development as assessed using serum sex hormone levels.

Discussion

Key results

In this systematic review, we found that four out of six studies indicated a relationship between increased exposure to air pollution and earlier onset of puberty [8,9,14,17]. One study produced inconclusive results [15], while another suggested that exposure to air pollutants might delay the onset of thelarche [16].

There has been no prior systematic review examining the effects of air pollution exposure on pubertal development and precocious puberty. The studies we included employed various research designs, such as cohort, cross-sectional, and case-crossover studies, featured different sample sizes, and tracked participants over varying lengths of time. These studies also investigated different exposures and outcomes and utilized a range of methods to measure exposure. We recognize that this diversity in study design could account for the inconsistent results regarding the impact of air pollution on precocious puberty. Nevertheless, the authors are inclined to believe that air pollution adversely affects precocious puberty, based on the accumulating evidence that air pollutants influence pubertal development through various mechanisms, which are not yet fully understood.

Interpretation and comparison with previous studies

When chemicals and heavy metals with endocrine-disrupting properties are released into the air from industrial emissions, vehicle emissions, and waste combustion, they can bind to PM [18]. Polycyclic aromatic hydrocarbons and heavy metals in PM, particularly from fossil fuel combustion, are recognized as endocrine disruptors due to their ability to activate aryl hydrocarbon, androgen, or estrogen receptors [19]. PM can act on estrogen receptors, triggering the release of kisspeptin, which subsequently initiates the secretion of gonadotropin-releasing hormone, thus starting the onset of puberty [13]. Epigenetic disruption caused by PM is a potential mechanism for triggering puberty through neuroendocrine components [20]. Moreover, PM can induce oxidative stress and systemic inflammation upon entering the respiratory tract [21]. Endocrine disruptors attached to PM influence hormone synthesis in endocrine glands or disrupt hormone transport to target organs. Research using mixtures of EDCs found in indoor air samples has shown that these compounds exhibit estrogenic and androgenic activities when tested in *in vitro* assay systems [22]. There is increasing evidence that certain EDCs are associated with various human health issues, such as reproductive problems in both females and males, and precocious puberty in children, as indicated in several previously reported systematic reviews [23–28].

Although the exact mechanisms connecting air pollutants and puberty onset remain unclear, the presence of EDCs in air, including PM, and their potential impact on puberty are areas that require active research.

Strengths and limitations of the included studies

These studies stand out for their pioneering research into the relationship between exposure to air pollutants and pubertal development. McGuinn et al. [17] conducted the first study

to explore the link between early life proximity to traffic and pubertal development within a multiethnic cohort. Huang et al. utilized a population-representative birth cohort and gathered clinical data on pubertal stages. Jung et al. [9] analyzed data from the fifth Korea National Health and Nutrition Examination Study, a representative sample of the South Korean population, and found consistent results across various models after making adjustments. Wronka et al. [8] carried out the inaugural study in a European country investigating the connection between air pollutant exposure and early menarche. Zhao et al. [15] also drew on data from two relatively large birth cohorts, with numerous relevant covariates available for adjustment. Yang et al. [14] employed a case-crossover study design and a distributed nonlinear model to evaluate the association between PM_{10} and $PM_{2.5}$ levels and the risk of precocious puberty. Zhao et al. [15] and Jung et al. [9] considered body mass index as a confounding factor, acknowledging that nongenetic lifestyle factors such as adiposity can influence the onset of puberty.

Nonetheless, the included studies have several limitations. First, exposure to air pollutants was estimated using data averaged for specific geographic areas, and the limited number of air quality monitoring stations may not provide a precise representation of individual exposure. Consequently, exposure misclassification and a lack of individual-level data are possible issues. Second, some studies were susceptible to recall bias, particularly those relying on self-reported data such as age at menarche. Additionally, most studies focused on single pollutant exposure, whereas in reality, humans are exposed to a complex mixture of air pollutants [29]. Moreover, the study designs varied, encompassing different exposures to air pollutants, and the outcomes included the incidence of precocious puberty, age at menarche, and pubertal stage at a specific age. This heterogeneity in study design and outcomes made it infeasible to conduct a meta-analysis and evaluate the risk of bias.

Recommendations for future studies and health implications

Future researchers should consider conducting prospective birth cohort studies to assess the long-term consequences of prenatal and postnatal exposure to air pollution. This is crucial because EDCs can function as obesogens during pregnancy, potentially altering fetal programming [30], and air pollution may have similar effects. Furthermore, to accurately assess individual exposure to air pollution, it is feasible to use advanced technologies and methods, such as personal monitoring devices. Expanding the study population to include diverse socioeconomic groups and geographical locations can enhance the generalizability of the findings. *In vitro* research is also necessary to understand the biological mechanisms underlying the association between air pollution and pubertal development. Additionally, research should focus on developing prevention policies and interventions aimed at mitigating the impact of air pollution. Air pollution is a global concern, and international collaboration among researchers and institutions worldwide can lead to a more comprehensive understanding of its effects on health.

Strengths and limitations of this review

To the best of our knowledge, this is the first systematic review to provide evidence of the impact of air pollution exposure on precocious puberty and pubertal development. Additionally, all review processes underwent peer review, adhering to the PRISMA guidelines. However, due to the heterogeneity among the studies, it was not feasible to conduct a meta-analysis to evaluate the combined effect of air pollution on the risk of precocious puberty. Future research on this topic is necessary, and as more studies become available, we aim to gather sufficient evidence to conduct a meta-analysis.

Conclusion

The evidence suggests that exposure to air pollution may lead to an earlier onset of puberty, although the results of studies have been inconsistent. To address this, further longitudinal studies are needed that accurately assess individual exposure to multiple air pollutants over extended periods. It is crucial to promote policies aimed at reducing exposure to air pollution. Additionally, sharing international data and conducting collaborative studies could provide valuable insights for developing preventive policies concerning exposure to air pollutants.

ORCID
Rosie Lee: https://orcid.org/0000-0003-3285-3916
Jongmin Oh: https://orcid.org/0000-0002-2980-6943
Eunji Mun: https://orcid.org/0009-0008-8590-1057
Jung Eun Choi: https://orcid.org/0000-0001-8956-4192
Kyung Hee Kim: https://orcid.org/0000-0002-3795-4671
Ji Hyen Lee: https://orcid.org/0000-0002-2234-1055
Hae Soon Kim: https://orcid.org/0000-0002-6976-6878
Eunhee Ha: https://orcid.org/0000-0002-4224-3858

Authors' contributions
Project administration: Kim HS, Ha E
Conceptualization: Lee R, Oh J, Mun E, Choi JE, Kim KH, Lee JH, Kim HS, Ha E
Methodology & data curation: Kim HS, Ha E
Funding acquisition: Kim HS, Ha E
Writing – original draft: Lee R
Writing – review & editing: Lee R, Oh J, Mun E, Choi JE, Kim KH, Lee JH, Kim HS, Ha E

Conflict of interest
Eunhee Ha has been a dean of the Ewha Womans University College of Medicine since August 2021; however, she was not involved in the peer review process or decision-making. Otherwise, no potential conflict of interest relevant to this article was reported.

Funding
This study was supported by a project titled "Institute of Ewha-SCL for Environmental Health (IESEH)" and Research of Environmental Examination Model for Children and Women (No. 1-2022-0205-001-2).

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Supplementary materials are available from: https://doi.org/10.12771/emj.2024.e20.

Supplement 1. Keywords used for the systematic review
Supplement 2. Reasons for excluding studies from the systematic review

References

1. Kim YJ, Kwon A, Jung MK, Kim KE, Suh J, Chae HW, et al. Incidence and prevalence of central precocious puberty in Korea: an epidemiologic study based on a national database. *J Pediatr* 2019;208:221-228. https://doi.org/10.1016/j.jpeds.2018.12.022
2. Bräuner EV, Busch AS, Eckert-Lind C, Koch T, Hickey M, Juul A. Trends in the incidence of central precocious puberty and normal variant puberty among children in Denmark, 1998 to 2017. *JAMA Netw Open* 2020;3(10):e2015665. https://doi.org/10.1001/jamanetworkopen.2020.15665
3. Eckert-Lind C, Busch AS, Petersen JH, Biro FM, Butler G, Bräuner EV, et al. Worldwide secular trends in age at pubertal onset assessed by breast development among girls: a systematic review and meta-analysis. *JAMA Pediatr* 2020;174(4):e195881. https://doi.org/10.1001/jamapediatrics.2019.5881
4. Sørensen K, Mouritsen A, Aksglaede L, Hagen CP, Mogensen SS, Juul A. Recent secular trends in pubertal timing: implications for evaluation and diagnosis of precocious puberty. *Horm Res Paediatr* 2012;77(3):137-145.

https://doi.org/10.1159/000336325
5. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr Rev* 2003;24(5):668-693. https://doi.org/10.1210/er.2002-0019
6. Fisher MM, Eugster EA. What is in our environment that effects puberty? *Reprod Toxicol* 2014;44:7-14. https://doi.org/10.1016/j.reprotox.2013.03.012
7. Lucaccioni L, Trevisani V, Marrozzini L, Bertoncelli N, Predieri B, Lugli L, et al. Endocrine-disrupting chemicals and their effects during female puberty: a review of current evidence. *Int J Mol Sci* 2020;21(6):2078. https://doi.org/10.3390/ijms21062078
8. Wronka I, Kliś K. Effect of air pollution on age at menarche in polish females, born 1993–1998. *Sci Rep* 2022;12(1):4820. https://doi.org/10.1038/s41598-022-08577-3
9. Jung EM, Kim HS, Park H, Ye S, Lee D, Ha EH. Does exposure to PM10 decrease age at menarche? *Environ Int* 2018;117:16-21. https://doi.org/10.1016/j.envint.2018.04.020
10. Darbre PD. Overview of air pollution and endocrine disorders. *Int J Gen Med* 2018;11:191-207. https://doi.org/10.2147/IJGM.S102230
11. Anderson JO, Thundiyil JG, Stolbach A. Clearing the air: a review of the effects of particulate matter air pollution on human health. *J Med Toxicol* 2012;8(2):166-175. https://doi.org/10.1007/s13181-011-0203-1
12. Beelen R, Raaschou-Nielsen O, Stafoggia M, Andersen ZJ, Weinmayr G, Hoffmann B, et al. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet* 2014;383(9919):785-795. https://doi.org/10.1016/S0140-6736(13)62158-3
13. Mouritsen A, Aksglaede L, Sørensen K, Sloth Mogensen S, Leffers H, Main KM, et al. Hypothesis: exposure to endocrine-disrupting chemicals may interfere with timing of puberty. *Int J Androl* 2010;33(2):346-359. https://doi.org/10.1111/j.1365-2605.2010.01051.x
14. Yang H, Ge A, Xie H, Li W, Qin Y, Yang W, et al. Effects of ambient air pollution on precocious puberty: a case-crossover analysis in Nanjing, China. *J Clin Med* 2023;12(1):282. https://doi.org/10.3390/jcm12010282
15. Zhao T, Triebner K, Markevych I, Standl M, Altug H, de Hoogh K, et al. Outdoor air pollution and hormone-assessed pubertal development in children: results from the GINIplus and LISA birth cohorts. *Environ Int* 2021;152:106476. https://doi.org/10.1016/j.envint.2021.106476
16. Huang JV, Leung GM, Mary Schooling C. The association of air pollution with pubertal development: evidence from Hong Kong's "Children of 1997" birth cohort. *Am J Epidemiol* 2017;185(10):914-923. https://doi.org/10.1093/aje/kww200
17. McGuinn LA, Voss RW, Laurent CA, Greenspan LC, Kushi LH, Windham GC. Residential proximity to traffic and female pubertal development. *Environ Int* 2016;94:635-641. https://doi.org/10.1016/j.envint.2016.06.031
18. Kampa M, Castanas E. Human health effects of air pollution. *Environ Pollut* 2008;151(2):362-367. https://doi.org/10.1016/j.envpol.2007.06.012
19. Hombach-Klonisch S, Pocar P, Kietz S, Klonisch T. Molecular actions of polyhalogenated arylhydrocarbons (PAHs) in female reproduction. *Curr Med Chem* 2005;12(5):599-616. https://doi.org/10.2174/0929867310504050599
20. Rzeczkowska PA, Hou H, Wilson MD, Palmert MR. Epigenetics: a new player in the regulation of mammalian puberty. *Neuroendocrinology* 2014;99(3-4):139-155. https://doi.org/10.1159/000362559
21. Chuang KJ, Chan CC, Su TC, Lee CT, Tang CS. The effect of urban air pollution on inflammation, oxidative stress, coagulation, and autonomic dysfunction in young adults. *Am J Respir Crit Care Med* 2007;176(4):370-376. https://doi.org/10.1164/rccm.200611-1627OC
22. Oziol L, Alliot F, Botton J, Bimbot M, Huteau V, Levi Y, et al. First characterization of the endocrine-disrupting potential of indoor gaseous and particulate contamination: comparison with urban outdoor air (France). *Environ Sci Pollut Res* 2017;24(3):3142-3152. https://doi.org/10.1007/s11356-016-8045-7
23. Uldbjerg CS, Koch T, Lim YH, Gregersen LS, Olesen CS, Andersson AM, et al. Prenatal and postnatal exposures to endocrine disrupting chemicals and timing of pubertal onset in girls and boys: a systematic review and meta-analysis. *Hum Reprod Update* 2022;28(5):687-716. https://doi.org/10.1093/humupd/dmac013
24. Wen Y, Liu SD, Lei X, Ling YS, Luo Y, Liu Q. Association of PAEs with precocious puberty in children: a systematic review and meta-analysis. *Int J Environ Res Public Health* 2015;12(12):15254-15268. https://doi.org/10.3390/ijerph121214974
25. Lee YJ, Jung HW, Kim HY, Choi YJ, Lee YA. Early-life exposure to per- and poly-fluorinated alkyl substances and growth, adiposity, and puberty in children: a systematic review. *Front Endocrinol* 2021;12:683297. https://doi.org/10.3389/fendo.2021.683297
26. Bigambo FM, Sun H, Yan W, Wu D, Xia Y, Wang X, et al. Association between phenols exposure and earlier puberty in children: a systematic review and meta-analysis. *Environ Res* 2020;190:110056.

https://doi.org/10.1016/j.envres.2020.110056
27. Castiello F, Freire C. Exposure to non-persistent pesticides and puberty timing: a systematic review of the epidemiological evidence. *Eur J Endocrinol* 2021;184(6):733-749. https://doi.org/10.1530/EJE-20-1038
28. Golestanzadeh M, Riahi R, Kelishadi R. Association of phthalate exposure with precocious and delayed pubertal timing in girls and boys: a systematic review and meta-analysis. *Environ Sci Process Impacts* 2020;22(4):873-894. https://doi.org/10.1039/C9EM00512A
29. Vedal S, Kaufman JD. What does multi-pollutant air pollution research mean? *Am J Respir Crit Care Med* 2011;183(1):4-6. https://doi.org/10.1164/rccm.201009-1520ED
30. Roth CL, DiVall S. Consequences of early life programing by genetic and environmental influences: a synthesis regarding pubertal timing. *Endocr Dev* 2015;29:134-152. https://doi.org/10.1159/000438883

Review

Ewha Med J 2024;47(2):e21
https://doi.org/10.12771/emj.2024.e21
eISSN 2234-2591

Return to sports following arthroscopic Bankart repair: a narrative review

Shafira Widya Utami[1], **Savina Rifky Pratiwi**[1], **Mitchel**[2], **Karina Sylvana Gani**[2], **Erica Kholinne**[1,3]

[1]Faculty of Medicine, Universitas Trisakti, Jakarta, Indonesia
[2]Faculty of Medicine and Health Sciences, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
[3]Department of Orthopaedic and Traumatology, Gatam Institute Eka Hospital, BSD, Indonesia

Received Feb 19, 2024
Revised Apr 7, 2024
Accepted Apr 11, 2024

Corresponding author
Erica Kholinne
Gatam Institute, Eka Hospital, Faculty of Medicine, Universitas Trisakti, Jl. Kyai Tapa No.1, Jakarta 11440, Indonesia
E-mail: erica@trisakti.ac.id

Keywords
Bankart lesions; Joint instability; Proprioception; Return to sport; Shoulder joint

A Bankart lesion is a tear of the labrum, the ring of cartilage that encircles the shoulder joint socket, that can occur when the shoulder is dislocated. This injury frequently affects young athletes and is associated with shoulder instability. This review was performed to provide an overview of anterior shoulder instability, with an emphasis on rehabilitation and the return to sports following arthroscopic Bankart repair. We searched the Google Scholar and PubMed academic databases through February 18th, 2024, utilizing keywords including "arthroscopic Bankart repair" and "return to sports". Our findings indicate that athletes who undergo arthroscopic Bankart repair exhibit higher rates of returning to sports compared to those who receive other anterior shoulder stabilization procedures. Several factors are considered when determining readiness to return to athletics, including time elapsed since surgery, type of sport, strength, range of motion, pain, and proprioception. Surgeons typically advise athletes to wait approximately 6 months after surgery before resuming sports activities. They also recommend that athletes regain at least 80% of the strength of the uninjured shoulder or achieve strength levels comparable to those prior to the injury. Additionally, patients are expected to attain a full range of motion without pain, which should be symmetrical to the uninjured side, and demonstrate improved proprioception in the shoulder. The sport in which an athlete participates can also influence the timeline for return. Those involved in overhead sports, like baseball or tennis, often experience lower success rates in returning to their sport compared to athletes from other disciplines.

Introduction

Background

An efficiently functioning glenohumeral joint depends on the integrity and coordinated interaction of both static and dynamic components. The structures essential for maintaining normal shoulder function are particularly susceptible to injury and dislocation. Such dislocations frequently involve the glenoid labrum, bony rim, ligaments, capsule, and humeral head [1]. The incidence of anterior shoulder instability ranges from eight to 17 dislocations per 1,000 person-years. Anterior shoulder dislocation rates are notably high among young athletes, particularly in contact sports such as football and rugby [1–5]. Anterior shoulder instability has multiple causes; however, the capsulolabral complex and Bankart lesion are commonly observed in young patients. A Bankart lesion is characterized by an anterior and inferior detachment of the labrum from the glenoid, along with capsuloligamentous injury below the equator of the glenoid [6].

Arthroscopic techniques for anterior shoulder stabilization have advanced considerably over the past two decades [7]. The outcomes of arthroscopic Bankart repair (ABR) are comparable to those of open repair in terms of recurrence rates, range of motion (ROM), and complications [8–11]. Recent studies have indicated that athletes undergoing ABR exhibit a higher rate of return to sport (RTS) compared to those treated with other anterior shoulder stabilization methods [12]. However, the rate at which athletes experience RTS following ABR varies widely among individual studies [13].

Objectives

This review was conducted to summarize anterior shoulder instability, focusing on rehabilitation and RTS following an ABR procedure.

Methods

Ethics statement

The present study was a review based on a literature search; consequently, neither institutional review board approval nor informed consent was necessary.

Study design

This study was a narrative review based on a search of academic databases.

Setting

The study involved a literature search of the Google Scholar and PubMed databases through February 18th, 2024. Keywords and terms like "arthroscopy Bankart repair" and "return to sports" were employed. The inclusion criteria specified that articles must be written in English and assess the relationship between ABR and RTS.

Results

The search yielded 11 relevant studies that satisfied the inclusion criteria (Table 1). These articles covered the timeframe from surgery to the resumption of athletic activities. Most studies suggest that athletes typically experience RTS approximately 6 months after surgery.

Discussion

Bankart lesion

A Bankart lesion is characterized by an anterior and inferior detachment of the labrum from the glenoid, accompanied by an injury to the capsuloligamentous structures below the equator of the glenoid (Fig. 1). This type of lesion commonly results from a traumatic anterior glenohumeral dislocation and is particularly prevalent among younger individuals [14]. Additionally, a traumatic anterior glenohumeral dislocation can lead to an avulsion fracture of the anterior glenoid rim, which is termed a bony Bankart lesion [15–17]. The extent of bone loss is a crucial determinant in the likelihood of recurrent glenohumeral instability following stabilization surgery [18].

Mechanism of injury

Shoulder instability manifests through the disruption of the dynamic and static stabilizing

Table 1. Characteristics of the included studies

No.	Study	Year	Sample (N)	Mean age (years)	Surgical technique	Return to sport
1	Hurley et al. [2]	2021	156	28±8	ABR	12 weeks for contact in training, 6 months for full contact and competition
2	Harada et al. [6]	2023	50	16.8±1.7	ABR	6.6±2.7 months (range, 3–18 months) for return to sport, 9.3±4.0 months (range, 6–24 months) for competitions, and 10.6±4.3 months (range, 8–24 months) for complete return
3	Porcellini et al. [15]	2002	25	25.6	ABR	12 weeks for non-contact sports, 5 months for contact sports
4	Kelley et al. [32]	2021	62	18± 7	ABR	6.5±0.7 months
5	Blonna et al. [33]	2016	30	>18	ABR	3–5 months for non-collision sports, 6 months for collision sports
6	Sedeek et al. [34]	2008	37	26.3	ABR	3 months for non-contact sports, 4 months for contact sports
7	Ide et al. [40]	2004	55	20	ABR	4 months for contact sports, 8.1 months for overhead sports, and 3.6 months for non-contact and non-overhead sports
8	Buckup et al. [41]	2018	20	27.75	ABR	6 months for non-contact and non-overhead sports, 7 months for overhead and contact sports, and 10 months for competition
9	Gibson et al. [42]	2016	34	23	ABR	11 weeks
10	Wilson et al. [43]	2020	43	18.1±3.7	Arthroscopic shoulder stabilization surgery	6 months
12	Plath et al. [50]	2015	66	29.3±10.4	ABR	3 months for specific training, 6 months for overhead and high-contact sports

No., number; ABR, arthroscopic Bankart repair.

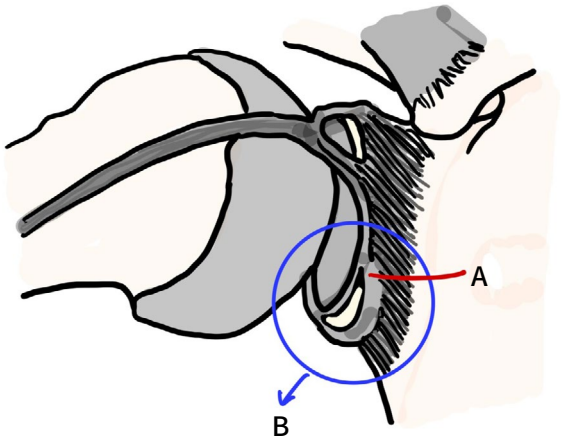


Fig. 1. Lesions of the shoulder. (A) Labral tear, (B) Bankart lesion.

elements of the glenohumeral joint, which can result in dislocation, subluxation, or a sensation of apprehension accompanied by pain. The stability of the shoulder is maintained by the glenoid labrum, the glenohumeral ligament complex, negative intra-articular pressure, and articular conformity. Furthermore, the rotator cuff and scapular stabilizers represent key dynamic contributors to shoulder restraint [1].

Anterior dislocation is the most common type of shoulder dislocation, accounting for approximately 97% of these injuries [19]. Anterior dislocation typically occurs when an individual falls with the arm abducted and externally rotated, causing the posterosuperior aspect of the humeral head to impact the anteroinferior aspect of the glenoid rim. This can result in damage to the humeral head, the glenoid labrum, or both (Fig. 2). Additionally, an indentation may develop on the humeral head due to a compression fracture, occurring when the humeral head is forced against the anterior glenoid rim during dislocation [20]. Rotator cuff injuries can arise in more than 50% of elderly patients [21].

Risk factors and recurrence rate of redislocation

Patients with a history of shoulder dislocation face an increased risk of recurrent dislocation. This often occurs due to inadequate tissue healing, laxity, and high levels of activity. Moreover, patients who have sustained rotator cuff tears or glenoid fractures are at a heightened risk of recurrent dislocation [19]. Another critical factor is glenoid bone loss exceeding 20%, which significantly contributes to recurrent anterior shoulder instability [22].

Regarding the recurrence rate of instability after ABR, research indicates a higher occurrence among younger patients [7]. In one study, patients aged 22 years or younger experienced

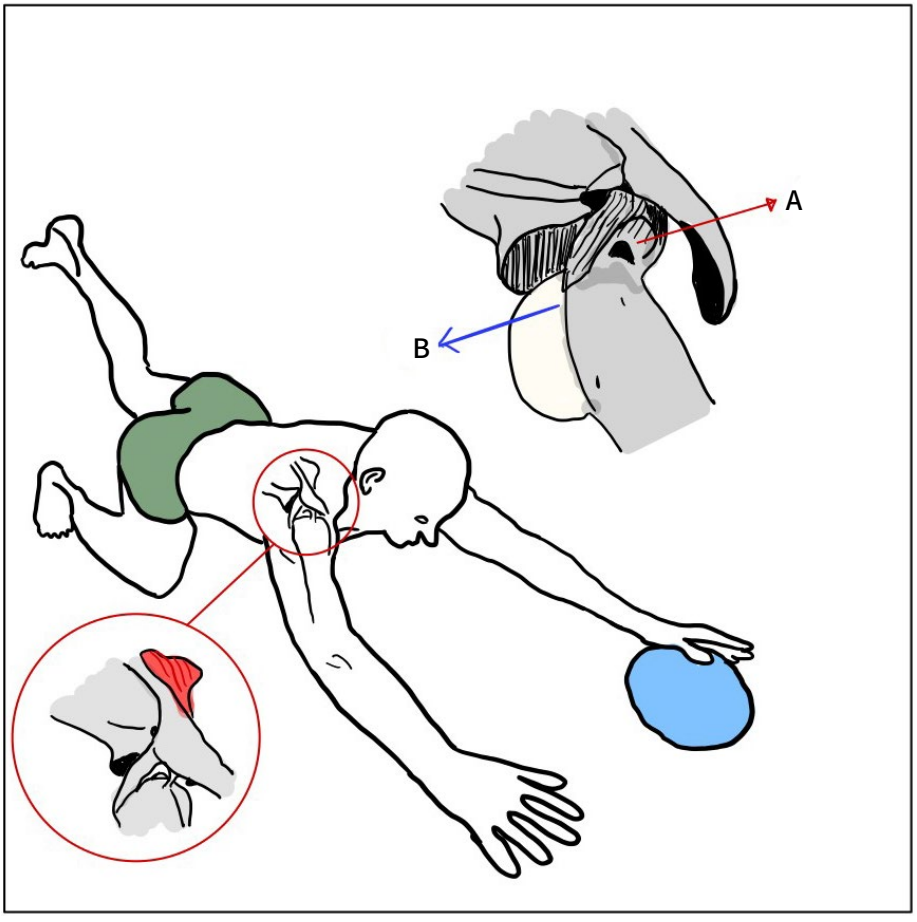


Fig. 2. Mechanism of injury in anterior shoulder dislocation resulting in (A) rotator cuff tear and (B) subluxed humerus.

https://doi.org/10.12771/emj.2024.e21

a recurrence rate after ABR of 13.3%, whereas older patients exhibited a rate of 6.3% [23]. Similarly, another study reported a recurrence rate of 51% among contact athletes aged 18 years or younger, compared to a 12% recurrence rate in a group of 25-year-old athletes [24]. Moreover, their findings indicated that the risk of recurrence among adolescent athletes was 2.2 times greater in athletes younger than 16 years old compared to those older than 16 years. However, the recurrence rate varies based on the type of sport, with contact and collision sports—such as rugby and American football—displaying exceptionally high recurrence rates [6]. In soccer, one study reported that goalkeepers have a recurrence rate more than eight times higher than field position players and experience worse functional outcomes. Goalkeepers often stop high-velocity shots with their hands, dive with outstretched arms, and forcefully throw balls, all actions that increase their risk of shoulder injury [25].

In a retrospective study of 271 patients who underwent primary ABR for anterior shoulder instability, researchers found that off-track Hill-Sachs lesions (HSL)—those that extend medially beyond the glenoid track—were associated with a higher risk of anterior engagement and instability compared to on-track HSL. The rate of surgical revision for patients with off-track HSL was 48% at an average follow-up of 53.5 months, while the rate for those with on-track HSL was 13% at an average follow-up of 42.3 months [22]. Another study, which included 100 recreational athletes who received ABR and were followed for an average of 12.7±2.1 years, revealed a 19% rate of subjective apprehension and a 19% rate of redislocation. Additionally, gradual declines were noted in clinical outcomes and sports activity levels over time. Surgeons are advised to carefully select candidates for ABR by considering risk factors such as the presence of off-track lesions, age under 20 years, and participation in contact sports [26].

The findings regarding follow-up procedures after primary anterior shoulder dislocation consistently support the use of ABR. Relative to ABR, a significantly higher recurrence rate of instability was observed after conservative treatment. Consequently, it is logical to anticipate the need for additional future procedures in patients initially treated conservatively. A key consideration is that instability frequently results in symptoms that can disrupt patients' engagement in sports activities [27].

Rehabilitation protocol

Postoperative rehabilitation therapy is essential for promoting the recovery of shoulder motion and strength, enabling patients to resume functional activities sooner and ultimately resulting in greater patient satisfaction [27–31]. The postoperative rehabilitation guidelines reported in the literature vary considerably, and broadly accepted guidelines for rehabilitation following ABR for anterior shoulder instability do not yet exist [31]. Kelley et al. presented a postoperative rehabilitation protocol for patients who have undergone ABR, including 2 years of follow-up. The specifics of this rehabilitation protocol are detailed in Tables 2, 3 [32].

Return to sport after arthroscopic Bankart repair

ABR was identified as having the highest rate of RTS across all age groups, surpassing other stabilization procedures such as open Bankart repair, open Latarjet, and arthroscopic Latarjet procedures [6,33]. A cohort study by Blonna et al. compared 30 participants undergoing ABR with 30 participants undergoing the open Bristow-Latarjet procedure, resulting in a higher Subjective Patient Outcome for Return to Sports score in the ARB group [33]. A systematic review of 16 articles evaluated the RTS rate after various surgical anterior shoulder stabilization techniques, revealing the highest RTS rate among athletes who underwent ABR (97.5%). Other

Table 2. Rehabilitation program goals

Week (phase)	Goal
1 to 4 (immediate postoperative)	- Protect repair - Mitigate consequences of immobilization - Promote dynamic stability and proprioception - Reduce pain and inflammation - Avoid stretching - Avoid active external rotation, abduction, or extension
5 to 12 (intermediate)	- Gradually restore full ROM - Preserve repair integrity - Restore muscular strength and balance - Enhance neuromuscular control
13 to 21 (minimal protection)	- Maintain full ROM - Improve muscular control, strength, power, and endurance - Practice core stabilization and conditioning - Weekly functional testing begins at week 16 - Weekly TSK-11 administration begins at week 16 - Sport-specific training begins at week 20
22 to 26 (advance to strengthening)	- Maintain full ROM - Improve strength, power, and endurance - Advance functional activities
26 to 32 (return-to-sports)	- Enhance strength, power, and endurance - Pass all functional assessments (Table 3) - Maintain mobility

ROM, range of motion; TSK-11, Tampa Scale of Kinesiophobia–11.

Table 3. Functional assessment test

Test	Goal	Pass
a. Overhand band reach	Demonstrate functional rotator cuff activity throughout multiplanar range of motion while avoiding trapezius dominance, trunk lean, and pelvic tilt	Maintain stability
b. Closed kinetic chain extremity stability test (CKCUEST)	Measure speed, agility, and power	21 touches (male) or 23 touches (female) in 15 seconds
c. Upper extremity Y balance	Using the operative arm as a stabilizer, test mobility and stability of the extremity and core; combines scapular stability and functional range of motion with core stabilization and thoracic rotation	3 consecutive progressions
d. One-arm hop test	Focus on stable core, maximum assessment of strength, and neuromuscular coordination	5 repetitions
e. Posterior Shoulder Endurance Test (PSET)	Assess posterior rotator cuff and deltoid strength	85% of contralateral arm strength
f. Trunk stability push-up	Stabilize spine and hips in sagittal plane during upper body symmetrical motion	3 repetitions with control
g. Long arm plank ball tap	Assess stability, proprioception, and endurance	10 bidirectional taps with body control
h. Plank weight stacking	Using the operative arm as a stabilizer, assess both proprioception and stability of the core and scapula	4 repetitions×1 lb

procedures examined included open Bankart repair (86.1%), open Latarjet procedure (83.6%), minimally invasive Latarjet procedure (94.0%), and ABR with remplissage (95.5%) [12].

Goals for ABR in young athletes include restoring shoulder function and enabling RTS at pre-injury levels [6]. Shoulder stabilization for Bankart lesions can be achieved through two methods:

arthroscopic surgery or open surgery. Both treatments involve reattaching the torn labrum to the glenoid [34]. A review focusing on the RTS in teenagers following surgical stabilization reported an overall return rate of 95%, with 77% of patients reaching pre-injury levels of performance [6]. Various criteria were used to assess the athletes' readiness to RTS, such as time elapsed since surgery, type of sport, strength, ROM, pain, and proprioception [35– 37]. The type of sport played was linked to outcomes such as RTS failure or complete RTS.

Time from surgery

The most common criterion for return to play (RTP) was the time elapsed since surgery, indicating a minimum duration between the surgical procedure and the athlete's capability to RTP [36]. A retrospective study of 50 teenage athletes who underwent ABR revealed that the average time for RTS was 6.6±2.7 months, with a range of 3 to 18 months. The time to return to competitive play averaged 9.3±4.0 months (range, 6 to 24 months), while achieving a complete return to pre-injury levels took 10.6±4.3 months (range, 8 to 24 months) [6]. A systematic review encompassing 58 studies reported that the timeframe for RTS post-surgery varied from 1.5 to 12 months, with a return after 6 months being the most cited duration [37]. Another systematic review, which included 34 studies, found that patients were typically allowed to RTS after a mean of 5.7 months (range, 1.9 to 32 months) following surgery [13]. More recently, a survey study involving 317 surgeons from the United States and Europe indicated that the most frequently recommended time for athletes to resume sports was 4 months after surgery. However, most of these surgeons advised waiting an additional period, most often 2 months, before granting athletes clearance to RTS [38].

Type of sports

The type of shoulder sport played can influence the likelihood of RTS. Allain et al. categorized sports that place strain on the shoulder into four distinct groups, as shown in Table 4 [39].

A study by Ide et al. reported that overhead athletes exhibited the lowest rate of complete RTS at 68%, compared to contact athletes and non-contact/non-overhead athletes, who had respective return rates of 86% and 100% [40]. Another study suggested that athletes should only return to overhead sports after 7 months post-surgery, and they should wait until 10 months after surgery before returning to competitive sports. Additionally, it is expected that athletes who participate in overhead sports will fully recover their external rotation capacity following ABR. Failure to achieve this recovery could negatively affect sports-related outcomes [41]. Gibson et al. found that ABR, combined with an accelerated rehabilitation program, allows professional football players to RTP relatively quickly, with an average time of 11 weeks [42].

Table 4. Types of sports involving shoulder activity

Group	Shoulder sport
G1	Non-collision/non-overhead
G2	High-impact/collision
G3	Overhead
G4	Martial arts

https://doi.org/10.12771/emj.2024.e21

Strength

Strength is a challenging parameter to measure objectively due to the influence of various factors. A total of 25 studies have incorporated muscle strength within RTS criteria, including achievement of complete strength restoration, pre-injury strength levels, at least 80% of the strength of the contralateral side, strength comparable to the contralateral side, symmetric strength in abduction and external rotation as determined by manual testing, grade 5 strength in all intrinsic and extrinsic shoulder muscles, and strength that equal to or exceeding baseline values [36]. A retrospective study assessed strength recovery post-surgery using isokinetic and isometric devices to provide an objective evaluation. However, the findings in the literature are inconsistent, and functional goals were more frequently achieved than strength criteria [43]. A systematic review investigated the strength criteria for RTS. However, the results were inconsistent, and the studies did not uniformly assess strength with the same type of device [37].

Range of motion

The assessment of shoulder ROM involves evaluating both active and passive movements to ensure that the athlete demonstrates symmetrical, full, and sport-specific ROM without experiencing pain or apprehension [36,44].

Pain

Another important factor for RTS is the assessment of pain following ABR. Pain is considered a criterion for RTS, but it always appears in conjunction with other criteria. In this context, pain has been defined as the presence of "non-painful ROM" and being "pain-free" during physical examination or participation in sports. Here, key distinctions must be made. For the general population, a complete absence of pain is not a prerequisite. However, for athletes who aim to return to their sports activities and achieve pre-injury performance levels, being pain-free is essential [37,45].

Proprioception

To date, few studies have included shoulder proprioception as a criterion for RTP. Tambe et al. noted an improvement in shoulder proprioception as an RTP criterion, yet the study did not detail the specific assessment modality employed [46].

Rate of return to sport

The rate of RTS at pre-injury levels varies widely according to the studies available, with figures ranging from 31% to 100% [13,37,47]. Memon et al. assessed the RTS in 1,866 patients following ABR. Their study found that 82% of competitive athletes accomplished RTS, and 88% of those returned to their pre-injury levels [13]. Abdul et al. examined RTS rates after shoulder stabilization surgery and reported a 97.5% RTS rate, with an average time of 5.9 months post-ABR [12]. Harada et al. observed high RTS rates in a cohort of 50 young athletes, with nearly all participants resuming sports; 96% returned to competitive play, and 76% fully regained their pre-injury performance levels without any complaints [6]. A systematic review that included 11 studies with 392 adolescent athletes who underwent ABR revealed a 79.8% return rate to sports at pre-injury levels [48]. In the sport of soccer, the RTS rate at the same level was significantly lower for goalkeepers compared to field players [25].

Reason for failing to return to sport

Athletes who experience injuries often face negative psychological responses, including

depression, anxiety, irritability, and a lack of confidence. These psychological reactions can affect a patient's decision to RTS, even after ABR [49]. A retrospective study evaluated patients who underwent ABR and did not RTS over a 24-month follow-up period. The study reported that 51.9% of patients harbored a persistent fear of re-injury, 25.0% believed their injury signified the natural conclusion of their athletic career, 15.4% felt that their lifestyle had changed, 11.5% experienced persistent pain, and 7.7% were unable to RTS due to other injuries [2]. In research by Tjong et al. involving 25 patients, several reasons were identified for not returning to sport after ABR, including fear of re-injury, a shift in priorities, mood disturbances, social support, and a lack of motivation [49]. Plath et al. reported that among athletes who did not RTS after ABR, the primary reasons were non–shoulder-related factors, followed by concerns about potential re-injury [50]. A recent study underscored kinesiophobia—fear of movement—as a prevalent factor affecting patients' psychological readiness to RTS. Psychological interventions, such as cognitive-behavioral therapy and mindfulness, have been proposed to potentially improve RTS rates in these patients [51].

Conclusion

ABR results in a high percentage of athletes returning to athletic activities, leading to the development of various criteria to support the RTP. Most surgeons advise athletes to wait 6 months after surgery before resuming sports, to regain at least 80% of the strength in the contralateral limb or a level of strength comparable to that prior to the injury, to achieve a full or symmetrical ROM without pain, and to demonstrate improved shoulder proprioception. However, the type of sport also influences the rate of RTS, with overhead sports displaying the lowest return rates. While many athletes successfully return to their previous level of competition, some may experience adverse psychological responses during the process.

ORCID
Shafira Widya Utami: https://orcid.org/0009-0003-6525-4597
Savina Rifky Pratiwi: https://orcid.org/0009-0003-7616-575X
Mitchel: https://orcid.org/0009-0001-3904-0232
Karina Sylvana Gani: https://orcid.org/0000-0001-7306-9752
Erica Kholinne: https://orcid.org/0000-0002-4326-8205

Authors' contributions
Project administration: Utami SW
Conceptualization: Mitchel
Methodology & data curation: Pratiwi SR, Kholinne E
Funding acquisition: not applicable
Writing – original draft: Utami SW, Pratiwi SR, Mitchel, Gani KS
Writing – review & editing: Utami SW, Pratiwi SR, Mitchel, Gani KS, Kholinne E

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Funding
Not applicable.

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Not applicable.

References

1. Galvin JW, Ernat JJ, Waterman BR, Stadecker MJ, Parada SA. The epidemiology and natural history of anterior shoulder instability. *Curr Rev Musculoskelet Med* 2017;10(4):411-424. https://doi.org/10.1007/s12178-017-9432-5
2. Hurley ET, Davey MS, Mojica ES, Montgomery C, Gaafar M, Jazrawi LM, et al. Analysis of patients unable to return to play following arthroscopic Bankart repair. *Surgeon* 2022;20(4):e158-e162. https://doi.org/10.1016/j.surge.2021.06.005
3. Owens BD, Duffey ML, Nelson BJ, DeBerardino TM, Taylor DC, Mountcastle SB. The incidence and characteristics of shoulder instability at the United States Military Academy. *Am J Sports Med* 2007;35(7):1168-1173. https://doi.org/10.1177/0363546506295179
4. Krøner K, Lind T, Jensen J. The epidemiology of shoulder dislocations. *Arch Orthop Trauma Surg* 1989;108(5):288-290. https://doi.org/10.1007/BF00932317
5. Simonet WT, Melton LJ, Cofield RH, Ilstrup DM. Incidence of anterior shoulder dislocation in Olmsted County, Minnesota. *Clin Orthop Relat Res* 1984;186:186-191. https://doi.org/10.1097/00003086-198406000-00030
6. Harada Y, Iwahori Y, Kajita Y, Takahashi R, Yokoya S, Sumimoto Y, et al. Return to sports after arthroscopic Bankart repair in teenage athletes: a retrospective cohort study. *BMC Musculoskelet Disord* 2023;24(1):64. https://doi.org/10.1186/s12891-023-06145-y
7. Stone GP, Pearsall AW. Return to play after open Bankart repair: a systematic review. *Orthop J Sports Med* 2014;2(2):1-5. https://doi.org/10.1177/2325967114522960
8. Friedman LGM, Griesser MJ, Miniaci AA, Jones MH. Recurrent instability after revision anterior shoulder stabilization surgery. *Arthroscopy* 2014;30(3):372-381. https://doi.org/10.1016/j.arthro.2013.11.019
9. Hobby J, Griffin D, Dunbar M, Boileau P. Is arthroscopic surgery for stabilisation of chronic shoulder instability as effective as open surgery? A systematic review and meta-analysis of 62 studies including 3044 arthroscopic operations. *J Bone Joint Surg Br* 2007;89-B(9):1188-1196. https://doi.org/10.1302/0301-620X.89B9.18467
10. Ng C, Bialocerkowski A, Hinman R. Effectiveness of arthroscopic versus open surgical stabilisation for the management of traumatic anterior glenohumeral instability. *Int J Evid Based Healthc* 2007;5(2):182-207. https://doi.org/10.1111/j.1479-6988.2007.00064.x
11. Pulavarti RS, Symes TH, Rangan A. Surgical interventions for anterior shoulder instability in adults. *Cochrane Database Syst Rev* 2009;(4):CD005077. https://doi.org/10.1002/14651858.CD005077.pub2
12. Abdul-Rassoul H, Galvin JW, Curry EJ, Simon J, Li X. Return to sport after surgical treatment for anterior shoulder instability: a systematic review: response. *Am J Sports Med* 2019;47(3):NP24-NP27. https://doi.org/10.1177/0363546519825642
13. Memon M, Kay J, Cadet ER, Shahsavar S, Simunovic N, Ayeni OR. Return to sport following arthroscopic Bankart repair: a systematic review. *J Shoulder Elb Surg* 2018;27(7):1342-1347. https://doi.org/10.1016/j.jse.2018.02.044
14. Clavert P. Glenoid labrum pathology. *Orthop Traumatol Surg Res* 2015;101(1):S19-S24. https://doi.org/10.1016/j.otsr.2014.06.028
15. Porcellini G, Campi F, Paladini P. Arthroscopic approach to acute bony Bankart lesion. *Arthroscopy* 2002;18(7):764-769. https://doi.org/10.1053/jars.2002.35266
16. Godin JA, Altintas B, Horan MP, Hussain ZB, Pogorzelski J, Fritz EM, et al. Midterm results of the bony Bankart bridge technique for the treatment of bony Bankart lesions. *Am J Sports Med* 2019;47(1):158-164. https://doi.org/10.1177/0363546518808495
17. Millett PJ, Horan MP, Martetschläger F. The "bony Bankart bridge" technique for restoration of anterior shoulder stability. *Am J Sports Med* 2013;41(3):608-614. https://doi.org/10.1177/0363546512472880
18. Nolte PC, Elrick BP, Bernholt DL, Lacheta L, Millett PJ. The bony Bankart: clinical and technical considerations. *Sports Med Arthrosc Rev* 2020;28(4):146-152. https://doi.org/10.1097/JSA.0000000000000286
19. Abrams R, Akbarnia H. Shoulder dislocations overview [Internet]. Treasure Island (FL): StatPearls; c2023 [cited 2024 Jan 20]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459125/
20. Pak T, Kim AM. Anterior glenohumeral joint dislocation [Internet]. Treasure Island (FL): StatPearls; c2023 [cited 2024 Jan 20]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK557862/
21. Cunningham NJ. Techniques for reduction of anteroinferior shoulder dislocation. *Emerg Med Australas* 2005;17(5-6):463-471. https://doi.org/10.1111/j.1742-6723.2005.00778.x
22. Schwihla I, Wieser K, Grubhofer F, Zimmermann SM. Long-term recurrence rate in anterior shoulder instability after Bankart repair based on the on- and off-track concept. *J Shoulder Elb Surg* 2023;32(2):269-275. https://doi.org/10.1016/j.jse.2022.07.025
23. Porcellini G, Campi F, Pegreffi F, Castagna A, Paladini P. Predisposing factors for recurrent shoulder dislocation after

arthroscopic treatment. *J Bone Joint Surg* 2009;91(11):2537-2542.
https://doi.org/10.2106/JBJS.H.01126
24. Torrance E, Clarke CJ, Monga P, Funk L, Walton MJ. Recurrence after arthroscopic labral repair for traumatic anterior instability in adolescent rugby and contact athletes. *Am J Sports Med* 2018;46(12):2969-2974.
https://doi.org/10.1177/0363546518794673
25. Pasqualini I, Rossi LA, Brandariz R, Tanoira I, Fuentes N, Denard PJ, et al. Effect of playing position on return to sport, functional outcomes, and recurrence after arthroscopic Bankart repair in soccer players. *Orthop J Sports Med* 2022;10(11):23259671221138106.
https://doi.org/10.1177/23259671221138106
26. Kim JS, Kim SC, Park JH, Kim HG, Kim DY, Lee SM, et al. Long-term effectiveness and outcome-determining factors of arthroscopic Bankart repair for recreational sports population: an assessment of 100 patients with a mean follow-up of 12.7 years. *Am J Sports Med* 2024;52(3):594-602.
https://doi.org/10.1177/03635465231220838
27. Hu B, Hong J, Zhu H, Yan S, Wu H. Arthroscopic Bankart repair versus conservative treatment for first-time traumatic anterior shoulder dislocation: a systematic review and meta-analysis. *Eur J Med Res* 2023;28(1):260.
https://doi.org/10.1186/s40001-023-01160-0
28. DeFroda SF, Mehta N, Owens BD. Physical therapy protocols for arthroscopic Bankart repair. *Sports Health* 2018;10(3):250-258.
https://doi.org/10.1177/1941738117750553
29. McIsaac W, Lalani A, Silveira A, Chepeha J, Luciak-Corea C, Beaupre L. Rehabilitation after arthroscopic Bankart repair: a systematic scoping review identifying important evidence gaps. *Physiotherapy* 2022;114:68-76.
https://doi.org/10.1016/j.physio.2021.03.014
30. Matache BA, Hurley ET, Wong I, Itoi E, Strauss EJ, Delaney RA, et al. Anterior shoulder instability part III—revision surgery, rehabilitation and return to play, and clinical follow-up—an international consensus statement. *Arthroscopy J Arthrosc Rel Surg* 2022;38(2):234-242.E6.
https://doi.org/10.1016/j.arthro.2021.07.019
31. Kim K, Saper MG. Postoperative management following arthroscopic Bankart repair in adolescents and young adults: a systematic review. *Arthrosc Sports Med Rehabil* 2020;2(6):E839-E845.
https://doi.org/10.1016/j.asmr.2020.05.016
32. Kelley TD, Clegg S, Rodenhouse P, Hinz J, Busconi BD. Functional rehabilitation and return to play after arthroscopic surgical stabilization for anterior shoulder instability. *Sports Health* 2022;14(5):733-739.
https://doi.org/10.1177/19417381211062852
33. Blonna D, Bellato E, Caranzano F, Assom M, Rossi R, Castoldi F. Arthroscopic Bankart repair versus open Bristow-Latarjet for shoulder instability: a matched-pair multicenter study focused on return to sport. *Am J Sports Med* 2016;44(12):3198-3205.
https://doi.org/10.1177/0363546516658037
34. Sedeek SM, Tey IK, Tan AHC. Arthroscopic Bankart repair for traumatic anterior shoulder instability with the use of suture anchors. *Singapore Med J* 2008;49(9):676-681.
35. Rossi LA, Pasqualini I, Tanoira I, Ranalletta M. Factors that influence the return to sport after arthroscopic Bankart repair for glenohumeral instability. *Open Access J Sports Med* 2022;13:35-40.
https://doi.org/10.2147/OAJSM.S340699
36. Griffith R, Fretes N, Bolia IK, Murray IR, Meyer J, Weber AE, et al. Return-to-sport criteria after upper extremity surgery in athletes: a scoping review, part 1: rotator cuff and shoulder stabilization procedures. *Orthop J Sports Med* 2021;9(8):23259671211021827.
https://doi.org/10.1177/23259671211021827
37. Ciccotti MC, Syed U, Hoffman R, Abboud JA, Ciccotti MG, Freedman KB. Return to play criteria following surgical stabilization for traumatic anterior shoulder instability: a systematic review. *Arthroscopy* 2018;34(3):903-913.
https://doi.org/10.1016/j.arthro.2017.08.293
38. Hurley ET, Matache BA, Colasanti CA, Mojica ES, Manjunath AK, Campbell KA, et al. Return to play criteria among shoulder surgeons following shoulder stabilization. *J Shoulder Elb Surg* 2021;30(6):E317-E321.
https://doi.org/10.1016/j.jse.2021.01.026
39. Allain J, Goutallier D, Glorion C. Long-term results of the Latarjet procedure for the treatment of anterior instability of the shoulder. *J Bone Joint Surg* 1998;80(6):841-852.
https://doi.org/10.2106/00004623-199806000-00008
40. Ide J, Maeda S, Takagi K. Arthroscopic Bankart repair using suture anchors in athletes: patient selection and postoperative sports activity. *Am J Sports Med* 2004;32(8):1899-1905.
https://doi.org/10.1177/0363546504265264
41. Buckup J, Welsch F, Gramlich Y, Hoffmann R, Roessler PP, Schüttler KF, et al. Back to sports after arthroscopic revision Bankart repair. *Orthop J Sports Med* 2018;6(2):2325967118755452.
https://doi.org/10.1177/2325967118755452
42. Gibson J, Kerss J, Morgan C, Brownson P. Accelerated rehabilitation after arthroscopic Bankart repair in professional footballers. *Shoulder Elb* 2016;8(4):279-286.
https://doi.org/10.1177/1758573216647898
43. Wilson KW, Popchak A, Li RT, Kane G, Lin A. Return to sport testing at 6 months after arthroscopic shoulder stabilization reveals residual strength and functional deficits. *J Shoulder Elb Surg* 2020;29(7):S107-S114.

https://doi.org/10.1016/j.jse.2020.04.035
44. Wilk KE, Bagwell MS, Davies GJ, Arrigo CA. Return to sport participation criteria following shoulder injury: a clinical commentary. *Int J Sports Phys Ther* 2020;15(4):624-642. https://doi.org/10.26603/ijspt20200624
45. Bravi M, Fossati C, Giombini A, Macaluso A, Lazzoli JK, Santacaterina F, et al. Criteria for return-to-play (RTP) after rotator cuff surgery: a systematic review of literature. *J Clin Med* 2022;11(8):2244. https://doi.org/10.3390/jcm11082244
46. Tambe A, Badge R, Funk L. Arthroscopic rotator cuff repair in elite rugby players. *Int J Shoulder Surg* 2009;3(1):8-12. https://doi.org/10.4103/0973-6042.50876
47. Stein T, Linke RD, Buckup J, Efe T, von Eisenhart-Rothe R, Hoffmann R, et al. Shoulder sport-specific impairments after arthroscopic Bankart repair: a prospective longitudinal assessment. *Am J Sports Med* 2011;39(11):2404-2414. https://doi.org/10.1177/0363546511417407
48. Kasik CS, Rosen MR, Saper MG, Zondervan RL. High rate of return to sport in adolescent athletes following anterior shoulder stabilisation: a systematic review. *J ISAKOS Joint Disord Orthop Sports Med* 2019;4(1):33-40. https://doi.org/10.1136/jisakos-2018-000224
49. Tjong VK, Devitt BM, Lucas Murnaghan M, Ogilvie-Harris DJ, Theodoropoulos JS. A qualitative investigation of return to sport after arthroscopic Bankart repair: beyond stability. *Am J Sports Med* 2015;43(8):2005-2011. https://doi.org/10.1177/0363546515590222
50. Plath JE, Feucht MJ, Saier T, Minzlaff P, Seppel G, Braun S, et al. Sporting activity after arthroscopic Bankart repair for chronic glenohumeral instability. *Arthroscopy* 2015;31(10):1996-2003. https://doi.org/10.1016/j.arthro.2015.04.087
51. Owusu-Ansah GE, Anudu EE, Ross PP, Ierulli VK, Mulcahey MK. Psychological readiness to return to sport after shoulder instability. *JBJS Rev* 2023;11(9):e23.00022. https://doi.org/10.2106/JBJS.RVW.23.00022

Review

Ewha Med J 2024;47(2):e22
https://doi.org/10.12771/emj.2024.e22
eISSN 2234-2591

What is the role of artificial intelligence in general surgery?

Seung Min Baik[1], **Ryung-Ah Lee**[2]

[1]Division of Critical Care Medicine, Department of Surgery, Ewha Womans University Mokdong Hospital, Ewha Womans University College of Medicine, Seoul, Korea
[2]Division of Colorectal Surgery, Department of Surgery, Ewha Womans University College of Medicine, Seoul, Korea

Received Mar 12, 2024
Accepted Apr 18, 2024

Corresponding author
Ryung-Ah Lee
Division of Colorectal Surgery, Department of Surgery, Ewha Womans University College of Medicine, 260 Gonghang-daero, Gangseo-gu, Seoul 07804, Korea
E-mail: ralee@ewha.ac.kr

Keywords
Artificial intelligence; General surgery; Machine learning; Deep learning; Algorithms

The capabilities of artificial intelligence (AI) have recently surged, largely due to advancements in deep learning inspired by the structure and function of the neural networks of the human brain. In the medical field, the impact of AI spans from diagnostics and treatment recommendations to patient engagement and monitoring, considerably improving efficiency and outcomes. The clinical integration of AI has also been examined in specialties, including pathology, radiology, and oncology. General surgery primarily involves manual manipulation and includes preoperative, intraoperative, and postoperative care, all of which are critical for saving lives. Other fields have strived to utilize and adopt AI; nonetheless, general surgery appears to have retrogressed. In this review, we analyzed the published research, to understand how the application of AI in general surgery differs from that in other medical fields. Based on previous research in other fields, the application of AI in the preoperative stage is nearing feasibility. Ongoing research efforts aim to utilize AI to improve and predict operative outcomes, enhance performance, and improve patient care. However, the use of AI in the operating room remains significantly understudied. Moreover, ethical responsibilities are associated with such research, necessitating extensive work to gather evidence. By fostering interdisciplinary collaboration and leveraging lessons from AI success stories in other fields, AI tools could be specifically tailored for general surgery. Surgeons should be prepared for the integration of AI into clinical practice to achieve better outcomes; therefore, the time has come to consider ethical and legal implications.

Introduction

Artificial intelligence (AI) has emerged as a revolutionary force in modern medicine, significantly reshaping diagnostics and treatment planning across various specialties [1,2]. In fields such as radiology and oncology, AI has had an unmistakable impact on improving diagnostic accuracy, enabling early disease detection, and optimizing treatment protocols [3,4]. For instance, in radiology, AI algorithms have revolutionized image analysis, facilitating more accurate interpretations and aiding in the early detection of illnesses [5].

The scope of AI integration ranges from diagnostics to patient management and care. Predictive analytics utilizing sophisticated machine learning (ML) algorithms are increasingly being employed to identify high-risk patients, predict complications, and personalize care plans [6]. This approach has ushered in a new era of proactive, patient-centric healthcare.

Moreover, AI is paving the way for precision medicine. By analyzing large datasets that include

genetic profiles and patient histories, AI systems can provide treatments specifically tailored to the needs of individual patients. This approach significantly improves therapeutic effectiveness and minimizes side effects [7].

Despite these advances, general surgery lags significantly behind other medical fields in both AI research and clinical applications. The volume of medical articles published on the use of AI is markedly lower in the field of surgery, especially in general surgery (Fig. 1).

While specialized fields such as neurosurgery and cardiology are increasingly incorporating AI to improve surgical planning and robotic assistance, general surgery has been notably slower in adopting these advanced technologies [8,9].

The reasons for this delay are multifactorial. One of these reasons relates to the diversity and spontaneity of surgical procedures. General surgery is a dynamic field where some operations are predictable and can be scheduled in advance, while others are unpredictable and often rely on real-time decision-making in the operating room [10]. The variability in case types within general surgery complicates the collection of the extensive and consistent data necessary to train AI systems [10]. This issue is further exacerbated by the relative scarcity of focused research efforts aimed at integrating AI into general surgical workflows [10]. Therefore, locating studies on the use of AI in general surgery within databases like PubMed proves challenging.

This article summarizes the current state of research on the application of AI in medicine and explores the future direction of general surgery as it adapts to a rapidly changing medical environment. It includes a discussion on how AI can be integrated into various aspects of general surgery, ranging from preoperative analytics to postoperative care, as well as the steps required to overcome existing challenges.

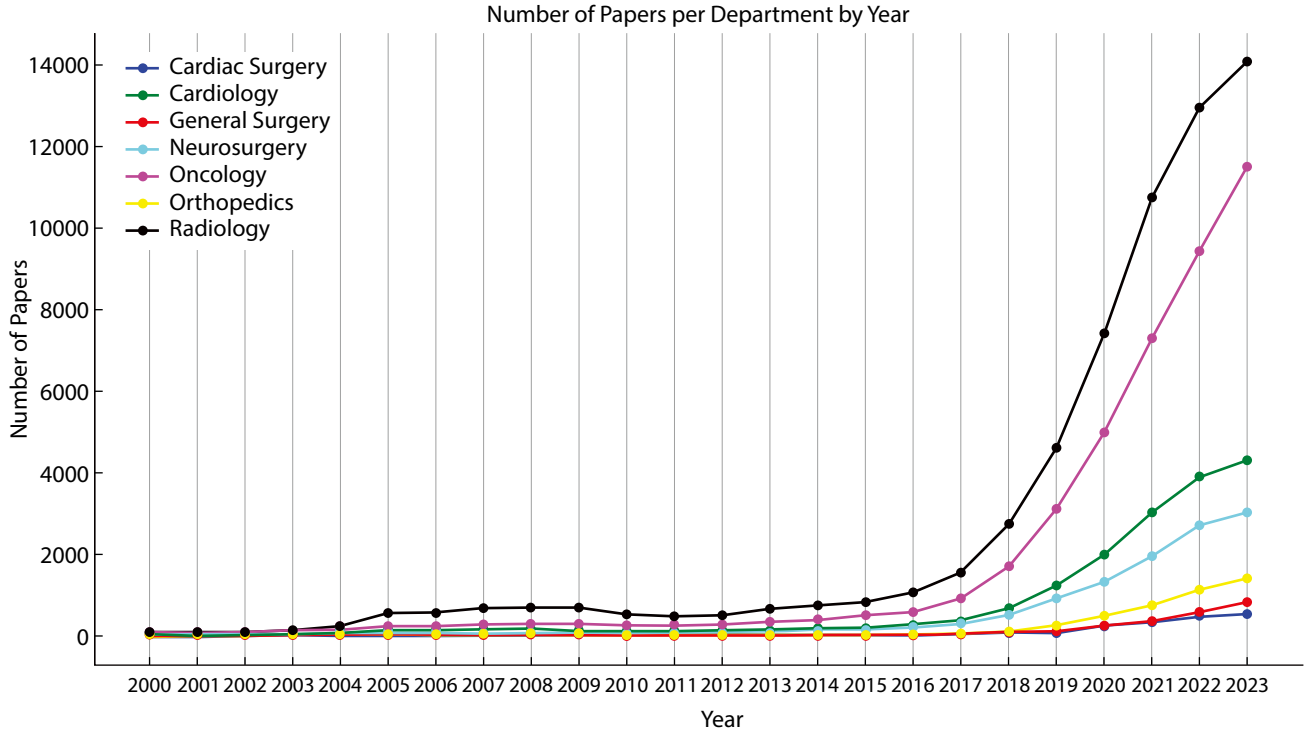


Fig. 1. Number of artificial intelligence papers published by each specialty from 2000 to 2023 is depicted.

https://doi.org/10.12771/emj.2024.e22

Ethics statement

It is a literature database-based review; therefore, neither approval by the institutional review board nor obtainment of informed consent was required.

Pioneering artificial intelligence in the medical specialties of radiology, oncology, and cardiology

In modern medicine, the integration of AI has been particularly pronounced in specialties such as radiology, oncology, and cardiology. In radiology, AI algorithms have revolutionized diagnostic processes and enhanced the accuracy of image interpretation, which is crucial for early disease detection and treatment planning.

Recent studies on AI in radiology have produced important findings. Lång et al. compared the clinical safety of an AI-assisted screen reading protocol to that of the conventional double reading method used in mammography screening. The study involved 80,000 women and assessed early screening outcomes, including cancer detection rates, retest rates, false-positive rates, positive predictive values of the retests, and the types of cancers detected. In the intervention group, 244 tumors were detected, comprising 184 invasive tumors and 60 *in situ* tumors. Meanwhile, in the control group, 203 tumors were identified, with 165 being invasive and 38 *in situ* [11].

A randomized controlled trial conducted by Nam et al. demonstrated that AI-based, computer-aided design software enhances the detection rate of actionable lung nodules in chest radiographs of health-screening participants. The AI group exhibited a higher detection rate of actionable nodules compared to the non-AI group (0.59% vs. 0.25%). Additionally, the detection rate of malignant lung nodules was also higher in the AI group than in the non-AI group (0.15% vs. 0.0%). The rates of misdiagnosis and positive reporting were similar between the AI and non-AI groups [12].

Sachpekidis et al. have demonstrated that a deep learning (DL)-based tool for automatically assessing bone marrow metabolism in patients with multiple myeloma is feasible and correlates with clinically relevant disease parameters. There is a significant positive correlation between the visual analysis of PET/CT scans and the metabolic tumor volume (MTV) and total lesion glycolysis (TLG) values, following the application of all six ^{18}F-fluorodeoxyglucose (FDG) uptake thresholds. Additionally, significant differences in MTV and TLG values were observed between patient groups across all applied thresholds.

The DL-based approach has demonstrated significant, moderate, positive correlations between bone marrow plasma cell infiltration and plasma β2-microglobulin levels, as well as with the automated quantitative PET/CT parameters, MTV and TLG [13].

Similarly, oncology has benefited from the use of AI, especially in the realm of personalized medicine. AI algorithms are employed to analyze patient data and predict responses to treatment, which allows oncologists to customize therapies based on the specific needs of individual patients. Clift et al. developed a clinically useful model that estimates the 10-year risk of breast cancer-related mortality for women at all stages of the disease. Additionally, they compared the outcomes of regression analyses with those of ML approaches. The final Cox model demonstrated good discriminatory power, evidenced by a Harrell's C-index of 0.858 (95% CI, 0.853–0.864), and showed moderate calibration. The model's performance varied across ethnic groups, exhibiting the highest discriminatory power in Chinese women (Harrell's

C-index=0.931) and the lowest in Bangladeshi women (Harrell's C-index=0.794). Moreover, the model generally performed well across various cancer stages, though its discriminatory power decreased as the cancer stage advanced [14].

Alaimo et al. have developed and validated a ML model to predict the early recurrence of intrahepatic cholangiocarcinoma following hepatectomy. The model, trained using 14 clinicopathological characteristics, demonstrates promising accuracy in predicting recurrences occurring within 12 months after surgery. It identifies tumor burden score as the most significant predictor of early recurrence, followed by perineural involvement. Additionally, the model's predictions of early recurrence strongly correlate with 3-year overall survival rates. Patients predicted to experience early recurrence exhibit significantly lower 3-year overall survival rates compared to those without such predictions [15]. A meta-analysis utilizing a substantial volume of recent data has been conducted to assess the effectiveness of AI in diagnosing lung cancer. The findings indicate that AI-assisted diagnostic systems achieve a sensitivity and specificity of 0.87, with a missed diagnosis rate and misdiagnosis rate each at 13%. The systems also show a positive likelihood ratio of 6.5, a negative likelihood ratio of 0.15, a diagnostic ratio of 43, and a combined sum of areas under the target operating characteristic curve of 0.93 [16].

Cardiology has kept pace with the AI revolution. AI systems in cardiology have been crucial in predicting cardiac events, thereby improving preventive cardiac care. A review has underscored the potential of AI for data interpretation and automated analysis in interventional cardiology procedures. ML techniques are employed in interventional cardiology for image reconstruction, interpretation, and analysis. ML models, including the lasso-penalized Cox proportional hazards regression model and the k-means clustering algorithm, have been utilized for predicting mortality and detecting the QRS complex, respectively.

ML algorithms have been developed for angiographic recognition, coronary angiographic interpretation, and intravascular ultrasonographic image segmentation. These algorithms have demonstrated promising outcomes in terms of recall, precision, accuracy, and agreement with expert analysts [17].

Another review has found that wearable devices, such as smartwatches and activity trackers, can collect and analyze long-term, continuous data on behavioral or physiological functions, providing healthcare providers with a more comprehensive picture of a patient's health compared to the traditional, sporadic measurements obtained through office consultations and hospitalizations. Wearable devices have numerous clinical applications, including screening for arrhythmias in high-risk populations and the remote management of chronic conditions like heart failure or peripheral artery disease [18].

Ishii et al. have developed and validated an ML-based model to predict future adverse events in patients with atrial fibrillation and stable coronary artery disease. Using randomized survival forest and Cox regression models, they created an integer-based risk score for all-cause mortality, myocardial infarction, stroke, and major bleeding, collectively defined as net adverse clinical events. This scoring system categorizes patients into three risk groups: low-risk (0–4 points), intermediate (5–8 points), and high-risk (≥9 points). The integer-based risk score has demonstrated strong performance in both the development and validation cohorts, exhibiting good discriminatory and calibration power. Decision curve analysis has shown a significant net benefit associated with this score [19].

The widespread adoption of AI in these specialties stands in stark contrast to its integration into general surgery, underscoring a significant gap in both research and clinical applications.

Advancements in artificial intelligence across various surgical departments

AI in neurosurgery has led to significant advancements in tumor identification and surgical planning. ML algorithms are employed to delineate tumors precisely, enhancing surgical accuracy and improving patient outcomes. Additionally, AI assists in predicting risks and developing personalized treatment plans.

Njiwa et al. investigated whether increased preoperative white matter (WM) ^{18}F-FDG uptake can be used to predict surgical outcomes and compared the predictive performance of ^{11}C-flumazenil (FMZ) and ^{18}F-FDG–PET, using advanced ML techniques. They showed that at the group level, patients who were non-seizure-free (NSF) had more pronounced periventricular ^{11}C-FMZ and ^{18}F-FDG signal increases than patients who were seizure-free (SF). Five out of eight patients who were NSF had a periventricular WM signal increase in both ^{11}C-FMZ and ^{18}F-FDG, whereas only one out of eight patients who were SF had a periventricular WM signal increase in ^{11}C-FMZ; and four out of eight had a periventricular WM signal increase in ^{18}F-FDG, at the optimized threshold. Random forest classification correctly identified seven out of eight SF patients and seven out of eight NSF patients using ^{11}C-FMZ images, but only four out of eight SF patients and six out of eight NSF patients, using ^{18}F-FDG. The presence of ipsilateral medial temporal lobe hypometabolism predicted SF outcome status, while the absence thereof predicted NSF; nonetheless, ^{11}C-FMZ-based methods performed better than ^{18}F-FDG-based methods [20].

Ma et al. developed a noninvasive ML model to assist in identifying the grade and mutational status of molecular markers in intramedullary gliomas. This development is significant, as invasive biopsies for histopathological analyses carry a high risk of tissue damage. The results indicated that the Swin transformer-based model achieved high accuracy and dice similarity coefficients in the automatic segmentation of lesions during both the sagittal (SAG) and transverse (TRA) phases, with values of 0.9929 and 0.8697 for the SAG phase and 0.9978 and 0.8738 for the TRA phase. The neural network, based on the proposed multimodal fusion (SAG–TRA–clinical) features, demonstrated superior performance in predicting the grade and mutational status of molecular markers in intramedullary gliomas. The area under the receiver operating characteristic curve (AUC) was 0.8431 for grade prediction, 0.7622 for alpha thalassemia/mental retardation syndrome, X-linked (ATRX) status prediction, and 0.7954 for tumor protein p53 status prediction. The WHO-Mind model achieved the highest AUC, with a value of 0.8431 in the test task; both the WHO-Mind and ATRX-Mind models recorded the highest accuracy, each with a value of 0.8889 [21].

AI, particularly robot-assisted surgery, increases the precision of cardiac surgery. Liu et al. compared the clinical outcomes of robot-assisted cardiac surgery (RACS), utilizing the da Vinci robotic surgery system, with those of traditional open-heart surgery (TOHS). There were no statistically significant differences between the RACS and TOHS groups in terms of reoperation rates due to postoperative bleeding, mortality numbers, and treatment interruptions. The RACS group had shorter operative times and intensive care unit stays, fewer postoperative hospital days, and a quicker return to normal daily activities post-discharge than the TOHS group [22].

Fujita et al. compared minimally invasive direct mitral valve replacement via right thoracotomy with robotic mitral valve replacement to determine the feasibility of using robotic techniques for more complex lesions. They found that the mean complexity score for robotic repairs was significantly higher than that for thoracotomy. Additionally, the robotic group underwent a

greater number of mitral valve replacements using polytetrafluoroethylene and performed fewer ablations. The overall cure rate was 100%, with no early mortalities or strokes observed in either group. In both groups, the mean postoperative residual mitral regurgitation was 0.3. The mean pressure gradient across the mitral valve was 2.4 mmHg in the robotic group and 2.7 mmHg in the thoracotomy group [23].

Another review article examined 27 studies that applied AI and big data to cardiac transplantation, categorizing them into four areas: etiology, diagnosis, prognosis, and treatment. AI-based algorithms demonstrated potential in predicting patterns and determining survival rates. However, the studies selected exhibited a significant risk of bias. The accuracy of AI-based models in predicting survival following cardiopulmonary transplantation and prognosis in thoracic organ transplantation was found to surpass that of traditional statistical methods. ML and DL techniques have improved diagnostic tools for detecting allograft rejection and predicting post-transplant survival. Additionally, ML has been employed to monitor the therapeutic levels of immunosuppressive drugs [24].

The role of AI in orthopedic surgery has been demonstrated in areas such as joint replacement and outcome prediction. AI is utilized for the customization of prosthetics and early diagnosis, thereby improving the success rates and effectiveness of orthopedic interventions.

Houserman et al. assessed the viability of an AI prediction model for knee arthroplasty, utilizing three-view radiography to determine if patients with knee pain required total knee arthroplasty (TKA), unicompartmental knee arthroplasty (UKA), or no arthroplasty at all. The AI model achieved an accuracy of 87.8% and a quadratic-weighted Cohen's kappa score of 0.811 in the holdout test set. It performed exceptionally well in determining whether a patient was a candidate for surgery, reaching an accuracy of 93.8%. The multiclass AUC scores for the three categories—TKA, UKA, and no surgery—were all above 0.95, specifically 0.974, 0.957, and 0.98, respectively. The AI/ML model, as well as AI models in general, demonstrated potential in predicting whether patients are suitable candidates for UKA, TKA, or no surgery [25].

Jang et al. utilized DL to automate the measurement of leg length discrepancy (LLD) using pelvic radiographs and to compare the LLD based on different anatomical landmarks. The DL algorithm has successfully measured LLD on pelvic radiographs by utilizing various combinations of landmarks, achieving intraclass correlation coefficients (ICCs) ranging from 0.73 to 0.98. Measurements of LLD using the teardrop and greater trochanter landmarks have shown an acceptable level of agreement, with an ICC of 0.72 [26].

Advancing artificial intelligence in general surgery: current research landscape and future directions

Research on AI in general surgery is expanding into numerous areas, reflecting the diverse applications of AI in this multifaceted field. The integration of AI into laparoscopic surgery enhances visualization, accuracy, and decision-making during procedures.

In robotic surgery, AI has been leveraged to improve the precision and autonomy of robotic systems, marking a significant shift toward more advanced surgical techniques.

Endo et al. discussed the impact of an AI system on identifying anatomical landmarks associated with reduced bile duct injury during laparoscopic cholecystectomy. After viewing a 20-second video where the AI highlighted landmarks, 26.9% of the images were annotated differently, primarily along the gallbladder line of the extrahepatic bile and cystic ducts. Of these changes, 70% were considered safe. The AI system assisted both novices and experts

in identifying landmarks such as the Rouviere sulcus and the inferior border of the liver, S4. It encouraged changing perspectives in 70% of cases, in a way that was considered safe [27].

Zhang et al. explored the feasibility of conditional autonomy in robotic surgery, specifically focusing on robotic appendectomy. This approach involved using demonstration data gathered from a human operator performing appendectomies in a simulated robotic environment to teach the system the movements and trajectories of the robotic instruments. Extensive validation in a simulated environment, utilizing the da Vinci research kit, demonstrated that the proposed method can perform appendectomies semi-automatically. A framework based on this method could decrease the total working path length, completion time, and appendix stump length, while preserving a high similarity to the demonstrated trajectories [28].

In addition, AI models for surgical risk assessment are currently being developed. These models use patient data and preoperative indicators to predict postoperative complications, aiming to tailor surgical approaches to the specific risks of individual patients. Additionally, AI plays a crucial role in surgical planning, especially in complex procedures. Here, AI-driven image interpretation aids surgeons in making informed decisions.

El Moheb et al. demonstrated that the AI risk calculator, Predictive OpTimal Trees in Emergency Surgery Risk (POTTER), surpassed the surgeon's gestalt in predicting postoperative mortality and outcomes for patients undergoing emergency surgery, except in cases of septic shock. Risk prediction for mortality, bleeding, and pneumonia improved when surgeons used POTTER, although there was no significant improvement for septic shock or ventilator dependence. The AUC was calculated to evaluate the predictive performance of surgeons who used POTTER compared to those who did not [29].

The postoperative phase has also benefited from AI, particularly in the areas of wound analysis and care. AI applications here concentrate on analyzing images of wounds and predicting healing outcomes, potentially leading to more personalized and effective postoperative care strategies.

Tomé et al. highlighted the necessity of AI by demonstrating the challenges in predicting postoperative infections using only correlated data. According to their research, postoperative infections occurred in 24 out of 349 operations, which accounts for 6.89% of all surgeries in their database. Correlation tests employing Pearson and Spearman coefficients indicated a weak correlation between the risk factors and the incidence of infection. An artificial neural network designed for pattern recognition successfully predicted infections in 77.3% of cases, achieving an AUC of 0.9050. Among the misclassifications, seven cases were incorrectly identified as having an infection when none was present, representing 2.0% of the data. Conversely, five cases were incorrectly identified as not having an infection when one was present, representing 1.4% of the data [30].

Overall, these diverse areas of AI application in general surgery underscore the potential of AI to transform various aspects of surgical practice, from preoperative planning to postoperative care [31,32]. As research progresses, the role of AI in general surgery is anticipated to grow, setting the stage for more innovative and effective surgical practices.

Bridging the artificial intelligence gap in general surgery

The integration of AI into specialties like radiology and cardiology has significantly improved diagnostic accuracy and patient care. This stands in stark contrast to its use in general surgery. The disparity underscores the unique challenges faced in general surgery, which include the

variability of surgical procedures and the difficulty in capturing comprehensive datasets for AI training.

Understanding the challenges and successful strategies used in other specialties can provide valuable insights for adapting AI applications in general surgery, suggesting a more focused approach to research and development in this area. The primary issue is the relative scarcity of research directed toward implementing AI in general surgical environments. The inherent variability and complexity of general surgical procedures pose significant challenges in standardizing AI applications, which in turn complicates the integration of AI. Additionally, constructing comprehensive and uniform datasets, crucial for training AI, continues to be a major hurdle in this field [33].

Despite these challenges, there are significant opportunities in general surgery where AI can make substantial contributions, such as in risk assessment and surgical planning [34–36]. Success stories from other medical and surgical fields offer a blueprint and valuable insights for integrating AI into general surgery. By drawing on these experiences, general surgery can tailor AI tools to meet its unique needs, potentially transforming patient care and surgical outcomes [36–38].

Promoting research on AI and the application of AI in general surgery requires fostering interdisciplinary collaboration across various fields, establishing standardized data collection and sharing protocols, securing dedicated funding, and integrating AI education into medical training. It is necessary to address ethical considerations and provide regulatory support to build trust in AI applications. Pilot projects and clinical trials are essential to demonstrate the efficacy and safety of AI technologies in clinical settings, paving the way for their integration into general surgery to enhance outcomes and patient care.

Conclusion

The future of AI in general surgery is poised for transformative growth, driven by emerging technologies. Surgical robotics are increasing precision and safety, virtual reality simulations are providing unparalleled training experiences, and predictive analytics are improving postoperative care.

Focusing research on these areas could significantly advance the field of general surgery, aligning it with the successes observed in other medical fields and opening new avenues for enhancing patient care. In radiology, oncology, and cardiology, AI has already begun to transform patient care by improving diagnostic accuracy, providing predictive analytics, and facilitating personalized treatment plans.

However, the field of general surgery stands at the threshold of a significant technological evolution, facing unique challenges that hinder the integration of AI. To effectively incorporate AI into general surgery and address delays in current research and development, interdisciplinary collaboration is essential. This requires forming partnerships among medical practitioners, AI technologists, data scientists, and policymakers. These collaborative efforts are vital for managing the complexities of general surgical procedures, standardizing AI applications, and constructing the comprehensive datasets required for AI training.

By leveraging diverse expertise, AI tools can be tailored to meet the unique requirements of general surgery, thereby improving surgical outcomes, procedural efficiency, and patient care.

The path forward requires a concerted effort to bridge this gap, focusing on the development of AI tools tailored to the specific needs of general surgery, from preoperative planning to

postoperative care. Embracing AI in general surgery not only promises to improve surgical outcomes and efficiency but also represents a critical step toward a future where healthcare fully leverages technology, marking a new chapter in the quest for enhanced patient care.

ORCID
Seung Min Baik: https://orcid.org/0000-0003-1051-6775
Ryung-Ah Lee: https://orcid.org/0000-0003-1146-3839

Authors' contributions
Project administration: not applicable
Conceptualization: Lee RA
Methodology & data curation: Baik SM
Funding acquisition: not applicable
Writing – original draft: Baik SM
Writing – review & editing: Baik SM, Lee RA

Conflict of interest
Ryung-Ah Lee has been an associate editor of the *Ewha Medical Journal* since August 2023. However, she was not involved in the review process. No other potential conflict of interest relevant to this review was reported.

Funding
Not applicable.

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Not applicable.

References

1. Mehta V. Artificial intelligence in medicine: revolutionizing healthcare for improved patient outcomes. *J Med Res Innov* 2023;7(2):e000292. https://doi.org/10.32892/jmri.292
2. Manfred D. Artificial intelligence (AI): what are the impacts for medicine? *J Artif Intell Cloud Comput* 2019;2(2):1-3. https://doi.org/10.47363/JAICC/2023(2)117
3. Paudyal R, Shah AD, Akin O, Do RKG, Konar AS, Hatzoglou V, et al. Artificial intelligence in CT and MR imaging for oncological applications. *Cancers* 2023;15(9):2573. https://doi.org/10.3390/cancers15092573
4. Derevianko A, Pizzoli SFM, Pesapane F, Rotili A, Monzani D, Grasso R, et al. The use of artificial intelligence (AI) in the radiology field: what is the state of doctor–patient communication in cancer diagnosis? *Cancers* 2023;15(2):470. https://doi.org/10.3390/cancers15020470
5. Ram S, Bodduluri S. Implementation of artificial intelligence–assisted chest X-ray interpretation: it is about time. *Am Thorac Soc* 2023;20(5):641-642. https://doi.org/10.1513/AnnalsATS.202303-195ED
6. Harry A. The future of medicine: harnessing the power of AI for revolutionizing healthcare. *Int J Multidiscip Sci Arts* 2023;2(1):36-47. https://doi.org/10.47709/ijmdsa.v2i1.2395
7. Demetriou DD, Hull R, Kgoebane-Maseko M, Lockhat Z, Dlamini Z. AI-enhanced digital pathology and radiogenomics in precision oncology. In: Dlamini Z, editor. Artificial intelligence and precision oncology: bridging cancer research and clinical decision support. Cham: Springer; 2023. p.93-113.
8. Zeineldin RA, Junger D, Mathis-Ullrich F, Burgert O. Development of an AI-driven system for neurosurgery with a usability study: a step towards minimal invasive robotics. *at - Automatisierungstechnik* 2023;71(7):537-546. https://doi.org/10.1515/auto-2023-0061
9. Vidal-Perez R, Vazquez-Rodriguez JM. Role of artificial intelligence in cardiology. *World J Cardiol* 2023;15(4):116-118. https://doi.org/10.4330/wjc.v15.i4.116
10. Voskens FJ, Abbing JR, Ruys AT, Ruurda JP, Broeders IAMJ. A nationwide survey on the perceptions of general surgeons on artificial intelligence. *Artif Intell Surg* 2022;2(1):8-17.

https://doi.org/10.20517/ais.2021.10
11. Lång K, Josefsson V, Larsson AM, Larsson S, Högberg C, Sartor H, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *Lancet Oncol* 2023;24(8):936-944. https://doi.org/10.1016/S1470-2045(23)00298-X
12. Nam JG, Hwang EJ, Kim J, Park N, Lee EH, Kim HJ, et al. AI improves nodule detection on chest radiographs in a health screening population: a randomized controlled trial. *Radiology* 2023;307(2):e221894. https://doi.org/10.1148/radiol.221894
13. Sachpekidis C, Enqvist O, Ulén J, Kopp-Schneider A, Pan L, Jauch A, et al. Application of an artificial intelligence-based tool in [18F]FDG PET/CT for the assessment of bone marrow involvement in multiple myeloma. *Eur J Nucl Med Mol Imaging* 2023;50(12):3697-3708. https://doi.org/10.1007/s00259-023-06339-5
14. Clift AK, Dodwell D, Lord S, Petrou S, Brady M, Collins GS, et al. Development and internal-external validation of statistical and machine learning models for breast cancer prognostication: cohort study. *BMJ* 2023;381:e073800. https://doi.org/10.1136/bmj-2022-073800
15. Alaimo L, Lima HA, Moazzam Z, Endo Y, Yang J, Ruzzenente A, et al. Development and validation of a machine-learning model to predict early recurrence of intrahepatic xholangiocarcinoma. *Ann Surg Oncol* 2023;30(9):5406-5415. https://doi.org/10.1245/s10434-023-13636-8
16. Liu M, Wu J, Wang N, Zhang X, Bai Y, Guo J, et al. The value of artificial intelligence in the diagnosis of lung cancer: a systematic review and meta-analysis. *PLOS ONE* 2023;18(3):e0273445. https://doi.org/10.1371/journal.pone.0273445
17. Subhan S, Malik J, Haq A, Qadeer MS, Zaidi SMJ, Orooj F, et al. Role of artificial intelligence and machine learning in interventional cardiology. *Curr Probl Cardiol* 2023;48(7):101698. https://doi.org/10.1016/j.cpcardiol.2023.101698
18. Hughes A, Shandhi MMH, Master H, Dunn J, Brittain E. Wearable devices in cardiovascular medicine. *Circ Res* 2023;132(5):652-670. https://doi.org/10.1161/CIRCRESAHA.122.322389
19. Ishii M, Kaikita K, Yasuda S, Akao M, Ako J, Matoba T, et al. Risk prediction score for clinical outcome in atrial fibrillation and stable coronary artery disease. *Open Heart* 2023;10(1):e002292. https://doi.org/10.1136/openhrt-2023-002292
20. Yankam Njiwa J, Gray KR, Costes N, Mauguiere F, Ryvlin P, Hammers A. Advanced [^{18}F]FDG and [^{11}C]flumazenil PET analysis for individual outcome prediction after temporal lobe epilepsy surgery for hippocampal sclerosis. *Neuroimage Clin* 2015;7:122-131. https://doi.org/10.1016/j.nicl.2014.11.013
21. Ma C, Wang L, Song D, Gao C, Jing L, Lu Y, et al. Multimodal-based machine learning strategy for accurate and non-invasive prediction of intramedullary glioma grade and mutation status of molecular markers: a retrospective study. *BMC Med* 2023;21(1):198. https://doi.org/10.1186/s12916-023-02898-4
22. Liu Z, Zhang C, Ge S. Efficacy and safety of robotic-assisted versus median sternotomy for cardiac surgery: results from a university affiliated hospital. *J Thorac Dis* 2023;15(4):1861-1871. https://doi.org/10.21037/jtd-23-197
23. Fujita T, Kakuta T, Kawamoto N, Shimahara Y, Yajima S, Tadokoro N, et al. Benefits of robotically-assisted surgery for complex mitral valve repair. *Interact Cardiovasc Thorac Surg* 2021;32(3):417-425. https://doi.org/10.1093/icvts/ivaa271
24. Palmieri V, Montisci A, Vietri MT, Colombo PC, Sala S, Maiello C, et al. Artificial intelligence, big data and heart transplantation: actualities. *Int J Med Inform* 2023;176:105110. https://doi.org/10.1016/j.ijmedinf.2023.105110
25. Houserman DJ, Berend KR, Lombardi AV Jr, Fischetti CE, Duhaime EP, Jain A, et al. The viability of an artificial intelligence/machine learning prediction model to determine candidates for knee arthroplasty. *J Arthroplasty* 2023;38(10):2075-2080. https://doi.org/10.1016/j.arth.2022.04.003
26. Jang SJ, Kunze KN, Bornes TD, Anderson CG, Mayman DJ, Jerabek SA, et al. Leg-length discrepancy variability on standard anteroposterior pelvis radiographs: an analysis using deep learning measurements. *J Arthroplasty* 2023;38(10):2017-2023.E3. https://doi.org/10.1016/j.arth.2023.03.006
27. Endo Y, Tokuyasu T, Mori Y, Asai K, Umezawa A, Kawamura M, et al. Impact of AI system on recognition for anatomical landmarks related to reducing bile duct injury during laparoscopic cholecystectomy. *Surg Endosc* 2023;37(7):5752-5759. https://doi.org/10.1007/s00464-023-10224-5
28. Zhang R, Chen J, Wang Z, Yang Z, Ren Y, Shi P, et al. A step towards conditional autonomy - robotic appendectomy. *IEEE Robot Autom Lett* 2023;8(5):2429-2436. https://doi.org/10.1109/LRA.2023.3254859
29. El Moheb M, Gebran A, Maurer LR, Naar L, El Hechi M, Breen K, et al. Artificial intelligence versus surgeon gestalt in predicting risk of emergency general surgery. *J Trauma Acute Care Surg* 2023;95(4):565-572. https://doi.org/10.1097/TA.0000000000004030
30. Tomé F, Michelin L, Lins RS, Bringmann DR, Corso LL. Using artificial neural networks for pattern recognition of post-surgical infections. *Braz J Health Rev* 2023;6(1):3329-3339. https://doi.org/10.34119/bjhrv6n1-260

31. Kokkinakis S, Kritsotakis EI, Lasithiotakis K. Artificial intelligence in surgical risk prediction. *J Clin Med* 2023;12(12):4016. https://doi.org/10.3390/jcm12124016
32. Watanabe A, Wiseman SM. A new era in surgical research: the evolving role of artificial intelligence. *Am J Surg* 2023;226(6):923-925. https://doi.org/10.1016/j.amjsurg.2023.06.040
33. Rimmer L, Howard C, Picca L, Bashir M. The automaton as a surgeon: the future of artificial intelligence in emergency and general surgery. *Eur J Trauma Emerg Surg* 2021;47(3):757-762. https://doi.org/10.1007/s00068-020-01444-8
34. Zhou XY, Guo Y, Shen M, Yang GZ. Application of artificial intelligence in surgery. *Front Med* 2020;14(4):417-430. https://doi.org/10.1007/s11684-020-0770-0
35. Mangano A, Valle V, Dreifuss NH, Aguiluz G, Masrur MA. Role of artificial intelligence (AI) in surgery: introduction, general principles, and potential applications. *Surg Technol Int* 2020;38:17-21. https://doi.org/10.52198/21.STI.38.SO1369
36. McCartney J. AI is poised to "revolutionize" surgery [Internet]. Chicago (IL): American College of Surgeons; c2023 [cited 2024 Jan 10]. Available from: https://www.facs.org/for-medical-professionals/news-publications/news-and-articles/bulletin/2023/june-2023-volume-108-issue-6/ai-is-poised-to-revolutionize-surgery/
37. Egert M, Steward JE, Sundaram CP. Machine learning and artificial intelligence in surgical fields. *Indian J Surg Oncol* 2020;11(4):573-577. https://doi.org/10.1007/s13193-020-01166-8
38. Bar O, Neimark D, Zohar M, Hager GD, Girshick R, Fried GM, et al. Impact of data on generalization of AI for surgical intelligence applications. *Sci Rep* 2020;10(1):22208. https://doi.org/10.1038/s41598-020-79173-6

Original Article

Ewha Med J 2024;47(2):e23
https://doi.org/10.12771/emj.2024.e23
eISSN 2234-2591

An accurate pediatric bone age prediction model using deep learning and contrast conversion

Dong Hyeok Choi[1,2,3], **So Hyun Ahn**[4,5], **Rena Lee**[6]

[1]Department of Medicine, Yonsei University College of Medicine, Seoul, Korea
[2]Medical Physics and Biomedical Engineering Lab (MPBEL), Yonsei University College of Medicine, Seoul, Korea
[3]Department of Radiation Oncology, Heavy Ion Therapy Research Institute, Yonsei Cancer Center, Yonsei University College of Medicine, Seoul, Korea
[4]Ewha Medical Research Institute, Ewha Womans University College of Medicine , Seoul, Korea
[5]Ewha Medical Artificial Intelligence Research Institute, Ewha Womans University College of Medicine, Seoul, Korea
[6]Department of Biomedical Engineering, Ewha Womans University College of Medicine, Seoul, Korea

Received Mar 3, 2024
Revised Apr 17, 2024
Accepted Apr 17, 2024

Corresponding author
So Hyun Ahn
Ewha Medical Research Institute, Ewha Womans University College of Medicine, 25, Magokdong-ro 2-gil, Gangseo-gu, Seoul 07804, Korea
E-mail: mpsohyun@ewha.ac.kr

Keywords
Bone age measurement; X-ray image; Deep learning

Objectives: This study aimed to develop an accurate pediatric bone age prediction model by utilizing deep learning models and contrast conversion techniques, in order to improve growth assessment and clinical decision-making in clinical practice.

Methods: The study employed a variety of deep learning models and contrast conversion techniques to predict bone age. The training dataset consisted of pediatric left-hand X-ray images, each annotated with bone age and sex information. Deep learning models, including a convolutional neural network , Residual Network 50 , Visual Geometry Group 19, Inception V3, and Xception were trained and assessed using the mean absolute error (MAE). For the test data, contrast conversion techniques including fuzzy contrast enhancement, contrast limited adaptive histogram equalization (HE) , and HE were implemented. The quality of the images was evaluated using peak signal-to-noise ratio (SNR), mean squared error, SNR, coefficient of variation, and contrast-to-noise ratio metrics. The bone age prediction results using the test data were evaluated based on the MAE and root mean square error, and the t-test was performed.

Results: The Xception model showed the best performance (MAE=41.12). HE exhibited superior image quality, with higher SNR and coefficient of variation values than other methods. Additionally, HE demonstrated the highest contrast among the techniques assessed, with a contrast-to-noise ratio value of 1.29. Improvements in bone age prediction resulted in a decline in MAE from 2.11 to 0.24, along with a decrease in root mean square error from 0.21 to 0.02.

Conclusion: This study demonstrates that preprocessing the data before model training does not significantly affect the performance of bone age prediction when comparing contrast-converted images with original images.

Introduction

Background

Heterogeneity in skeletal maturation is influenced by a complex interplay of factors, including genetic predispositions, the nutritional and growth status of the child, the onset of precocious puberty, hormonal variations, conditions related to pediatric endocrinology and metabolic disorders, and ailments affecting the musculoskeletal system [1–3]. The assessment

of bone age, especially through methods that examine growth plates, is crucial not only for identifying precocious puberty and providing benchmarks for growth trajectories and future height predictions but also for managing conditions such as adolescent idiopathic scoliosis and determining the appropriate timing for orthopedic interventions in children with skeletal anomalies [3–6]. Thus, the appraisal of bone age using standardized methods is paramount for diagnosing, managing, and developing effective therapeutic strategies for these conditions.

Conventional methods for determining bone age in children, such as cervical vertebral maturation, the Roche-Wainer-Thissen criterion for knee assessment, and Risser's sign for evaluating the iliac crest apophysis, are supplemented by more commonly used techniques like the Greulich and Pyle (GP) and Tanner Whitehouse (TW3) methods, which utilize radiographic images of the left hand. The GP method provides a straightforward way to estimate bone age by comparing the bony structures of the hand and wrist with a sex-specific collection of images that depict various stages of skeletal maturity. However, its accuracy can be compromised in cases of significant skeletal deformities, with the range of evaluative intervals in the image collection spanning from six months to a year. In contrast, the Tanner-Whitehouse (TW) approach assigns grades from A to I to each bone in the targeted area, comparing them to a standard dataset and aggregating these maturity scores to predict bone age. While the TW method is known for its complexity and precision, offering enhanced reliability, it also requires a more substantial time commitment [1,7].

The reliability of both GP and TW assessments depends on the subjective interpretation by radiologists, which can lead to variability in outcomes based on the evaluator's expertise [2,6]. This highlights the clinical need for more accurate and time-efficient methods for determining bone age. Recent advancements have led to the development of automated bone age assessment techniques that utilize AI technology, with commercial AI-based software solutions like BoneXpert and VUNO now available for clinical use. These innovations represent a significant shift towards more precise and dependable bone age assessment protocols.

Despite these technological advances, challenges remain, particularly in analyzing images affected by suboptimal quality or unusual skeletal structures. Furthermore, there is a significant lack of discussion concerning the effectiveness of post-processing techniques in conventional growth plate analyses.

Objectives

This study aims to address these gaps by exploring methods to enhance image contrast, thereby improving the accuracy of region of interest classification and contributing to the advancement of bone age assessment technologies.

Methods

Ethics statement

This study is based on publicly available, anonymous X-ray image data; therefore, approval by the institutional review board and the requirement for informed consent were exempted.

Study design

This was a methodological study to predict values in model training for bone age.

Study procedure

In this study, we used training and validation data that had been preprocessed and normalized,

utilizing Light hand X-ray images and comma-separated values (CSV) file-type labels for model training. We employed several models, including a convolutional neural network (CNN), Residual Network 50 (ResNet 50), Visual Geometry Group (VGG) 19, Inception V3, and Xception. To derive the predicted values for the test images, we stored the weight value corresponding to the smallest validation loss observed during the model training. The resulting values were saved in a comma-separated values file, and we evaluated each model by comparing the root mean square error (RMSE) values.

Data sources

The data used in this study were obtained from the dataset released during the 2017 RSNA AI Pediatric Bone Age Challenge (Dataset 1), which was created by Stanford University and the University of Colorado and annotated by multiple expert observers. This dataset includes a total of 126,111 pediatric left-hand X-ray images, each labeled with the subject's sex and bone age. The age range of the subjects in these images spans from 1 month to 228 months and comprises 6,833 male and 5,778 female subjects. All data feature normalized resolution and have not been processed. Additionally, the data were collected in a multi-institutional setting, with labeling performed collaboratively by two pediatric radiologists from each institution. Table 1 shows specific details regarding the 2017 RSNA AI Pediatric Bone Age Challenge dataset. Data generated and/or analyzed during the current study are available in Dataset 2.

Preprocessing and augmentation

For the model training phase, 100,888 images, representing 80% of the total 126,111 images in the dataset, were used for training. The remaining 20%, or 25,123 images, were set aside for validation. The testing procedures utilized a subset of 100 images. Additionally, all images were resized to a resolution of 256×256 pixels in RGB format, and processing was carried out in batches of 32, using a random seed of 42 to ensure consistency. The training images underwent augmentation through vertical flipping, a technique used to increase data diversity and improve the model's generalization performance.

Table 1. Description of the 2017 RSNA pediatric bone age challenge dataset

Items	Description
Imaging modality	X-ray Preferred name: digital radiography RadLex ID: RID10351
Annotation pattern	Whole study label
Annotation methodology and structure	Method of annotation - Manual Annotation output - Spreadsheet (alphanumeric) Storage, Portability, Interoperability - Downloadable ZIP file (RSNA website)
Imaging file/structure set format	Portable Network Graphic (PNG)
Image characteristics	Resolution - Normalized Preprocessing - None
Labeler demographics	Scope of annotation: multi-institutional - Two pediatric radiologists from each institution clinical report

https://doi.org/10.12771/emj.2024.e23

Training and evaluation

The Adam optimizer was used as the optimization function, and mean absolute error (MAE) served as the evaluation metric. The model underwent 50 epochs, each consisting of 300 steps, and it was subjected to both training and validation processes. These processes were essential for monitoring validation loss to determine the model's optimal performance, which was achieved when the loss value was at its minimum. The loss value and MAE from the validation phase confirmed the learning verification for each model on a monthly basis.

The formula for MAE is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{N}\left|y_i - y_i^{'}\right| \quad (1)$$

n is the number of samples or data points, y_i represents the actual or observed value, and y_i' represents the predicted value.

The comparison of performance evaluations across models was shown as the distribution of differences between the labeled age and the predicted age.

Image contrast conversion and quantitative analysis

Contrast conversion procedures were conducted on 100 test datasets. Three distinct algorithms were employed for contrast adjustment: fuzzy contrast enhancement (FCE), histogram equalization (HE), and contrast limited adaptive histogram equalization (CLAHE). The FCE algorithm enhances image contrast by applying principles of fuzzy logic. This method involves fuzzifying the pixel intensities and then defuzzifying the resulting fuzzy set. The formal expression for the FCE algorithm is articulated as follows:

$$FCE\ (x,\ mean\ ,\ std\ ) = e^{-0.5} \times (stdx -\ mean\ ) \quad (2)$$

The conventional HE technique employs histogram equalization to enhance contrast. This algorithm involves calculating the histogram of the input image, followed by deriving the cumulative distribution function. Afterward, histogram normalization is performed, and the cumulative distribution function is used to adjust the pixel values in the image.

Conversely, the CLAHE algorithm utilizes a contrast-constrained adaptive HE approach to enhance image contrast. This method divides the image into discrete, small blocks, applying HE independently to each one. Contrast constraints are applied to improve the contrast within each image segment. Subsequently, all blocks are combined to produce the final image.

To assess the image quality of the contrast-transformed image, we analyzed several metrics, including the peak signal-to-noise ratio (PSNR), mean squared error (MSE), signal-to-noise ratio (SNR), coefficient of variation (COV), and contrast-to-noise ratio (CNR). The formulas for each metric are as follows.

$$PSNR = 10 \times log_{10}\left(\frac{M_p}{MSE}\right) \quad (3)$$

M_p is the maximum possible pixel value, and MSE is the mean squared error between the original and distorted images.

$$MSE = \frac{1}{AB}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[I(i,j) - K(i,j)]^2 \quad (4)$$

https://doi.org/10.12771/emj.2024.e23

A and B are the dimensions of the image. $I(i,j)$ and $K(i,j)$ are the pixel intensities of the original and distorted images, respectively.

$$SNR = 10 \times log_{10}\left(\frac{SP}{NP}\right) \quad (5)$$

SP represents the strength of the desired information in the image. NP represents the level of unwanted background noise in the image.

$$COV = 100 \times \frac{SD}{M} \quad (6)$$

M represents the average contrast level in the image. SD denotes the variability or dispersion of noise within the image.

$$CNR = \frac{MC}{SD} \quad (7)$$

MC represents the average contrast level in the image.

Image contrast conversion and quantitative analysis

A comprehensive assessment was conducted using 100 test sets to calculate the MAE and RMSE, thereby evaluating the accuracy of bone age estimation for each contrast-converted image. MAE was calculated according to Equation (1), and RMSE according to Equation (8).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{N}\left|y_i - y_i'\right|^2} \quad (8)$$

n is the number of samples or data points, y_i represents the actual or observed value, and y_i' represents the predicted value.

Statistical methods

The statistical significance of the findings was assessed using the t-test, with the predetermined threshold for statistical significance established at $P<0.05$.

Results

Subjects' characteristics

The sex distribution of subjects and the monthly age distribution for males and females in this study are presented in Fig. 1.

Model performance

The RMSE values for the predicted bone age relative to the actual age, used as metrics to assess model performance in the study, were 50.91 for CNN, 55.29 for ResNet 50, 50.29 for VGG 19, 48.74 for Inception V3, and 41.12 for Xception. A graphical representation illustrating the outcomes of bone age prediction in relation to chronological age is shown in Fig. 2.

X-ray image contrast conversion

CLAHE, FCE, and HE were individually applied to the test data for model evaluation to perform contrast transformation. An example of a contrast-enhanced image is shown in Fig. 3.

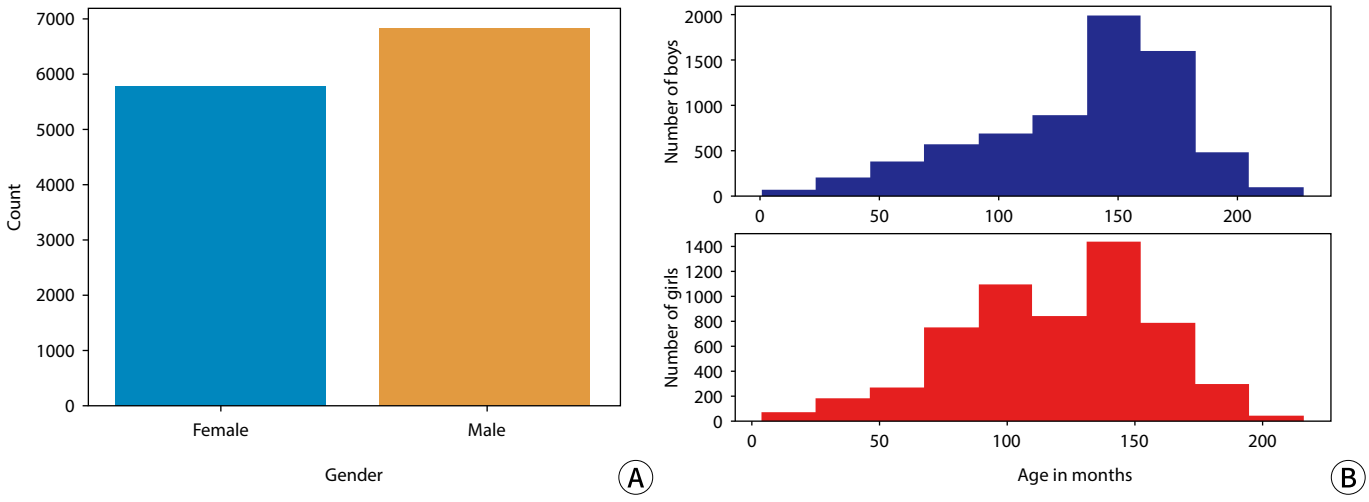


Fig. 1. Histograms depicting (A) the sex distribution and (B) the monthly age distribution for males and females.

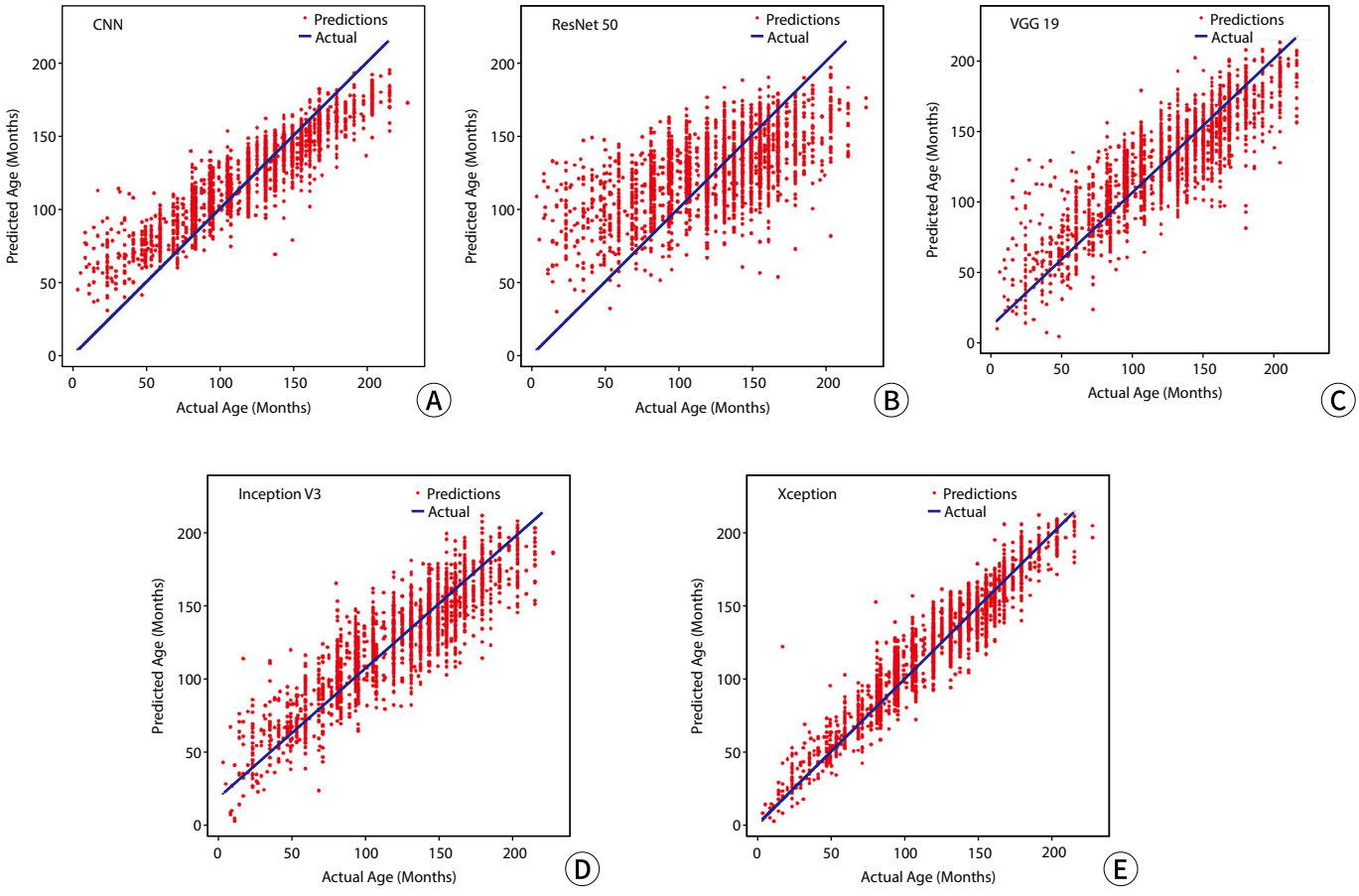


Fig. 2. Comparison of bone age and model predictions. (A) CNN, (B) ResNet 50, (C) VGG 19, (D) Inception V3, and (E) Xception. The blue line represents the actual bone age, while the red dot represents the predicted result. CNN, convolutional neural network; ResNet 50, Residual Network 50; VGG 19, Visual Geometry Group 19.

https://doi.org/10.12771/emj.2024.e23

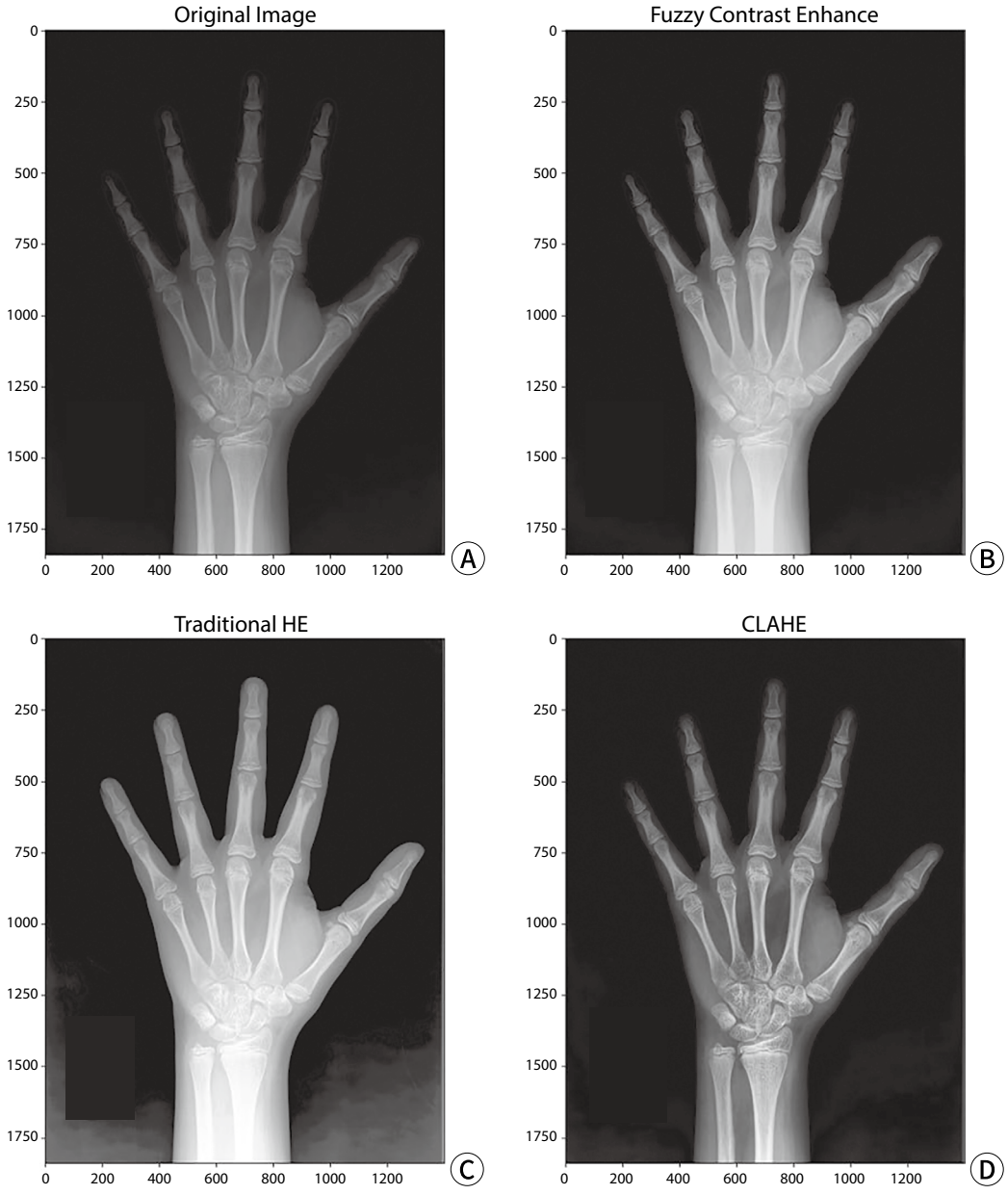


Fig. 3. The original left-hand X-ray image and the image after applying each contrast conversion algorithm. (A) Original image, (B) FCE algorithm applied, (C) HE algorithm applied, (D) CLAHE algorithm applied. HE, histogram equalization; CLAHE, contrast limited adaptive histogram equalization; FCE, fuzzy contrast enhancement.

The quantitative assessment of each image utilized PSNR, MSE, SNR, and CNR (Fig. 4). In terms of PSNR and MSE values, image quality was ranked from highest to lowest as follows: FCE, CLAHE, and HE. Regarding factors evaluating noise and signal intensity in the images, SNR and COV exhibited higher values in the order of HE, FCE, and CLAHE. Specifically, for HE, SNR and COV were notably higher at 1.83 and 1.31, respectively, representing more than a sevenfold and threefold difference compared to other algorithms, respectively. In assessing contrast, CNR values were highest for HE, followed in descending order by FCE and CLAHE, with HE demonstrating the highest contrast at 1.29.

https://doi.org/10.12771/emj.2024.e23

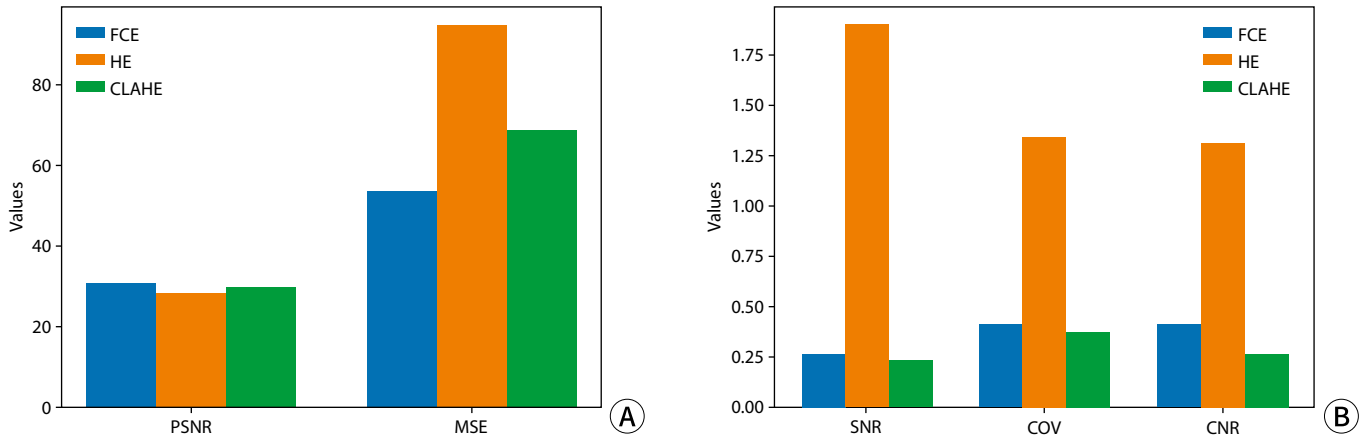


Fig. 4. Quantitative analysis results of images obtained using the contrast conversion algorithm. (A) PSNR and MSE results; (B) SNR, COV, and CNR results. FCE, fuzzy contrast enhancement; HE, histogram equalization; CLAHE, contrast limited adaptive histogram equalization; PSNR, peak signal-to-noise ratio; MSE, mean squared error; SNR, signal-to-noise ratio; COV, coefficient of variation; CNR, contrast-to-noise ratio.

Bone age prediction

A total of 100 original and contrast-enhanced images were used as test data for bone age prediction in each model. Table 2 presents the MAE, RMSE, and P-value of the bone age prediction results across various models and contrast conversion algorithms. To facilitate comparison of bone age prediction performance using each contrast algorithm, evaluation results for the original images were also included. The accuracy of bone age prediction has improved, with statistically significant enhancements observed when using CLAHE in the CNN model, HE in the Inception V3 model, and HE in the VGG 19 model. In the Xception model, although the application of CLAHE and FCE algorithms led to better accuracy in bone age prediction, the improvements were not statistically significant.

Improvements in bone age prediction led to a reduction in MAE from 2.11 to 0.24 and a decrease in RMSE from 0.21 to 0.02.

Discussion

Key results

In this study, we implemented various bone age prediction models using identical parameters and evaluated the results by modifying the contrast of the test data. The Xception model demonstrated the most accurate bone age predictions. After adjusting the contrast, the PSNR and MSE metrics revealed that the FCE algorithm delivered the highest quality results. Furthermore, the quantitative assessments of SNR, COV, and CNR indicated that the HE algorithm produced the highest values. The prediction of bone age with contrast-adjusted images showed improved performance in 5 out of 15 cases compared to the original images. However, two of these five cases did not achieve statistical significance.

Interpretation

The primary cause of these outcomes was linked to the use of unprocessed images in the training dataset. The original images, obtained from various institutions, showed variations in how much of the left hand was captured, with some images featuring the left hand in non-

Table 2. MAE, RMSE, and t-test results for each model, comparing the actual bone age to the predicted bone age for both the original images and the images subjected to the contrast conversion algorithm

Model	Items	Original	Image contrast conversion methods		
			CLAHE	FCE	HE
CNN	MAE	26.05	25.81	32.21	31.74
	RMSE	2.60	2.58	3.22	3.17
	P-value		<0.05	0.22	<0.05
ResNet 50	MAE	43.69	46.58	44.63	54.35
	RMSE	4.37	4.66	4.46	5.43
	P-value		<0.05	<0.05	<0.05
VGG 19	MAE	34.26	36.09	34.86	33.08
	RMSE	3.43	3.61	3.49	3.31
	P-value		<0.05	<0.05	<0.05
Inception V3	MAE	34.54	36.03	34.94	33.29
	RMSE	3.45	3.60	3.49	3.33
	P-value		0.25	<0.05	<0.05
Xception	MAE	32.24	30.13	31.64	34.13
	RMSE	3.22	3.01	3.06	3.41
	P-value		0.49	0.19	0.05

MAE, mean absolute error; RMSE, root mean square error; CLAHE, contrast limited adaptive histogram equalization; FCE, fuzzy contrast enhancement; HE, histogram equalization; CNN, convolutional neural network; ResNet 50, Residual Network 50; VGG 19, Visual Geometry Group 19.

horizontal positions. Although training the model with diverse datasets might enhance its applicability across different institutions, it could also negatively affect the model's performance. Future efforts will focus on acquiring preprocessed training data, which will involve adjusting the image contrast and ensuring that each image is horizontally aligned at the wrist bone through image registration. Additionally, in this study, the training and validation sets were separated in only one instance for individual model training. Future plans include the use of k-fold learning during model training to facilitate integrated learning and validation across the entire dataset.

Racial and ethnic disparities, along with variations in nutritional status and overall health, may affect bone age measurements. This suggests that applying bone age criteria directly to contemporary children and adolescents may not be appropriate [8]. Previous studies have developed deep learning-based bone age prediction models specifically optimized for Korean children and adolescents. These models use hand and wrist radiographs and have been evaluated for their validity compared to conventional methods [9].

This study demonstrated that the deep learning-based Korean model achieved superior bone age prediction accuracy compared to conventional methods, marking a significant advancement in precise growth assessment and clinical decision-making. The Korean bone age model reduces prediction biases and delivers more accurate age predictions across different age groups. Therefore, it is imperative to develop bone age prediction models that are customized for various racial groups.

Limitations

This study does not have any limitations that warrant discussion.

Suggestion for further studies

Future research directions include preprocessing training data to ensure consistency in image quality and registration, implementing k-fold training to enhance model robustness, and fine-tuning models using datasets specific to Korean populations. These endeavors aim to enhance the overall accuracy and applicability of bone age prediction models in clinical practice, ultimately improving growth assessment and clinical decision-making for pediatric patients.

Conclusion

This study shows that when model learning is performed using non-preprocessed data, there is no significant difference in bone age prediction performance between contrast-converted images and original images. Rather than applying post-processing to the test dataset to improve predictions, it will be necessary to preprocess the training dataset.

ORCID
Dong Hyeok Choi: https://orcid.org/0000-0002-7065-6115
So Hyun Ahn: https://orcid.org/0000-0002-0116-3325
Rena Lee: https://orcid.org/0009-0003-3630-7813

Authors' contributions
Project administration: Ahn SH
Conceptualization: Ahn SH
Methodology & data curation: Choi DH, Lee R
Funding acquisition: Ahn SH, Lee R
Writing – original draft: Choi DH, Ahn SH, Lee R
Writing – review & editing: Choi DH, Ahn SH, Lee R

Conflict of interest
So Hyun Ahn has been an assistant editor of the *Ewha Medical Journal* since August 2023. However, she was not involved in the review process. No other potential conflict of interest relevant to this review was reported.

Funding
This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2023-00240003 and RS-2023-00257618).

Data availability
Data files are available from Harvard Dataverse: https://doi.org/10.7910/DVN/NMUR8X

Dataset 1. 2017 RSNA AI Pediatric Bone Age Challenge. Available from: https://www.rsna.org/rsnai/ai-image-challenge/RSNA-Pediatric-Bone-Age-Challenge-2017
Dataset 2. The datasets generated during and/or analyzed during the current study

Acknowledgments
Not applicable.

Supplementary materials
Not applicable.

References

1. Cavallo F, Mohn A, Chiarelli F, Giannini C. Evaluation of bone age in children: a mini-review. *Frontiers Pediatr* 2021;9:580314. https://doi.org/10.3389/fped.2021.580314
2. Wang S, Wang X, Shen Y, He B, Zhao X, Cheung PWH, et al. An ensemble-based densely-connected deep learning system for assessment of skeletal maturity. *IEEE Trans Syst Man Cybern Syst* 2020;52(1):426-437. https://doi.org/10.1109/TSMC.2020.2997852
3. Ferrillo M, Curci C, Roccuzzo A, Migliario, M, Invernizzi M, de Sire A. Reliability of cervical vertebral maturation compared to hand-wrist for skeletal maturation assessment in growing subjects: a systematic review. *J Back Musculoskelet Rehabil* 2021;34(6):925-936. https://doi.org/10.3233/BMR-210003

4. Satoh M, Hasegawa Y. Factors affecting prepubertal and pubertal bone age progression. *Front Endocrinol* 2022;13:967711. https://doi.org/10.3389/fendo.2022.967711
5. Ahn KS, Bae B, Jang WY, Lee JH, Oh S, Kim BH, et al. Assessment of rapidly advancing bone age during puberty on elbow radiographs using a deep neural network model. *Eur Radiol* 2021;31(12):8947-8955. https://doi.org/10.1007/s00330-021-08096-1
6. Maratova K, Zemkova D, Sedlak P, Pavlikova M, Amaratunga SA, Krasnicanova H, et al. A comprehensive validation study of the latest version of BoneXpert on a large cohort of Caucasian children and adolescents. *Front Endocrinol* 2023;14:1130580. https://doi.org/10.3389/fendo.2023.1130580
7. Son SJ, Song Y, Kim N, Do Y, Kwak N, Lee MS, et al. TW3-based fully automated bone age assessment system using deep neural networks. *IEEE Access* 2019;7:33346-33358. https://doi.org/10.1109/ACCESS.2019.2903131
8. Gilsanz V, Ratib O. Hand bone age: a digital atlas of skeletal maturity. Berlin: Springer; 2005.
9. Kim PH, Yoon HM, Kim JR, Hwang JY, Choi JH, Hwang J, et al. Bone age assessment using artificial intelligence in Korean pediatric population: a comparison of deep-learning models trained with healthy chronological and Greulich-Pyle ages as labels. *Korean J Radiol* 2023;24(11):1151-1163. https://doi.org/10.3348/kjr.2023.0092

Original Article

Ewha Med J 2024;47(2):e24
https://doi.org/10.12771/emj.2024.e24
eISSN 2234-2591

Using an influenza epidemic threshold different from those in the United States and Europe caused longer epidemic periods in Korea during the 2018–2019, 2019–2020, and 2022–2023 seasons: a comparative study

Joowon Lee, Sooyoung Huh, Haesook Seo

Infectious Disease Research Center, Citizen's Health Bureau, Seoul Metropolitan Government, Seoul, Korea

Received Jan 26, 2024
Revised Feb 16, 2024
Accepted Mar 7, 2024

Corresponding author
Joowon Lee
Infectious Disease Research Center, Citizen's Health Bureau, Seoul Metropolitan Government, 110 Sejong-daero, Jung-gu, Seoul 04524, Korea
E-mail: joowon.lee@seoul.go.kr

Keywords
Influenza, human; Sentinel surveillance; Epidemics; Seasons

Objectives: During the COVID-19 pandemic, the first seasonal influenza epidemic was declared in the 37th week of 2022 in Korea and has continued through the winter of 2023–2024. However, this finding has not been observed in the United States and Europe. The present study aimed to determine whether the prolonged influenza epidemic in Korea from 2022 to 2023 was caused by using a different influenza epidemic threshold compared to the thresholds used in the United States and Europe.
Methods: Korea, the United States, and Europe use different methods to set seasonal influenza epidemic thresholds. First, we calculated the influenza epidemic thresholds for influenza seasons using the different methods of those three regions. Using these epidemic thresholds, we then compared the duration of influenza epidemics for the most recent three influenza seasons.
Results: The epidemic thresholds estimated by the Korean method were lower than those by the other methods, and the epidemic periods defined using the Korean threshold were estimated to be longer than those defined by the other regions' thresholds.
Conclusion: A low influenza epidemic threshold may have contributed to the prolonged influenza epidemic in Korea, which was declared in 2022 and has continued until late 2023. A more reliable epidemic threshold for seasonal influenza surveillance needs to be established in Korea.

Introduction

Background

Influenza is a communicable disease primarily caused by influenza A or B viruses. It is a common acute respiratory illness that tends to spread during the winter season in Korea. Transmission occurs through respiratory droplets emitted from infected subjects. The basic reproduction number, which is defined as the average number of secondary cases per case in a totally susceptible population, has been observed to range from 1.27 to 1.8 in the four pandemics since the 20th century and during seasonal influenza epidemics [1]. Influenza poses a high risk of complications and can result in serious clinical outcomes in vulnerable populations, such as those aged 65 and above, children, and people with chronic diseases. It also increases absenteeism

rates at workplaces and schools, and rapidly increases the demand for medical care, significantly impacting society from a public health perspective.

Influenza surveillance involves collecting data on influenza transmission trends and circulating virus types to predict the timing and intensity of epidemics. Surveillance helps maintain an appropriate level of preparedness, with the goal of minimizing the socioeconomic impact during the influenza season. In Korea, surveillance measures include monitoring outpatient illness, hospitalizations, and virological factors to determine the onset and end of epidemics, analyze their progression, and manage seasonal influenza based on pathogen characteristics. Outpatient illness surveillance is carried out at approximately 200 sentinel sites nationwide as of October 2023, with 87 of these sites also participating in virological surveillance. Additionally, influenza hospitalizations and deaths are surveilled at secondary and tertiary hospitals to monitor the severity of seasonal influenza [2].

In Korea and the United States (U.S.), influenza surveillance among outpatients collects information on influenza-like illness (ILI). The proportion of visits with ILI among all outpatient visits is estimated and used as a monitoring tool for influenza surveillance [2,3]. While the case definition of ILI varies across countries and agencies, it generally consists of fever and respiratory infection symptoms. Since ILI is a common clinical presentation of various respiratory infections, it is used as a surrogate indicator for tracking influenza epidemics even though it does not accurately estimate the incidence of influenza [4,5]. The positivity rate of respiratory specimens for influenza viruses is also used as a monitoring indicator, with the European Centre for Disease Prevention and Control (ECDC) defining the start and end of influenza epidemics based on a 10% positivity rate in respiratory samples collected from sentinel sites [6].

In Korea, unlike the U.S. and Europe, an influenza alert that was issued in the 37th week of 2022 has been in effect for 69 consecutive weeks as of December 2023 [7]. This is the first occurrence of such a phenomenon since the establishment of the influenza surveillance system in Korea in 2000. It also deviates from the well-known epidemiological characteristic that influenza typically spreads in the winter season in temperate regions [8–10]. Therefore, a rigorous assessment is necessary to determine whether this phenomenon can be interpreted as reflecting an actual prolonged increase in influenza activity.

Objectives

This study aimed to compare the duration of seasonal influenza epidemics calculated using different influenza epidemic thresholds used in Korea, the U.S., and Europe. The implications of the study findings will be discussed in the context of the ongoing influenza epidemic in Korea that has lasted since 2022.

Methods

Ethics statement

This study was based on public data. Neither approval by the institutional review board nor obtainment of informed consent was required.

Study design

This was a comparative study.

Setting

The epidemic periods for the three most recent seasons (2018–2019, 2019–2020, 2022–2023) were analyzed, excluding the two seasons (2020–2021, 2021–2022) that did not experience influenza epidemics. This analysis applied the epidemic thresholds established by Korea, the U.S., and Europe.

Data sources and measurement

National influenza surveillance data in Korea were collected from the weekly reports published by the Korea Disease Control and Prevention Agency (KDCA) [11]. Weekly data on ILI rates, the number of respiratory specimens, and the number of influenza-positive specimens from the 36th week of 2015 to the 35th week of 2023 were used. In Korea, ILI is defined as the presence of fever of 38°C or higher and respiratory infection symptoms, such as cough and sore throat. Accordingly, this study applied the Korean ILI definition to data to estimate the epidemic threshold for Korea using the Korean and the U.S. methods [2]. Data from the 2020–2021 and 2021–2022 seasons were excluded from the analysis, as there was no influenza outbreak during these seasons. The influenza season was defined as the period from the 36th week of each year to the 35th week of the following year. Each week was defined as starting on Sunday and ending on Saturday. For a week that spanned two years, the week number was assigned based on the year in which the Sunday fell.

The influenza epidemic thresholds from the KDCA, the U.S. Centers for Disease Control and Prevention (CDC), and the ECDC were chosen for a comparative analysis. Korea and the U.S. use the sum of the mean and two standard deviations of weekly ILI rates from non-influenza time periods as the epidemic threshold. In Korea, data from the past 3 seasons are used for the calculation, whereas data from the past 2 seasons are used in the U.S. [2,3]. Europe, in contrast, uses a 10% influenza positivity rate in respiratory specimens as the epidemic threshold (Table 1) [6]. For the Korean and U.S. methods, the earlier of two consecutive weeks when the weekly ILI rate exceeds the epidemic threshold is defined as the start of the influenza epidemic, and the week prior to two consecutive weeks when the weekly ILI rate does not reach the epidemic threshold is defined as the end of the epidemic. In the European method, the epidemic period is determined based on the 10% influenza positivity rate.

Bias

There was no bias in selecting target data.

Table 1. Epidemic thresholds for seasonal influenza

Country/region	Epidemic threshold
Korea (KDCA)	- [Mean rate of ILI (per 1,000) during non-influenza weeks for the most recent three seasons+(2×SDs)] - A non-influenza time period is defined as two or more consecutive weeks in which influenza positivity among respiratory laboratory samples is lower than 2%
United States (CDC)	- [Mean percentage of ILI during non-influenza weeks for the most recent two seasons+(2×SDs)] - A non-influenza time period is defined as two or more consecutive weeks in which each week accounted for less than 2% of the season's total number of specimens that tested positive for influenza
Europe (ECDC)	- 10% influenza positivity among respiratory laboratory samples

KDCA, Korea Disease Control and Prevention Agency; CDC, Centers for Disease Control and Prevention; ECDC, European Centre for Disease Prevention and Control; ILI, influenza-like illness.

Study size

There was no need to calculate the study size because known data were used.

Statistical methods

Descriptive statistics were used.

Results

Epidemic trends in Korea according to the Korean influenza epidemic threshold

In total, 312 weeks of data from six influenza seasons were collected from the KDCA's weekly reports. Despite using the same influenza surveillance data, differences were observed in the epidemic thresholds and the durations of epidemics depending on the applied method. In the 2018–2019 season, two influenza epidemic peaks were observed (Fig. 1, Dataset 1), and in the 2019–2020 season, the epidemic ended early due to the measures taken in response to the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with no spring outbreak observed (Fig. 2, Dataset 2). In the 2022–2023 season, elevated ILI rates were observed in general (Fig. 3, Dataset 3).

Non-influenza time periods

To calculate the epidemic threshold, the Korean method used ILI rates from non-influenza periods in 156 weeks of the past 3 matching seasons, while the U.S. method used ILI rates from non-influenza periods in 104 weeks of the past 2 matching seasons. For the Korean method, the non-influenza period for the epidemic threshold calculations for the 2018–2019, 2019–2020, and 2022–2023 seasons comprised an average of 70.7 weeks (45.3%), while the non-influenza period for the same seasons using the U.S. method was an average of 74.7 weeks (71.8%). Since the non-influenza period defined by the Korean method was shorter, lower ILI rates were used in the epidemic threshold calculation than with the U.S. method. Accordingly, the mean and SD

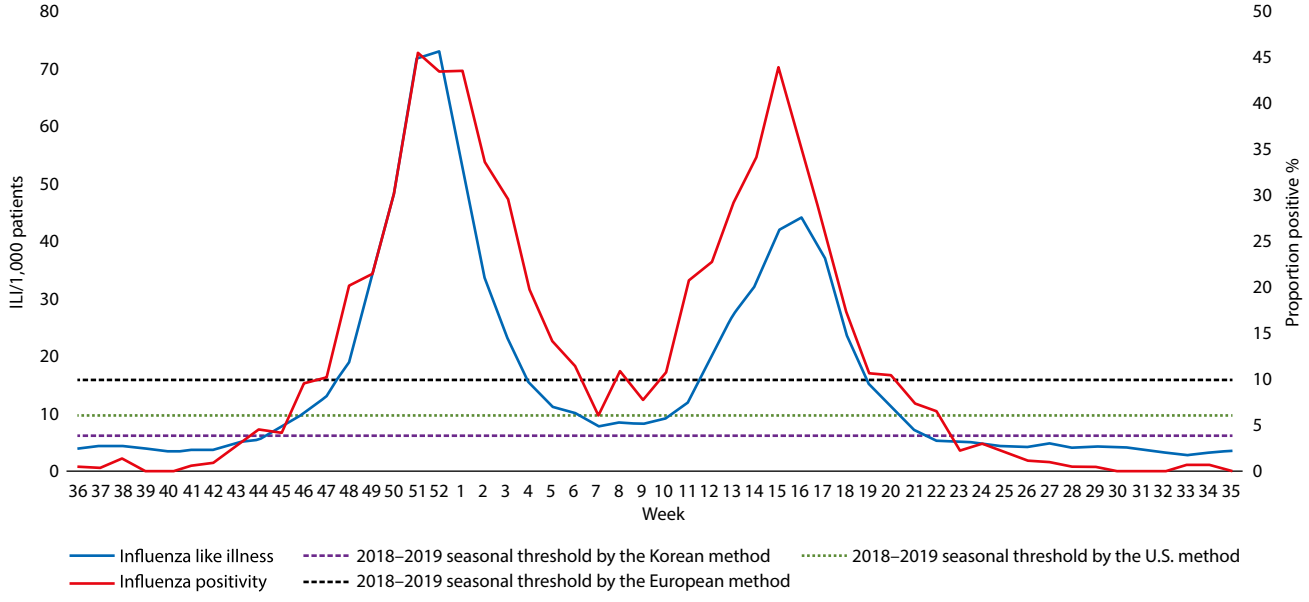


Fig. 1. 2018–2019 season influenza activity. ILI, influenza-like illness.

https://doi.org/10.12771/emj.2024.e24

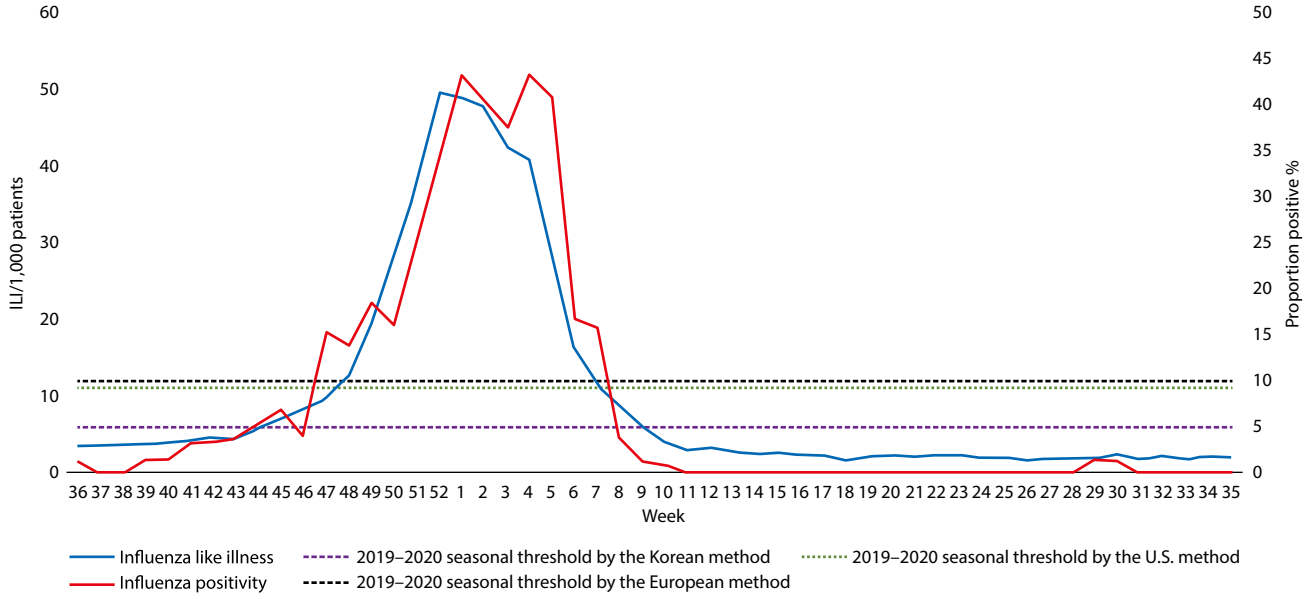


Fig. 2. 2019–2020 season influenza activity. ILI, influenza-like illness.

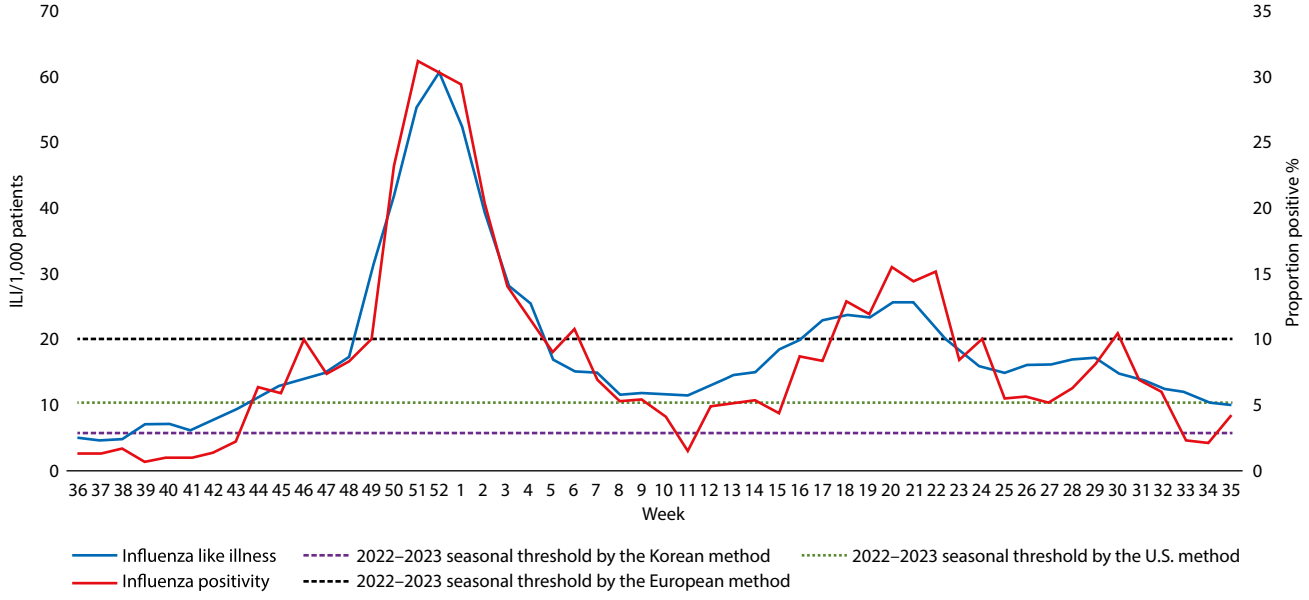


Fig. 3. 2022–2023 season influenza activity. ILI, influenza-like illness.

values derived from the ILI rates for the non-influenza time periods used in the Korean method were lower than those for the U.S. method (Table 2).

Influenza epidemic thresholds

The influenza epidemic thresholds estimated using the Korean method were 6.3 ILI cases per 1,000 patients (2018–2019 season), 6.0 ILI cases per 1,000 patients (2019–2020 season), 5.8 ILI cases per 1,000 patients (2022–2023 season), and 6.3 ILI cases per 1,000 patients (2023–2024 season). The thresholds calculated using the U.S. method were 9.7 ILI cases per 1,000 patients

https://doi.org/10.12771/emj.2024.e24

Table 2. Characteristics of the influenza seasons used for epidemic threshold calculation

Influenza season	Korean method				U.S. method			
	Seasons used for the calculation (length*)	Length* of non-influenza period	Mean rate of ILI during non-influenza weeks	SD of ILI during non-influenza weeks	Seasons used for the calculation (length*)	Length* of non-influenza period	Mean rate of ILI during non-influenza weeks	SD of ILI during non-influenza weeks
2018–2019	2015–2016, 2016–2017, 2017–2018 (156)	75	4.63	0.85	2016–2017, 2017–2018 (104)	76	5.70	2.01
2019–2020	2016–2017, 2017–2018, 2018–2019 (156)	64	4.38	0.79	2017–2018, 2018–2019 (104)	74	5.87	2.62
2022–2023	2017–2018, 2018–2019, 2019–2020 (156)	73	3.51	1.15	2018–2019, 2019–2020 (104)	74	4.64	2.91

ILI, influenza-like illness.
*Unit: weeks.

(2018–2019 season), 11.1 ILI cases per 1,000 patients (2019–2020 season), 10.5 ILI cases per 1,000 patients (2022–2023 season), and 18.7 ILI cases per 1,000 patients (2023–2024 season). The Korean method yielded lower epidemic thresholds than the U.S. method in all seasons. The largest difference was observed for the 2023–2024 season, which reflected the 2022–2023 season data during the COVID-19 pandemic (Table 3).

The epidemic period estimated using the thresholds of Korea, U.S., and Europe

The estimated duration of the epidemic period using the Korean threshold was longer than those calculated with the U.S. and European thresholds in all seasons. The durations based on the Korean thresholds were 29 weeks (2018–2019 season), 17 weeks (2019–2020 season), and 49 weeks (2022–2023 season). In contrast, the durations based on the U.S. thresholds were 23 weeks (2018–2019 season), 12 weeks (2019–2020 season), and 43 weeks (2022–2023 season). Durations calculated with the European thresholds were the shortest, at 24 weeks (2018–2019 season), 13 weeks (2019–2020 season), and 16 weeks (2022–2023 season; Table 4). The mean epidemic duration for three seasons based on the Korean thresholds was 5.7 weeks (22%) longer per season compared to that calculated using the U.S. method, and 14 weeks (79%) longer compared to that calculated using the European method. The difference was particularly prominent in the 2022–2023 season, where the duration based on the Korean thresholds was 33 weeks longer than that based on the European threshold.

Discussion

Key results

This study compared the influenza epidemic periods in Korea during the 2018–2019, 2019–2020, and 2022–2023 seasons, utilizing epidemic thresholds from Korea, the U.S., and Europe. The influenza epidemic threshold calculated with the Korean method was approximately 49% lower than that calculated using the U.S. method. Additionally, the epidemic durations based on the Korean thresholds were longer than those based on the U.S. and European thresholds.

Table 3. Comparison of influenza epidemic thresholds according to different methods

Season	Threshold according to the Korean method*	Threshold according to the U.S. method*	European threshold
2018–2019	6.3	9.7	10% influenza positivity
2019–2020	6.0	11.1	
2022–2023	5.8	10.5	
2023–2024	6.3	18.7	

ILI, influenza-like illness.
*Unit: ILI cases per 1,000 patients.

Table 4. Comparison of epidemic period durations according to different thresholds

Season	Duration according to the Korean threshold (weeks)	Duration according to the U.S. threshold (weeks)	Duration according to the European threshold (weeks)
2018–2019	29	23	24
2019–2020	17	12	13
2022–2023	49	43	16

Interpretation

The differences in epidemic thresholds between the Korean and the U.S. methods primarily stemmed from the different definitions of the non-influenza time period in both countries. In Korea, this period is defined as when the weekly influenza positivity rate falls below 2%, a figure that is influenced by the activity levels of other respiratory viruses. In contrast, the U.S. defines the non-influenza period as the time when the number of weekly influenza-positive specimens is less than 2% of the season's total count of influenza-positive specimens, a measure that remains unaffected by the activity of other respiratory viruses. It is important to note that the 2% influenza positivity rate used in Korea's definition of the non-influenza period is significantly lower than the European epidemic threshold of 10%. This discrepancy results in the use of lower ILI rates from periods of relatively low influenza activity to calculate the Korean epidemic threshold. Consequently, this could lead to the earlier issuance and later lifting of influenza epidemic alerts.

The largest difference between the epidemic durations determined by ILI rate-oriented thresholds and those determined by influenza positivity rate-oriented thresholds was observed in the 2022–2023 season. The epidemic duration based on the Korean method was 49 weeks, which was roughly 3 times longer than the 16-week epidemic duration based on the European threshold. This discrepancy can be partially attributed to changes in the incidence levels of overall respiratory infectious diseases. During the COVID-19 pandemic, the implementation of stringent social distancing measures and enhanced personal hygiene practices led to a reduced incidence of respiratory infections. However, this low incidence began to increase as public health measures were relaxed in the latter half of 2022. Subsequently, a resurgence of various respiratory infections elevated the overall ILI rates, which likely contributed to the extended epidemic period by interacting with the lower epidemic threshold in the 2022–2023 season [12,13].

According to project reports from the Korea Respiratory Virus Integrated Surveillance System, which tests respiratory samples from respiratory infection patients, including cases of ILI, at sentinel sites, the influenza virus detection rates in 2022 and 2023 were 5.5% and 16.1%, respectively. These rates are similar to or lower than the rates of 17% and 14% observed in 2018 and 2019, before the COVID-19 pandemic. In contrast, the detection rates of SARS-CoV-2 in

2022 and 2023 were 9.4% and 9.8%. The detection rates of other respiratory viruses have also increased since 2021 (Table 5) [11,14,15]. This finding is consistent with a report that the first post-pandemic influenza epidemic was not considered unexpected in terms of extent and severity in most countries [16]. Therefore, the increase in ILI rates in the 2022–2023 season is likely due to the increased activity of SARS-CoV-2 and other respiratory viruses, rather than a significant increase in influenza activity.

ILI is a widely used medical concept and a reliable indicator in influenza surveillance [17]. However, since the clinical features of various respiratory infections often overlap, a rise in influenza activity alone may not fully explain the increase in ILI rates. When analyzing a rise in ILI rates, it is crucial to consider all circulating respiratory pathogens collectively, in conjunction with laboratory test results, to determine the extent to which each pathogen contributes to the increase [18]. The ongoing influenza epidemic in Korea since the 37th week of 2022 appears to be a phenomenon resulting from the combination of the lower influenza epidemic threshold in Korea and the overall increase in ILI rates due to the increased activity of SARS-CoV-2 and other respiratory viruses.

Limitations

The study did not explore the characteristics of each epidemic threshold, such as the sensitivity and specificity of its application. Merely comparing the durations of epidemic periods does not determine which method should be recommended for a specific country. The approach to setting the influenza epidemic threshold should be assessed differently, taking into account each country's health system capacity and disease burden. In this study, the influenza epidemic thresholds used in the U.S. and Europe were only compared with those in Korea. Additional comparisons with other countries may increase the generalizability of the study findings.

Conclusion

Influenza surveillance systems are designed to minimize the socioeconomic losses caused by influenza epidemics, making the application of an appropriate epidemic threshold crucial. This study noted variations in the duration of the epidemic period based on the threshold applied. A low influenza epidemic threshold may have contributed to the prolonged epidemic period, which was declared in the 37th week of 2022 and continued until the end of 2023 in Korea. The optimal epidemic threshold should be examined from various perspectives, including the

Table 5. Virological surveillance results of respiratory specimens

Year	Total respiratory virus positivity [% (n*)]	Influenza virus positivity [% (n*)]	SARS-CoV-2 virus positivity [% (n*)]	Other respiratory virus† positivity [% (n*)]
2023	81.4 (Na)	16.1 (Na)	9.8 (Na)	55.5 (Na)
2022	72.7 (Na)	5.5 (491)	9.4 (Na)	57.8 (5,205)
2021	65.1 (3,009)	0 (0)	NA	65.1 (3,009)
2020	48.6 (2,830)	12 (701)	NA	36.6 (2,129)
2019	60.2 (7,311)	14 (1,702)	NA	46.2 (5,609)
2018	63.0 (Na)	17 (Na)	NA	46.0 (Na)

SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; Na, not available; NA, not applicable.
*n: number of positive samples.
†Other respiratory viruses: respiratory syncytial virus, parainfluenza virus, bocavirus, adenovirus, human coronavirus, metapneumovirus, rhinovirus.

evolving epidemiological characteristics of respiratory infectious diseases since the emergence of SARS-CoV-2. Adopting multiple indicators could enable the issuance of more reliable flu alerts and the implementation of more effective countermeasures.

ORCID
Joowon Lee: https://orcid.org/0009-0005-7015-8238
Sooyoung Huh: https://orcid.org/0009-0009-1113-1659
Haesook Seo: https://orcid.org/0000-0002-3554-6172

Authors' contributions
Project administration: Lee J
Conceptualization: Lee J
Methodology & data curation: Lee J
Funding acquisition: not applicable
Writing – original draft: Lee J
Writing – review & editing: Lee J, Huh S, Seo H

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Funding
Not applicable.

Data availability
The data used in this study are retrieved from the infectious diseases portal of the KDCA available from: https://dportal.kdca.go.kr/pot/is/st/influ.do
Data files are available from Harvard Dataverse: https://doi.org/10.7910/DVN/9ZYE8N

Dataset 1. Research data used to draw Fig. 1, which shows seasonal influenza activity during the 2018–2019 season in Korea according to different influenza epidemic thresholds in Korea, the United States, and Europe
Dataset 2. Research data used to draw Fig. 2, which shows seasonal influenza activity during the 2019–2020 season in Korea according to different influenza epidemic thresholds in Korea, the United States, and Europe
Dataset 3. Research data used to draw Fig. 3, which shows seasonal influenza activity during the 2022–2023 season in Korea according to different influenza epidemic thresholds in Korea, the United States, and Europe

Acknowledgments
Not applicable.

Supplementary materials
Not applicable.

References

1. Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC Infect Dis* 2014;14:480. https://doi.org/10.1186/1471-2334-14-480
2. Korea Disease Control and Prevention Agency. Management strategies for seasonal influenza, 2023-24 season. Cheongju: Korea Disease Control and Prevention Agency; 2023.
3. Centers for Disease Control and Prevention [CDC]. U.S. influenza surveillance: purpose and methods [Internet]. Atlanta (GA): CDC; c2024 [cited 2024 Jan 5]. Available from: https://www.cdc.gov/flu/weekly/overview.htm
4. Brammer L, Budd A, Cox N. Seasonal and pandemic influenza surveillance considerations for constructing multicomponent systems. *Influenza Other Respir Viruses* 2009;3(2):51-58. https://doi.org/10.1111/j.1750-2659.2009.00077.x
5. Spencer JA, Shutt DP, Moser SK, Clegg H, Wearing HJ, Mukundan H, et al. Distinguishing viruses responsible for influenza-like illness. *J Theor Biol* 2022;545:111145. https://doi.org/10.1016/j.jtbi.2022.111145
6. European Centre for Disease Prevention and Control [ECDC]. Seasonal influenza 2022-2023: annual epidemiological report for 2023. Stockholm: ECDC; 2023.
7. Korea Disease Control and Prevention Agency. Announcement of 2023-2024 influenza season initiation [Internet]. Cheongju (KR): Korea Disease Control and Prevention Agency; c2023 [cited 2024 Jan 5]. Available from: https://www.kdca.go.kr/board/board.es?mid=a20501010000&bid=0015&list_no=723470&cg_code=&act=view&nPage=1

8. Tamerius JD, Shaman J, Alonso WJ, Bloom-Feshbach K, Uejio CK, Comrie A, et al. Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLOS Pathog* 2013;9(3):e1003194. https://doi.org/10.1371/journal.ppat.1003194
9. Korea Disease Control and Prevention Agency. 2022 Annual report on infectious disease reports [Internet]. Cheongju (KR): Korea Disease Control and Prevention Agency; c2022 [cited 2024 Feb 16]. Available from: https://dportal.kdca.go.kr/pot/bbs/BD_selectBbs.do?q_bbsSn=1010&q_bbsDocNo=20230908669355443&q_clsfNo=1
10. Lee JW. Another flu season declaration... The first season continued for more than a year [Internet]. Seoul (KR): The Dong-a Ilbo; c2023 [cited 2023 Sep 15]. Available from: https://www.donga.com/news/Health/article/all/20230915/121190131/1
11. Korea Disease Control and Prevention Agency. Infectious diseases, sentinel surveillance report [Internet]. Cheongju (KR): Korea Disease Control and Prevention Agency; c2024 [cited 2024 Jan 5]. Available from: https://dportal.kdca.go.kr/pot/bbs/BD_selectBbsList.do?q_bbsSn=1010&q_bbsDocNo=&q_clsfNo=2&q_searchKeyTy=&q_searchVal=&q_currPage=1&q_sortName=&q_sortOrder=
12. Cha J, Seo Y, Kang S, Kim I, Gwack J. Sentinel surveillance results for influenza and acute respiratory infections during the coronavirus disease 2019 pandemic. *Public Health Wkly Rep* 2023;16(20):597-612. https://doi.org/10.56786/PHWR.2023.16.20.1
13. Kim IH, Kang SK, Cha JO, Seo YJ, Gwack J, Lee NJ, et al. Changes in patterns of respiratory virus since the coronavirus disease 2019 pandemic (until April 2023). *Public Health Wkly Rep* 2023;16(20):621-631. https://doi.org/10.56786/PHWR.2023.16.20.3
14. Lee NJ, Woo S, Lee J, Rhee JE, Kim EJ. 2021-2022 Influenza and respiratory viruses laboratory surveillance report in the Republic of Korea. *Public Health Wkly Rep* 2023;16(3):53-65. https://doi.org/10.56786/PHWR.2023.16.3.1
15. Woo SH, Lee NJ, Lee JH, Rhee JE, Kim EJ. Korea 2022-2023 influenza and respiratory viruses laboratory surveillance report. *Public Health Wkly Rep* 2024;17(2):455-469. https://doi.org/10.56786/PHWR.2024.17.12.1
16. de Jong SPJ, Felix Garza ZC, Gibson JC, van Leeuwen S, de Vries RP, Boons GJ, et al. Determinants of epidemic size and the impacts of lulls in seasonal influenza virus circulation. *Nat Commun* 2024;15(1):591. https://doi.org/10.1038/s41467-023-44668-z
17. Fleming DM, Elliot AJ. Lessons from 40 years' surveillance of influenza in England and Wales. *Epidemiol Infect* 2008;136(7):866-875. https://doi.org/10.1017/S0950268807009910
18. Kelly H, Birch C. The causes and diagnosis of influenza-like illness. *Aust Fam Physician* 2004;33(5):305-309.

Original Article

Ewha Med J 2024;47(2):e25
https://doi.org/10.12771/emj.2024.e25
eISSN 2234-2591

The use of the bicipital groove as an intraoperative landmark for proximal humeral rotation during fracture fixation

Hyun-Joo Lee[1], Sanghyun Joung[2], Erica Kholinne[3], Suk-Joong Lee[4], Jong Pil Yoon[1], Jun Tan[5], In-Ho Jeon[6]

[1]Department of Orthopaedic Surgery, School of Medicine, Kyungpook National University, Kyungpook National University Hospital, Daegu, Korea
[2]AIRS, Daegu, Korea
[3]Department of Orthopedic Surgery, St. Carolus Hospital, Faculty of Medicine, Universitas Trisakti, Jakarta, Indonesia
[4]Department of Orthopaedic Surgery, Gyeongsang National University Changwon Hospital, Changwon, Korea
[5]Department of Hand Surgery, Affiliated Hospital of Nantong University, Nantong, China
[6]Department of Orthopaedic Surgery, Asan Medical Center, School of Medicine, University of Ulsan, Seoul, Korea

Received Jan 1, 2024
Revised Jan 31, 2024
Accepted Feb 5, 2024

Corresponding author
Hyun-Joo Lee
Department of Orthopaedic Surgery, School of Medicine, Kyungpook National University, Kyungpook National University Hospital, 130 Donguk-ro, Jung-gu, Daegu 41944 , Korea
E-mail: lidmania@daum.net

Keywords
Humerus; Anatomic models; Anatomic landmarks; Rotation

Objectives: This study aimed to quantify the relationship between proximal humeral rotation and the lateral border of the bicipital groove on fluoroscopic imaging.
Methods: A composite normal humerus with a marker placed on the lateral border of the bicipital groove was affixed to a custom rotation device at the proximal cut segment. Consecutive fluoroscopic images were captured from −60° to 60° in 5° increments and from −15° to 15° in 1° increments. The index value was calculated by taking the ratio of the distance from the medial boundary of the proximal humerus to the lateral border of the bicipital groove to the distance between the medial and lateral boundaries of the proximal humerus. The correlation between the humeral rotation and the index value was determined.
Results: The index value showed a strong positive linear correlation position during internal rotation of the humerus across the entire range (r=0.998, P<0.001), as well as when the humerus was externally rotated, ranging from 15° of internal rotation to 15° of external rotation (r=0.991, P<0.001).
Conclusion: The lateral border of the bicipital groove may serve as a useful intraoperative landmark for assessing proximal humeral rotation. This could potentially enhance the outcomes of humeral fracture repair and upper arm arthroplasty.

Introduction

Background

Restoration of the original anatomy is a crucial aspect of fracture treatment. During surgery, coronal and sagittal angulations in long bones can be aligned using two-dimensional (2D) fluoroscopic imaging. However, assessing rotational deformity with conventional 2D fluoroscopic images is subjective. While the extent of fragment rotation can be estimated by comparing the affected side to the contralateral normal side, this method necessitates images of the contralateral side and relies entirely on its condition. Additionally, bilateral bone morphology is not symmetrical. In cases involving both sides, comparison is not possible, making quantitative estimation of rotation extent unattainable. The intraoperative three-dimensional (3D) estimation

of bone position using 2D images has garnered interest in orthopedic surgery, including computer-assisted techniques. Rotation assessment in the lower extremity, such as the femur or tibia, is typically performed using specific landmarks. For instance, the size of the lesser trochanter is used as an indicator of the femur's internal rotation status [1,2]. In contrast, studies on humeral rotation are less common than those for the lower extremity. To date, only a handful of studies have investigated the measurement of humeral rotation without the use of landmarks [3–5]. Therefore, we developed a method to estimate and evaluate the rotational alignment of the proximal humerus using a specific landmark.

Ojbectives

The aim of the current study was to quantify the relationship between proximal humeral rotation and the lateral border of the bicipital groove as seen on fluoroscopic imaging. We hypothesized that the lateral border of the bicipital groove could act as a practical landmark for assessing humeral rotation.

Methods

Ethics statement

It is not a human population study; therefore, neither approval by the institutional review board nor obtainment of informed consent was required.

Experimental setup and acquisition of fluoroscopic images

A composite sawbone model of a humerus (#3404, Sawbones, Vashon Island, WA, USA) was sectioned at the midpoint of the shaft. Prior to sectioning, a longitudinal line was drawn on the anterior surface to ensure that the proximal half retained half of this line. The proximal segment of the humeral model was then secured to a custom rotation device, aligning the longitudinal line with the 0° rotation mark on the device. Consequently, a 0° rotation on the device corresponded to a neutral alignment of the proximal segment relative to the distal segment of the humerus. For precise control and high accuracy, we employed a modular actuator with 0.1-mm precision (Dynamixel Pro, ROBOTIS, Seoul, Korea) as the custom rotation device. A metal dot was affixed to the lateral edge of the bicipital groove at the point corresponding to the largest diameter of the humeral head in preparation for fluoroscopic imaging. Since the humeral head is spherical, any point on its surface can serve as a rotational reference through geometric calculation. We chose the lateral edge of the bicipital groove as this reference due to its relative ease of identification on imaging. The location for the metal dot was specifically chosen because the maximum circumference of the hypothetical sphere would exhibit the greatest change with each degree of rotation, thus providing the highest sensitivity to rotational changes. The assembly was positioned on a radiolucent table beneath an image intensifier for imaging purposes. The rotation device was set up to maintain the distal segment stationary while the proximal segment was rotated from −60° to 60° in 5° increments and from −15° to 15° in 1° increments, achieving an accuracy of 0.1°. A fluoroscopic image was captured at each incremental position of rotation (Fig. 1).

Data analysis

We used a specialized program to calculate an index value indicative of humeral rotation. The user identified a rectangle and three points on fluoroscopic images, as depicted in Fig. 2. The rectangle was placed over the diaphysis to establish the humerus's long axis, which was

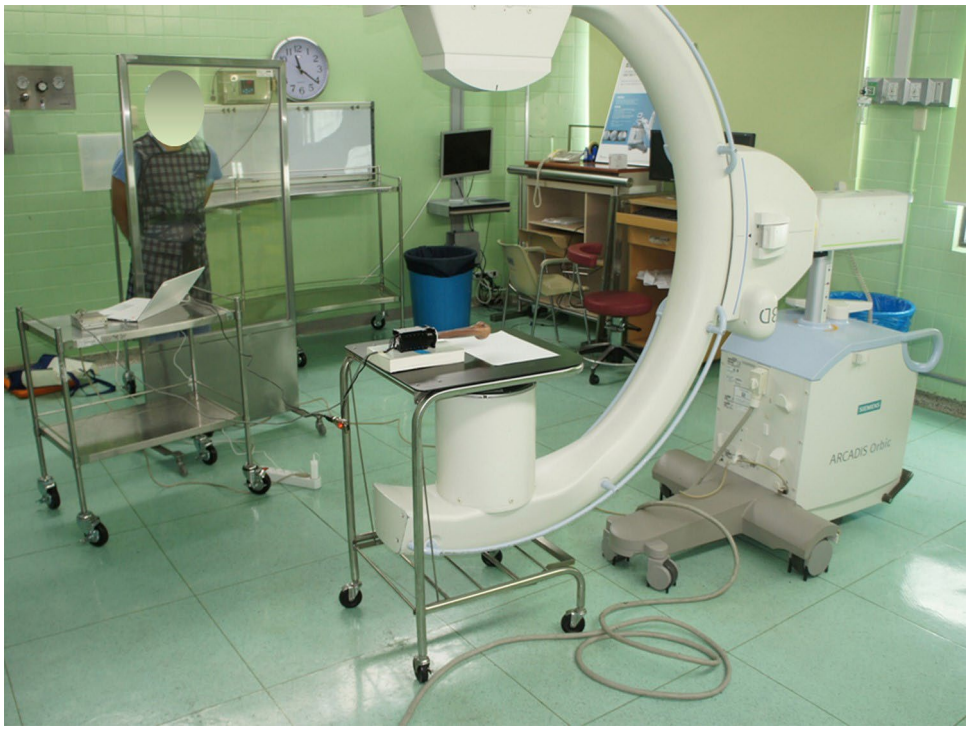


Fig. 1. Experimental setting. The motor–humerus complex was positioned on a radiolucent table under a C-arm. Using the rotation device, the proximal part was rotated, while the cut distal part was fixed. A fluoroscopic image was taken at each consecutive rotational position.

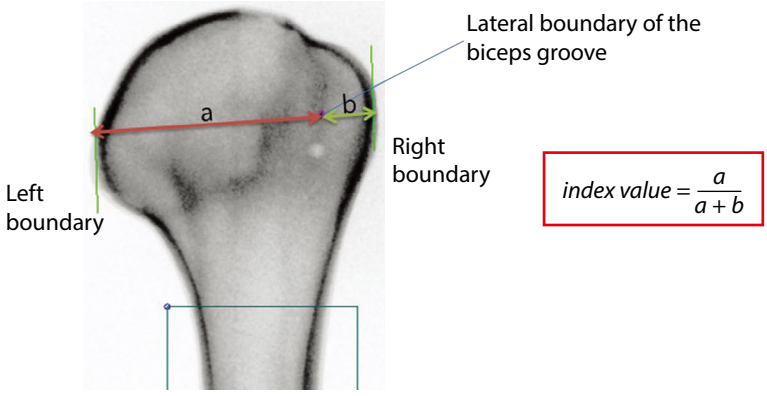


Fig. 2. A custom program. After making a block, two lines (green lines) are automatically generated. Lines from the medial and lateral boundary of the humeral head to the lateral border of the bicipital groove can be drawn perpendicularly to the green lines. The program calculated the index value.

determined from the selected area through principal component analysis. The medial and lateral boundaries of the humerus were marked, ensuring that both were tangential to its long axis. Additionally, a point was marked at the lateral edge of the bicipital groove. The value "a" was the distance between the medial boundary of the humeral head, "b" was the distance between the lateral boundary of the bicipital groove, and the value of "a+b" was the distance from the medial to the lateral boundary. The index value was the ratio of "a" to "a+b" in equation 1 (Fig. 1), with

https://doi.org/10.12771/emj.2024.e25

the assumption that the index value correlates with humeral rotation.

Statistical analysis

The Kolmogorov–Smirnov test was used to assess the normality of the distribution. The dataset of index values followed a normal distribution; therefore, Pearson's correlation coefficient was employed to examine the relationship between the index value and the angle of humeral rotation. A regression equation was also derived. The threshold for statistical significance was established at $P<0.05$. Both descriptive and analytical analyses were performed using SPSS version 15.0 (SPSS, Chicago, IL, USA).

Results

The index value showed a strong positive linear correlation with position during internal rotation of the humerus (correlation [IR]=0.998; $P<0.001$). Similarly, a moderate positive linear correlation was observed with position during external rotation of the humerus (correlation [ER]=0.693; $P<0.001$). Notably, within the range of 15° internal rotation to 15° external rotation, the correlation remained strongly positive (correlation [IR15–ER15]=0.991; $P<0.001$; Fig. 3).

The regression equations for internal and external rotation were as follows:

$$\text{Index value} = 0.00926 \times (\text{angle}) + 0.83604 \text{ (internal rotation)} \quad (1)$$

$$\text{Index value} = 0.00217 \times (\text{angle}) + 0.86511 \text{ (external rotation)} \quad (2)$$

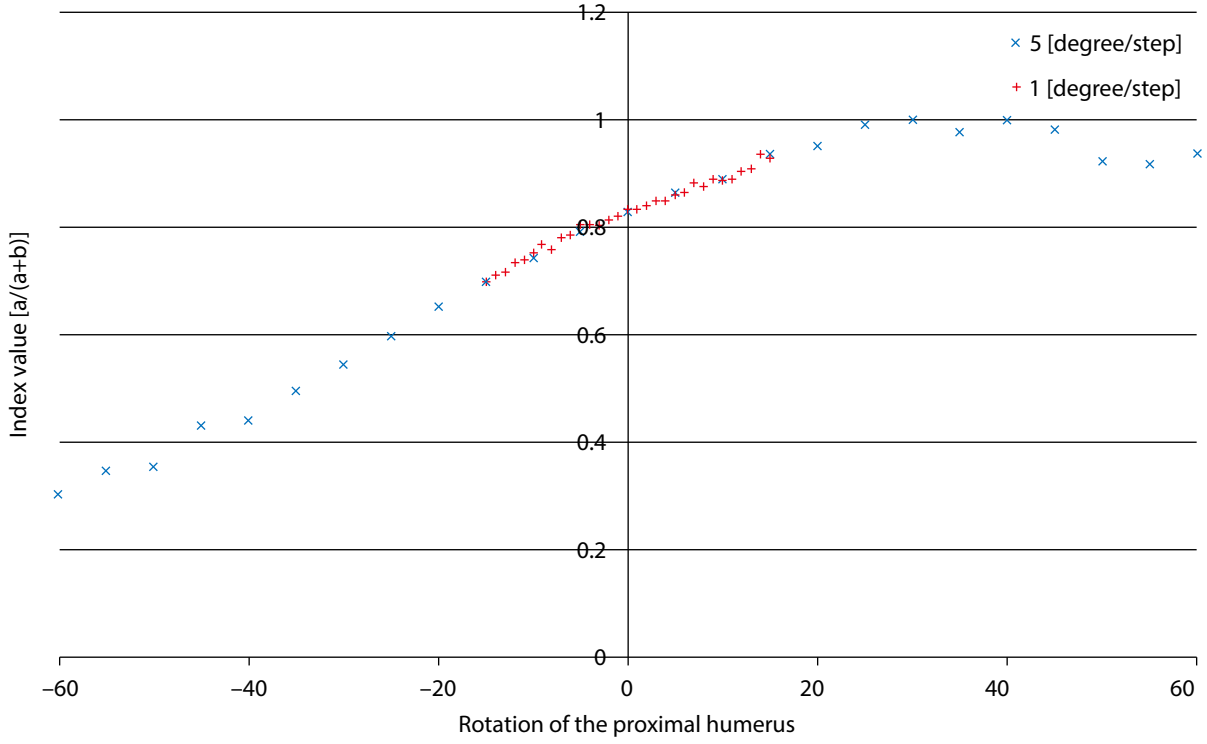


Fig. 3. Linear correlation between the index value and humeral rotation. A strongly positive correlation was observed within the range from 15° of internal rotation to 15° of external rotation.

https://doi.org/10.12771/emj.2024.e25

The regression equation for internal and external rotation between −15° and 15° was as follows:

Index value = 0.00727 × (angle) + 0.82225

Discussion

Key results

We found that the index value of the lateral border of the bicipital groove exhibits a moderate-to-strong correlation with the rotational angle of the humerus as seen on fluoroscopic imaging. These findings could prove beneficial for minimally invasive plate osteosynthesis (MIPO), a technique that has recently become more popular.

Interpretation/comparison with previous studies

MIPO offers the advantage of preserving periosteal blood supply; however, it is often associated with rotational malalignment due to the lack of direct visualization of the fracture site [6]. Accordingly, we established a linear correlation between the index landmark and the rotation angle. The clinical significance of our study is that it provides a method for estimating the rotation angle of the proximal humerus. When the distal humerus is positioned neutrally on fluoroscopy, variations in the lateral border of the bicipital groove can indicate the degree of rotation relative to the distal part. This allows for the assessment of rotation without the need for repeated fluoroscopic examination of the distal part.

The acceptable limit for rotational malalignment is generally considered to be 20 degrees. The degree of malrotation is directly related to a reduction in the range of motion [7]. While anatomical rotational alignment is possible using open reduction and internal fixation, achieving correct humeral alignment during MIPO surgery can be more challenging. In acute cases, palpating the epicondyles may be difficult due to traumatic edema or in patients with obesity. Utilizing an index value for measurement provides a quantitative assessment of the reduction and alignment of the fractured fragments. Variations in humeral anatomy among different patients necessitate this approach. By measuring the index value, surgeons can customize the treatment to accommodate individual differences in bone structure and alignment. This is critical for attaining optimal anatomical alignment, which is a key factor in ensuring functional recovery.

Malrotation in the humerus is generally considered more acceptable than in the lower extremities, which has led to limited research on humeral rotation. Consequently, precise criteria or landmarks for assessment using plain radiographs have yet to be established. However, studies by Itoi et al. [8] and Sabo et al. [9] have reported that humeral malrotation leads to malunion, whereas Li et al. [4] found that it had a negative effect on shoulder function. Moreover, recent advances in shoulder and elbow arthroplasty have demonstrated that the sequelae of humeral malrotation are caused by altered kinematics [8,9]. A study on humeral shaft fracture repair assessed rotation during surgery using the cortical step sign [10]. Boileau et al. [11] used the shape of the bicipital groove to assess rotation by comparing the ipsilateral and contralateral sides. However, neither method was able to provide quantitative measurements. CT is highly reliable and accurate for evaluating humeral rotation, but its feasibility during surgery is questionable. Tan et al. [3] used the cortical density of the lesser tuberosity as a landmark for humeral rotation and showed its validity in a cadaver study. In contrast, our study was able to measure rotation in 1° increments using a custom device, thus offering superior accuracy. We utilized the lateral border of the bicipital groove as a landmark because it provides a clear reference point when the lesser tuberosity is not visible,

particularly during ranges of internal rotation. A linear correlation was found between the position of this landmark and the degree of humeral rotation. This relationship may need adjustment if the proximal humerus obscures the medial line of the greater tuberosity due to the position of the lesser tuberosity. Nevertheless, within the clinically relevant range of humeral rotation for computer-assisted fracture surgery (internal rotation 15° to external rotation 15°), we observed a strong positive linear correlation with humeral rotation. The accurate estimation of humeral rotation using a landmark is crucial for both conventional MIPO and fracture surgery. The bicipital groove has also been suggested as a reliable intraoperative landmark for restoring humeral retrotorsion during shoulder replacement or for reconstructing the premorbid anatomy of the proximal humerus [12]. Our study confirmed the reliability of using the bicipital groove and found a linear correlation between the landmark and the humeral rotation.

Future trends in orthopedic surgery will rely on robot-assisted or computer-assisted techniques, which can reduce soft tissue damage and increase the accuracy of reduction by targeting the exact point for incision and manipulation [13]. In robot-assisted surgery, exact data points are needed, such as for computer-assisted arthroscopic subscapularis repair. The inability to visualize the subscapularis tendon footprint on arthroscopy is generally accepted. However, careful registration of the palpable lateral border of the bicipital groove allows the surface registration of an anatomical landmark of the proximal humerus. This improves accuracy when inserting an anchor to the lesser tuberosity. We aimed to define accurate landmarks rather than intuitively relying on comparison with the contralateral side. Because the anatomy of the humerus varies and a fragmented or distorted humerus anatomy may hinder the use of this landmark, other complementary 3D methods should be performed to determine the exact position of the proximal humerus. The lateral border of the bicipital groove can be used as a clinically important guide to evaluate humeral rotational alignment for fracture reduction or other computer-assisted surgical procedures, particularly in the range between −15° and 15°, where measurement errors often occur.

Limitations

Our estimation method has certain limitations. First, the lateral border of the bicipital groove may be obliterated in a comminuted fracture, severe osteoporosis, or the presence of implant-related materials. Improperly positioned shoulder images can also interfere with accurate imaging of the landmark. Thus, the placement of the metal dot may not align with what is observed in fluoroscopic images. However, the lateral border of the bicipital groove becomes more discernible when the humerus is internally rotated. The lesser tubercle can also serve as an intraoperative landmark for humeral rotation. Second, we cannot generalize the data to all patients because of variations in anatomy, such as in the degree of humeral anteversion or anatomical variation of bicipital groove. However, our study demonstrates the value of objective data for estimating humeral rotation, which could be used in a practical clinical setting. Therefore, future studies are needed to determine more generalized or normative data. Finally, estimating the rotation angle using a shoulder image alone assumes that the elbow joint is in its neutral position, which is relatively easy to achieve because of the wide posterior surface. Thus, we assumed that the effect of elbow positioning is minimal.

Conclusion

The lateral border of the bicipital groove can serve as an intraoperative landmark for the quantitative estimation of proximal humeral rotation. This landmark proves beneficial in minimally

invasive or robotic surgeries targeting the proximal humerus. Assessing humeral rotation during surgery can enhance the results of humeral fracture repairs and upper arm arthroplasty procedures.

ORCID
Hyun-Joo Lee: https://orcid.org/0000-0003-2837-3434
Sanghyun Joung: https://orcid.org/0000-0001-5958-3907
Erica Kholinne: https://orcid.org/0000-0002-4326-8205
Suk-Joong Lee: https://orcid.org/0000-0002-6769-6474
Jong Pil Yoon: https://orcid.org/0000-0001-6446-6254
Jun Tan: https://orcid.org/0000-0003-3426-8454
In-Ho Jeon: https://orcid.org/0000-0002-9289-9193

Authors' contributions
Project administration: Lee HJ
Conceptualization: Lee HJ, Joung S, Tan J, Jeon IH
Methodology & data curation: Lee HJ, Joung S, Tan J
Funding acquisition: Lee HJ
Writing – original draft: Lee HJ, Joung S
Writing – review & editing: Lee HJ, Joung S, Kholinne E, Lee SJ, Yoon JP, Tan J, Jeon IH

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Funding
This work was supported by a grant from the Biomedical Research Institute, Kyungpook National University Hospital (2015).

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Not applicable.

References

1. Zhang Q, Liu H, Chen W, Li X, Song Z, Pan J, et al. Radiologic measurement of lesser trochanter and its clinical significance in Chinese. *Skelet Radiol* 2009;38(12):1175-1181. https://doi.org/10.1007/s00256-009-0662-5
2. Kim JJ, Kim E, Kim KY. Predicting the rotationally neutral state of the femur by comparing the shape of the contralateral lesser trochanter. *Orthopedics* 2001;24(11):1069-1070. https://doi.org/10.3928/0147-7447-20011101-18
3. Tan J, Lee HJ, Aminata I, Chun JM, Kekatpure AL, Jeon IH. Radiographic landmark for humeral head rotation: a new radiographic landmark for humeral fracture fixation. *Injury* 2015;46(4):666-670. https://doi.org/10.1016/j.injury.2014.10.059
4. Li Y, Wang C, Wang M, Huang L, Huang Q. Postoperative malrotation of humeral shaft fracture after plating compared with intramedullary nailing. *J Shoulder Elbow Surg* 2011;20(6):947-954. https://doi.org/10.1016/j.jse.2010.12.016
5. Park SJ, Kim E, Jeong HJ, Lee J, Park S. Prediction of the rotational state of the humerus by comparing the contour of the contralateral bicipital groove: method for intraoperative evaluation. *Indian J Orthop* 2012;46(6):675-679. https://doi.org/10.4103/0019-5413.104210
6. Wang C, Li J, Li Y, Dai G, Wang M. Is minimally invasive plating osteosynthesis for humeral shaft fracture advantageous compared with the conventional open technique? *J Shoulder Elbow Surg* 2015;24(11):1741-1748. https://doi.org/10.1016/j.jse.2015.07.032
7. Rommens PM, Blum J, Runkel M. Retrograde nailing of humeral shaft fractures. *Clin Orthop Relat Res* 1998;350:26-39. https://doi.org/10.1097/00003086-199805000-00004
8. Itoi E, King GJW, Niebur GL, Morrey BF, An KN. Malrotation of the humeral component of the capitellocondylar total elbow replacement is not the sole cause of dislocation. *J Orthop Res* 1994;12(5):665-671. https://doi.org/10.1002/jor.1100120509

9. Sabo MT, Athwal GS, King GJW. Landmarks for rotational alignment of the humeral component during elbow arthroplasty. *J Bone Joint Surg* 2012;94(19):1794-1800. https://doi.org/10.2106/JBJS.J.01740
10. Lee HJ, Oh CW, Oh JK, Apivatthakakul T, Kim JW, Yoon JP, et al. Minimally invasive plate osteosynthesis for humeral shaft fracture: a reproducible technique with the assistance of an external fixator. *Arch Orthop Trauma Surg* 2013;133(5):649-657. https://doi.org/10.1007/s00402-013-1708-7
11. Boileau P, Bicknell RT, Mazzoleni N, Walch G, Urien JP. CT scan method accurately assesses humeral head retroversion. *Clin Orthop Relat Res* 2008;466(3):661-669. https://doi.org/10.1007/s11999-007-0089-z
12. Vlachopoulos L, Carrillo F, Dunner C, Gerber C, Székely C, Fürnstahl P. A novel method for the approximation of humeral head retrotorsion based on three-dimensional registration of the bicipital groove. *J Bone Joint Surg* 2018;100(15):e101. https://doi.org/10.2106/JBJS.17.01561
13. Micic I, Kholinne E, Hong H, Choi H, Kwak JM, Sun Y, et al. Navigation-assisted suture anchor insertion for arthroscopic rotator cuff repair. *BMC Musculoskelet Disord* 2019;20(1):633. https://doi.org/10.1186/s12891-019-3021-2

Original Article

Ewha Med J 2024;47(2):e26
https://doi.org/10.12771/emj.2024.e26
eISSN 2234-2591

OxyMask is not superior to a non-rebreathing oxygen mask for oxygen supply in a post-anesthesia care unit in Korea: a comparative study

Seung Hee Yoo[1,2], **In-Young Yoon**[2], **Dong Yeon Kim**[1,2], **Sooyoung Cho**[1,2]

[1]Department of Anesthesiology and Pain Medicine, Ewha Womans University College of Medicine, Seoul, Korea
[2]Department of Anesthesiology and Pain Medicine, Ewha Womans University Mokdong Hospital, Ewha Womans University College of Medicine, Seoul, Korea

Received Dec 1, 2023
Revised Jan 20, 2024
Accepted Apr 17, 2024

Corresponding author
Sooyoung Cho
Department of Anesthesiology and Pain Medicine, Ewha Womans University Mokdong Hospital, Ewha Womans University College of Medicine, 1071 Anyangcheon-ro, Yangcheon-gu, Seoul 07985, Korea
E-mail: sooyoung.cho@ewha.ac.kr

Keywords
General anesthesia; Hypoxia; Oxygen saturation; Recovery room; Respiratory insufficiency

Objectives: OxyMask, a novel product, has recently been used to administer oxygen postoperatively to patients who have undergone general anesthesia. This study aimed to evaluate the incidence of hypoxia in patients under general anesthesia upon arrival to the post-anesthesia care unit (PACU) using arterial blood gas analysis, and to compare the effectiveness of OxyMask with a non-rebreathing oxygen mask for oxygen administration.

Methods: We retrospectively investigated anesthesia-related data from the electronic medical records of 460 patients treated from April to November 2021. We analyzed patients aged 20 years or older who had undergone general anesthesia and whose perioperative arterial blood gas analysis results were available upon arrival to the PACU. These patients were grouped into the non-rebreathing oxygen mask (n=223) and OxyMask (n=237) groups, and statistical analysis was performed utilizing their anesthesia records.

Results: No patients exhibited hypoxia upon arrival to the recovery room. The oxygen concentration increased after oxygen administration; its concentration during the recovery room period ($\Delta 2\ PaO_2$) was 10.7±42.3 and 13.9±38.5 mmHg in the non-rebreathing oxygen mask and OxyMask groups, respectively. This difference was not statistically significant. Moreover, the arterial oxygen saturation between the end of surgery and upon arrival to the PACU ($\Delta 1\ SaO_2$) and the arterial oxygen saturation 20 minutes after oxygen administration at the PACU ($\Delta 2\ SaO_2$) did not significantly differ between the groups.

Conclusion: OxyMask was not superior to a non-rebreathing oxygen mask in terms of the effectiveness of oxygen supply.

Introduction

Background

The immediate postoperative period is a risky time when hypoxia is highly likely to occur. Should respiratory complications arise during post-anesthesia management, they can lead to serious consequences. Thus, comprehensive management is essential to ensure thorough monitoring of the patient. Postoperative hypoxia, which develops immediately after surgery, is significantly associated with factors such as anesthesia duration, surgical incision site, age, obesity, and pain [1–5].

All general patients are transferred from the operating room (OR) to the post-anesthesia care

unit (PACU) after surgery, breathing room air; at this point, there is a risk of developing hypoxia. Upon arrival in the PACU, patients receive oxygen through various devices such as masks and cannulas. Previously, our institution administered oxygen using a non-rebreathing oxygen mask (Teleflex, Morrisville, NC, USA; Fig. 1); however, we have recently begun using the OxyMask (Southmedic, Barrie, ON, Canada), a novel product (Fig. 2).

A non-rebreathing oxygen mask is a closed-type mask, which leads to air re-circulation at the lower part of the mask, specifically around the chin area. This can result in the accumulation of carbon dioxide (CO_2) in the expiratory gas [6]. The flow of oxygen in this mask is almost parallel to the face and directed toward the nose, which may not effectively support oxygen inhalation through the mouth. In contrast, the OxyMask is an open-type mask designed to prevent the accumulation of CO_2 inside the mask by allowing oxygen to flow from the center of the mask towards both the nasal and oral cavities [6]. This design could facilitate oxygen inhalation and improve CO_2 removal from the mask. However, there is a concern that the open structure of the mask might lead to oxygen dispersion.

Objectives

We assessed the incidence and severity of hypoxia during patient transfer from the OR to the PACU in individuals who had undergone general anesthesia and were extubated. Additionally, this study compared the effectiveness of a non-rebreathing oxygen mask and OxyMask in the

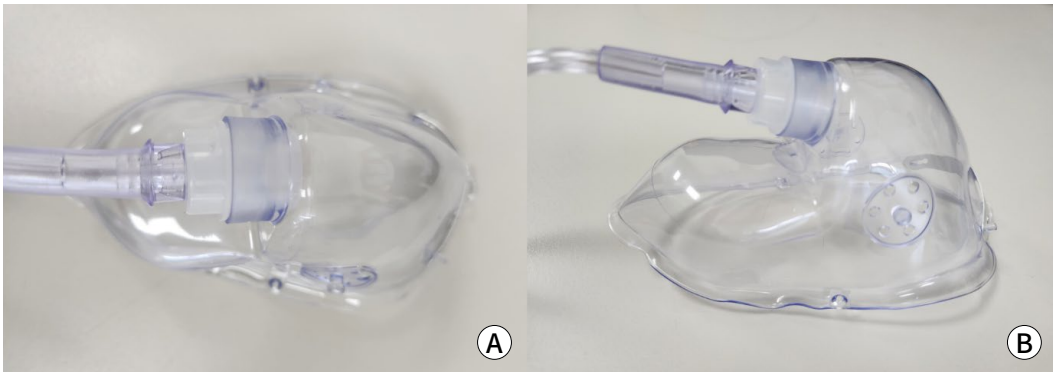


Fig. 1. Hudson RCI non-rebreathing oxygen mask. (A) Top view, (B) side view.

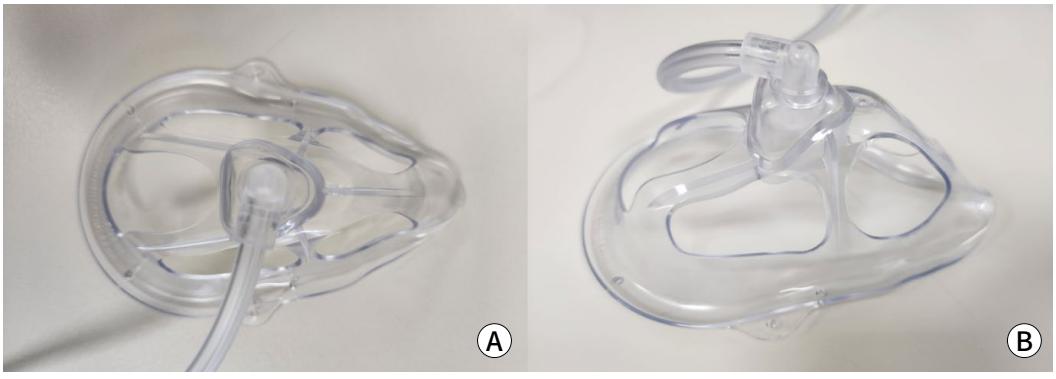


Fig. 2. OxyMask. (A) Top view, (B) side view.

https://doi.org/10.12771/emj.2024.e26

PACU by utilizing arterial blood gas analysis (ABGA).

Methods

Ethics statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (IRB) of Ewha Womans University Mokdong Hospital (IRB number: 2021-12-048-001). The requirement to obtain written informed consent was waived by the IRB since this study was performed retrospectively.

Study design

This was a comparative study. The manuscript was described according to the STROBE statement (https://www.strobe-statement.org/).

Setting

We reviewed and analyzed patients' data from their electronic medical records from April to November 2021 at Ewha Womans Medical Center. A non-rebreathing oxygen mask was used from May 1, 2021 to July 31, 2021. The OxyMask was used from August 1, 2021 to November 30, 2021.

Participants

We investigated 460 patients aged 20 years or older who had undergone general anesthesia in non-cardiac surgery, were extubated at the end of anesthesia, and had perioperative ABGA results available upon arrival to the PACU and before leaving the PACU. Patients with severe cardiopulmonary disease who experienced hypoxia and dyspnea and were therefore administered supplemental oxygen from the OR were excluded. The general protocol for anesthesia at our institution was as follows: Perioperatively, the management of anesthesia was at the discretion of the attending anesthesiologist. All patients were extubated in the OR and then transferred to the PACU after confirming spontaneous breathing with an oxygen supply via a bag-mask at 6 L/min for at least 5 minutes. During the transfer to the PACU, all patients breathed room air (FiO_2 0.21) spontaneously. Upon arrival to the PACU, all patients received supplemental oxygen at 6 L/min (FiO_2 0.44) using a non-rebreathing oxygen mask or an OxyMask, and an ABGA was performed (the first ABGA in the PACU). After at least 20 minutes of oxygen administration in the PACU, a second ABGA was performed for all patients. The attending anesthesiologist determined whether to discharge patients from the PACU based on the discharge criteria, following the Aldrete score system, which evaluates activity, respiration, circulation, consciousness, and skin color (Supplement 1) after discontinuation of the oxygen supply.

Variables (study outcomes)

The primary outcome was a comparison of the effectiveness of oxygen delivery between the OxyMask and the non-rebreathing oxygen mask during the stay in the PACU, as determined by ABGA results. Postoperative hypoxia was defined as an oxyhemoglobin saturation (SpO_2) of 90% or less for at least 2 minutes, or an SpO_2 of 85% or less at any time point. The secondary outcome focused on the incidence of postoperative hypoxia upon arrival to the PACU, which was also assessed using ABGA results.

Data sources and measurement

The collected data included the patients' demographic and clinical characteristics, as follows: sex, age, American Society of Anesthesiologists physical status, height, weight, body mass index, the presence of pulmonary disease, operation time, and anesthesia time. Intraoperative and postoperative vital signs and ABGA results were collected, and the following vital signs were recorded at the time of ABGA sampling: systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate, SpO_2, respiratory pattern, level of consciousness, and numerical pain scores were recorded every 5 minutes in all patients. From the ABGA results, pH, arterial oxygen partial pressure (PaO_2), arterial carbon dioxide partial pressure, and arterial oxygen saturation (SaO_2) were also recorded. Additionally, we calculated the difference in PaO_2 and SaO_2 between the last ABGA in the OR and the first ABGA in the PACU (Δ_1 PaO_2 and Δ_1 SaO_2, respectively). Moreover, we calculated the difference in PaO_2 and SaO_2 between the first and second PACU measurements (Δ_2 PaO_2 and Δ_2 SaO_2, respectively).

Bias

There was no selection bias reportable in this study.

Study size

Sample size estimation was not performed because this study included all target patients, with the exclusion of those who received supplemental oxygen in the OR. Additionally, the authors did not allocate participants to specific groups.

Statistical methods

Statistical analyses were performed using the anesthesia records of 460 patients, divided into two groups: the OxyMask group (n=237) and the non-rebreathing oxygen mask group (n=223; Fig. 3). Continuous variables were analyzed using either the Student t-test or the Mann–Whitney U test following a normality assessment with the Shapiro–Wilk test. Results were presented as

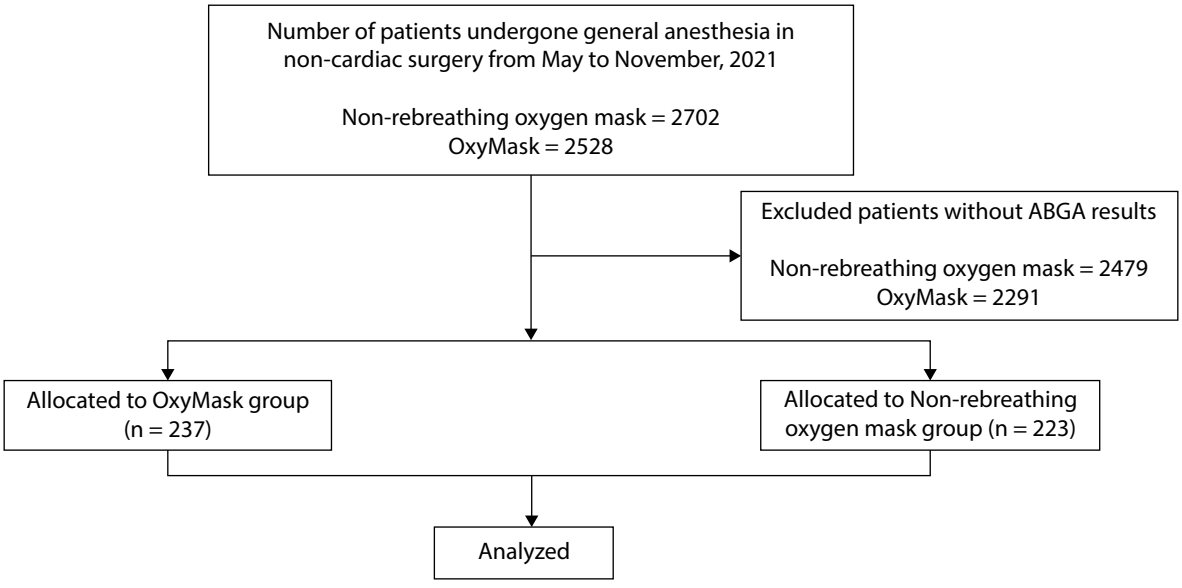


Fig. 3. Flow diagram of the participants.

means±SDs or as medians (interquartile ranges). Categorical variables were analyzed using the chi-square test or Fisher exact test, which was applied when more than 20% of the expected frequencies were fewer than 5. These results were presented as percentages (%). All statistical analyses were carried out using SPSS version 25 (IBM, Armonk, NY, USA). A P-value of less than 0.05 was considered statistically significant.

Results

Participants' demographic and clinical characteristics in the non-rebreathing oxygen mask and OxyMask groups

We analyzed 460 patients, who were divided into the OxyMask group (n=237) and non-rebreathing oxygen mask group (n=223). The two groups did not differ significantly in terms of demographic or clinical characteristics, except the operation time (167.9±125.2 vs. 192.9±139.8 min, P=0.044; Table 1). Patients who had pulmonary disease were comparable between the OxyMask and non-rebreathing oxygen mask groups (Table 1).

Incidence of postoperative hypoxia upon arrival to the post-anesthesia care unit

No hypoxia episodes occurred among the patients. One patient in the OxyMask group had a minimum PaO_2 value of 111.8 mmHg upon arrival to the PACU, with spontaneous breathing of room air after surgery.

Arterial blood gas analysis during the operation and the post-anesthesia care unit period

Upon arrival to the PACU, PaO_2 was significantly lower in the OxyMask group than in the non-rebreathing oxygen mask group (162.1±50.3 vs. 181.9±62.0 mmHg, respectively, P<0.001). Similarly, upon discharge from the PACU, PaO_2 was significantly lower in the OxyMask group than in the non-rebreathing oxygen mask group. (176.0±49.2 vs. 192.6±64.5 mmHg, respectively, P=0.002).

Table 1. Demographic characteristics of both groups

Demographic characteristic	Non-rebreathing oxygen mask (n=223)	OxyMask (n=237)	P-value
Age (years)	60.6±17.3	62.3±15.4	0.275
Sex (M/F)	91/132	105/132	0.448
Height (cm)	160.6±9.9	162.2±9.1	0.064
Weight (kg)	60.6±11.9	63.4±11.0	0.010
BMI (kg/m^2)	23.5±3.8	24.1±3.7	0.073
ASA physical classification (1/2/3/4)	41/137/45/0	27/154/55/1	0.141
Pulmonary disease (yes/no)	28/195	39/198	0.236
Operation time (min)	167.9±125.2	192.9±139.8	0.044
Anesthesia time (min)	215.4±130.3	237.8±141.8	0.079

Values are presented as mean±SDs.
The Student t-test was performed for continuous variables and the chi-square test for categorical variables.
P<0.05 is regarded as indicating statistical significance.
BMI, body mass index; ASA, American Society of Anesthesiologists.

https://doi.org/10.12771/emj.2024.e26

Δ_1 PaO_2, the gradient of arterial oxygen partial pressure between the end of surgery and upon arrival to the PACU, was significantly higher in the OxyMask group than in the non-rebreathing oxygen mask group (40.5±48.7 vs. 19.9±52.8 mmHg, respectively, P<0.001; Table 2). However, Δ_2 PaO_2, the gradient of arterial oxygen partial pressure 20 minutes after the administration of oxygen in the PACU, was not significantly different between the OxyMask and the non-rebreathing oxygen mask groups (13.9±38.5 vs. 10.7±42.3 mmHg, respectively, P=0.393; Table 2).

Δ_1 SaO_2, the gradient of arterial oxygen saturation between the end of surgery and arrival to the PACU, was not significantly different between the OxyMask and non-rebreathing oxygen mask groups (0.13±0.70 vs. 0.14±0.77 mmHg, respectively, P=0.912; Table 2). Moreover, Δ_2 SaO_2, the gradient of arterial oxygen saturation 20 minutes after the administration of oxygen supply in the PACU, was also not significantly different between the OxyMask and non-rebreathing oxygen mask groups (0.51±4.93 vs. 0.32±1.02 mmHg, respectively, P=0.544; Table 2).

Δ_1 $PaCO_2$, the gradient of arterial carbon dioxide partial pressure between the end of surgery and upon arrival to the PACU, was not significantly different between the OxyMask and non-rebreathing oxygen mask groups (40.0±5.4 vs. 39.3±5.4 mmHg, respectively, P=0.153; Table 2). Moreover, Δ_2 $PaCO_2$, the gradient of arterial carbon dioxide partial pressure 20 minutes after the administration of oxygen supply in the PACU, was also not significantly different between the OxyMask and non-rebreathing oxygen mask groups (38.1±5.3 vs. 37.4±4.8 mmHg, respectively,

Table 2. Arterial blood gas analysis during operation and the PACU period

Variable	During the operation (OR)			Arrival at the PACU (PACU[1])			Discharge from the PACU (PACU[2])		
	Non-rebreathing oxygen mask (n=223)	OxyMask (n=237)	P-value	Non-rebreathing oxygen mask (n=223)	OxyMask (n=237)	P-value	Non-rebreathing oxygen mask (n=223)	OxyMask (n=237)	P-value
SBP	117.8±17.1	119.4±16.9	0.310	129.8±8.3	133.5±21.3	0.049	129.6±18.6	134.6±21.7	0.009
DBP	66.6±12.0	66.3±11.4	0.807	77.1±14.3	79.3±16.3	0.115	75.7±11.6	77.8±14.2	0.095
HR	79.7±14.1	80.3±14.7	0.662	85.4±13.9	85.0±15.1	0.768	79.7±13.4	80.9±14.3	0.362
PACU	13.2±4.4	12.9±4.3	0.471	17.7±4.1	17.6±6.8	0.769	16.6±6.2	16.4±3.7	0.686
SpO_2	100.0±0.3	100.0±0.4	0.941	100.0±0.0	100.0±0.2	0.207	100.0±0.0	100.0±0.0	-
pH	7.40±0.05	7.41±0.05	0.211	7.38±0.05	7.38±0.05	0.719	7.39±0.06	7.40±0.05	0.699
pCO_2	38.3±3.6	37.8±3.4	0.106	40.0±5.4	39.3±5.4	0.153	38.1±5.3	37.4±4.8	0.155
pO_2	201.8±48.4	202.6±47.6	0.850	181.9±62.0	162.1±50.3	<0.001*	192.6±64.5	176.0±49.2	0.002*
Δ_1PaO_2	-	-	-	19.9±52.8	40.5±48.7	<0.001*	-	-	-
Δ_2PaO_2	-	-	-	-	-	-	10.7±42.3	13.9±38.5	0.393
SaO_2	98.7±0.7	98.6±0.8	0.256	98.0±5.0	98.1±1.2	0.736	98.5±0.7	98.5±0.7	0.193
Δ_1SaO_2	-	-	-	0.13±0.70	0.14±0.77	0.912	-	-	-
Δ_2SaO_2	-	-	-	-	-	-	0.51±4.93	0.32±1.02	0.544

Values are presented as mean±SD.

The Student t-test was performed for continuous variables.

P<0.05 is regarded as indicating statistical significance, and * significant results are shown.

OR, operating room; PACU, post-anesthesia care unit; SBP, systolic blood pressure; DBP, diastolic blood pressure; HR, heart rate; PACU, respiratory rate; SpO_2, oxyhemoglobin saturation; $PaCO_2$, arterial carbon dioxide partial pressure; PaO_2, arterial oxygen partial pressure; Δ_1PaO_2, PaO_2 at OR–PaO_2 at PACU[1]; Δ_2PaO_2, PaO_2 at PACU[2]–PaO_2 at PACU[1]; SaO_2, arterial oxygen saturation; Δ_1SaO_2, SaO_2 at OR–SaO_2 at PACU[1]; Δ_2SaO_2, SaO_2 at PACU[2]–SaO_2 at PACU[1].

https://doi.org/10.12771/emj.2024.e26

P=0.155; Table 2).

Discussion

Key results

Upon arrival to the PACU, there were no cases of postoperative hypoxia; furthermore, there was no difference in the effects of oxygen delivery between the OxyMask and the non-rebreathing oxygen mask groups during their stay in the PACU, as indicated by ABGA results.

Interpretation

Postoperative hypoxia during the early recovery period after general anesthesia is primarily due to respiratory depression caused by residual anesthetics. Therefore, it is crucial to provide appropriate and prompt oxygen supply to all patients during this time. Typically, healthy patients are transferred from the OR to the PACU breathing room air. However, breathing room air immediately after surgery can pose a risk of hypoxia, as lung function tends to deteriorate. This deterioration is characterized by a decrease in functional residual capacity, an increase in airway closure, and the development of both a ventilation/perfusion mismatch and atelectasis [7]. Further, CO_2 retention caused by hypoventilation can bring about hypoxia by replacing the oxygen from the alveoli; this is particularly important when the inhaled air is not oxygen-enriched [8].

Daley et al. investigated the incidence of hypoxemia in the PACU among adult patients who had undergone general anesthesia for elective surgery. They monitored SpO_2 levels using continuous, non-invasive pulse oximetry [9]. Their findings indicated that 41% of patients experienced hypoxemia after the oxygen supply, which had been administered for 30 minutes during their PACU stay, was discontinued. However, the condition rapidly improved with the reintroduction of oxygen, suggesting that supplemental oxygen is necessary following general anesthesia. Tyler et al. continuously monitored SaO_2 using pulse oximetry [8]. They reported that hypoxemia, which was defined as $SaO_2 \leq 85\%$ for patients who were breathing room air during their transfer from the OR to the PACU after general anesthesia and after discontinuation of oxygen supply, occurred in 35% of all patients (33 of 95 patients). The mean time interval taken for SaO_2 to decrease from 100% to 85% following the discontinuation of oxygen was 155 ±74 s. They reported that postoperative hypoxemia was not related to the anesthetic agents, age, anesthesia time, or level of consciousness. In their study, all patients were transferred from the OR to the PACU in 5 minutes with breathing room air and did not experience hypoxemia, as PaO_2 was 111.8 mmHg upon arrival to the PACU. In our study, hypoxemia did not occur during the immediate transfer to the PACU. This was likely due to the very short elapsed time following the discontinuation of oxygen for transfer and the maintenance of PaO_2 at 150 mm Hg or higher, facilitated by oxygen administration during surgery.

Oxygen was discovered centuries ago and has been administered to patients using various devices, such as the conventional simple mask or cannula. The FiO_2 range delivered to patients depends on individual patient factors and the choice of oxygen delivery device. The simple mask typically used is a mostly closed-type mask that can cause air re-circulation at the lower part of the mask, near the chin, potentially leading to the accumulation of CO_2 due to the rebreathing of expired gases. Additionally, it may be unsuitable for oral oxygen inspiration because the direction of the oxygen flow is almost parallel to the face and directed towards the nose.

OxyMask features an open design that enables oxygen to diffuse directly into the mouth and nose through a structure shaped like a pentagon with five arms extending from the base of the

mask [6]. This open design aims to minimize the buildup of expired CO_2 in the re-circulation area and offers several advantages. Additionally, it directs the flow of oxygen from the central area of the mask towards the middle of the nasal and oral cavities. However, our results did not show a difference in $PaCO_2$ levels. At PACU discharge, the $PaCO_2$ was 38.1 mmHg in the simple mask group and 37.4 mmHg in the OxyMask group, indicating that conventional simple masks also do not cause CO_2 retention.

Lamb and Piper compared the effectiveness of the OxyMask and the non-rebreathing oxygen mask using a mannequin head. They reported that the OxyMask was superior, demonstrating higher inspired oxygen, lower inspired CO_2, and more efficient CO_2 clearance [10]. DeJuilio et al. retrospectively evaluated patients pre- and post-implementation of OxyMask and reported that the previously used simple mask could be switched to the OxyMask because the OxyMask was favorable in terms of safety and cost-effectiveness [11]. Paul et al. achieved a mean FiO_2 of 25.4%–80.1% using the OxyMask, delivering 1.5–15 L/min of oxygen in healthy volunteers [6]. The OxyMask is an open-system mask with an oxygen diffuser directed toward the nasal and oral cavities, allowing control over the flow rate and oxygen concentrations. Yanez et al. investigated whether FiO_2 ranges depend on the mask type when delivering oxygen to the lips and oropharynx [12]. For 10 healthy volunteers, two sampling lines were attached and FiO_2 was measured. One sampling catheter was attached to the patient's lips and the other one was attached to the oropharynx through a nostril. The FiO_2 levels were not significantly different between the lips and oropharynx with the simple mask. However, the measured FiO_2 at the lips was higher than that at the oropharynx with the OxyMask. This drop in FiO_2 at the oropharynx was attributed to the open design of the OxyMask, which lacks a perfect seal, and is considered a dilutional effect by nasal breathing or perioral room air entrainment [12]. We investigated the effectiveness of the oxygen supply between the simple mask and OxyMask by comparing the difference in measured PaO_2 during 20 minutes of oxygen administration in the PACU. The PaO_2 after 20 minutes of oxygen administration did not differ significantly between the non-rebreathing oxygen mask and the OxyMask in this study (10.7±42.3 vs. 13.9±38.5 mmHg, respectively; P=0.393), indicating that the OxyMask was non-superior compared to the non-rebreathing oxygen mask in delivering oxygen.

No hypoxia events occurred in any study patients during the period of transfer from the OR to the PACU following general anesthesia. Perioperative management was consistent in all patients in both groups, and no manipulation was conducted to assess the incidence of hypoxia under general conditions. The OxyMask group experienced a greater drop in PaO_2 levels than was observed in the non-rebreathing oxygen mask group, between the end of surgery and arrival to the PACU. The difference in measured PaO_2 levels from the end of surgery to arrival to the PACU in the OxyMask group (40.5±48.7 mmHg) was significantly greater than that in the non-rebreathing oxygen mask group (19.9±52.8 mmHg) (P<0.001). The reason for the greater drop in PaO_2 in the OxyMask group was difficult to determine because no significant between-group differences were observed in demographic characteristics and comorbidities such as pulmonary diseases, which could have affected the oxygen demand and could have increased the risk of postoperative pulmonary complications.

While there was a difference in weight (P=0.010) between the two groups, the effects of the two masks were considered insignificant, as there was no difference in BMI (P=0.073; Table 1). Although the operation time was significantly longer in the OxyMask group, there was no difference in the occurrence of hypoxia between the two groups (Table 1). Since this study did not demonstrate a significant difference in the occurrence of hypoxia between the two masks,

it suggests that the OxyMask is not superior to the non-rebreathing oxygen mask in providing suitable oxygen. Further research is needed to establish the OxyMask as an alternative to the non-rebreathing oxygen mask for patients undergoing longer operations.

Limitations

Our study is subject to several limitations. Since the FiO_2 in the OxyMask group was not measured, it is challenging to confirm which mask delivered oxygen more efficiently. However, it can be hypothesized that the FiO_2 of the OxyMask might be lower than that of the non-rebreathing oxygen mask. However, further research is required to investigate this possibility. Another limitation is that, due to the retrospective nature of the study, there may have been a time difference between oxygenation and ABGA after PACU arrival, despite institutional protocols that dictate they should be performed simultaneously. Therefore, the timing of ABGA may not have been consistent. To compensate for this, we compared the change in PaO_2 between the two groups rather than the absolute values. Third, we did not investigate the types of surgery between the two groups, although the ABGA results could have differed according to whether patients underwent laparoscopic or open surgery. Unlike open surgery, laparoscopic surgery involves CO_2 insufflation into the abdominal cavity, which may lead to differences in values of PaO_2 or $PaCO_2$ between the two groups. Therefore, to generalize our results, further prospective studies are required to clarify the differences between the OxyMask and the non-rebreathing oxygen mask, considering the influence of comorbidities.

Conclusion

No hypoxia events occurred upon arrival to the PACU in any of the patients in this study. Therefore, it is practicable for healthy adult patients to breathe room air without supplemental oxygen when being transferred from the OR to the PACU. The increase in PaO_2 levels following oxygen administration in the PACU did not differ significantly between the two types of masks. The OxyMask was not more effective in delivering oxygen than the non-rebreathing oxygen mask.

ORCID
Seung Hee Yoo: https://orcid.org/0000-0002-6811-7198
In-Young Yoon: https://orcid.org/0000-0002-7400-0797
Dong Yeon Kim: https://orcid.org/0000-0002-4414-5653
Sooyoung Cho: https://orcid.org/0000-0002-0232-766X

Authors' contributions
Project administration: Yoo SH, Yoon IY, Kim DY, Cho S
Conceptualization: Kim DY, Cho S
Methodology & data curation: Yoon IY, Kim DY
Funding acquisition: not applicable
Writing – original draft: Yoo SH, Yoon IY
Writing – review & editing: Yoo SH, Yoon IY, Kim DY, Cho S

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Funding
Not applicable.

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials

Supplementary materials are available from: https://doi.org/10.12771/emj.2024.e26.

Supplement 1. Aldrete score system comprising the evaluation of activity, respiration, circulation, consciousness, and skin color

References

1. Harte PJ, Courtney DF, O'Sullivan EG, Brady MP. Duration of anaesthesia and post-operative hypoxaemia. *Ir J Med Sci* 1982;151(6):169-174. https://doi.org/10.1007/BF02940173
2. Ali J, Khan TA. The comparative effects of muscle transection and median upper abdominal incisions on postoperative pulmonary function. *Surg Gynecol Obstet* 1979;148(6):863-866.
3. Kitamura H, Sawa T, Ikezono E. Postoperative hypoxemia: the contribution of age to the maldistribution of ventilation. *Anesthesiology* 1972;36(3):244-252. https://doi.org/10.1097/00000542-197203000-00009
4. Vaughan RW, Engelhardt RC, Wise L. Postoperative hypoxemia in obese patients. *Ann Surg* 1974;180(6):877-882. https://doi.org/10.1097/00000658-197412000-00014
5. Spence AA, Alexander JI. Mechanisms of postoperative hypoxaemia. *Proc R Soc Med* 1972;65(1):12-14.
6. Paul JE, Hangan H, Hajgato J. The OxyMask(™) development and performance in healthy volunteers. *Med Devices* 2009;2:9-17.
7. Craig DB. Postoperative recovery of pulmonary function. *Anesth Analg* 1981;60(1):46-52.
8. Tyler IL, Tantisira B, Winter PM, Motoyama EK. Continuous monitoring of arterial oxygen saturation with pulse oximetry during transfer to the recovery room. *Anesth Analg* 1985;64(11):1108-1112.
9. Daley MD, Norman PH, Colmenares ME, Sandler AN. Hypoxaemia in adults in the post-anaesthesia care unit. *Can J Anaesth* 1991;38(6):740-746. https://doi.org/10.1007/BF03008452
10. Lamb K, Piper D. Southmedic OxyMask™ compared with the Hudson RCI® Non-Rebreather Mask™: safety and performance comparison. *Can J Respir Ther* 2016;52(1):13-15.
11. DeJuilio PA, Jenkins MB, Huml JP. Evaluation of safety and cost of an open-design oxygen mask in a large community hospital. *Respir Care* 2018;63(4):412-416. https://doi.org/10.4187/respcare.05567
12. Yanez ND, Fu AY, Treggiari MM, Kirsch JR. Oropharyngeal oxygen concentration is dependent on the oxygen mask system and sampling location. *Respir Care* 2020;65(1):29-35. https://doi.org/10.4187/respcare.07027

Health Statistics

Ewha Med J 2024;47(2):e27
https://doi.org/10.12771/emj.2024.e27
eISSN 2234-2591

Drug-induced death statistics in Korea between 2011 and 2021

Seokmin Lee

Statistics Research Institute, Statistics Korea, Daejeon, Korea

Received Feb 27, 2024
Revised Apr 4, 2024
Accepted Apr 10, 2024

Corresponding author
Seokmin Lee
Statistics Research Institute, Statistics Korea, 713 Hanbatdaero, Seo-gu, Daejeon 35220, Korea
E-mail: leon32@korea.kr

Keywords
Cause of death; Narcotics; Analgesics, opioid; International Classification of Diseases; Republic of Korea

Objectives: This study analyzed drug-induced death statistics in Korea between 2011 and 2021.
Methods: Cause-of-death statistics data from Statistics Korea were examined based on the Korean Standard Classification of Diseases and Causes of Death and the International Statistical Classification of Diseases and Related Health Problems, 10th revision.
Results: In 2021, there were 559 drug-induced deaths, marking a 172.7% increase compared to 2011, which recorded 205 deaths. The rate of drug-induced deaths per 100,000 people was 1.1 in 2021, up 153.6% from 0.4 in 2011. The mortality rate for men aged 25–34 years and women aged 35–44 years each increased fourfold from 2011 to 2021: from 0.3 to 1.2 for the former and 0.3 to 1.3 for the latter. Of the drug-induced deaths in 2021, 75.0% (419/559) were due to intentional self-harm, and 10.4% (58/559) were accidental. The number of deaths attributed to medical narcotics in 2021 was 169, a 5.5-fold increase from 2011. The most commonly implicated drugs in these deaths were sedative-hypnotic drugs, benzodiazepines, and opioids. Sedative-hypnotic drugs and benzodiazepines were frequently involved in cases of intentional self-harm, while opioids and psychostimulants were more often associated with accidental deaths.
Conclusion: The death rate from drug-induced causes is considerably lower in Korea than in the United States (1.1 vs. 29.2). However, the number of such deaths has increased recently. Since these deaths occur predominantly among younger age groups and are often the result of intentional self-harm, there is a clear need for systematic management and the implementation of targeted policies.

Introduction

Background

Deaths caused by drugs are both preventable and avoidable. Furthermore, drug overdose represents a significant issue that requires policy intervention, as it can escalate into larger social problems. Recently, drug abuse has emerged as a major global concern. According to the United States Centers for Disease Control and Prevention, there were 106,699 drug overdose deaths in 2021, marking a sharp increase since 2000. Notably, there has been a significant rise in deaths attributed to opioids such as fentanyl [1,2]. The impact of drugs on mortality involves both direct and indirect factors. Direct causes refer to cases where the primary cause of death is drug-related, as classified by the World Health Organization (WHO) in the International Standard Classification of Diseases (ICD-10). Indirect factors involve drug use increasing the risk of deaths

from other causes, such as intentional self-harm, liver disease, hepatitis, and heart disease. The Global Burden of Disease study reported that drug use is responsible for approximately 114,000 indirect deaths and 350,000 direct deaths annually [3].

Objectives

This study analyzed the characteristics of drug-related deaths, aiming to inform and support drug-related policies. Additionally, it sought to identify risk factors associated with drug-related deaths to aid in the development of strategies to reduce such fatalities. The report specifically focused on the demographic characteristics, types of deaths, and the various drugs involved in drug-induced fatalities.

Methods

Ethics statement

This study involved an analysis of public data; therefore, neither approval by the institutional review board nor the obtainment of informed consent was required.

Study design

This descriptive study was based on public data from Statistics Korea, and it was described according to the STROBE Statement available from: https://www.strobe-statement.org/.

Setting, participants, data source, and measurement

This study analyzed microdata on cause of death statistics from Statistics Korea spanning from 2011 to 2021 to examine the characteristics of drug-related deaths. The cause of death statistics in Korea are compiled from death certificates. To enhance the accuracy of determining the underlying cause of death, Statistics Korea integrates 22 types of administrative data for each individual. The detailed administrative data includes health insurance information from the National Health Insurance Service, cancer registry data from the National Cancer Center, criminal investigation records and traffic accident investigation data from the National Police Agency, autopsy records from the National Forensic Service, emergency records from the National Emergency Medical Center, among others. Notably, drug-related deaths are reliably documented, reflecting data from police investigations and autopsy reports provided by the National Forensic Service. The ICD-10 code list for causes of death due to drugs is provided in Supplement 1. Deaths due to drugs were categorized by cause of death into disease, accident, intentional self-harm, and homicide, and further analyzed by classifying the drugs involved into opioids, sedatives, and psychotropic agents.

Bias

There was no bias in data collection and analysis.

Study size

The entire population of the Republic of Korea was included. No sample size estimation was required.

Statistical methods

Descriptive statistics were applied to present the results of the data analysis.

Results

Drug-induced death

In 2021, there were 559 drug-induced deaths in Korea, marking a 172.7% increase from the 205 deaths recorded in 2011 (Table 1, Fig. 1). The average number of drug-induced deaths per day was 1.5. The mortality rate was 1.1 per 100,000 population, and the age-standardized mortality rate was 1.1 per 100,000 standardized population. Deaths due to drugs steadily increased throughout the study period and predominantly occurred in relatively young age groups (Fig. 2). While the highest percentage of all deaths in 2021 occurred in individuals aged 80–84 (18.1%), a significant proportion of drug-related deaths occurred in those aged 64 or younger.

Compared to 2011, the number of deaths in 2021 increased across all age groups for both men and women, with a notable rise in the younger demographics. Specifically, the mortality rate for men aged 25 to 34 and for women aged 15 to 24 saw significant increases (Supplement 2). Of the deaths caused by drugs in 2021, 75.0% were intentional self-harm and 10.4% were unintended accidents. The number of deaths attributed to intentional self-harm involving drugs has increased since 2011. Over the past three years, the average age at death from drug-related causes has been consistently lower for women than for men (Fig. 3). When categorizing deaths by drug type since 2011, the three most prevalent drugs based on their effects were sedatives and sleeping pills, such as zolpidem and benzodiazepines; psychotropic drugs, including antidepressants and neuroleptics; and a combination of narcotics and psychotropic drugs, notably fentanyl (Supplement 3).

Table 1. The number of drug-induced deaths, death rate, and age-standardized death rate between 2011 and 2021

Year	No. of deaths (deaths)	Death rate (deaths per 100,000 population)	Age-standardized death rate (deaths per 100,000 standard population)
2011	205	0.41	0.37
2012	255	0.51	0.44
2013	269	0.53	0.46
2014	306	0.60	0.52
2015	300	0.59	0.48
2016	399	0.78	0.65
2017	321	0.63	0.51
2018	406	0.79	0.65
2019	434	0.85	0.7
2020	501	0.98	0.84
2021	559	1.09	0.94
Change (absolute)			
from 2011	354	0.7	0.57
from 2020	58	0.1	0.1
Change (proportional, %)			
from 2011	172.7	166.2	153.6
from 2020	10.4	10.4	10.8

https://doi.org/10.12771/emj.2024.e27

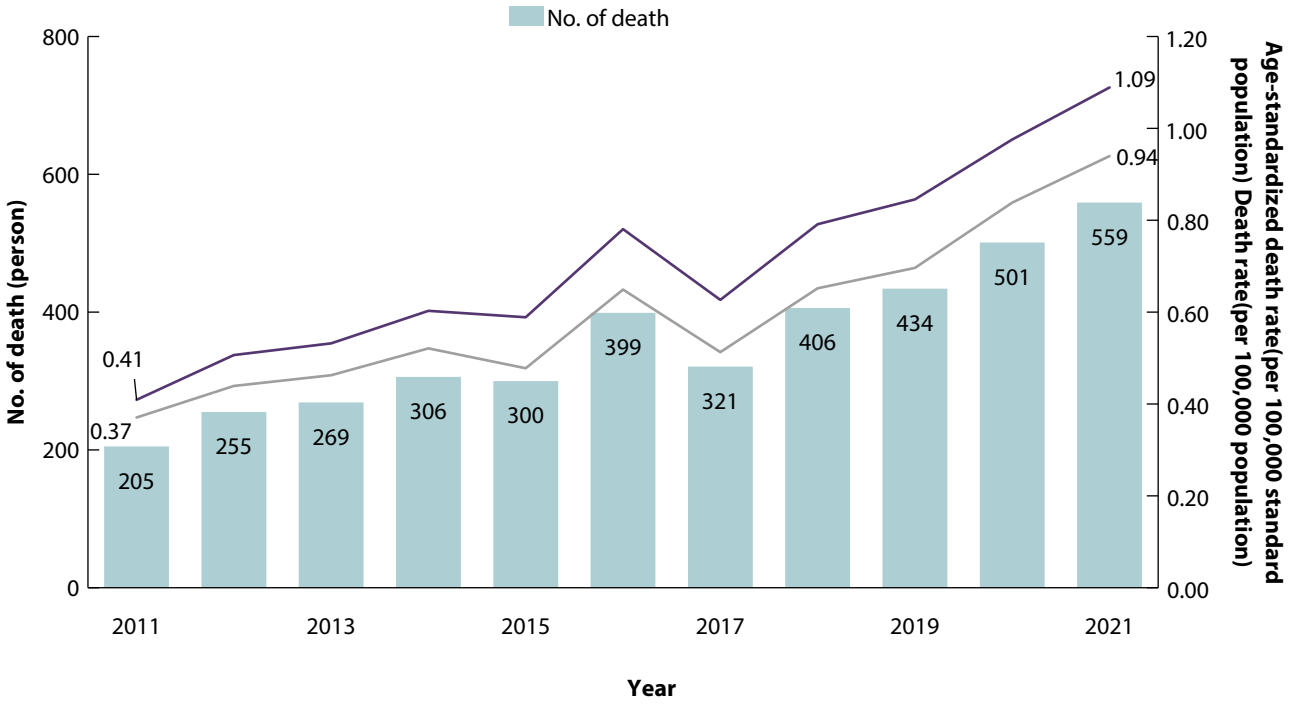


Fig. 1. Drug-induced deaths, death rate, and age-standardized death rate, 2011–2021 (units: people, per 100,000 people, and per 100,000 standard population).

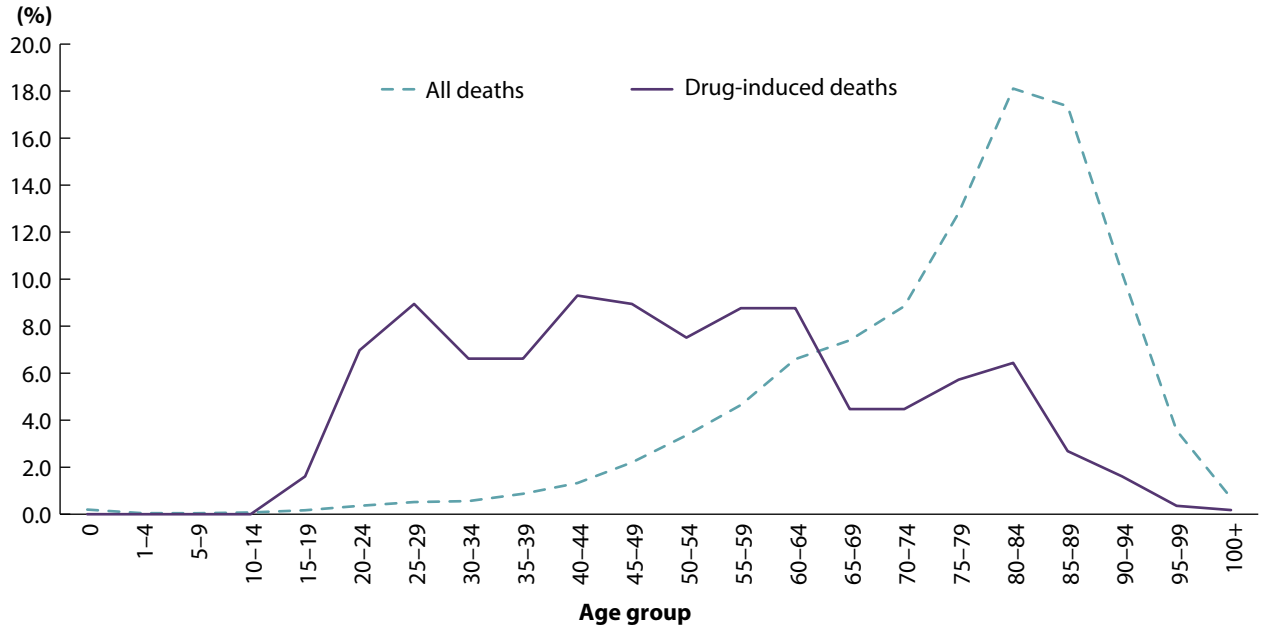


Fig. 2. Proportional age distribution for drug-induced deaths versus total deaths, 2021.

https://doi.org/10.12771/emj.2024.e27

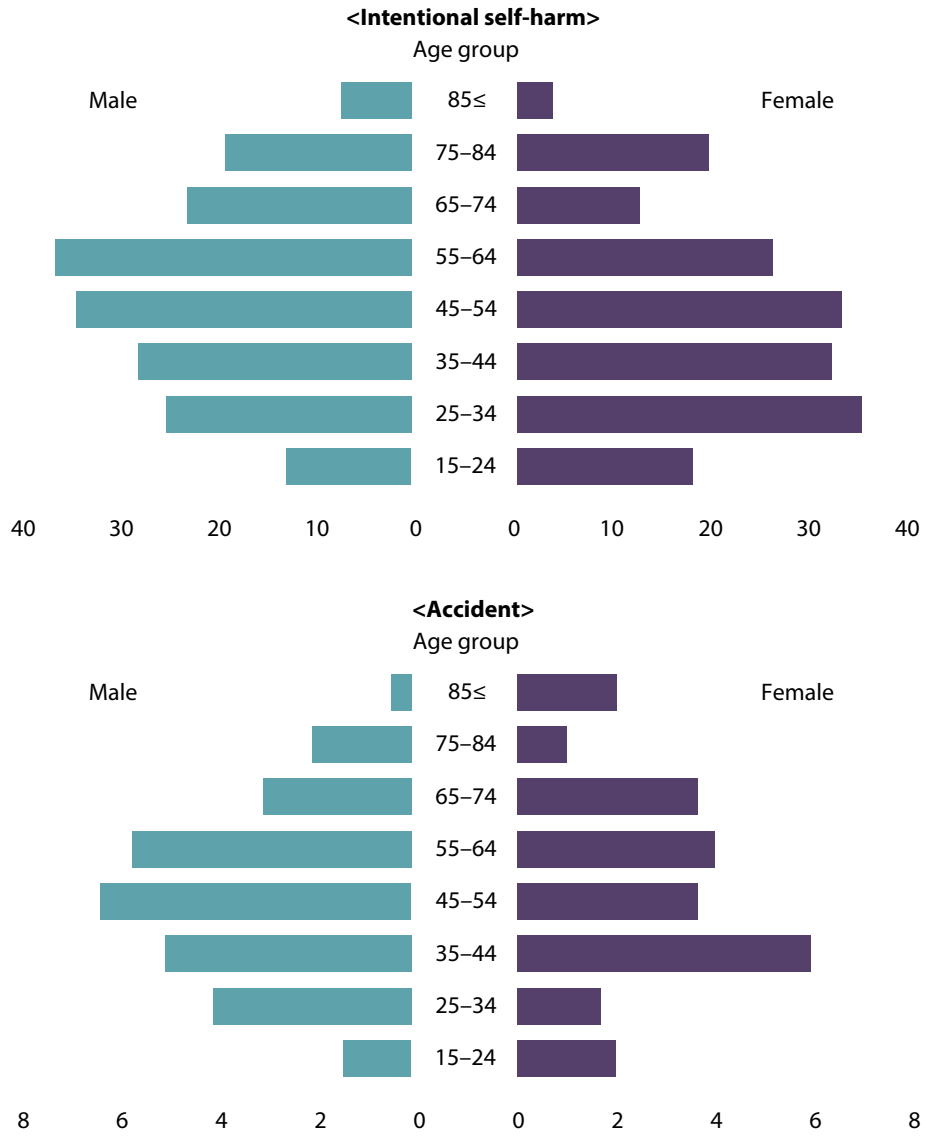


Fig. 3. Annual average number of drug-induced deaths by sex and age, 2019–2021.

Death due to medical narcotics

To analyze deaths specifically attributed to designated medical narcotics in Korea, narcotic drugs were categorized according to ICD-10 codes (Supplement 1). In 2021, there were 169 deaths due to medical narcotics, representing a 5.5-fold increase from the 31 deaths recorded in 2011. Although the number of deaths decreased from 127 in 2016 to 89 in 2019, there has been a rapid increase for two consecutive years (Fig. 4). A detailed breakdown of deaths by type of medical narcotic shows that sedative-hypnotic drugs account for the highest number, followed by benzodiazepines and opioids. Notably, the number of deaths associated with sedative-hypnotic drugs, such as zolpidem, and opioids, such as fentanyl, is on the rise (Fig. 5). Men had a higher proportion of deaths involving psychostimulants than women, and women had a higher proportion of deaths involving sedative-hypnotic drugs, general anesthetics, and appetite depressants than men. An analysis of medical narcotics deaths by age between 2019 and 2021

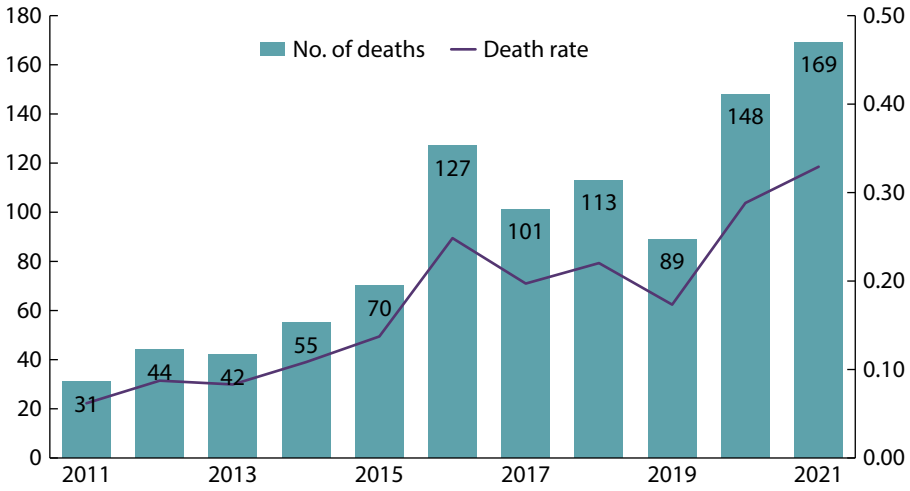


Fig. 4. Number of deaths and death rate due to medical narcotics, 2011–2021.

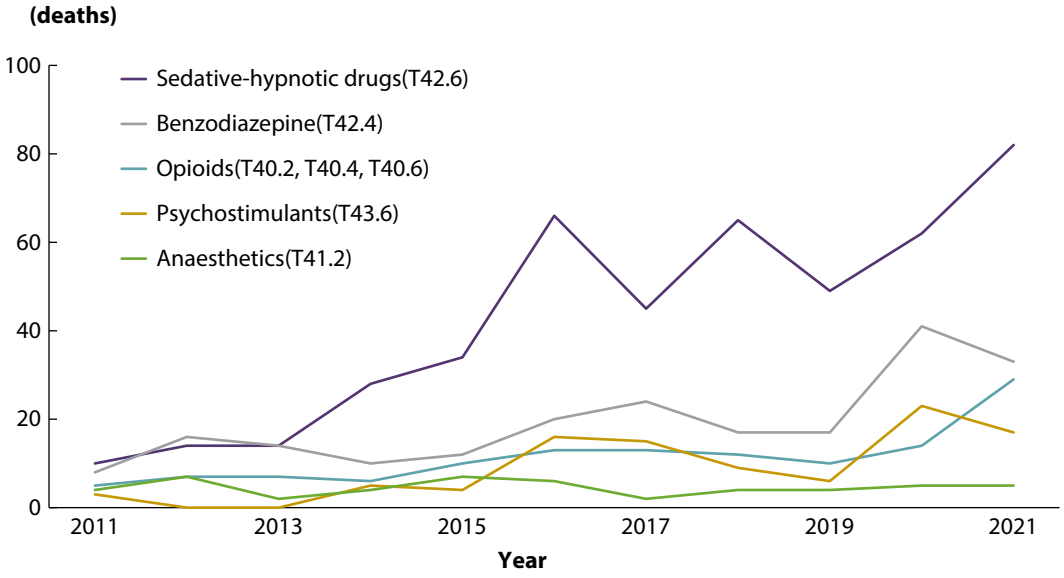


Fig. 5. Number of deaths due to medical narcotics, 2011–2021.

revealed that the risk of death from narcotics varied with age. Specifically, opioids accounted for a high proportion of deaths among individuals aged 25 to 54, benzodiazepines were involved in a large proportion of deaths among those aged 45 to 64, sedative-hypnotic drugs predominated among those aged 55 to 74, and deaths related to general anesthetics were most common among those aged 25 to 34 (Table 2).

Among deaths attributed to medical narcotics, psychostimulants and opioids represented a significant percentage of accidental fatalities. In instances of intentional self-harm, sedative-hypnotic drugs and benzodiazepines were commonly employed. Specifically, sedative-hypnotic drugs constituted 56.7% of drug-related intentional self-harm cases (Supplement 4).

To analyze the risk of death associated with the use of medical narcotics, the number of health insurance claims for narcotics was compared to the number of deaths. For both opioids

Table 2. The number of deaths due to medical narcotics by age group between 2019 and 2021 (unit: deaths)

Age group	Opioids (T40.2, T40.4, T40.6)	Anesthetics (T41.2)	Benzodiazepine (T42.4)	Sedative-hypnotic drugs (T42.6)	Others (T41.1, T42.3, T43.6, T48.3, T50.5)
15–24	1	2	7	3	8
25–34	12	6	7	16	7
35–44	14	4	8	19	15
45–54	11	2	17	31	16
55–64	8	0	22	42	6
65–74	4	0	13	42	3
75–84	0	0	12	29	-
≥85	3	0	5	11	-

and psychotropic drugs, the proportion of deaths relative to the number of claims is higher in younger age groups, indicating that the risk of death from narcotic drugs is comparatively high among the young. Specifically, the number of deaths relative to the number of opioid claims in the 25–44 age group represents a higher proportion compared to other age groups (Supplement 5).

Discussion

Key results

In 2021, there were 559 drug-induced deaths, marking a 172.7% increase from the 205 deaths recorded in 2011. The rate of drug-induced deaths per 100,000 people rose to 1.1 in 2021, up 153.6% from 0.4 in 2011. Of the drug-induced deaths in 2021, 75.0% were due to intentional self-harm, and 10.4% were accidental. Deaths attributed to medical narcotics reached 169 in 2021, a significant increase, up 5.5 times from 31 in 2011. The most commonly involved drugs in these fatalities were sedative-hypnotic drugs, benzodiazepines, and opioids.

Interpretation

While most deaths occur between the ages of 80 and 84, the majority of drug-related deaths took place in individuals under the age of 64 (Fig. 2). This suggests that deaths due to drugs often result in premature mortality compared to other causes, thereby disproportionately increasing the disease burden. Furthermore, there has been a significant rise in the risk of death among younger age groups over the past decade. Given that 75% of drug-related deaths are due to intentional self-harm (Fig. 3), it is evident that intentional self-harm involving drugs has significantly contributed to the increased mortality rates in this demographic. This trend highlights the emergence of drug-induced intentional self-harm as a pressing social issue.

Deaths due to medical narcotics have increased more rapidly than those due to other drugs (Fig. 4). Gender differences were observed in deaths from medical narcotics: men were more likely to die from drugs with stimulating effects, while women were more likely to die from drugs with sedative effects. The types of medical narcotics associated with the highest mortality rates also varied by age, reflecting the fact that the most commonly prescribed drugs and treatments differ across age groups (Supplement 5). Notably, benzodiazepines were disproportionately involved in deaths among the young age group of 15 to 24 years old (Table 2). Because opioids

are often used as painkillers for terminal cancer patients, there are limited medical applications of opioids in younger age groups. However, there are two potential reasons for the relatively high risk of opioid-related deaths among young people. The first is the misuse of narcotic drugs, where death results from intentional misuse without adhering to prescribed dosages or methods of administration. The second involves medication being obtained through illegal distribution or purchase, rather than being prescribed through a legitimate health insurance system. To conduct a thorough analysis, it is essential to prepare big data linking narcotic drug prescriptions to death data.

Among medical narcotics, drugs with sedative effects—including sedative-hypnotic drugs, anesthetics, and benzodiazepines—are frequently used for intentional self-harm (Dataset 1). Therefore, these drugs require special management. When comparing the number of health insurance claims to the number of deaths associated with medical narcotics, the ratio for individuals aged 25–44 was notably high (Table 2). This indicates an elevated risk of death in this younger age group, necessitating targeted management and policies to reduce drug-related deaths.

Comparison with previous studies

No previous articles have reported drug-induced death statistics in Korea. As drug addiction becomes an increasingly significant social issue in many countries, including the United States, the need for robust statistics to inform related policies is becoming more apparent. In the United States, the age-standardized drug-induced death rate for the total population increased by 29.4% from 22.8 in 2019 to 29.5 in 2020 [2]. The European Union has developed an estimation model to address the problem of undercounting drug-related deaths [4].

Limitations/suggestions

Lack of statistical indicators related to drug-induced deaths

The need for policy support to address drug-related deaths is growing, yet there is a significant shortage of statistical indicators that can determine the extent and risk factors associated with these fatalities. This scarcity of statistical indicators for drug-induced deaths stems from three primary factors.

The first issue is the incompleteness of the criteria used to classify deaths caused by drugs. Typically, the management of drug distribution and prescriptions is governed by the Anatomical Therapeutic Chemical Classification (ATC) codes, which are designated by the Collaborating Center for Pharmaceutical Statistics Methods (WHOCC), an affiliate of the WHO. Each ATC code is structured into five levels: drug application site, drug efficacy, drug characteristics, chemical properties, and individual ingredients. This detailed classification system facilitates the specific categorization of drugs, such as opioid-related drugs and benzodiazepines. However, when classifying causes of death, the ICD-10 from the International Standard Classification System (WHO-FIC), another affiliate of the WHO, is utilized. The data on the number of deaths derived from ICD-10 codes is not without its limitations. Due to inconsistencies in code ranges, deaths caused by drugs other than narcotics are inadvertently included. For instance, the T48.3 code, which denotes poisoning by cough medicine, encompasses drugs other than the medical narcotics dextromethorphan and zipeprol. Nonetheless, the risk of death and addiction is significantly higher with medical narcotics than with other general drugs. Additionally, the T43.6 code, which refers to intoxication by psychostimulants with abuse potential, primarily includes methylphenidate, a legal drug used to treat attention deficit–hyperactivity disorder,

and methamphetamine, an illegal substance. The ICD-10 coding system's limitations in terms of the details of drug classification suggest that there is potential for further subdivision in the upcoming revised ICD-11. Thus, the classification systems for drug prescriptions and causes of death differ significantly, particularly in that the cause-of-death codes do not adequately classify drugs in detail.

The second factor is the scarcity of data regarding drug-induced deaths. While drug prescriptions are well-documented, including details about the recipient, the dosage, and the specific medications prescribed, information about drug-related deaths can typically only be obtained through an autopsy or toxicology testing. Furthermore, elderly individuals often take various medications for multiple conditions, and it is not uncommon for younger people to intentionally consume multiple drugs.

Third, there is a lack of linked data spanning prescriptions, illnesses, and deaths. It is crucial to determine whether drug-related deaths are due to acute or chronic poisoning. Furthermore, the underlying diseases and health status of the deceased should be taken into account to accurately assess the impact of the drug on mortality. Therefore, analyzing data that connects prescriptions, illnesses, and deaths is essential. By examining linked data, it would be possible to empirically ascertain the risk of death associated with a drug by comparing its risk and efficacy against the number of people prescribed the drug or the dosage prescribed.

Because cause-of-death statistics must adhere to the standards set by the WHO, analyses involving multiple drugs or drug efficacy, which are necessary for drug death statistics, are not suitable. Therefore, new, separate statistics are required to accurately identify the characteristics of deaths caused by drugs.

How to improve drug-related death statistics

The target population for these statistics is defined as Korean citizens whose deaths are associated with drugs. Determining whether drugs are related to a death is only possible through autopsy results; therefore, the actual population is defined as those cases where drugs were detected in autopsies conducted by the National Forensic Service. To develop new statistics using autopsy data, it is essential to establish standards for classifying drugs. In the statistics for drug-induced deaths among Koreans, the statistical classification (ATC code) used for drug prescription and management does not align with the code that classifies the cause of death (ICD-10). Currently, the ICD-10 does not specify detailed drug types. Therefore, a linkage table between the ICD-11 and ATC codes must be developed. To ensure the stability of time series in this linkage table, ICD-10 codes can be additionally linked to construct statistics using a consistent classification system from the drug prescription stage through to the death stage. By developing statistics that classify autopsy data using both ATC codes and ICD-11 codes, it becomes possible to analyze not only the efficacy and type of drugs but also the risks associated with polypharmacy. With detailed classification of drug types, it would be feasible to construct linked big data that spans drug prescription, disease prevalence, and death.

Data sources for linked big data necessitate the integration of health insurance claim details, autopsy data, and cause-of-death statistics. These linked datasets can serve as empirical evidence for analyzing drug death risks and drug safety. Specifically, when multiple drugs are consumed, it is possible to analyze further that the risk of death may increase due to synergistic effects.

Among many countries, the method Australia uses to compile cause-of-death statistics is similar to that of Korea, making drug-induced death statistics from Australia highly relevant to

Korea. Generally, drug-related deaths are prone to underestimation; however, the reliability of these statistics is enhanced in Australia by incorporating the coroner's approval process, and in Korea by including both autopsy results and police investigations in the statistical compilation [5,6]. Australia has compiled statistics on drugs and opioids and has implemented targeted policies, which have led to a decrease in mortality rates.

Conclusion

As drug prices fall and online transactions on platforms like the dark web and cryptocurrency become more prevalent, making them difficult to trace, the risk of drug-related deaths has increased. This is compounded by a rise in the overseas inflow of drugs. Therefore, statistical indicators that can be used to establish and evaluate related policies are essential. It is anticipated that this study and further in-depth analysis will facilitate the development of future research that will identify population groups at risk for drug use, provide targeted educational support, and establish guidelines, ultimately helping to reduce the risk of premature death due to drugs.

ORCID
Seokmin Lee: https://orcid.org/0000-0001-6642-2677

Authors' contributions
All work was done by Seokmin Lee.

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Funding
Not applicable.

Data availability
Raw data are available from microdata on cause of death statistics from Statistics Korea from 2011 to 2021 at https://mdis.kostat.go.kr/index.do
Data files are available from Harvard Dataverse: https://doi.org/10.7910/DVN/D3CXBJ

Dataset 1. Drug-induced deaths, including deaths due to medical narcotics, from 2011 to 2021 by sex, age, type of death, marital status, and educational attainment provided by Statistics Korea

Acknowledgments
Not applicable.

Supplementary materials
Supplementary materials are available from: https://doi.org/10.12771/emj.2024.e27.

Supplement 1. ICD-10 codes for causes of death due to drugs
Supplement 2. The number of deaths and death rate due to drugs by sex and age
Supplement 3. The ranking of deaths due to drug characteristics between 2011 and 2021
Supplement 4. Number of deaths due to medical narcotics by type of death between 2019 and 2021
Supplement 5. Number of deaths compared to the number of medical narcotic prescriptions by age between 2019 and 2021
Supplement 6. Data used to generate Figs. 1–5

References

1. Kochanek KD, Murphy SL, Xu J, Arias E. Deaths: final data for 2020. *Natl Vital Stat Rep* 2023;72(10):1-92.
2. Spencer MR, Miniño AM, Warner M. Drug overdose deaths in the United States, 2001—2021. *NCHS Data Brief* 2022;(457):1-8. https://dx.doi.org/10.15620/cdc:122556
3. Organisation for Economic Co-operation and Development [OECD]. Addressing problematic opioid use in OECD countries [Internet]. Paris (FR): OECD; c2019 [cited 2023 Oct 4]. Available from: https://doi.org/10.1787/a18286f0-en
4. Giraudon I, Mathis F, Hedrich D, Vicente J, Noor A. Drug-related deaths and mortality in Europe: update from the EMCDDA

expert network [Internet]. Lisbon (PT): European Monitoring Centre for Drugs and Drug Addiction; c2021 [cited 2023 Oct 4]. Available from: https://doi.org/10.2810/777564
5. Australian Bureau of Statistics. Causes of death [Internet]. Canberra (AT): Australian Bureau of Statistics; c2021 [cited 2023 Oct 4]. Available from: https://www.abs.gov.au/statistics/health/causes-death
6. Australian Bureau of Statistics. Opioid-induced deaths in Australia [Internet]. Canberra (AT): Australian Bureau of Statistics; c2019 [cited 2023 Oct 4]. Available from: https://www.abs.gov.au/articles/opioid-induced-deaths-australia

Case Report

Ewha Med J 2024;47(2):e28
https://doi.org/10.12771/emj.2024.e28
eISSN 2234-2591

Gastric adenocarcinoma with enteroblastic differentiation in a 67-year-old man in Korea: a case report

Hae Rin Lee[1,2], **Gwang Ha Kim**[1,2,3], **Dong Chan Joo**[1,2], **Moon Won Lee**[1,2], **Bong Eun Lee**[1,2], **Kyung Bin Kim**[4]

[1]Division of Gastroenterology, Pusan National University Hospital, Busan, Korea
[2]Department of Internal Medicine, Pusan National University School of Medicine, Busan, Korea
[3]Biomedical Research Institute, Pusan National University Hospital, Busan, Korea
[4]Department of Pathology, Pusan National University Hospital, Busan, Korea

Received Mar 4, 2024
Revised Apr 19, 2024
Accepted Apr 19, 2024

Corresponding author
Gwang Ha Kim
Department of Internal Medicine, Pusan National University School of Medicine, and Biomedical Research Institute, Pusan National University Hospital, 179 Gudeok-ro, Seo-gu, Busan 49241, Korea
E-mail: doc0224@pusan.ac.kr

Keywords
Adenocarcinoma; Alpha-fetoproteins; Endoscopic mucosal resection; Gastrectomy; Transcription factor 4

We report a rare case of gastric adenocarcinoma with enteroblastic differentiation (GAED) that was treated with endoscopic submucosal dissection followed by additional distal gastrectomy with lymph node dissection. A 67-year-old man underwent endoscopic submucosal dissection for a gastric lesion, which was diagnosed as GAED with submucosal and lymphatic invasion. Histologically, GAED is characterized by a tubulopapillary growth pattern and clear cells that resemble those of the primitive fetal gut. Immunohistochemically, GAED variably expresses oncofetal proteins such as glypican-3, alpha-fetoprotein, and spalt-like transcription factor 4. Despite negative margins, additional gastrectomy with lymph node dissection was performed due to submucosal and lymphatic invasion. No residual tumor or metastasis was detected, and the patient remained disease-free for 2 years before dying from causes unrelated to GAED. Given its aggressive nature, frequent lymphovascular invasion, and high metastatic potential, clinicians should recognize the histopathological diagnosis of this rare tumor and its propensity for aggressiveness.

Introduction

Gastric adenocarcinoma with enteroblastic differentiation (GAED), also known as clear cell gastric carcinoma, is a rare and poorly documented malignancy, representing less than 1% of all gastric cancers [1,2]. While GAED represents a subtype of alpha-fetoprotein (AFP)-producing adenocarcinomas [1], the relationship between GAED and AFP production is not well understood [2]. Histologically, the tumor is characterized by an intestine-like structure composed of cuboidal or columnar neoplastic cells with clear cytoplasm. These cells test positive for oncofetal proteins, such as glypican-3, spalt-like transcription factor 4 (SALL4), and AFP [3]. GAED tends to be more aggressive than conventional adenocarcinoma, with a higher propensity for lymphovascular invasion and a greater likelihood of metastasis to the liver and lymph nodes (LNs) [4]. In this report, we present a rare case of GAED that was managed with endoscopic submucosal dissection and additional distal gastrectomy with LN dissection.

Case presentation

Ethics statement

This case report was granted an exemption from consent and review by the Pusan National University Hospital Research Ethics Review Committee (IRB No. 2402-023-136).

Patient information

A 67-year-old man visited our hospital seeking treatment for high-grade dysplasia in the stomach, which was identified during esophagogastroduodenoscopy at a health checkup. The patient was asymptomatic. His medical history included alcoholic hepatitis and chronic hepatitis B, along with heavy alcohol consumption and a 40-pack-year smoking history.

Clinical findings

The results of the physical examination were unremarkable.

Diagnostic assessment

Laboratory analysis indicated a slight elevation of liver function test results, suggestive of alcoholic hepatitis. Tumor markers, including serum AFP, carcinoembryonic antigen, and carbohydrate antigen 19–9, were within normal limits. Esophagogastroduodenoscopy revealed a 2-cm slightly depressed lesion with nodular mucosal changes on the anterior wall of the gastric prepylorus (Fig. 1A). Magnifying endoscopy with narrow-band imaging revealed a clear demarcation line and irregular patterns in microsurface (MS) and microvascular (MV) structures,

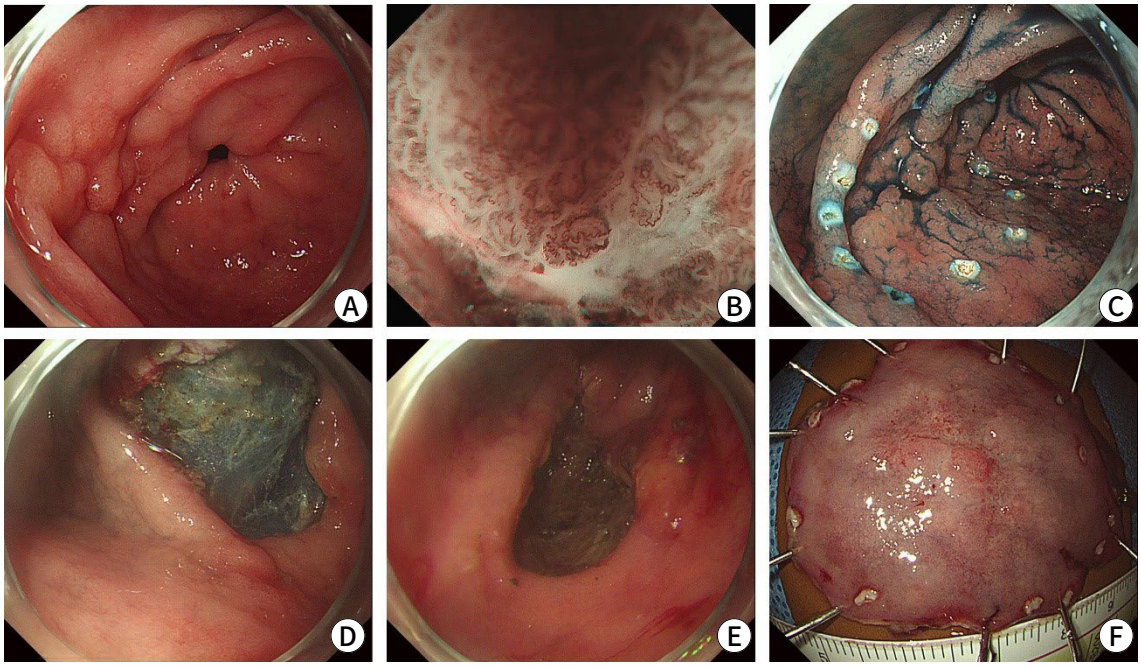


Fig. 1. Endoscopic submucosal dissection for early gastric cancer. (A) Conventional endoscopy and indigo carmine chromoendoscopy reveal a 2-cm slightly depressed lesion with nodular mucosal changes on the anterior wall of the gastric prepylorus. (B) Magnifying endoscopy with narrow-band imaging shows irregular microsurface and microvascular patterns. (C) Marking dots are placed around the lesion. (D) A circumferential incision and submucosal dissection are performed using an insulated-tip knife. (E) The lesion is completely excised. (F) The resected specimen is shown.

https://doi.org/10.12771/emj.2024.e28

including irregular oval/tubular MS and irregular loop MV patterns (Fig. 1B). Endoscopic ultrasonography indicated that the lesion was confined to the mucosal layer. Abdominal and chest computed tomography revealed no evidence of LN involvement or distant metastases.

Therapeutic intervention and final diagnosis

Endoscopic submucosal dissection was performed to achieve complete resection of the lesion (Fig. 1C–E). Assessment of the gross appearance of the resected specimen revealed a 19-mm, IIc lesion with an irregular mucosal surface (Fig. 1F). Based on microscopic examination, the tumor exhibited a tubulopapillary growth pattern along with submucosal invasion (Fig. 2A). The tumor was overlaid by conventional adenocarcinoma and was partially composed of cuboidal or columnar cells with clear cytoplasm, a feature reminiscent of the primitive fetal gut and indicative of enteroblastic differentiation (Fig. 2B). On immunohistochemical staining, the tumor cells tested negative for glypican-3 and AFP, which are recognized as oncofetal proteins (Fig. 2C, D). Although the horizontal and deep resection margins were tumor-free, the tumor had penetrated the deep submucosa (750 μm from the muscularis mucosa) and exhibited lymphatic invasion. Consequently, additional distal gastrectomy with LN dissection was performed. Examination of the surgical specimens revealed no residual tumor or LN metastasis.

Follow-up and outcomes

During the 2-year follow-up period, we observed no evidence of local or distant recurrence. However, 3 years after the intervention for GAED, the patient died of necrotizing pneumonia and uncontrolled alcoholic hepatitis.

Discussion

GAED, also known as clear cell gastric carcinoma, is a rare entity in the stomach. Clear cell carcinomas are more commonly found in the lower urinary tract and the female reproductive

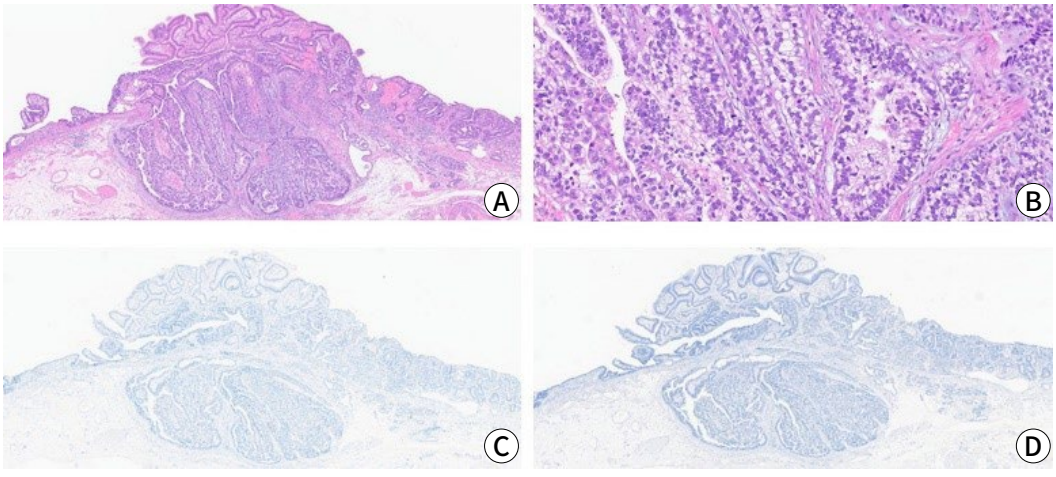


Fig. 2. Histopathological findings. (A) The tumor exhibits a tubulopapillary growth pattern and submucosal invasion (hematoxylin and eosin [H&E] stain, ×40). (B) Tumor cells display clear cytoplasm and a tubular pattern, indicative of enteroblastic adenocarcinoma (H&E stain, ×200). (C,D) Tumor cells test negative for glypican-3 (C) and alpha-fetoprotein (D) (immunohistochemical stain, ×40).

https://doi.org/10.12771/emj.2024.e28

system, specifically the endometrium and ovary. Due to the infrequency of GAED, its clinicopathologic and immunohistochemical features are not yet fully understood [2,5,6]. Histologically, enteroblastic adenocarcinoma often coexists with conventional well-differentiated or moderately-differentiated tubular adenocarcinoma, found in the upper portion of the tumor [2]. GAED is characterized by a tubulopapillary growth pattern, with predominantly clear cells and luminal eosinophilic secretions [7]. In the present case, magnifying endoscopy with narrow-band imaging revealed irregular oval/tubular MS and irregular loop MV patterns. These findings were in line with the histopathological characteristics of GAED, as described in previous reports [8,9].

On immunohistochemical staining, most GAEDs variably express three enteroblastic lineage markers, also known as oncofetal proteins. These markers include glypican-3 (associated with hepatoid gastric carcinoma), AFP (a marker for hepatocellular carcinoma and yolk sac tumor), and SALL4 (a marker for AFP-producing gastric carcinoma) [10–12]. For the diagnosis of GAED, glypican-3 is the most sensitive marker, followed by SALL4 and AFP [2]. In the present case, staining was negative for glypican-3 and AFP, and SALL4 staining could not be performed due to a lack of necessary testing equipment. When immunohistochemical staining reveals AFP production in gastric carcinoma cells, the lesion is classified as AFP-producing gastric carcinoma. This type of carcinoma is associated with a poor prognosis due to a high incidence of lymphovascular invasion and liver metastasis [13].

Clear cells are characterized by an abundant cytoplasm filled with substances such as glycogen, lipid, water, or mucin [6]. Gastric carcinomas with clear cell changes (GCCs) typically display a cytoplasmic accumulation of glycogen and mucin. Our previous research indicated that GCCs secondary to glycogen deposition are associated with the expression of AFP, glypican-3, and CD10. In contrast, GCCs with mucin deposition are linked to the expression of MUC5AC and MUC6 [14]. GAED and hepatoid adenocarcinoma are representative histologic subtypes of GCC. Hepatoid adenocarcinoma with clear cells is distinguished from GAED by its poor prognosis, diffuse and strong expression of oncofetal proteins, and intestinal mucin phenotype [7]. In contrast, GAED exhibits focally heterogenous expression of oncofetal proteins and frequently expresses CD10, CDX-2, and MUC6, but not MUC2 and MUC5AC. These features suggest a gastric antral/intestinal mucin phenotype with focal enteroblastic differentiation [7].

Similar to AFP-producing adenocarcinoma, the presence of clear cell changes in gastric cancer is associated with a poor prognosis compared to conventional gastric adenocarcinoma [14]. Research indicates that most patients with GAED (90%) exhibit lymphatic and/or vascular invasion [2]. LN metastasis is observed in 40% of early-stage cases and 84% of advanced cases, which exceeds the rates observed in conventional gastric adenocarcinoma (20%–45%).

In conclusion, GAED is a rare malignancy characterized by distinct histopathological features. It is more aggressive than conventional adenocarcinoma, with frequent lymphovascular invasion and metastasis to the liver and LNs; consequently, it has a poor prognosis. Clinicians should therefore recognize the histopathological diagnosis of this rare tumor and remain cognizant of its aggressive behavior.

ORCID
Hae Rin Lee: https://orcid.org/0009-0000-2921-9127
Gwang Ha Kim: https://orcid.org/0000-0001-9721-5734
Dong Chan Joo: https://orcid.org/0000-0001-8734-4938
Moon Won Lee: https://orcid.org/0000-0002-8411-6398
Bong Eun Lee: https://orcid.org/0000-0003-2734-2134
Kyung Bin Kim: https://orcid.org/0000-0001-5430-4235

Authors' contributions
Project administration: Kim GH
Conceptualization: Kim GH
Methodology & data curation: Joo DC, Lee MW, Lee BE, Kim KB
Funding acquisition: not applicable
Writing – original draft: Lee HR, Kim GH
Writing – review & editing: Lee HR, Kim GH, Joo DC, Lee MW, Lee BE, Kim KB

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Funding
Not applicable.

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Not applicable.

References

1. Kumar S, Jabbar K. Gastric adenocarcinoma with enteroblastic differentiation: a rare find. *Am J Clin Pathol* 2020;154(Suppl 1):S65. https://doi.org/10.1093/ajcp/aqaa161.141
2. Murakami T, Yao T, Mitomi H, Morimoto T, Ueyama H, Matsumoto K, et al. Clinicopathologic and immunohistochemical characteristics of gastric adenocarcinoma with enteroblastic differentiation: a study of 29 cases. *Gastric Cancer* 2016;19(2):498-507. https://doi.org/10.1007/s10120-015-0497-9
3. Dias E, Santos-Antunes J, Nunes AC, Rodrigues JA, Pinheiro J, Macedo G. Gastric adenocarcinoma with enteroblastic differentiation: an unexpected cause of upper gastrointestinal bleeding. *Acta Gastroenterol Belg* 2021;84(4):678-679. https://doi.org/10.51821/84.4.022
4. Abada E, Anaya IC, Abada O, Lebbos A, Beydoun R. Colorectal adenocarcinoma with enteroblastic differentiation: diagnostic challenges of a rare case encountered in clinical practice. *J Pathol Transl Med* 2022;56(2):97-102. https://doi.org/10.4132/jptm.2021.10.28
5. Afshar Ghotli Z, Serra S, Chetty R. Clear cell (glycogen rich) gastric adenocarcinoma: a distinct tubulo-papillary variant with a predilection for the cardia/gastro-oesophageal region. *Pathology* 2007;39(5):466-469. https://doi.org/10.1080/00313020701569972
6. Govender D, Ramdial PK, Clarke B, Chetty R. Clear cell (glycogen-rich) gastric adenocarcinoma. *Ann Diagn Pathol* 2004;8(2):69-73. https://doi.org/10.1053/j.anndiagpath.2004.01.002
7. Kwon MJ, Byeon S, Kang SY, Kim KM. Gastric adenocarcinoma with enteroblastic differentiation should be differentiated from hepatoid adenocarcinoma: a study with emphasis on clear cells and clinicopathologic spectrum. *Pathol Res Pract* 2019;215(9):152525. https://doi.org/10.1016/j.prp.2019.152525
8. Ishikawa A, Nakamura K. Gastric adenocarcinoma with enteroblastic differentiation resected through endoscopic submucosal dissection: a case report. *Case Rep Gastroenterol* 2024;18(1):68-73. https://doi.org/10.1159/000535954
9. Kato T, Hikichi T, Nakamura J, Takasumi M, Hashimoto M, Kobashi R, et al. Two cases of gastric adenocarcinoma with enteroblastic differentiation resected by endoscopic submucosal dissection. *Clin J Gastroenterol* 2021;14(3):736-744. https://doi.org/10.1007/s12328-021-01356-z
10. Kinjo T, Taniguchi H, Kushima R, Sekine S, Oda I, Saka M, et al. Histologic and immunohistochemical analyses of α-fetoprotein: producing cancer of the stomach. *Am J Surg Pathol* 2012;36(1):56-65. https://doi.org/10.1097/PAS.0b013e31823aafec
11. Ushiku T, Shinozaki A, Shibahara J, Iwasaki Y, Tateishi Y, Funata N, et al. SALL4 represents fetal gut differentiation of gastric cancer, and is diagnostically useful in distinguishing hepatoid gastric carcinoma from hepatocellular carcinoma. *Am J Surg Pathol* 2010;34(4):533-540. https://doi.org/10.1097/PAS.0b013e3181d1dcdd
12. Yamauchi N, Watanabe A, Hishinuma M, Ohashi KI, Midorikawa Y, Morishita Y, et al. The glypican 3 oncofetal protein is a promising diagnostic marker for hepatocellular carcinoma. *Mod Pathol* 2005;18(12):1591-1598. https://doi.org/10.1038/modpathol.3800436

13. Kono K, Amemiya H, Sekikawa T, Iizuka H, Takahashi A, Fujii H, et al. Clinicopathologic features of gastric cancers producing alpha-fetoprotein. *Dig Surg* 2002;19(5):359-365.
https://doi.org/10.1159/000065838
14. Kim JY, Park DY, Kim GH, Jeon TY, Lauwers GY. Does clear cell carcinoma of stomach exist? Clinicopathological and prognostic significance of clear cell changes in gastric adenocarcinomas. *Histopathology* 2014;65(1):90-99.
https://doi.org/10.1111/his.12372

Case Report

Ewha Med J 2024;47(2):e29
https://doi.org/10.12771/emj.2024.e29
eISSN 2234-2591

Nontuberculous mycobacterial infection in a sporotricoid distribution in Korea: a case report

Jin Ju Lee, Yoon Jin Choi, Ji Yeon Byun, You Won Choi, Joo Young Roh, Hae Young Choi

Department of Dermatology, Ewha Womans University College of Medicine, Seoul, Korea

Received Dec 20, 2023
Revised Mar 29, 2024
Accepted Apr 16, 2024

Corresponding author
Ji Yeon Byun
Department of Dermatology, Ewha Womans University College of Medicine, 1071 Anyangcheon-ro, Yangcheon-gu, Seoul 07985, Korea
E-mail: jybyun@ewha.ac.kr

Keywords
Anti-bacterial agents; Biopsy; Nontuberculous mycobacteria; Polymerase chain reaction; Republic of Korea

Nontuberculous mycobacterial infections, which are often acquired from environmental sources such as water and soil, exhibit a variety of cutaneous manifestations that frequently lead to misdiagnoses and delays in treatment. A 77-year-old woman presented with multiple skin lesions in a sporotricoid distribution on her right leg, which persisted despite standard antibiotic treatments. Based on the skin biopsy, revealing granulomatous inflammation with acid-fast bacilli, and PCR testing, a nontuberculous mycobacterial infection was diagnosed. Antimycobacterial drug combinations, including clarithromycin, isoniazid, and rifampicin for 4 months, complete the skin lesion's clearance. This case underscores the need for heightened suspicion and the use of appropriate diagnostic techniques, including tissue biopsies and molecular methods such as PCR.

Introduction

Nontuberculous mycobacterial infections are caused by mycobacteria other than *Mycobacterium tuberculosis* and *Mycobacterium leprae*. Nontuberculous mycobacteria are commonly found in the environment, particularly in water and soil, and are more frequently associated with skin diseases than *M. tuberculosis* [1]. The infections they cause present a broad spectrum of skin symptoms. Due to this diversity, these infections are often misdiagnosed, leading to delays in treatment [2].

Case presentation

Ethics statement

Informed consent for publication was obtained from the patient.

Patient information

A 77-year-old woman presented with multiple skin lesions on her right leg that had developed approximately 3 to 4 months previously. Aside from hypertension, she had no significant medical history and no known exposure to water or soil that might explain her condition.

Clinical findings

A physical examination revealed several erythematous to maroon-colored crusted deep

nodules arranged in a linear pattern on her right leg (Fig. 1).

Timeline

She was initially treated for cellulitis, but her condition did not improve. Therefore, she was referred for further investigation.

Diagnostic assessment

A skin biopsy demonstrated granulomatous inflammation extending deep into the subcutaneous tissue (Fig. 2A, B). Acid-fast bacilli (AFB) were identified with Ziehl-Neelsen staining (Fig. 2C). PCR analysis for mycobacteria was also performed on the tissue specimen, and the results were positive. We used AdvanSure TB/NTM real-time PCR (LG Chem, Seoul, Korea); however, this system cannot define the exact type of tuberculosis. Attempts to culture the nontuberculous mycobacteria, both in a mycobacteria growth indicator tube and on Lowenstein-Jensen medium, were unsuccessful.

Therapeutic intervention

Treatment began with minocycline (50 mg twice daily), leading to gradual improvement over 3 months, but was halted due to gray hyperpigmentation at the treated sites. A switch to clarithromycin (500 mg daily) led to moderate improvement, but new lesions appeared after 4 months. Therefore, the regimen was modified to include isoniazid (200 mg per day) and rifampicin (450 mg per day), leading to noticeable clinical improvement within a month.

Follow-up and outcomes

After 4 months on this regimen, the lesions completely cleared, and no relapse was noted

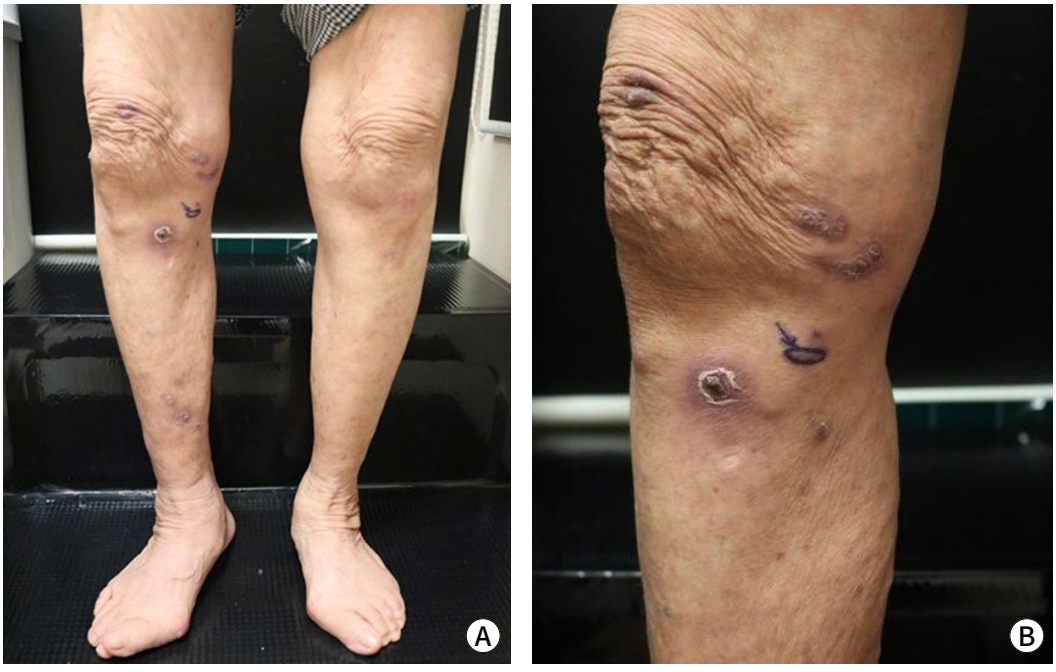


Fig. 1. Multiple erythematous to maroon-colored crusted deep nodules were arranged linearly on the right leg. Informed consent was obtained for the publication of this case report and accompanying images.

https://doi.org/10.12771/emj.2024.e29

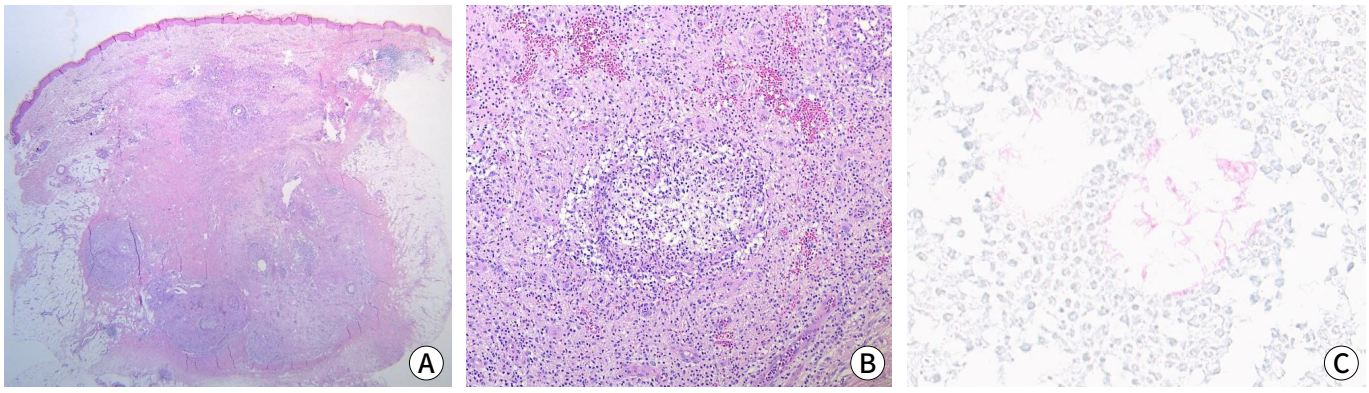


Fig. 2. Histological findings. (A,B) Granulomatous inflammation was observed in the dermis and subcutaneous tissue (hematoxylin and eosin: A, ×100; B, ×200). (C) The Ziehl-Neelsen stain revealed acid-fast bacilli (×400). Informed consent was obtained for the publication of this case report and the accompanying images.

during a 6-month follow-up.

Discussion

The prevalence of skin infections caused by nontuberculous mycobacteria appears to be increasing. These infections manifest with a range of cutaneous symptoms, such as abscesses, cellulitis, sporotrichoid nodules, ulcers, and panniculitis. The diverse nature of these symptoms makes diagnosis challenging, necessitating a high degree of suspicion in relevant clinical contexts to ensure timely identification. Nontuberculous mycobacterial infections should be suspected in patients whose skin infections are resistant to standard treatments [3].

Infections that present in a 'sporotrichoid' form are characterized by multiple lesions along the superficial lymphatic vessels, resembling the lymphangitis observed in sporotrichosis [4]. Various mycobacteria, including *Mycobacterium marinum*, *Mycobacterium kansasii*, *Mycobacterium avium* complex, and *Mycobacterium chelonae*, are known to exhibit this sporotrichoid pattern [5].

The diagnosis of mycobacterial infection necessitates tissue biopsies to evaluate the presence of AFB and to culture tissue specimens. Molecular techniques, such as PCR, are increasingly utilized to accurately identify mycobacterial pathogens in tissue samples [5]. In this instance, AFB were detected histologically, and nontuberculous mycobacteria were confirmed through PCR, although the specific organism could not be cultured.

Treatment guidelines recommend susceptibility testing of mycobacterial isolates to optimize the choice of specific antimycobacterial drug combinations [5]. Due to the inability to isolate the causative mycobacterium, empirical treatments were administered, assuming an *M. marinum* infection, which typically demonstrates a sporotricoid pattern. There is no standardized treatment regimen for nontuberculous mycobacterial infections, owing to the rarity of cases and the absence of controlled trials. Common regimens for *M. marinum* include tetracyclines, specifically minocycline and doxycycline, trimethoprim-sulfamethoxazole, rifampicin, and clarithromycin. For resistant cases, a combination of rifampicin and ethambutol may be employed. The duration of therapy varies based on clinical response and can last up to 1 year [6]. It is advised to continue medication for at least 4–8 weeks after lesions have disappeared [7].

In conclusion, we report a case of nontuberculous mycobacterial infection presenting with a sporotrichoid distribution. Obtaining histopathology and conducting appropriate culture or

molecular tests are essential for making the diagnosis.

ORCID
Jin Ju Lee: https://orcid.org/0000-0002-6188-6306
Yoon Jin Choi: https://orcid.org/0000-0002-0375-0312
Ji Yeon Byun: https://orcid.org/0000-0003-4519-9474
You Won Choi: https://orcid.org/0000-0001-6315-3889
Joo Young Roh: https://orcid.org/0000-0002-9878-6691
Hae Young Choi: https://orcid.org/0000-0003-3460-2539

Authors' contributions
Project administration: Byun JY
Conceptualization: Byun JY, Choi YW, Roh JY, Choi HY
Methodology & data curation: Choi YW, Roh JY, Choi HY
Funding acquisition: not applicable
Writing – original draft: Byun JY
Writing – review & editing: Lee JJ, Choi YJ, Byun JY, Choi YW, Roh JY, Choi HY

Conflict of interest
Ji Yeon Byun has been an associate editor of the *Ewha Medical Journal*; however, she was not involved in peer review process or decision making. No other potential conflict of interest relevant to this article was reported.

Funding
Not applicable.

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Not applicable.

References

1. Sethi A. Tuberculosis and infections with atypical mycobacteria. In: Kang S, Amagai M, Bruckner AL, Enk AH, Margolis DJ, McMichael AJ, editors. Fitzpatrick's dermatology in general medicine. 9th ed. New York: McGraw-Hill; 2019. p.2870-2871.
2. Jogi R, Tyring SK. Therapy of nontuberculous mycobacterial infections. *Dermatol Ther* 2004;17(6):491-498. https://doi.org/10.1111/j.1396-0296.2004.04051.x
3. Gonzalez-Santiago TM, Drage LA. Nontuberculous mycobacteria: skin and soft tissue infections. *Dermatol Clin* 2015;33(3):563-577. https://doi.org/10.1016/j.det.2015.03.017
4. Weedon D. Weedon's skin pathology. 3rd ed. London: Churchill Livingstone; 2010. p.560.
5. Franco-Paredes C, Marcos LA, Henao-Martínez AF, Rodríguez-Morales AJ, Villamil-Gómez WE, Gotuzzo E, et al. Cutaneous mycobacterial infections. *Clin Microbiol Rev* 2018;32(1):e00069-18. https://doi.org/10.1128/cmr.00069-18
6. Palenque E. Skin disease and nontuberculous atypical mycobacteria. *Int J Dermatol* 2000;39(9):659-666. https://doi.org/10.1046/j.1365-4362.2000.00821.x
7. Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, et al. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am J Respir Crit Care Med* 2007;175(4):367-416. https://doi.org/10.1164/rccm.200604-571ST

Case Report

Ewha Med J 2024;47(2):e30
https://doi.org/10.12771/emj.2024.e30
eISSN 2234-2591

Ultrasound-guided sciatic nerve block in a patient with sciatic neuropathy associated with uterine myoma: a case report

Bo Kyung Kang, Min Hyouk Beak, Won-joong Kim

Department of Anesthesiology and Pain Medicine, Ewha Womans University Mokdong Hospital, Ewha Womans University College of Medicine, Seoul, Korea

Received Mar 13, 2024
Revised Apr 15, 2024
Accepted Apr 16, 2024

Corresponding author
Won-joong Kim
Department of Anesthesiology and Pain Medicine, Ewha Womans University Mokdong Hospital, Ewha Womans University College of Medicine, 1071 Anyangcheon-ro, Yangcheon-gu, Seoul 07985, Korea
E-mail: ickypoo@ewha.ac.kr

Keywords
Diagnosis, differential; Ultrasonography, interventional; Magnetic resonance imaging; Myoma; Sciatica

Although sciatica is commonly associated with lumbar spinal issues, it is important to acknowledge that non-spinal factors can also play a significant role in this condition. This is particularly relevant for female patients, in whom gynecologic conditions can lead to secondary sciatic neuropathy. Herein, we report the case of a 66-year-old woman who experienced posterolateral right lower extremity radiating pain. We initially performed a lumbar transforaminal epidural steroid injection, but the pain persisted. Subsequently, hip MRI revealed sciatic neuropathy adjacent to the pedunculated portions of a uterine myoma. We then performed a sub-gluteal sciatic nerve block under ultrasound guidance, resulting in significant relief of her pain. In conclusion, hip MRI can be helpful for the differential diagnosis of sciatica, and ultrasound-guided sciatic nerve block can be considered an appropriate and effective treatment option.

Introduction

Sciatica refers to pain that travels downward from the buttocks along the path of the sciatic nerve, originating from the L4 through S2 nerve roots [1]. These roots converge in the lumbosacral plexus to give rise to the peroneal and tibial nerves, which jointly form the sciatic nerve [2]. Previous studies have reported that approximately 85% of sciatica cases are associated with lumbar disc disorders [3]. The pain typically affects one side but can be bilateral, depending on factors such as disc rupture, foraminal stenosis, lumbar stenosis, and spondylolisthesis [2]. Nevertheless, although lumbar disc diseases commonly contribute to sciatica, it is crucial for clinicians to keep in mind that non-spinal factors can also play a significant role. Instances of non-spinal causes of sciatica include piriformis syndrome, zoster sine herpete, pelvic and gynecologic conditions, diabetic neuropathy, and trauma at the gluteal injection site, among others [2]. These conditions can lead to sciatic nerve compression through various mechanisms. Notably, in female patients, gynecologic issues such as intrapelvic endometriosis and tumors, as well as leiomyomas, can result in secondary sciatic neuropathy. In this context, we present a case of sciatic neuropathy in a female patient with a uterine myoma who was treated with ultrasound-guided sciatic nerve block. Informed written consent was provided by the patient for the publication of this case report and the accompanying images.

Case presentation

Ethics statement

Informed consent for publication was obtained from the patient.

Patient information

A 66-year-old woman (height, 149 cm; weight, 51.4 kg; body mass index, 23.0 kg/m^2) presented to our pain clinic with a 5-month history of right hip pain and a limping gait. She reported experiencing radiating pain in the posterolateral aspect of her right lower extremity while walking, with a numerical rating scale (NRS) score of 9 (NRS: 0=no pain, 10=worst pain imaginable). The pain was exacerbated by walking on the ground or lying on her right side and persisted even at rest. Before her referral to our hospital, she had undergone nerve blocks several times at other clinics, but they had no effect. The patient had no past medical history and was currently taking gabapentin (300 mg) and limaprost (5 μg) three times a day without pain relief.

Clinical findings

On physical examination, the active and passive ranges of motion of the hip in internal rotation were slightly decreased, but the straight leg raise test was negative on both the right and left sides. Additionally, motor and sensory examinations were intact in both lower limbs. Lumbar MRI revealed multilevel bulging discs and L4–5 degenerative spondylolisthesis, but no significant neural foraminal stenosis or central canal stenosis (Fig. 1).

Timeline

We initially diagnosed the patient with radiculopathy caused by degenerative spondylolisthesis

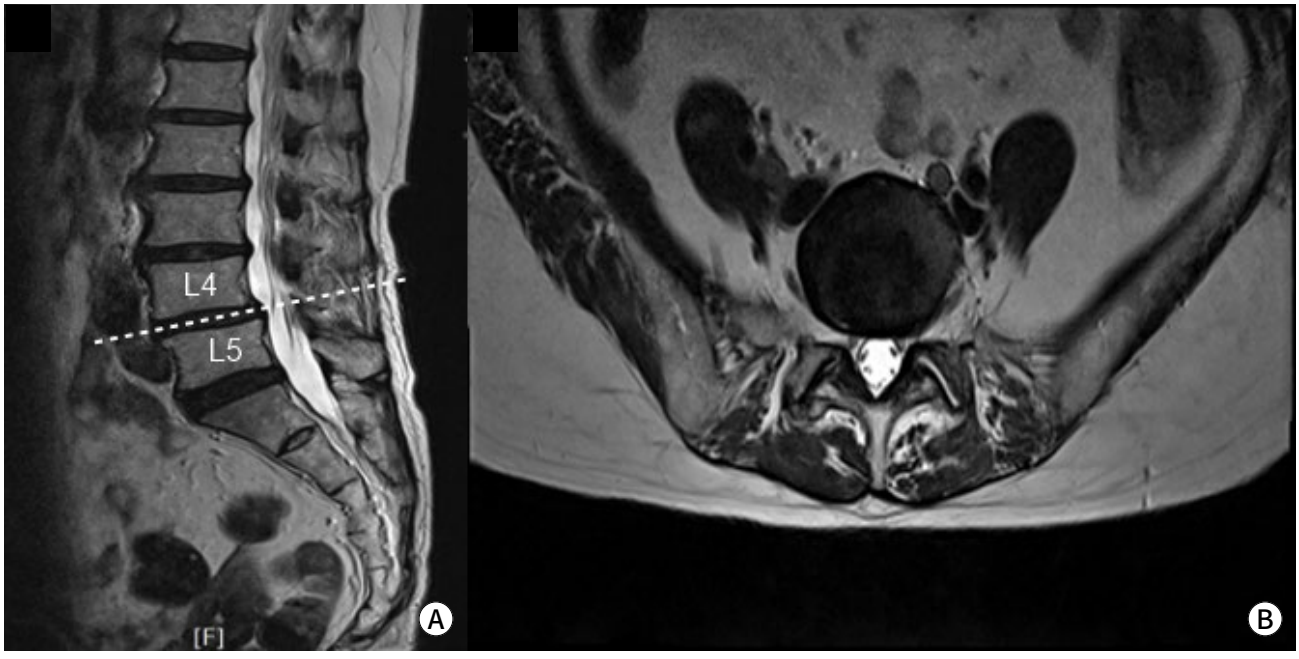


Fig. 1. Lumbar MRI. (A) Sagittal view, (B) axial view.

https://doi.org/10.12771/emj.2024.e30

at L4–5, and therefore, a right L5–S1 transforaminal epidural steroid injection was performed. At a 1-week follow-up, the patient's pain score decreased from NRS 9 to 7 on the right lateral thigh, but tingling and aching pain in the right buttock and posterior thigh persisted. As the patient's pain was not effectively relieved, we performed a piriformis muscle injection (0.1875% ropivacaine [5 mL], dexamethasone [5 mg]) under ultrasound guidance to rule out piriformis syndrome. After the procedure, she experienced gradual pain relief, with an NRS score of 5. We repeated the piriformis muscle injection 2 weeks later, but radiating pain in the buttock and posterior thigh remained.

Diagnostic assessment

For further evaluation, we decided to perform hip MRI and a needle EMG study. EMG did not reveal specific abnormal findings, but hip MRI showed thickening and T2 hyperintensity of the right sciatic nerve with perineural fat infiltrations extending to the sub-gluteal region. Additionally, this area was adjacent to the pedunculated portions of a uterine myoma (Fig. 2).

Therapeutic intervention

Based on the MRI findings, we planned to perform a sciatic nerve block under ultrasound guidance. The patient was in a prone position. A curvilinear transducer was then placed in the transverse plane at the lateral buttocks, and the procedure was performed slightly below the gluteal region. The transducer was placed between the greater trochanter and ischial tuberosity, and the sub-gluteal space was seen as a hypoechoic space between the hyperechoic borders of the gluteus maximus and quadratus femoris muscles, extending from the greater trochanter to the ischial tuberosity. The sciatic nerve was visualized as an oval hyperechoic structure between the greater trochanter and ischial tuberosity. Under ultrasound guidance, we performed a sub-gluteal sciatic nerve block with a mixture of 0.1875% ropivacaine (5 mL) and dexamethasone (5 mg; Fig. 3).

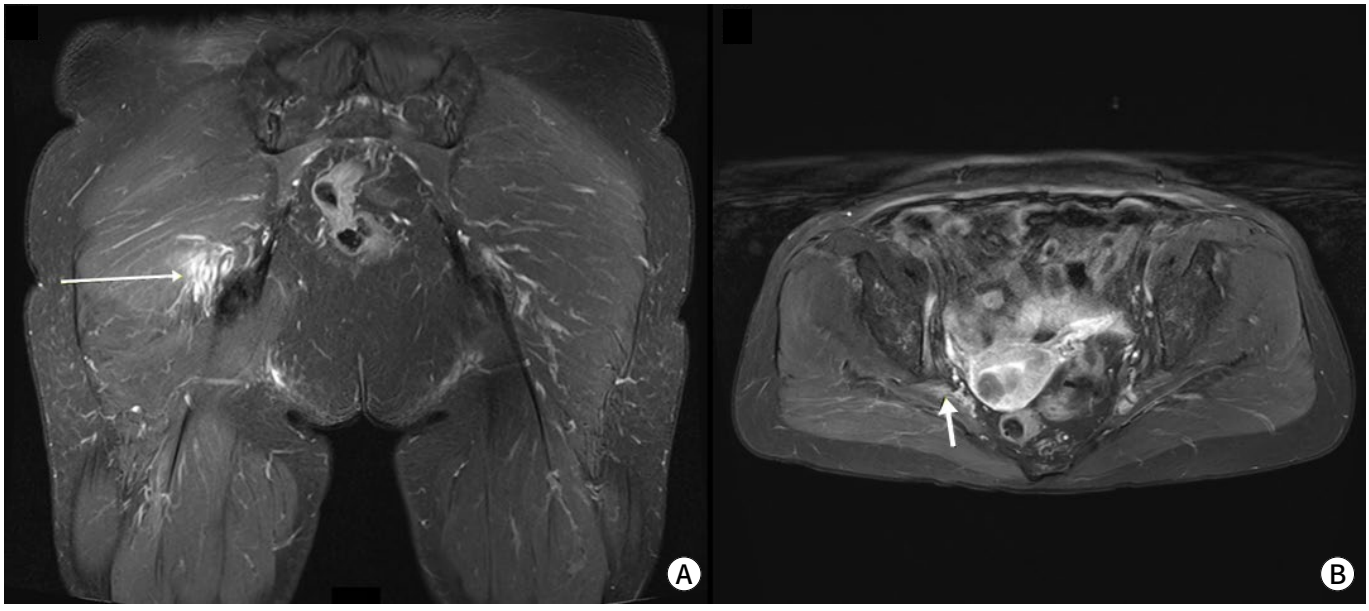


Fig. 2. Hip MRI. (A) Coronal view, (B) axial view. Arrows: sciatic nerve.

https://doi.org/10.12771/emj.2024.e30

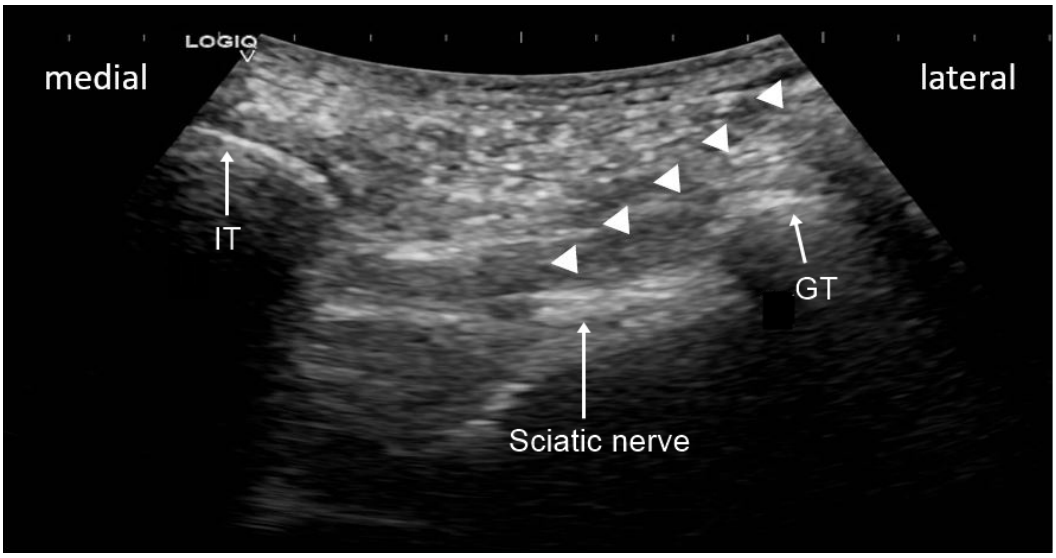


Fig. 3. Ultrasound-guided sub-gluteal sciatic nerve block. Arrowhead: needle trajectory. IT, ischial tuberosity; GT, greater trochanter.

Follow-up and outcomes

The patient's pain was substantially relieved, reaching an NRS score of 2. She expressed dramatic pain relief after the procedure and was totally satisfied. No more aching symptoms remained when lying down or walking. Following the procedure, she was prescribed a nonsteroidal anti-inflammatory drug and antiepileptic drug for 4 weeks. At a follow-up visit 3 months later, continued symptom improvement was noted. She was advised to return if her symptoms worsened, but she did not attend any subsequent visits.

Discussion

To the best of the authors' knowledge, this is the first report of successful treatment using ultrasound-guided sciatic nerve block in a patient with sciatic neuropathy objectively confirmed to have been caused by a uterine myoma based on hip MRI.

Several cases have been reported in which a uterine myoma was identified as the cause of sciatica [1,4–6]. In those cases, despite variations in patient age, the size, number, and location of the myomas, and proximity to menopause, the common treatment for sciatica was total hysterectomy. Additionally, sciatic neuropathy was not objectively confirmed by MRI in those cases. However, in our case, hysterectomy was not indicated, and we pursued an alternative method to relieve her pain—specifically, based on the findings from hip MRI, we performed an ultrasound-guided sciatic nerve block in the targeted region.

Sciatic nerve block is categorized into parasacral, subgluteal, anterior, and popliteal approaches based on the injection region. The target for the parasacral approach is the sciatic nerve at a point distal to the lateral edge of the sacrum and caudal to the sacroiliac joint [7]. However, this approach is essentially a sacral plexus block that targets branches of the entire sacral plexus before the true sciatic nerve is formed at the inferior edge of the piriformis muscle. It is performed at the level of the greater sciatic foramen [7]. The sub-gluteal approach targets the sciatic nerve as it traverses the sub-gluteal space, located between the greater trochanter

https://doi.org/10.12771/emj.2024.e30

and the ischial tuberosity. This space is found between the posterior aspect of the quadratus femoris and the anterior aspect of the gluteus maximus [8,9]. In our patient, we performed a sub-gluteal sciatic nerve block at the confirmed site under ultrasound guidance. The anterior approach aims at the sciatic nerve in the proximal thigh as it descends medially to the femur, situated between the adductor magnus anteriorly and the biceps femoris and semitendinosus posteriorly. The popliteal approach targets the sciatic nerve as it divides into the common peroneal nerve and the tibial nerve in the popliteal fossa region, typically 5–12 cm from the popliteal crease [9,10].

The sciatic nerve traverses a short intrapelvic course from the pelvis, passing through the greater sciatic foramen [11]. The sacral plexus is located on the posterior pelvic wall, anterior to the piriformis muscle, and posterior to the sigmoid colon, ureter, and internal iliac vessels. Due to its close proximity to the piriformis, the sciatic nerve is susceptible to irritation and entrapment [12]. According to a previous report, in cases of sciatica associated with obstetrical gynecological disorders, endometriosis is the most common cause, followed by factors related to pregnancy and labor, fibroids, sacral osteophytes, endosalpingiosis, needle interventions, pelvic metastasis, piriformis-related sciatica, and singular cases involving adenomyosis, intrauterine devices, hematocolpos, tubo-ovarian abscesses, and retroverted uterus [13].

The treatment approach for uterine myomas is influenced by various considerations. Relatively novel, less-invasive approaches are options, alongside pharmacological therapy, conventional surgical methods, and expectant management. Surgical procedures are considered if the patient exhibits abnormal uterine bleeding that fails to respond to conservative management, or if there is strong suspicion of pelvic malignancy, myoma growth after menopause, distortion of the endometrial cavity in infertile women, pain or pressure symptoms that diminish quality of life, or anemia resulting from chronic uterine blood loss [14]. In this case, our patient did not experience abnormal uterine bleeding, but did report radiating lower extremity pain. She was not a candidate for surgery; therefore, we planned an ultrasound-guided sciatic nerve block.

It is challenging to accurately explain why pain relief was achieved after just one sciatic nerve block without treating the causative disease. Even a single nerve block using a local anesthetic can provide lasting pain relief, which is attributed to neuroplasticity [15]. This likely applies to our case as well. Additionally, this patient underwent piriformis muscle injections twice before the sub-gluteal sciatic nerve block. During piriformis muscle injections, the medication can spread to the sciatic nerve, leading to a parasacral sciatic nerve block. This may explain why the patient's symptoms were somewhat alleviated after the piriformis muscle injections. In other words, although the patient only underwent one sub-gluteal sciatic nerve block, it can be considered that she experienced repeated sciatic nerve blocks as a result of the spread of medication during the previous injections.

In conclusion, when female patients experience radiating pain that does not respond as expected to lumbar treatment, it is crucial to consider the possibility of gynecologic problems in the differential diagnosis of sciatica. Hip MRI can be helpful for diagnosis, and if surgical treatment is not indicated, ultrasound-guided sciatic nerve block can be an appropriate and effective treatment option.

ORCID
Bo Kyung Kang: https://orcid.org/0000-0003-2744-1044
Min Hyouk Beak: http://orcid.org/0009-0005-8439-6552
Won-joong Kim: https://orcid.org/0000-0003-2046-8690

Authors' contributions
Project administration: Kim W
Conceptualization: Kim W
Methodology & data curation: Kang BK, Beak MH
Funding acquisition: not applicable
Writing – original draft: Kang BK, Beak MH, Kim W
Writing – review & editing: Kang BK, Beak MH, Kim W

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Funding
Not applicable.

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Not applicable.

References

1. Bodack MP, Cole JC, Nagler W. Sciatic neuropathy secondary to a uterine fibroid: a case report. *Am J Phys Med Rehabil* 1999;78(2):157-159. https://doi.org/10.1097/00002060-199903000-00015
2. Ropper AH, Zafonte RD. Sciatica. *N Engl J Med* 2015;372(13):1240-1248. https://doi.org/10.1056/NEJMra1410151
3. Porchet F, Wietlisbach V, Burnand B, Daeppen K, Villemure JG, Vader JP. Relationship between severity of lumbar disc disease and disability scores in sciatica patients. *Neurosurgery* 2002;50(6):1253-1259. https://doi.org/10.1097/00006123-200206000-00014
4. Murata Y, Takahashi K, Murakami M, Moriya H. An unusual cause of sciatic pain. *J Bone Joint Surg Br* 2001;83(1):112-113. https://doi.org/10.1302/0301-620x.83b1.10260
5. Felice KJ, Donaldson JO. Lumbosacral plexopathy due to benign uterine leiomyoma. *Neurology* 1995;45(10):1943-1944. https://doi.org/10.1212/wnl.45.10.1943
6. Murphy DR, Bender MI, Green G. Uterine fibroid mimicking lumbar radiculopathy: a case report. *Spine* 2010;35(24):E1435-E1437. https://doi.org/10.1097/BRS.0b013e3181e8ab84
7. Mansour NY. Reevaluating the sciatic nerve block: another landmark for consideration. *Reg Anesth* 1993;18(5):322-323.
8. Karmakar MK, Kwok WH, Ho AM, Tsang K, Chui PT, Gin T. Ultrasound-guided sciatic nerve block: description of a new approach at the subgluteal space. *Br J Anaesth* 2007;98(3):390-395. https://doi.org/10.1093/bja/ael364
9. Shevlin S, Johnston D, Turbitt L. The sciatic nerve block. *BJA Educ* 2020;20(9):312-320. https://doi.org/10.1016/j.bjae.2020.04.004
10. Sladjana UZ, Ivan JD, Bratislav SD. Microanatomical structure of the human sciatic nerve. *Surg Radiol Anat* 2008;30(8):619-626. https://doi.org/10.1007/s00276-008-0386-6
11. Salazar-Grueso E, Roos R. Sciatic endometriosis: a treatable sensorimotor mononeuropathy. *Neurology* 1986;36(10):1360-1363. https://doi.org/10.1212/WNL.36.10.1360
12. Beaton LE, Anson BJ. The sciatic nerve and the piriformis muscle: their interrelation a possible cause of coccygodynia. *J Bone Joint Surg* 1938;20(3):686-688.
13. Al-Khodairy AW, Bovay P, Gobelet C. Sciatica in the female patient: anatomical considerations, aetiology and review of the literature. *Eur Spine J* 2007;16(6):721-731. https://doi.org/10.1007/s00586-006-0074-3
14. Duhan N. Advances in management of uterine myomas. *Front Biosci* 2013;5(1):12-22. https://doi.org/10.2741/e592
15. Petersen-Felix S, Curatolo M. Neuroplasticity: an important factor in acute and chronic pain. *Swiss Med Wkly* 2002;132(21-22):273-278. https://doi.org/10.4414/smw.2002.09913

Guidelines

Ewha Med J 2024;47(2):e31
https://doi.org/10.12771/emj.2024.e31
eISSN 2234-2591

Strengthening the Reporting of Observational Studies in Epidemiology(STROBE): 국문판 설명문서

Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration — a Korean translation

Jan P. Vandenbroucke[1], Erik von Elm[2,3], Douglas G. Altman[4], Peter C. Gøtzsche[5], Cynthia D. Mulrow[6], Stuart J. Pocock[7], Charles Poole[8], James J. Schlesselman[9], Matthias Egger[2,10] for the STROBE Initiative

[1]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
[2]Institute of Social & Preventive Medicine (ISPM), University of Bern, Bern, Switzerland
[3]Department of Medical Biometry and Medical Informatics, University Medical Centre, Freiburg, Germany
[4]Cancer Research UK/NHS Centre for Statistics in Medicine, Oxford, UK
[5]Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
[6]University of Texas Health Science Center, San Antonio, TX, USA
[7]Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, UK
[8]Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, NC, USA
[9]Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, and University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA
[10]Department of Social Medicine, University of Bristol, Bristol, UK

Received Feb 29, 2024
Revised Apr 22, 2024
Accepted Apr 24, 2024

Corresponding author
Matthias Egger
Institute of Social & Preventive Medicine (ISPM), University of Bern, Bern, Switzerland
E-mail: hc.ebinu.mpsi@eborts

의학 연구의 대부분은 관찰 연구이다. 관찰 연구의 보고는 종종 불충분한 품질을 보이기도 한다. 부실한 보고는 연구의 강점과 약점을 평가하고 연구 결과의 일반화 가능성을 평가하는 데 방해가 된다. 방법론 전문가, 연구자, 편집자 그룹은 경험적 근거와 이론적 고려 사항을 고려하여 관찰 연구 보고의 질을 개선하기 위한 역학 STROBE 권고안을 개발했다. STROBE statement은 논문의 제목, 초록, 서론, 방법, 결과 및 토론 부분에 대한 22개 항목의 체크리스트로 구성되어 있다. 18개 항목은 코호트연구, 환자 대조군 연구, 단면연구에 공통으로 적용되며, 4개 항목은 세 가지 연구설계 각각에 따라 다르다. STROBE statement는 저자에게 관찰 연구에 대한 보고를 개선하는 방법에 대한 지침을 제공하고 심사자, 편집자 및 독자가 연구를 비판적으로 평가하고 해석하는 데 도움이 된다. 이 설명 문서는 STROBE statement의 사용, 이해 및 보급을 향상 시키는 것이 목적이다. 각 체크리스트 항목의 의미와 근거가 제시되어 있다. 각 항목에 대해 하나 또는 여러 개의 출판된 예시 논문과 가능한 경우 관련 경험적 연구 및 방법론 문헌에 대한 내용이 참고사항으로 제공된다. 유용한 흐름도의 예도 포함되어 있다. 본 문서 및 관련 웹사이트(http://www.strobe-statement.org/)는 관찰 연구 보고를 개선하는 데 유용한 자료가 될 것이다.

서론

* It is a Korean translation of the Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation

합리적인 의료 행위를 위해서는 질병의 원인과 발병 기전, 진단, 예후, 치료에 대한 지식이 필요하다. 무작위 배정 임상시험(randomized trials)은 치료법 및 기타 중재에 대한 귀중한 근거를 제공한다. 그러나 임상 또는 공중보건 지식의 상당수는 관찰 연구(observational study)에서 나온 것이다[1]. 임상 전문 학술지에 게재된 연구 논문 10편 중 약 9편이 관찰 연구이다[2,3].

and elaboration. *PLoS Med* 2007 Oct 16;4(10):e297. doi: 10.1371/journal.pmed.0040297. PMID: 17941715; PMCID: PMC2020496. Translation was done under the permission of the STROBE. Korean medical terminologies are based on the English-Korean Medical Terminology 6th edition, available from: https://term.kma.org/index.asp.

Translators
Soo Young Kim

Department of Family Medicine, Kangdong Sacred Heart Hospital, Hallym University College of Medicine, Seoul, Korea

Scientific reviewer
Hyeon Chang Kim[1,2]

[1]Cardiovascular and Metabolic Diseases Etiology Research Center, Yonsei University College of Medicine, Seoul, Korea
[2]Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Korea

ORCID
Soo Young Kim
https://orcid.org/0000-0002-3205-9408
Hyeon Chang Kim
https://orcid.org/0000-0001-7867-1240

Korean proofreading
Editorial office of the *Ewha Medical Journal*

STROBE statement

관찰 연구 보고는 조사의 강점과 약점을 평가할 수 있을 만큼 상세하지 않고, 명확하지 않은 경우가 많다[4,5]. 관찰 연구 보고를 개선하기 위해, 우리는 STROBE statement 체크리스트를 개발했다. 항목은 논문의 제목, 초록, 서론, 방법, 결과 및 토론과 관련이 있다. STROBE statement는 최근 여러 저널에 발표되었다[6]. 우리의 목표는 관찰 연구에서 계획, 수행 및 발견한 내용을 명확하게 제시하는 것이다. 이 권고사항은 연구 설계나 수행에서 반드시 따라야 할 방안은 아니며, 특정 방법론을 강요하거나 발표할 때 반드시 따라야 할 통일안으로 삼지 않는다.

STROBE는 노출과 건강 결과 사이의 연관성을 조사하는 관찰연구에 대한 일반적인 보고 권장사항(reporting recommendations)을 제공한다. STROBE는 관찰 연구의 세 가지 주요 유형인 코호트(cohort study), 환자 대조군(case-control study), 단면연구(cros-sectional study)를 다룬다. 저자는 이러한 연구 설계(study design)를 설명하기 위해 다양한 용어를 사용한다. 예를 들어, '추적 조사(follow-up study)'와 '추적 연구(longitudinal study)'는 '코호트연구'의 동의어로, '유병률 연구(prevalence study)'는 '단면연구'의 동의어로 사용되기도 한다. 하지만 일반적으로 코호트, 환자 대조군, 단면연구 용어를 가장 많이 사용하므로 이 용어를 선택했다. 안타깝게도 용어가 잘못 사용되거나[7] 부정확하게 사용되는 경우가 많다[8]. 박스 1에서는 세 가지 연구 설계의 특징을 설명한다.

관찰 연구의 범위

관찰 연구는 질병의 잠재적 원인에 대한 첫 번째 힌트를 보고하는 것부터 이전에 보고된 연관성의 규모를 확인하는 것까지 다양한 목적으로 수행된다. 연구에 대한 아이디어는 임상 관찰이나 생물학적 통찰력에서 나

박스 1. STROBE에서 다루는 주요 연구 설계

코호트, 환자 대조군, 단면 설계는 특정 집단과 기간에 건강 관련 사건의 발생을 조사하는 다양한 접근 방식이다. 이러한 연구는 질병 또는 질병 완화(disease remission), 장애, 합병증, 사망, 생존, 위험 요인(risk factors) 발생 등 다양한 유형의 건강 관련 사안을 다룰 수 있다.

코호트연구에서는 연구자가 사람들을 장기간 추적한다. 연구자는 기준 시점에 사람과 노출(exposures)에 대한 정보를 수집하고 시간이 경과한 후 결과 발생을 평가한다. 연구자는 일반적으로 노출된 개인과 노출되지 않은 개인을 비교하거나 노출 범주가 다른 개인 그룹을 비교한다. 연구자는 여러 가지 다른 결과를 평가하고 추적 관찰 중 여러 시점에서 노출 및 결과 변수를 조사할 수 있다. 폐쇄형 코호트(closed cohorts, 예: 출생 코호트)는 연구 시작 시점에 정해진 수의 참가자를 등록하고 그 시점부터 정해진 종료 날짜까지 정해진 간격으로 추적한다. 개방형 코호트에서는 사람들이 서로 다른 시점(예: 한 마을의 주민)에 연구 집단에 들어오고 나가는 등 연구 집단이 역동적이다. 개방형 코호트(open cohorts)는 사망, 출생, 이동으로 인해 변화하지만 연령, 성별과 같은 변수와 관련된 인구 구성은 특히 단기간에 걸쳐 거의 일정하게 유지될 수 있다. 폐쇄형 코호트에서는 누적 발생률(cumulative incidences) (위험도[risk])과 발생률(incidence rates)을 추정할 수 있으며, 노출된 그룹과 노출되지 않은 그룹을 비교하면 위험비(risk ratio) 또는 비율비(rate ratio)를 추정할 수 있다. 공개 코호트는 발생률(incidence rates)과 비율비(rate ratios)를 추정한다.

환자대조군연구에서 연구자는 특정 질병 결과가 발생한 사람(환자)과 해당 결과가 발생하지 않은 사람(대조군) 간의 노출을 비교한다. 연구자는 기본 코호트(underlying cohort) 또는 인구의 단면을 대표하는 환자와 대조군을 수집하는 것을 목표로 한다. 이러한 인구는 지리적으로 정의할 수도 있지만, 의료 시설과 같이 더 느슨하게 정의할 수도 있다. 환자 표본은 100% 또는 가능한 환자의 일부일 수 있지만, 대조군 표본은 일반적으로 관련 결과가 없는 사람들 중 일부에 불과하다. 대조군은 환자가 발생한 코호트 또는 사람들의 집단을 나타낸다. 조사자는 환자와 대조군 사이에서 질병의 추정 원인에 노출될 확률의 비율을 계산한다(박스 7 참조). 환자와 대조군에 대한 추출 전략과 연구 대상 인구의 특성에 따라 환자대조군연구에서 얻은 오즈비(odds ratio)는 위험비(risk ratio), 비율비(rate ratio) 또는 (유병률[prevalence]) 오즈비로 해석된다[16,17]. 발표된 대부분의 환자대조군연구는 공개 코호트를 표본으로 하므로 비율비를 직접 추정할 수 있다.

단면연구에서 연구자는 노출, 위험 요인 또는 질병의 유병률을 조사하기 위해 종종 같은 시점에 표본의 모든 개인을 평가한다. 일부 단면연구는 분석적이며 노출과 질병 사이의 잠재적 인과 관계를 정량화하는 것을 목표로 한다. 이러한 연구는 노출 그룹 간의 질병 유병률을 비교하여 코호트연구처럼 분석할 수 있다. 또한 질병이 있는 그룹과 없는 그룹 간의 노출 확률을 비교하여 환자 대조군 연구처럼 분석할 수도 있다. 모든 설계에서 발생할 수 있지만 단면연구에서 특히 어려운 점은 노출과 결과의 시간 순서가 때때로 명확할 수 있지만 노출이 질병에 선행했음을 입증하는 것이다. 예를 들어 노출 변수가 선천적이거나 유전적인 연구에서는 두 가지를 동시에 측정하더라도 노출이 질병에 선행했다고 확신할 수 있다.

올 수 있다. 또한 비공식적으로 자료를 살펴보다가 추가 탐색으로 이어지는 아이디어가 떠오를 수도 있다. 수천 명의 환자를 진료한 임상의가 눈에 띄는 환자 한 명을 메모하는 것처럼, 연구자는 자료에서 특별한 것을 발견할 수 있다. 자료를 여러 각도에서 살펴보는 것은 불가능하거나 바람직하지 않을 수 있지만[9], 초기 관찰을 확인하거나 반박하기 위한 추가 연구가 필요한 경우가 많다[10]. 기존 자료는 잠재적 인과 요인에 대한 새로운 아이디어를 조사하는 데 사용될 수 있으며, 그 자체만으로 거부 또는 확증에 충분할 수 있다. 하지만 이전 보고서의 잠재적 문제를 극복하기 위해 특별히 고안된 연구가 뒤따르는 경우도 있으며, 이 경우 새로운 자료를 수집하는 목적을 위해 계획된다. 이를 통해 다양한 견해, 예를 들어 하위 그룹을 살펴보는 것의 장점이나 미리 정해진 표본 크기(pre-determined sample size)의 중요성과 같은 다양한 관점으로 이어지기도 한다. STROBE는 발견, 이에 대한 반박 또는 확증에 이르기까지 관찰 연구의 다양한 용도를 수용하려고 노력한다. 필요한 경우 특정 권장 사항이 적용되는 상황을 명시한다.

이 문서 사용 방법

이 문서는 여러 학술지[6]에 체크리스트 항목을 소개한 짧은 STROBE 논문과 연결되어 있으며, STROBE의 필수적인 부분을 구성한다. 저자의 의도는 연구를 어떻게 해야 하는지가 아니라 연구를 잘 보고하는 방법을 설명하는 것이다. 각 체크리스트 항목에 대한 자세한 설명을 제공한다. 각 설명 앞에는 적절한 예시가 제시된다. 이는 예시를 든 연구가 일률적으로 잘 보고되었거나 잘 수행되었다는 의미도 아니며, 나중에 다른 사람들에 의해 확인되었다는 의미에서 해당 연구 결과가 신뢰할 만하다는 의미도 아니다. 설명과 예시 외에도 박스 1-8에는 보충 정보가 포함되어 있다. 이는 이론적 요점에 대한 기억을 되살리거나 기술적 배경에 대한 세부 정보를 빠르게 얻고자 하는 독자를 위한 것이다. 이러한 요점을 완전히 이해하려면 인용된 교과서나 방법론 논문을 공부해야 할 수도 있다.

STROBE는 유전적 연관성 연구(genetic linkage studies), 감염병 모델링(infectious disease modelling) 또는 증례보고(case reports) 및 환자군연구(case series)와 같은 연구설계에는 적용되지 않는다[11,12]. 그렇지만 STROBE의 많은 핵심 요소가 이러한 설계에 적용되므로, 이러한 연구를 보고하는 저자는 현재의 권고안을 유용하게 활용할 수 있다. 진단 검사(diagnostic tests), 종양 표지자(tumour markers)와 유전적 연관성(genetic associations)을 구체적으로 다루는 관찰 연구 저자의 경우, STARD[13], REMARK[14], STREGA[15] 권고사항이 유용할 수 있다.

STROBE 체크리스트의 항목

이제 STROBE 체크리스트의 22개 항목(표 1)에 대해 설명하고, 각 항목에 대해 공개된 예시를 제시한다.

일부 예시는 인용문을 삭제하거나 약어를 수정하여 편집했다. 18개 항목은 세 가지 연구 설계에 모두 적용되며, 4개 항목은 설계에 따라 다르다. 별표 제시된 항목(예: 항목 8*)은 환자대조군연구에서 환자군과 대조군, 코호트 및 단면연구에서 노출군과 비노출군에 대해 정보를 별도로 제공해야 함을 나타낸다. 저자는 모든 항목을 논문의 어딘가에 언급할 것을 권장하지만, 정확한 위치나 순서를 규정하지는 않는다. 예를 들어, 여러 개의 개별 항목으로 결과를 보고하는 것에 대해 논의하지만, 저자가 단일 텍스트 파트 또는 표에서 여러 항목을 다룰 수도 있다.

항목

제목과 초록

1 (a). 제목 또는 초록에 일반적으로 사용되는 용어를 사용하여 연구 설계를 제시한다.

예시

"신발 및 부츠 제조 산업 근로자의 백혈병 발생률: 환자 대조 연구"[18].

표 1. STROBE Statement—checklist of items that should be included in reports of observational studies

	No	권고사항
제목, 초록	1	(a) 제목 또는 초록에 일반적으로 사용되는 용어를 사용하여 연구 설계를 제시한다.
		(b) 초록에 수행된 연구와 발견한 내용에 대한 유익하고 균형 잡힌 요약을 제공한다.
서론		
배경/정당성	2	연구의 과학적 배경과 근거를 제시한다.
목적	3	사전 설정한 가설을 포함해서 연구목적을 제시한다.
방법		
연구디자인	4	논문 초반부에 연구 설계의 핵심 요소를 제시한다.
세팅	5	연구 환경, 장소와 주요 일시(모집, 노출측정, 추적관찰, 자료 수집기간 등을 포함)를 제시한다.
참여자	6	(a) *코호트연구*—포함기준과 대상자참여자 선정방법을 제시한다. *환자대조군연구*—포함기준과 환자 확인과 대조군 선정 방법을 제시한다. 환자와 대조군 선택의 정당성을 제시한다. *단면연구*—포함기준과 참여자 선정방법을 제시한다.
		(b) *코호트연구*—짝지은 연구(matched studies)의 경우 짝짓기 기준과 노출군과 비노출군의 수를 제시한다 *환자대조군연구*—짝지은 연구의 경우 짝짓기 기준과 환자군 당 대조군의 수를 제시한다.
변수	7	모든 결과, 노출, 예측인자(predictors), 잠재적 교란요인(potential confounders) 및 효과변경인자(effect mdoifiers)를 명확히 정의한다. 해당되는 경우 진단 기준을 제시한다.
자료원/측정	8*	개별 변수에 대해, 자료원과 평가(측정) 방법을 제시한다. 두 군 이상이 있으면 평가방법의 비교성에 대해 기술한다.
바이어스	9	잠재적 바이어스(bias)를 해결하기 위한 노력에 대해 기술한다.
연구규모	10	연구규모(study size)가 어떻게 결정되었는지 설명한다.
양적 변수	11	분석에서 양적변수(quantitative variales)를 어떻게 다루었는지 설명한다. 그룹화한 경우, 그 이유와 방법을 기술한다.
통계적 방법	12	(a) 교란 통제 방법(control for confounding)을 포함한 모든 통계분석 방법을 기술한다.
		(b) 하위그룹과 상호작용을 평가하는 데 사용한 모든 방법을 기술한다.
		(c) 결측치(missing data)를 어떻게 다루었는지 기술한다.
		(d) *코호트연구*—추적 실패를 어떻게 다루었는지 기술한다. *환자대조군연구*—환자군과 대조군의 짝짓기 방법에 대해 설명한다. *단면연구*—해당되는 경우, 표본 추출 전략을 고려한 분석방법을 기술한다.
		(e) 모든 민감도(sensitivity) 분석에 대해 기술한다.
결과		
참여자	13*	(a) 각 단계별 인원수를 보고한다. – 잠재적 적격자 수, 적격기준을 평가한 수, 적격 확인된 수, 연구 포함된 수, 추적 완료자 수, 분석자 수
		(b) 각 단계에서 제외된 이유를 밝힌다.
		(c) 흐름도(flow diagram) 사용을 고려한다.
자료기술	14*	(a) 연구참여자의 일반적 특성(인구학적, 임상적, 사회적), 노출변수 및 잠재적 교란변수에 대한 정보를 제시한다.
		(b) 관심 변수별 자료가 결측된 사람 수를 제시한다.
		(c) *코호트연구*—추적관찰시간 정보를 요약한다(평균, 총관찰기간).
결과자료	15*	*코호트연구*—시간에 따른 결과 발생 수 혹은 요약 정보를 제시한다. *환자대조군연구*—노출 범주별 수 혹은 요약 정보를 제시한다. *단면연구*—결과 발생 수 혹은 요약정보를 제시한다.
주요 결과	16	(a) 보정하지 않은 추정치(unadjusted estimate)를 제시한다. 해당되는 경우, 교란변수를 보정한 추정치와 그 정밀도(precision, 예, 95% 신뢰구간)를 제시한다. 보정한 교란변수들과 선택한 이유를 밝힌다.
		(b) 연속변수(continuous variables)를 범주화 했으면 범주의 범위(category boundaries)를 제시한다.
		(c) 적절하다면, 상대위험도(relative risk)를 의미 있는 기간 동안의 절대위험도(absolute risk)로 변환하는 것을 고려한다.
기타 분석	17	하위그룹 및 상호작용 분석과 민감도 분석 등 기타 수행된 분석을 보고한다.
고찰		
주요 결과	18	연구 목적에 비추어 주요 결과를 요약한다.
제한점	19	잠재적 바이어스나 오류의 가능성을 고려하여 연구의 제한점을 고찰한다. 잠재적 바이어스에 대해서는 그 방향성과 크기에 대해서도 고찰한다.
해석	20	연구목적, 제한점, 유사한 연구 결과와 다른 관련 근거들을 고려하여 결과에 대한 신중한 전반적 해석을 제시한다.
일반화가능성	21	연구 결과의 일반화 가능성(generalizability, 외적타당도[external validity])에 대해 고찰한다.
다른 정보		
자금원	22	본 연구의 연구비 출처와 연구비 제공자의 역할을 보고한다. 해당되는 경우, 본 논문이 기반한 원래 연구에 대해서도 설명한다.

설명

독자가 제목이나 초록에서 사용된 디자인을 쉽게 식별할 수 있어야 한다. 연구 설계에 대해 명시적이고 일반적으로 사용되는 용어는 전자 데이터베이스에서 논문의 올바른 색인에 도움이 된다[19,20].

1 (b). 초록에 수행한 작업과 발견한 내용에 대한 유익하고 균형 잡힌 요약을 제공한다.

예시

"배경: HIV에 감염된 환자의 예상 생존율은 공중 보건의 주요 관심사이다.
목적: 일반 인구와 비교하여 HIV에 감염된 인구의 생존 기간과 연령별 사망률을 추정한다.
디자인: 인구 기반 코호트연구.
설정: 1995년부터 2005년까지 덴마크에서 치료를 받은 모든 HIV 감염인.
환자: 전국적인 덴마크 HIV 코호트연구에 참여한 각 구성원은 성별, 생년월일, 거주 지역에 따라 일반 인구에서 최대 99명과 짝짓기되었다.
측정: 저자들은 나이를 시간 척도로 하여 카플란-마이어 수명표를 계산하여 연령에 따른 생존율을 25년으로 추정했다. HIV 감염 환자와 이에 해당하는 일반 인구는 환자의 HIV 진단일로부터 사망, 이민 또는 2005년 5월 1일까지 관찰되었다.
결과: 3,990명의 HIV 감염 환자와 37만 9,872명의 일반 인구가 연구에 참여하여 2만 2,744명(중앙값, 5.8세/인)과 268만 9,287명(중앙값, 8.4년/인)의 관찰 기간을 기록했다. 참가자의 3%가 추적 조사에서 손실되었다. 25세부터의 생존기간 중앙값은 HIV 감염 환자의 경우 19.9년(95% CI, 18.5–21.3), 일반 인구의 경우 51.1년(CI, 50.9–51.5)이었다. HIV 감염 환자의 경우 2000–2005년 기간 동안 생존 기간이 32.5년(CI, 29.4–34.7)으로 증가했다. C형 간염 동시 감염이 확인된 환자(16%)를 제외한 하위 그룹에서는 같은 기간 동안 생존 중앙값이 38.9년(CI, 35.4–40.1)으로 나타났다. 일반 인구에 비해 HIV 감염 환자의 상대 사망률은 연령이 증가할수록 감소한 반면, 초과 사망률은 연령이 증가할수록 증가했다.
제한점: 관찰된 사망률은 현재 최대 관찰 기간인 10년 이후에도 적용되는 것으로 가정한다.
결론: 고강도 항레트로바이러스 치료 후기에 HIV 감염 진단을 받은 젊은이의 예상 생존 중앙값은 35년 이상이다. 그러나 일반 인구에 비해 이러한 사람들의 사망률을 더욱 낮추기 위해서는 여전히 지속적인 노력이 필요하다"[21].

설명

초록은 독자가 연구를 이해하고 논문을 읽을지 여부를 결정할 수 있도록 핵심 정보를 제공한다. 일반적인 구성 요소에는 연구 질문, 방법 및 결과에 대한 간략한 설명, 결론이 포함된다[22]. 초록은 연구의 주요 세부사항을 요약해야 하며 논문에서 제공되는 정보만 제시해야 한다. 주요 결과는 참여자 수, 연관성 추정치, 변동성 및 불확실성에 대한 적절한 측정치(예: 신뢰 구간이 있는 오즈비)를 포함하는 수치 형식으로 제시하는 것이 좋다. 노출이 결과와 유의미하게 연관되어 있거나 연관되어 있지 않다는 것만 명시하는 것은 불충분하다. 연구의 배경, 설계, 수행 및 분석과 관련된 일련의 제목은 독자가 필수 정보를 빠르게 습득하는 데 도움이 될 수 있다[23]. 많은 학술지에서는 이러한 구조화된 초록을 요구하며, 이는 구조화되지 않은 요약보다 더 높은 품질과 더 많은 정보를 제공하는 경향이 있다[24,25].

서론

서론에서는 연구를 수행한 이유와 연구 질문 및 가설을 설명해야 한다. 다른 사람들이 연구의 맥락(context)을 이해하고 현재 지식에 대한 잠재적 기여도를 판단할 수 있도록 해야 한다.

2. 배경/근거: 현재 연구의 과학적 배경과 정당성을 제시한다.

예시

"아동과 청소년의 비만 유병률 증가에 대한 우려는 아동기 비만과 성인기 심혈관 질환 위험 및 사망률 증가

사이의 잘 문서화된 연관성에 초점을 맞추고 있다. 소아 비만은 아동기와 청소년기에 상당한 사회적, 심리적 영향을 미치지만 성인기의 사회적, 사회경제적, 심리적 영향에 대해서는 알려진 바가 거의 없다. 최근의 체계적 문헌고찰에 따르면 신체적 건강 결과 외에 아동기 비만의 결과에 대한 추적 연구는 없었으며, 청소년기 비만의 사회경제적 영향에 대한 추적 연구는 단 두 건에 불과했다. Gortmaker 등은 1981년 청소년기 후반에 비만이었던 미국 여성은 과체중이 아니었던 여성에 비해 7년 후 결혼할 확률이 낮고 소득이 낮은 반면, 과체중이었던 남성은 결혼할 확률이 낮다는 것을 발견했다. Sargent 등은 1974년 16세 때 비만이었던 영국 여성(남성은 제외)이 23세 때 비만이 아닌 또래보다 소득이 7.4% 적다는 사실을 발견했다. (...) 우리는 1970년 영국 출생 코호트의 추적 자료를 사용하여 아동 비만의 성인 사회경제적, 교육적, 사회적, 심리적 결과를 조사했다"[26].

설명

이 연구의 과학적 배경은 독자들에게 중요한 맥락을 제공한다. 연구의 무대를 설정하고 연구의 초점을 설명한다. 주제에 대해 알려진 내용과 현재 지식의 격차에 대한 개요를 제공하고 연구에서 어떤 부분을 다루고 있는지 설명한다. 배경 자료에는 최근의 관련 연구와 관련 연구에 대한 체계적인 검토가 포함되어야 한다.

3. 목적: 미리 정해진 가설을 포함하여 구체적인 연구목표를 제시한다.

예시

"우리의 주요 목적은 1) 다양한 사회 경제적 배경을 가진 환자들에게 서비스를 제공하는 4개의 지역사회 기반 일차 진료 성인 의학 진료소에 내원하는 여성 환자들 사이에서 가정 폭력의 유병률을 확인하고, 2) 현재 학대받는 환자와 현재 학대받지 않는 환자 사이의 인구통계학적 및 임상적 차이를 확인하는 것이었다"[27].

설명

연구목적은 연구의 세부적인 목표이다. 잘 만들어진 목적은 모집단, 노출 및 결과, 추정할 매개변수를 명시한다. 목적은 특정 가설 또는 연구가 해결하도록 설계된 질문으로 공식화될 수 있다. 초기 발견 단계와 같이 일부 상황에서는 목적이 덜 구체적일 수 있다. 그럼에도 연구자의 의도를 명확하게 보고해야 한다. 예를 들어, 중요한 하위 그룹이나 추가 분석이 원래 연구의 목적이 아니었지만 자료 분석 중에 발생한 경우, 이에 대해 적절하게 기술해야 한다(항목 4, 17, 20 참조).

방법

방법 파트에서는 독자들이 연구의 본질적인 측면을 이해할 수 있도록 계획된 내용과 수행된 내용을 충분히 상세히 기술하고, 방법이 신뢰할 수 있고 유효한 답변을 얻는 데 적절한지 판단하고 신뢰할 수 있고 유효한 답변을 제공하는 데 적절한지 판단하고, 원래 계획에서 벗어난 부분이 합리적인지 평가할 수 있어야 한다.

4. 연구 설계: 각 논문의 초반부에 연구디자인의 핵심적인 요소를 제시한다.

예시

"우리는 환자-대조군 설계의 변형인 환자-교차 설계(case-crossover design)를 사용했는데, 이는 짧은 노출(운전자의 휴대폰 사용)이 드문 결과(충돌 사고)의 위험을 일시적으로 높힐 때 적절하다. 우리는 충돌이 발생한 것으로 추정되는 시간에 운전자가 휴대전화를 사용한 것과 다른 적절한 시간대에 같은 운전자가 휴대전화를 사용한 것을 비교했다. 운전자는 스스로 제어하기 때문에 충돌 위험에 영향을 미칠 수 있지만 단기간에 변경되지 않는 운전자의 특성을 설계에서 제어한다. 통제기간 동안의 위험과 충돌 위험이 유사해야 하므로, 위험 구간(충돌 직전 시간)의 휴대폰 활동과 전주 통제 구간(참가자가 운전 중이었지만 충돌하지 않은 동등한 시간) 동안의 휴대폰 활동을 비교했다"[28].

설명

독자가 연구의 기본 사항을 이해할 수 있도록 방법 초반(또는 서론 끝 부분)에 연구 설계의 핵심 요소를 제

시하는 것이 좋다. 예를 들어, 저자는 연구가 일정 기간 동안 추적관찰한 코호트연구였다는 점을 명시하고 코호트별 노출 상태를 설명해야 한다. 마찬가지로, 환자-대조군 설계인 경우, 환자군과 대조군과 모집단을 설명해야 한다. 단면 조사인 경우, 인구집단과 횡단면 조사 시점을 언급해야 한다. 연구가 세 가지 주요 연구 유형의 변형인 경우 명확성이 추가로 필요하다. 예를 들어, 환자대조군연구 설계의 변형 중 하나인 환자 교차 연구의 경우 예시와 같이 간결한 설명이 필요할 수 있다[28].

이러한 연구 설계 용어는 정의가 명확하지 않으므로 저자는 연구를 단순히 '전향적' 또는 '후향적'이라고 부르는 것을 자제해야 한다[29]. 일부 문헌에서 코호트와 전향적 연구를 동의어로 간주하고 환자 대조군 연구를 후향연구라고 표현한다[30]. 다른 문헌에서는 연구 아이디어가 개발된 시점을 기준으로 전향적 코호트연구와 후향적 코호트연구를 구분한다[31]. 세 번째 용법은 환자를 선정할 때 관심 있는 노출에 대한 자료가 존재했는지 여부에 따라 전향적 및 후향적 환자 대조군 연구를 구분한다[32]. 일부에서는 이러한 용어를 사용하지 않거나[33], 코호트연구를 설명할 때 '동시적' 및 '역사적'이라는 대안을 채택할 것을 권고하기도 한다[34]. STROBE에서는 전향적 및 후향적이라는 단어와 동시적 및 역사적이라는 대체 용어를 사용하지 않다. 저자가 이러한 단어를 사용할 때마다 그 의미를 정의할 것을 권장한다. 가장 중요한 것은 저자가 자료 수집이 언제 어떻게 이루어졌는지 정확히 설명하는 것이 좋다.

방법 파트의 첫 번째 부분에는 해당 보고서가 여러 연구 중 하나인지 여부를 언급할 수도 있다. 새로운 보고서가 원래의 연구 목표와 일치하는 경우, 일반적으로 이전 출판물을 참조하고 연구의 주요 특징을 간략하게 다시 설명함으로써 이를 제시한다. 그러나 연구의 목적은 시간이 지남에 따라 변화할 수도 있다.

연구자들은 주로 관리 목적으로 수집된 공식적인 생체 통계, 원래는 완전성을 위해서만 포함된 설문지의 항목, 다른 목적으로 수집된 혈액 샘플 등 원래 의도하지 않은 용도로 자료를 사용하는 경우가 많다. 예를 들어, 아스피린과 카로틴에 대한 무작위 대조 시험인 의사 건강 연구는 나중에 응고인자 V 유전자의 점 돌연변이가 정맥 혈전증 위험 증가와 관련이 있지만 심근경색이나 뇌졸중과는 관련이 없음을 입증하는 데 사용되었다[35]. 기존 자료를 이차적으로 사용하는 것은 관찰 연구의 창의적인 부분이며 반드시 결과의 신뢰도가 떨어지거나 중요도가 낮아지는 것은 아니다. 그러나 원래의 연구 목적을 간략하게 다시 설명하는 것은 독자가 연구의 맥락과 자료의 가능한 한계를 이해하는 데 도움이 될 수 있다.

5. 세팅: 모집기간, 노출, 추적관찰, 자료 수집을 포함한 세팅, 위치, 적절한 일시를 명시한다.

예시

"파시토스 코호트연구(Pasitos Cohort Study)는 1998년 4월부터 2000년 10월까지 텍사스 엘패소 카운티의 소코로와 산 엘리자리오에 있는 여성, 유아 및 아동 클리닉과 멕시코 시우다드 후아레즈에 있는 멕시코 사회보장연구소의 모자 클리닉에서 임산부를 모집했다. 등록한 코호트 아동이 출생하기 전인 기준 시점에 연구진은 어머니를 대상으로 가정 환경에 관한 인터뷰를 실시했다. 현재 진행 중인 이 코호트연구에서는 생후 6개월부터 6개월 간격으로 후속 검사를 실시하였다"[36].

설명

독자는 연구 결과의 맥락과 일반화 가능성을 평가하기 위해 환경 요인과 장소에 대한 정보가 필요하다. 환경 요인 및 치료법과 같은 노출은 시간이 지남에 따라 변할 수 있다. 또한 연구 방법도 시간이 지남에 따라 발전할 수 있다. 연구가 언제 진행되었는지, 어떤 기간 동안 참가자를 모집하고 추적 관찰했는지 아는 것은 연구를 역사적 맥락에서 파악하고 결과를 해석하는 데 중요하다.

세팅 정보에는 모집 장소 또는 출처(예: 선거인 명부, 외래 진료소, 암 등록 센터 또는 3차 진료 센터)가 포함된다. 위치에 대한 정보는 조사가 이루어진 국가, 도시, 병원 등을 나타낼 수 있다. 기간만 설명하지 말고 날짜를 명시하는 것이 좋다. 노출, 질병 발생, 모집, 추적 관찰 시작과 종료, 자료 수집 날짜는 서로 다를 수 있다. 주목할 점은 생존 분석을 사용한 종양학 저널의 132개 보고서 중 약 80%가 환자 발생 시작일과 종료일을 포함했지만 추적 관찰이 종료된 날짜를 보고한 경우는 24%에 불과했다[37].

6. 참가자

6 (a). 코호트연구: 포함기준과 참여자 선정방법을 제시한다. 후속 조치 방법을 설명한다.

예시

"아이오와 여성 건강 연구(Iowa Women's Health Study)의 참가자는 1985년 아이오와 주 자동차 운전 면허증 목록에서 추출한 55–69세의 모든 여성 중 무작위 표본으로, 해당 연령대의 아이오와 여성 중 약 94%를 차지했다. (...) 1987년 10월과 1989년 8월에 후속 설문지를 우편으로 발송하여 생체 상태와 주소 변경 사항을 평가했다. (...) 비흑색종 피부암을 제외한 발생 암은 아이오와 주 보건 등록부에서 확인했다. (...) 아이오와 여성 건강 연구 코호트는 이름, 성, 처녀 이름, 우편 번호, 생년월일 및 사회 보장 번호의 조합으로 레지스트리와 일치했다 "[38].

6 (a). *환자대조군연구*—포함기준과 환자 확인과 대조군 선정 방법을 제시한다. 환자와 대조군 선택의 정당성을 제시한다.

예시

"1999년과 2000년에 진단된 피부 흑색종 환자는 아이오와 암 등록부를 통해 확인되었다. (...) 아이오와 암 등록부를 통해 확인된 대조군은 같은 기간에 진단된 대장암 환자이다. 대장암은 흔하고 생존 기간이 비교적 길며 비소 노출이 대장암 발생과 결정적으로 연관되지 않았기 때문에 대장암 대조군을 선택했다[39].

6 (a). 단면연구: 포함기준과 참여자 선정방법을 제시한다.

예시

"국제질병분류 9차 개정판에 따른 심근경색(코드 410)을 주 진단으로 하는 환자를 퇴원 진단 코드에서 후향적으로 식별했으며, 다섯 번째 자리가 2인 코드는 후속 치료 에피소드를 지정하였다. (...) 1994년 2월부터 1995년 7월까지 심근경색이 발생한 전체 메디케어(Medicare) 코호트에서 무작위 표본이 선정되었다. (...) 적격 환자는 최소 30분 이상 12시간 미만의 흉통이 발생한 후 병원에 내원해야 했으며, 초기 심전도에서 2개의 연속된 리드에서 최소 1 mm의 ST 분절 상승이 있어야 했다"[40].

설명

연구 참여자에 대한 자세한 설명은 독자가 연구 결과의 적용 가능성을 이해하는 데 도움이 된다. 연구자는 일반적으로 적격 참여자(eligible participants)의 임상적, 인구통계학적 및 기타 특성을 정의하여 연구 모집단을 제한한다. 일반적인 자격 기준은 연령, 성별, 진단 및 동반 질환과 관련이 있다. 이러한 중요성에도 불구하고 자격 기준이 적절하게 보고되지 않는 경우가 많다. 뇌졸중 관찰 연구에 대한 조사에서 49건의 보고서 중 17건(35%)이 자격 기준을 명시하지 않았다[5].

적격성 기준은 포함 및 제외 기준으로 제시될 수 있지만, 이러한 구분이 항상 필요하거나 유용한 것은 아니다. 어쨌든 저자는 모든 자격 기준을 보고하고 연구 모집단을 선정한 그룹(예: 지역 또는 국가의 일반 인구) 및 모집 방법(예: 광고를 통한 추천 또는 자체 선정)을 설명하는 것이 좋다.

후속 조사 절차가 무응답과 후속 조사 손실을 최소화했는지, 모든 참가자에게 유사한 절차를 적용했는지 등 후속 조사 절차에 대한 세부 정보를 파악하면 결과의 타당성을 판단하는 데 도움이 된다. 예를 들어, 급성 감염을 감지하기 위해 IgM 항체를 사용한 연구에서 독자들은 혈액 검사 간격이 너무 길어 일부 감염을 놓쳤을 가능성이 있는지 판단할 수 있도록 IgM 항체 검사 간격을 알아야 했다[41]. 노출된 그룹과 노출되지 않은 그룹 간에 추적관찰 절차가 달랐던 다른 연구들에서, 독자들은 사건의 불균등한 확인 또는 추적관찰에 대한 무응답 또는 손실의 차이로 인해 상당한 바이어스를 인지할 수 있다[42]. 따라서 연구자는 참가자를 추적하는 데 사용된 방법과 그 방법이 모든 참가자에게 동일하게 적용되었는지, 변수 확인의 완전성을 설명하는 것이 좋다(14번 항목 참조).

환자대조군연구에서 환자와 대조군의 선택은 결과를 해석하는 데 매우 중요하며, 그 선택 방법은 연구의 타당성에 큰 영향을 미친다. 일반적으로 대조군은 환자가 발생한 집단에서 나와야 한다. 일반 인구에서 발생한

환자의 경우 인구 명단 추출, 무작위 전화 걸기, 이웃 또는 친구 대조군 등 장단점이 있는 다양한 방법이 대조군 표본 추출에 사용된다. 이웃 또는 친구 대조군은 노출에 내재적 짝짓기가 있을 수 있다[17]. 다른 질병을 가진 대조군은 인구 기반 대조군, 특히 병원 기반 환자의 경우 병원의 모집단을 더 잘 반영하고 리콜의 비교가능성이 높으며 모집이 용이하기 때문에 이점이 있을 수 있다. 그러나 관심 노출이 대조군 질환의 발병 또는 입원 위험에 영향을 미치는 경우 문제가 발생할 수 있다[43,44]. 이 문제를 해결하기 위해 종종 최상의 방어 가능한 대조 질병(defensible control diseases)이 사용된다[45].

6 (b). 코호트연구—짝지은 연구의 경우 짝짓기 기준과 노출군과 비노출군의 수를 제시한다.

예시

"처음에 스타틴을 투여받은 각 환자에 대해 성향 기반 짝짓기를 사용하여 다음 프로토콜에 따라 스타틴을 투여받지 않은 대조군 1명을 식별했다. 첫째, 스타틴 사용 또는 패혈증 위험과 잠재적으로 관련된 광범위한 요인 목록을 기반으로 전체 코호트의 각 환자에 대한 성향 점수를 계산했다. 둘째, 각 스타틴 사용자는 성별, 연령(1년 이상 또는 미만), 기준일(3개월 이상 또는 미만)에 따라 더 작은 규모의 비스타틴 사용자 풀에 짝짓기되었다. 셋째, 각 스타틴 사용자와 가장 가까운 성향 점수(0.2 SD 이내)를 가진 대조군을 1:1 방식으로 선택하고 나머지 대조군은 폐기했다"[46].

6 (b). 환자 대조군 연구: 짝지은 연구의 경우 짝짓기 기준과 환자 당 대조군의 수를 제시한다.

예시

"우리는 일반 진료 기록에 자폐증 또는 기타 만연성 발달 장애(other pervasive developmental disorders, PDD) 진단이 기록되어 있지 않고, 환자의 PDD 진단일에 참여 진료소에 등록되어 있고 생존해 있는 연구 집단 내 개인 중에서 모든 환자에 대해 5명의 대조군을 선정하는 것을 목표로 삼았다. 대조군은 생년월일(최대 1세 이하), 성별, 일반 진료에 따라 환자와 개별적으로 짝짓기되었다. 300건의 환자 각각에 대해 모든 짝짓기 기준을 충족하는 대조군 5명을 식별할 수 있었다. 나머지 994건의 경우 하나 이상의 대조군이 제외되었다..."[47].

설명

짝짓기는 환자대조군연구에서 훨씬 더 일반적이지만, 때때로 연구자가 코호트연구에서 짝짓기를 사용하는 경우도 있는데 그 이유는 추적 관찰을 시작할 때 그룹을 비교 가능하게 해주기 때문이다. 코호트연구에서 짝짓기를 사용하면 잠재적 교란 요인에 대해 그룹을 직접 비교할 수 있으며 환자 대조군 연구보다 복잡한 문제가 덜 발생한다. 예를 들어, 상대 위험도 추정을 위해 짝짓기를 고려할 필요가 없다[48]. 코호트연구에서의 짝짓기는 통계적 정밀도를 높일 수 있으므로 연구자는 분석에서 짝짓기를 허용하여 더 좁은 신뢰 구간을 얻을 수 있다.

환자대조군연구에서 짝짓기는 환자와 대조군 간의 변수 분포, 특히 잠재적 교란 변수의 분포의 유사성을 보장하여 연구의 효율성을 높이기 위해 수행된다[48,49]. 짝짓기는 환자당 하나 이상의 대조군을 사용하여 다양한 방법으로 수행할 수 있으므로, 짝짓기 변수 선택의 근거와 사용된 방법의 세부 사항을 설명해야 한다. 일반적으로 사용되는 짝짓기 형태는 빈도 짝짓기(그룹 짝짓기이라고도 함)과 개별 짝짓기이다. 빈도 짝짓기에서는 조사자가 대조군을 선택하여 짝짓기 변수의 분포가 환자의 분포와 동일하거나 유사하도록 한다. 개별 짝짓기는 각 환자에 하나 또는 여러 개의 대조군을 짝짓기하는 것이다. 직관적으로 매력적이고 때로는 유용하지만, 환자대조군연구에서의 짝짓기에는 여러 가지 단점이 있으며 항상 적절한 것은 아니므로 분석에서 이를 고려해야 한다(박스 2 참조).

간단한 짝짓기 절차조차 제대로 보고되지 않을 수 있다. 예를 들어, 저자는 대조군을 '5년 이내' 또는 '5년 연령대'를 사용하여 환자와 일치시켰다고 명시할 수 있다. 즉, 환자가 54세인 경우 해당 대조군은 50세에서 54세 사이 또는 54세에서 5년 이내인 49세에서 59세 사이에 있어야 한다는 의미인가? 넓은 연령대(예: 10년)를 선택하면 대조군이 평균적으로 환자보다 젊을 수 있기 때문에 연령에 따른 잔존 교란의 위험이 있다(박스 4 참조).

박스 2. 환자대조군연구에서의 짝짓기

환자대조군연구에서는 환자와 대조군의 짝짓기를 사용할지, 사용한다면 어떤 변수를 짝짓기할지, 정확한 짝짓기 방법과 적절한 통계 분석 방법을 현명하게 선택해야 한다. 전혀 일치시키지 않으면 일부 주요 잠재적 교란 변수(예: 연령, 성별)의 분포가 환자와 대조군 간에 근본적으로 다르다는 것을 의미할 수 있다. 분석에서 이를 보정할 수 있지만 통계적 효율성에 큰 손실이 있을 수 있다.

환자대조군연구에서의 짝짓기 사용과 그 해석은 특히 여러 위험 요인에 대해 짝짓기를 시도하는 경우, 그 중 일부가 주요 관심사 노출과 관련이 있을 수 있는 경우 어려움이 따른다[50,51]. 예를 들어, 잠재적 대조군으로 사용할 수 있는 수천 명의 여성에 대한 정보가 있는 대규모 약물 역학 자료 베이스에 중첩된 심근경색 및 경구 피임약에 대한 환자 대조군 연구에서 연구자는 각 심근경색 환자와 유사한 수준의 위험 요소를 가진 일치하는 대조군을 선택하고 싶을 수 있다. 한 가지 목적은 경구 피임약 처방에 영향을 미칠 수 있는 요인을 보정하여 적응증에 따른 교란을 통제하는 것이다. 그러나 심근경색 환자는 나이가 많은 경향이 있기 때문에 대조군은 더 이상 경구 피임약 복용을 대표하지 않는 대조군이 될 것이다. 이는 몇 가지 시사점이 있다. 자료를 조잡하게 분석하면 짝짓기 요인이 노출과 연관된 경우 일반적으로 단일성에 편향된 오즈비가 산출된다. 이에 대한 해결책은 짝짓기 또는 계층화 분석을 수행하는 것이다(항목 12d 참조). 또한, 짝짓기된 대조군은 전체 모집단을 대표하지 않기 때문에 대조군 간의 노출 분포를 더 이상 모집단 기여 비율을 추정하는 데 사용할 수 없다(박스 7 참조)[52]. 또한 짝짓기 요인의 효과를 더 이상 연구할 수 없으며, 잘 짝짓기된 대조군을 찾는 것이 번거로울 수 있으므로 짝짓기되지 않은 대조군을 구하기가 더 쉽고 대조군 규모가 더 클 수 있으므로 짝짓기되지 않은 대조군을 사용한 설계가 더 바람직한다. 과도한 짝짓기는 또 다른 문제로, 일치하는 환자대조군연구의 효율성을 떨어뜨리고, 바이어스를 유발할 수도 있다. 짝짓기 변수가 노출과 밀접한 관련이 있는 경우 정보가 손실되고 연구의 힘이 감소한다. 그러면 동일한 짝짓기 세트의 많은 개인이 동일하거나 유사한 수준의 노출을 갖는 경향이 있으므로 관련 정보를 제공하지 못한다. 짝짓기 변수가 교란 변수가 아니라 노출과 질병 사이의 인과 경로에 있는 경우 짝짓기는 수정할 수 없는 바이어스를 유발할 수 있다. 예를 들어, 체외 수정은 다태아 및 저체중아 출산 증가로 인한 주산기 사망 위험 증가와 관련이 있다[53]. 다태아 또는 출생 체중을 기준으로 짝짓기를 수행하면 결과가 null로 편향되며, 이는 분석에서 해결할 수 없다.

짝짓기는 직관적으로 매력적이지만 관련된 복잡성으로 인해 방법론가들은 환자 대조군 연구에서 일상적인 짝짓기를 사용하지 말 것을 권고하고 있다. 대신 각 잠재적 짝짓기 요인을 신중하게 고려하고, 짝짓기를 하지 않고 보정 변수로 측정하여 사용할 수 있다는 점을 인식할 것을 권장한다. 이에 따라 사용되는 짝짓기 요인의 수가 줄어들고, 위에서 논의한 몇 가지 문제를 피할 수 있는 빈도 짝짓기의 사용이 증가하고 있으며, 짝짓기를 전혀 사용하지 않는 환자 대조군 연구도 증가하고 있다[54]. 짝짓기는 교란 요인(예: 연령)의 분포가 일치하지 않는 비교 그룹 간에 근본적으로 다를 수 있는 경우에 가장 바람직하거나 심지어 필수적이다[48,49].

7. 변수: 적용가능 하다면, 모든 결과, 노출, 예측인자, 잠재적 교란인자, 효과변경자를 명확히 정의한다.

예시

"주요 선천성 기형만 분석에 포함되었다. 경미한 기형은 유럽 선천성 기형 등록(EUROCAT)의 제외 목록에 따라 제외되었다. 한 아이에게 한 장기 계통의 주요 선천성 기형이 두 개 이상 있는 경우, 해당 기형은 장기 계통별 분석에서 하나의 결과로 처리되었다 (...) 통계 분석에서 잠재적 교란 요인으로 고려된 요인은 분만 시 산모의 연령과 이전 동등성 수였다. 잠재적 효과 변경 요인으로 고려된 요인은 항간질제 환급 시 산모의 나이와 분만 시 산모의 나이였다"[55].

설명

저자는 결과, 노출, 예측인자, 잠재적 교란변수, 잠재적 효과 변경변수 등 분석에 고려되고 분석에 포함된 모든 변수를 정의해야 한다. 질병 결과에는 진단 기준에 대한 적절하고 상세한 설명이 필요하다. 이는 환자대조군연구의 환자군, 코호트연구의 추적 관찰 중 질병 발생, 단면연구의 유행성 질병에 대한 기준에 적용된다. 명확한 정의와 이를 준수하기 위해 취한 조치는 연구에서 주요 관심사인 질병 상태에 대해 특히 중요하다.

일부 연구의 경우 노출 변수는 '결정요인(determinant) ' 또는 '예측요인(predictor)'으로, 결과는 '평가변수(endpoints)'로 부르는 것이 적절할 수 있다. 다변량 모델에서 저자는 결과를 '종속 변수'로, 노출 및 교란 변수에 대해 '독립 변수' 또는 '설명 변수'라는 용어를 사용하는 경우가 있다. 후자는 노출과 교란 변수를 구분하지 않기 때문에 정확하지 않다.

초기 발견 단계에서 많은 변수를 측정하여 탐색 분석에 포함시킨 경우, 부록, 추가 표 또는 별도의 출판물에 각 변수에 대한 세부 정보가 포함된 목록을 제공하는 것이 좋다. 주목할 만한 점은 최근 국제 역학 저널에서 특정 연구에서 여러 시점에 측정된 항목에 대한 자세한 정보를 포함하는 '코호트 프로필' 파트를 신설했다는 점이다[56,57]. 마지막으로, 저자는 최종 모델에 포함된 변수만 선택적으로 보고하지 말고 통계 분석을 위해 고

려한 모든 '후보 변수'를 보고할 것을 권장한다(항목 16a 참조)[58,59].

8. 자료원/측정: 개별 변수에 대해, 자료원과 평가(측정) 방법을 제시한다. 두 군 이상이면 평가방법의 비교성에 대해 기술한다.

예시 1

"총 카페인 섭취량은 주로 미국 농무부 식품 성분 자료를 사용하여 계산했다. 카페인 함량은 커피 한 잔당 137 mg, 차 한 잔당 47 mg, 콜라 음료 한 캔 또는 한 병당 46 mg, 초콜릿 캔디 1개 7 mg이라고 가정했다. 이 (카페인) 섭취량 측정 방법은 NHS I 코호트와 남성 건강 전문가를 대상으로 한 유사한 코호트연구 모두에서 유효한 것으로 나타났다. (...) 자가 보고된 고혈압 진단은 NHS I 코호트에서 신뢰할 수 있는 것으로 밝혀졌다"[60].

예시 2

"짝지은 환자와 대조군 샘플은 항상 동일한 배치에서 함께 분석되었으며 실험실 직원은 환자와 대조군을 구분할 수 없었다"[61].

설명

노출, 교란 요인 및 결과를 측정하는 방식은 연구의 신뢰성과 타당성에 영향을 미친다. 측정 오류와 노출 또는 결과의 잘못된 분류는 인과 관계를 감지하기 어렵게 만들거나 가짜 관계를 생성할 수 있다. 잠재적 교란요인에 대한 측정 오류는 잔류 교란의 위험을 증가시킬 수 있다[62,63]. 따라서 평가 또는 측정의 유효성 또는 신뢰성에 대한 연구 결과를 보고할 때 사용된 참조 표준에 대한 세부 정보를 포함하면 도움이 된다. 첫 번째 예에서처럼 단순히 타당도 연구를 인용하기보다는 저자가 측정 오차 보정 또는 민감도 분석에 사용할 수 있는 추정 타당도 또는 신뢰도를 제시하는 것이 좋다(12e 및 17번 항목 참조).

또한 비교 대상 그룹이 자료 수집 방식과 관련하여 차이가 있는지 파악하는 것도 중요하다. 이는 실험실 검사(두 번째 예시에서와 같이) 등에서 중요할 수 있다. 예를 들어, 면접관이 먼저 모든 환자에 대해 질문한 다음 대조군에 대해 질문하거나 그 반대의 경우 학습 곡선으로 인해 바이어스가 발생할 수 있으며, 면접 순서를 무작위로 지정하는 등의 방법으로 이 문제를 방지할 수 있다. 비교 그룹에 동일한 진단 검사를 제공하지 않거나 한 그룹이 다른 그룹보다 같은 종류의 검사를 더 많이 받는 경우에도 정보 바이어스가 발생할 수 있다(9번 항목 참조).

9. 바이어스: 잠재적 바이어스를 다룬 노력에 대해 기술한다.

예시 1

"자살에 대한 대부분의 환자 대조 연구에서는 대조군이 살아있는 개인으로 구성되지만, 우리는 다른 원인으로 사망한 사람들로 구성된 대조군을 갖기로 했다. (...) 사망한 개인으로 구성된 대조군의 경우, 위험 요인을 평가하는 데 사용되는 정보 출처는 최근에 가족이나 가까운 동료의 죽음을 경험한 정보원이므로 살아있는 대조군을 사용할 때보다 자살 그룹의 정보 출처와 더 비슷하다"[64].

예시 2

"제2형 당뇨병이 있는 여성이 당뇨병이 없는 여성보다 더 면밀한 안과 감시를 받는 경우, 감지(detection) 바이어스가 제2형 당뇨병(T2DM)과 원발성 개방각 녹내장(POAG) 사이의 연관성에 영향을 미칠 수 있다. 당뇨병이 있는 여성과 없는 여성이 보고한 평균 안과 검진 횟수를 비교했다. 또한 보다 면밀한 안과 감시와 관련된 공변량(백내장, 황반변성, 안과 검사 횟수, 신체 검사 횟수에 대한 자가 보고)을 추가로 통제하여 POAG의 상대적 위험도를 다시 계산했다"[65].

설명

편향된 연구는 사실과 체계적으로 다른 결과를 초래한다(박스 3 참조). 독자는 바이어스의 가능성을 줄이기 위해 연구를 수행하는 동안 어떤 조치가 취해졌는지 아는 것이 중요하다. 이상적으로 연구자는 연구를 계

박스 3. 바이어스

바이어스는 연구 결과가 실제 값과 체계적으로 벗어나는 것을 말한다. 일반적으로 바이어스는 연구 설계 또는 실행 중에 발생하며 나중에 수정할 수 없다. 바이어스와 교란은 동의어가 아니다. 바이어스는 잘못된 정보 또는 참여자 선택으로 인해 잘못된 연관성이 발견될 때 발생한다. 교란은 사실적으로는 맞지만 설명되지 않은 근본적인 요인이 노출과 결과 모두와 연관되어 있기 때문에 인과적으로 해석할 수 없는 관계를 만든다(박스 5 참조). 또한 바이어스는 측정된 자료의 통계적 변동(어느 방향이든)으로 인해 발생하는 실제 값과의 편차, 즉 무작위 오류와 구별해야 한다. 가능한 많은 바이어스의 원인이 설명되어 있으며 다양한 용어가 사용된다[68,69]. 정보 바이어스와 선택 바이어스라는 두 가지 간단한 범주가 도움이 될 수 있다.

정보 바이어스는 자료의 완전성 또는 정확성의 체계적 차이로 인해 노출 또는 결과와 관련하여 개인을 차별적으로 잘못 분류할 때 발생한다. 예를 들어, 당뇨병 여성이 더 정기적이고 철저한 안과 검사를 받는다면 당뇨병이 없는 여성보다 녹내장 확인이 더 많이 일어날 것이다(항목 9 참조)[65]. 비특이적 위장 불편을 유발하는 약물을 복용하는 환자는 약물이 더 많은 궤양을 유발하지 않더라도 약물을 복용하지 않는 환자보다 위 내시경 검사를 더 자주 받고 더 많은 궤양을 발견할 수 있다. 이러한 유형의 정보 바이어스는 '탐지 바이어스' 또는 '의료 감시 바이어스'라고도 한다. 그 영향을 평가하는 한 가지 방법은 여러 연구 그룹에서 의료 감시의 강도를 측정하고 통계 분석에서 이를 보정하는 것이다. 환자대조군연구에서 정보 바이어스는 환자군에서 자가 해당 질병이 없는 대조군보다 과거 노출을 어느 정도 정확하게 회상하거나 보고할 의향이 있는 경우 발생한다('회상 바이어스'라고도 함). '면접관 바이어스'는 면접관이 연구 가설을 인지하고 무의식적 또는 의식적으로 자료를 선택적으로 수집하는 경우 발생할 수 있다[70]. 따라서 연구 참여자와 연구자를 어떤 형태로든 눈가림 처리하는 것이 유용할 수 있다.

환자대조군연구에서 환자 또는 대조군을 포함할 확률이 노출과 관련이 있는 경우 선택 바이어스가 발생할 수 있다. 예를 들어, 심부정맥 혈전증 연구를 위해 참가자를 모집하는 의사가 다리에 불편함이 있고 경구 피임약을 복용하는 여성에게서 이 질환을 진단할 수 있다. 그러나 피임약을 복용하지 않는 비슷한 증상을 가진 여성에게는 심부정맥 혈전증을 진단하지 않을 수도 있다. 이러한 바이어스는 진단 서비스에 동일한 방식으로 의뢰된 환자와 대조군을 사용하여 대응할 수 있다[71]. 마찬가지로, 질병 등록부를 사용하면 노출과 질병 사이의 가능한 관계가 알려진 경우, 의심되는 원인 물질에 노출된 환자가 등록부에 제출될 가능성이 더 높을 수 있다[72]. '응답 바이어스'는 연구에 응답한 사람과 참여를 거부한 사람 사이의 특성 차이가 유병률, 발병률, 일부 상황에서는 연관성 추정치에 영향을 미치는 경우 발생하는 또 다른 유형의 선택 바이어스이다. 일반적으로 선택 바이어스는 연구의 내부 타당도에 영향을 미친다. 이는 일반적으로 연구 참여자를 선정할 때 발생할 수 있는 문제와는 다르며, 이는 연구의 내적 타당성보다는 외적 타당성에 영향을 미친다(21번 항목 참조).

획할 때 잠재적인 바이어스의 원인을 신중하게 고려한다. 보고 단계에서 저자는 항상 관련 바이어스의 가능성을 평가할 것을 권장한다. 특히 바이어스의 방향과 규모를 논의하고 가능하면 추정해야 한다.

예를 들어, 환자대조군연구에서 정보 바이어스가 발생할 수 있지만, 첫 번째 예에서와 같이 적절한 대조군을 선택하면 바이어스를 줄일 수 있다[64]. 두 번째 예에서는 참가자 의료 감시의 차이가 문제가 되었다[65]. 따라서 이 문제를 해결하기 위해 수집한 추가 자료에 대해 더 자세히 설명해야 한다. 연구자가 추적 연구에서 변수 측정에서 발생할 수 있는 '드리프트'(drift)에 대응하거나 관찰자가 다수인 경우 변동성을 줄이기 위한 자료 수집 품질 관리 프로그램을 설정한 경우, 이를 설명해야 한다.

안타깝게도 결과를 보고할 때 중요한 바이어스를 언급하지 않는 경우가 많다. 암 병력이 있는 환자의 두 번째 암 발생 위험을 조사한 1990년부터 1994년까지 발표된 43건의 환자 대조군 및 코호트연구 중 의료 감시 바이어스가 언급된 논문은 5건에 불과했다[66]. 1998년에 3개의 정신의학 저널에 발표된 정신 건강 연구 보고서를 조사한 결과, 392개 논문 중 13%만이 응답 바이어스에 대해 언급했다[67]. 뇌졸중 연구의 코호트연구 조사에 따르면 1999년부터 2003년까지 발표된 49편의 논문 중 14편(28%)이 연구 참여자 모집 시 잠재적인 선택 바이어스에 대해 언급했으며 35편(71%)은 모든 유형의 바이어스가 결과에 영향을 미칠 수 있는 가능성을 언급했다[5].

10. 연구 규모: 어떻게 연구규모에 도달했는지를 설명한다.

예시 1

"연구 기간 동안 해당 지역의 환자 수가 표본 크기를 결정했다"[73].

예시 2

"해당 지역의 산후 우울증에 대한 설문 조사에서 19.8%의 유병률을 기록했다. 우울증이 있다고 가정할 때

정상 체중의 자녀를 둔 산모의 우울증을 20%로 가정하고 영양실조 아동을 둔 산모의 우울증에 대한 오즈비를 3으로 가정하면 80%의 검정력과 5%의 유의도를 가진 72개의 환자-대조군 세트(한 환자 당 한 명의 대조군)가 필요하다."[74].

설명

연구는 연구 질문에 의미 있게 답하기 위해 충분히 좁은 신뢰 구간으로 점 추정치를 얻을 수 있을 만큼 충분히 커야 한다. 작은 연관성과 연관성이 없는 것을 구별하려면 큰 표본이 필요하다. 소규모 연구는 종종 가치 있는 정보를 제공하지만, 신뢰 구간이 넓으면 신뢰 구간이 좁은 추정치를 제공하는 연구와 비교하여 현재 지식에 기여하는 바가 적을 수 있다. 또한 '흥미로운' 또는 '통계적으로 유의미한' 연관성을 보여주는 소규모 연구가 '유의미한' 결과가 없는 소규모 연구보다 더 자주 발표된다. 이러한 연구는 발견의 맥락에서 초기 신호를 제공할 수 있지만, 독자에게 잠재적인 약점에 대해 알려야 한다.

관찰 연구에서 표본 크기 결정의 중요성은 상황에 따라 다르다. 다른 목적으로 이미 사용 가능한 자료를 분석하는 경우, 자료 분석이 문헌에 실질적으로 기여할 수 있는 충분한 통계적 정밀도를 가진 결과를 산출할 수 있는지 여부가 주요 질문이며, 표본 크기 고려는 비공식적으로 이루어질 것이다. 새로운 연구를 계획할 때 공식적이고 선험적인 표본 크기 계산이 유용할 수 있다[75,76]. 이러한 계산은 일반적으로 생성되는 단일 숫자가 암시하는 것보다 더 많은 불확실성과 관련이 있다. 예를 들어, 관심 사건의 비율 또는 계산의 중심이 되는 기타 가정에 대한 추정치는 일반적으로 추측이 아니라면 부정확하다[77]. 최종 분석에서 얻은 정밀도는 다변량 분석에서 교란 변수를 포함하거나[78], 주요 변수를 측정할 수 있는 정밀도[79], 일부 개인을 제외함으로써 감소하기 때문에 사전에 결정할 수 없는 경우가 많다.

역학 연구에서 표본 크기에 대한 내용을 자세히 설명한 경우는 거의 없다[4,5]. 조사자가 적절한 공식 표본 크기 계산을 수행한 경우 이를 보고할 것을 권장한다. 다른 상황에서는 연구 규모를 결정한 고려 사항(예: 위의 첫 번째 예에서와 같이 사용 가능한 표본이 고정되어 있음)을 명시해야 한다. 통계적 유의성에 도달하기 전에 관찰 연구를 조기에 중단한 경우, 독자에게 이 사실을 알려야 한다. 연구 규모 또는 후향적 검정력 계산에 대한 사후 정당화로 독자를 귀찮게 하지 말아야 한다[77]. 독자의 관점에서 신뢰 구간은 궁극적으로 얻은 통계적 정밀도를 나타낸다. 신뢰 구간은 통계적 불확실성에만 영향을 미치며 연구에 존재할 수 있는 모든 불확실성을 반영하지는 않는다는 점을 인식해야 한다(20번 항목 참조).

11. 양적 변수: 분석에서 양적변수를 어떻게 다루었는지 설명한다. 가능하다면 어떤 그룹화를 선택했고 그 이유가 무엇인지 기술한다.

예시

"글래스고 코마 척도가 8 미만인 환자는 심각한 부상을 입은 것으로 간주되고, 9 이상이면 뇌 손상이 덜 심각한 것으로 간주된다. 이 두 범주의 GCS와 부상 후 12개월 이내 사망 발생의 연관성을 조사했다."[80].

설명

연구자는 노출, 효과 변경인자 및 교란 요인에 대한 정량적 자료를 수집하고 분석하는 방법을 선택한다. 예를 들어, 연속 노출 변수를 그룹화하여 새로운 범주형 변수를 만들 수 있다(박스 4 참조). 그룹화 선택은 이후 분석에 중요한 영향을 미칠 수 있다[81,82]. 범주 수, 분절점, 범주 평균 또는 중앙값 등 정량 자료를 그룹화한 이유와 방법을 설명하는 것이 좋다. 자료를 표 형식으로 보고할 때는 각 범주별로 환자 수, 대조군 수, 고위험군 수, 위험에 노출된 시간 등을 제시해야 한다. 표는 효과 측정 추정치 또는 모델 피팅 결과만으로 구성되어서는 안된다.

조사자는 모든 정보를 유지하기 위해 노출을 연속적인 것으로 모델링할 수 있다. 모든 정보를 유지하기 위해 이러한 선택을 할 때는 노출과 결과의 관계의 특성을 고려해야 한다. 선형 관계를 자동으로 가정하는 것은 잘못된 것일 수 있으므로 선형성에서 벗어날 수 있는 가능성을 조사해야 한다. 작성자는 분석 중에 탐색한 대체 모델(예: 로그 변환, 이차 항[quadratic terms] 또는 스플라인 함수[spline functions) 사용)을 언급할 수 있다. 노출과 결과 사이의 비선형 관계를 추정하는 방법에는 여러 가지가 있다[82–84]. 또한 주요 관심사인

박스 4. 그룹화

연속 자료를 그룹화할 수 있는 몇 가지 이유가 있다[86]. 자료를 수집할 때 몇 년에 걸친 회상을 기반으로 노출을 인위적으로 연속 측정값을 구하는 것보다 서수 변수를 사용하는 것이 더 나을 수 있다. 범주는 모든 변수를 유사한 스타일로 제시하거나 용량-반응 관계를 제시하는 등 프레젠테이션에도 유용할 수 있다.

그룹화는 분석을 단순화하기 위해(예: 선형성 가정을 피하기 위해) 수행될 수도 있다. 그러나 그룹화하면 정보가 손실되고 통계적 검정력이 떨어질 수 있다[87]. 특히 이분화를 사용할 경우 더욱 그렇다[82,85,88]. 연속적인 교란 변수를 그룹화할 경우, 변수의 교란 효과 중 일부가 보정되지 않은 채로 남아있는 잔존 교란이 발생할 수 있다(박스 5 참조)[62,89]. 범주 수를 늘리면 검정력 손실과 잔존 교란을 줄일 수 있으며, 특히 대규모 연구에 적합하다. 소규모 연구에서는 제한된 수로 인해 그룹을 거의 사용할 수 없다.

연구자는 진단 또는 예후와 관련이 있거나 실용성 또는 통계적 근거에 따라 일반적으로 사용되는 값을 기준으로 그룹화를 위한 분절점을 선택할 수 있다. 사분위수를 사용하여 각 그룹에 동일한 수의 개인을 선택할 수도 있다[90]. 반면에 더 극단적인 외부 그룹을 선택하고 중간 그룹을 외부 그룹보다 크게 설정하여 결과와의 연관성에 대한 더 많은 통찰력을 얻을 수 있다[91]. 환자대조군연구의 경우, 대조군에서 분포를 도출하는 것이 모집단을 반영하기 위한 것이므로 선호된다. 여러 대안 중에서 사후에 분절점를 선택하는 경우 독자에게 알려야 한다. 특히, P값을 최소화하기 위해 분절점를 선택한 경우 실제 연관성의 강도가 과장될 수 있다[81].

그룹화된 변수를 분석할 때는 변수의 기본 연속성을 인식하는 것이 중요하다. 예를 들어, 정렬된 그룹에 걸쳐 위험의 가능한 추세를 조사할 수 있다. 일반적인 접근 방식은 그룹의 순위를 연속 변수로 모델링하는 것이다. 그룹 점수의 이러한 선형성은 그룹 간 간격이 동일한 경우(예: 10세 연령 그룹) 실제 선형 관계에 가까워지지만 그렇지 않은 경우에는 그렇지 않다. Il'yasova 등은[92] 메타분석을 용이하게 하고 용량-반응에 대한 본질적으로 가치 있는 정보를 제공하기 위해 표준 오차와 함께 범주형 및 연속형 효과 추정치를 모두 공개할 것을 권장한다. 한 분석이 다른 분석에 정보를 제공할 수 있으며 어느 쪽도 가정이 없는 것은 아니다. 저자는 종종 순서를 무시하고 참조 범주와 비교하여 각 범주에 대한 추정치(및 P값)를 개별적으로 고려한다. 이 방법은 설명에는 유용할 수 있지만 그룹 간 위험의 실제 추세를 감지하지 못할 수 있다. 추세가 관찰되는 경우 기울기에 대한 신뢰 구간은 관찰의 강도를 나타낼 수 있다.

정량적 노출에 대해 연속 분석과 그룹 분석을 모두 제시하는 것이 유익할 수 있다.

최근 조사에 따르면, 역학 연구 논문의 3분의 2가 정량적 노출 변수를 연구했다[4]. 50개 논문 중 42개(84%) 논문에서 노출은 몇 가지 정렬된 범주로 분류되었지만, 선택에 대한 근거가 명시되지 않은 경우가 많았다. 15개의 논문에서 선형 연관성을 사용하여 연속 노출을 모델링했지만, 선형성을 확인했다고 보고한 논문은 2개에 불과했다. 또 다른 조사에서는 심리학 문헌 중 이분화가 정당화된 논문은 110개 중 22개(20%)에 불과했다[85].

12. 통계적 방법:

12 (a). 교란 통제 방법을 포함해서 모든 통계적 방법을 밝힌다.

예시

"비교 그룹에 연령 또는 성별에 따른 교란이 존재하는지 평가하기 위해 만텔-헨젤 기법을 사용하여 보정 상대 위험도를 계산했다. 95% 신뢰 구간은 Greenland과 Robins 등에 따른 분산값으로 보정 상대위험도를 계산하였다"[93].

설명

일반적으로 올바른 통계 분석은 하나만 있는 것이 아니며, 같은 질문을 다루지만 서로 다른 가정을 전제로 하는 여러 가지 가능성이 존재한다고 할 수 있다. 그럼에도 불구하고, 연구 프로토콜에서 최소한 주요 연구 목표에 대한 분석 방법을 미리 결정해야 한다. 종종 원래 예상했던 분석 대신 또는 추가 분석이 필요한 경우가 있으며, 이러한 분석은 자료의 내용에 따라 이루어진다. 연구를 보고할 때 저자는 독자에게 특정 분석이 자료를 보고하게 되었는지 여부를 알려야 한다. 사전 지정 분석과 탐색적 분석의 구분이 때때로 모호할 수 있지만, 저자는 특정 분석에 대한 이유를 명확히 밝혀야 한다.

비교 대상 그룹이 일부 특성과 관련하여 유사하지 않은 경우, 계층화 또는 다변량 회귀분석을 통해 가능한 교란 변수를 보정해야 한다(박스 5 참조)[94]. 종종 연구 설계에 따라 어떤 유형의 회귀 분석을 선택할지 결정된다. 예를 들어, 코호트연구에서는 콕스 비례 위험 회귀가 일반적으로 사용되는 반면[95], 환자 대조군 연구

박스 5. 교란

교란은 말 그대로 효과의 혼란을 의미한다. 어떤 연구에서 노출과 질병 위험 사이에 연관성이 있거나 연관성이 없는 것처럼 보일 수 있다. 연관성이 있거나 없는 것처럼 보이는 것은 실제로 그럴 수도 있지만, 노출과 연관된 다른 요인이 있기 때문일 수 있다. 이러한 다른 요인을 교란 요인 또는 교란자(confounder)라고 한다. 따라서 교란 요인은 노출의 잠재적인 '인과적' 연관성에 대한 잘못된 평가를 제공한다. 예를 들어, 중년이 되어 혈압이 높아지는 여성이 경구 피임약을 덜 처방받는 경우, 피임약을 사용하는 여성과 그렇지 않은 여성의 심혈관 질환 발생 빈도를 단순 비교하면 피임약이 심장 질환을 예방한다는 잘못된 인상을 줄 수 있다.

잠재적인 교란 요인에 대해 미리 고려해야 한다. 이를 통해 연구 설계에 정보를 제공하고 자세한 정보를 찾아야 하는 교란 요인을 식별하여 적절한 자료를 수집할 수 있다. 제한 또는 짝짓기를 사용할 수 있다. 위의 예에서 교란 요인인 혈압이 높지 않은 여성으로 제한할 수 있다. 혈압을 기준으로 짝짓기하는 것도 가능하지만 반드시 바람직한 것은 아니다(박스 2 참조). 분석 단계에서 연구자는 층화 또는 다변량 분석을 사용하여 교란 변수의 영향을 줄일 수 있다. 층화 분석은 자료를 교란 요인에 대한 계층(예: 혈압 계층)으로 나누고, 각 계층 내에서 연관성 추정치를 평가하고, 모든 계층에 대한 가중 평균으로 결합된 연관성 추정치를 계산하는 것으로 구성된다. 다변량 분석은 동일한 결과를 얻을 수 있지만 더 많은 변수를 동시에 고려할 수 있다. 다변량 분석은 더 유연하지만 노출과 질병 사이의 관계에 대한 수학적 형태에 대한 추가적인 가정을 포함할 수 있다.

관찰 연구에서는 교란변수를 고려하는 것이 중요하지만, 교란변수를 보정한 분석이 연관성의 '인과적 부분'을 확립한다고 가정해서는 안 된다. 잔존 교란(교란을 통제하려는 시도가 실패한 후에도 남아있는 교란[102]), 무작위 추출 오류, 선택 바이어스, 정보 바이어스로 인해 결과가 여전히 왜곡될 수 있다(박스 3 참조).

에서는 로지스틱 회귀가 자주 선택되는 방법이다[96,97]. 최종 모형의 결과만 제시할 것이 아니라 변수 선택에 대한 구체적인 절차를 충분히 설명해야 한다[98,99]. 최종 모델에 포함할 잠재적 교란변수 목록을 좁히기 위해 모델을 비교하는 경우, 이 과정을 설명해야 한다. 하나 또는 두 개의 공변량이 자료 분석에서 명백한 교란의 상당 부분을 차지하는지 독자에게 알려주는 것이 도움이 된다. 결측값 대체 절차, 자료 변환, 기여 위험 계산과 같은 기타 통계 분석에 대해서도 설명해야 한다. 표준이 아니거나 새로운 접근법을 참조하고 사용된 통계 소프트웨어를 보고해야 한다. 기본 원칙으로, 통계적 방법은 '원본 자료에 접근할 수 있는 지식이 있는 독자가 보고된 결과를 확인할 수 있도록 충분히 상세하게 설명하는 것이 좋다'[100].

한 실증 연구에서 교란 보정을 보고한 169개 논문 중 93개(55%)만이 연속형 및 다범주 변수를 통계 모델에 입력하는 방법을 명확하게 명시했다[101]. 또 다른 연구에 따르면 교란 변수에 대한 통계 분석이 보정된 67건의 논문에서 교란 변수가 어떻게 선택되었는지 대부분 불분명했다[4].

12 (b). 하위그룹과 교호성을 평가한 모든 방법에 대해 기술한다.

예시

"연구된 3가지 생활습관 관련 위험 요인에 대한 감수성의 성별 차이는 Rothman에 따라 생물학적 교호성을 테스트하여 탐색했다. 4가지 범주 $\left(a^{-}b^{-}, a^{-}b^{+}, a^{+}b^{-}, \text{and } a^{+}b^{+}\right)$의 새로운 복합 변수를 성별과 관심 있는 이분법적 노출에 대해 재정의했다. a–, b–는 노출이 없음을 나타낸다. 연령에 대한 보정 후 범주별로 상대위험도를 계산하였다. 교호성 효과는 절대 효과의 가산성(additivity)에서 벗어난 것으로 정의하고, 교호성으로 인한 초과 RR(RERI)을 계산했다:

$$\text{RERI} = \text{RR}\left(a^{+}b^{+}\right) - \text{RR}\left(a^{-}b^{+}\right) - \text{RR}\left(a^{+}b^{-}\right) - 1$$

여기서 RR(a+b+)는 두 요인에 모두 노출된 사람의 상대위험도이며, RR(a–b–)는 참조범주였다(RR=1.0). 95% CI는 호스머와 르메쇼가 제안한 대로 계산했다. RERI가 0이면 교호성이 없음을 의미한다[103].

설명

17번 항목에서 자세히 설명한 것처럼, 연구 모집단의 하위 그룹에 국한된 분석의 사용과 가치에 대해 많은 논쟁이 있다[4,104]. 그럼에도 불구하고 하위 그룹 분석은 종종 수행된다[4]. 독자는 어떤 하위 그룹 분석이 사전에 계획되었고, 어떤 하위 그룹 분석이 자료를 분석하는 동안 발생했는지 알아야 한다. 또한 그룹 간에 효

과 또는 연관성이 다른지 여부를 조사하기 위해 어떤 방법을 사용했는지 설명하는 것도 중요하다(17번 항목 참조).

교호성은 한 요인이 다른 요인의 효과를 조절하는 상황과 관련이 있다(따라서 '효과 변경'이라고도 함). 두 요인의 공동 작용은 두 가지 방식으로 특성화할 수 있다. 가산 척도에서는 위험 차이로, 승수 척도에서는 상대 위험도로 특성화할 수 있다(박스 8 참조).

12 (c). 결측치를 어떻게 다루었는지 기술한다.

예시

"결측 자료 분석 절차에서는 무작위 결측(MAR) 가정을 사용했다. STATA에서 다중 다변량 결측값 대체의 MICE(연쇄 방정식에 의한 다변량 결측값 대체) 방법을 사용했다. 다변량 로지스틱 회귀 분석에서 결측값이 적절히 결측값 대체된 자료 사본 10개를 각각 독립적으로 분석했다. 변수 추정치의 평균을 구하여 단일 평균 추정치를 구하고 루빈의 규칙에 따라 표준 오차를 보정했다"[106].

설명

자료 결측은 관찰 연구에서 흔히 발생한다. 연구 참여자에게 보내는 설문지가 항상 완전하게 작성되는 것은 아니며, 참여자가 모든 후속 방문에 참석하지 않을 수도 있고, 일상적인 자료원과 임상 데이터베이스가 불완전한 경우가 많다. 자료 결측의 보편성과 중요성에도 불구하고 자료 결측 문제에 대해 자세히 보고한 논문은 거의 없다[5,107]. 연구자는 결측된 자료를 해결하기 위해 여러 접근법 중 하나를 사용할 수 있다. 박스 6에서 다양한 접근법의 장점과 한계를 설명한다. 저자는 각 관심 변수(노출, 결과, 교란 요인)와 분석의 각 단계별로 결측된 값의 수를 보고할 것을 권장한다. 저자는 가능한 경우 결측값에 대한 이유를 제시하고, 연구 참여자의 흐름을 설명할 때 자료 결측으로 인해 제외된 개인 수를 명시해야 한다(13번 항목 참조). 결측된 자료를 설명하는 분석의 경우, 저자는 분석의 특성(예: 다중 결측값 대체)과 가정(예: 무작위 결측, 박스 6 참조)을 설명해야 한다.

박스 6. 결측된 자료: 문제점 및 가능한 해결 방법

결측된 자료를 처리하는 일반적인 접근 방식은 특정 분석에 필요한 모든 변수에 대한 완전한 자료를 가진 개인으로 분석을 제한하는 것이다. 이러한 '완전한 환자' 분석은 많은 상황에서 편향되지 않지만, 편향될 수 있으며 항상 비효율적이다[108]. 자료가 결측된 개인이 전체 표본의 전형이 아닐 경우 바이어스가 발생한다. 비효율성은 분석을 위한 표본 크기가 줄어들기 때문에 발생한다.

반복 측정에 마지막 관측치를 이월하여 사용하면 결과의 예고를 경험한 사람이 선택적으로 탈락할 경우 시간이 지남에 따라 추세가 왜곡될 수 있다[109]. 교란 요인에 대해 결측된 범주 지표를 삽입하면 잔존 교란이 증가할 수 있다[107]. 각 결측값을 가정 또는 추정값으로 대체하는 추정은 관심 있는 연관성을 약화시키거나 과장할 수 있으며, 아래에 설명된 정교한 방법을 사용하지 않으면 표준 오차가 너무 작아질 수 있다.

루빈은 관측값이 결측될 확률에 대한 모델을 기반으로 결측 자료 문제의 유형학을 개발했다[108,110]. 특정 관측값이 결측될 확률이 관측 가능한 변수의 값에 의존하지 않는 경우 자료는 완전 무작위 결측(missing completely at random[MCAR])으로 설명된다. 관찰된 자료가 주어졌을 때 관찰이 결측될 확률이 결측된 자료의 실제 값과 무관한 경우 자료는 무작위 결측(missing at random[MAR])이라고 한다. 예를 들어, 어린 아이일수록 폐활량 측정값이 결측되기 쉽지만, 연령을 고려한 후 결측 확률이 실제 관찰되지 않은 폐 기능과는 무관하다고 가정한다. 그러면 연령을 포함한 모델에서 결측된 폐 기능 측정값은 MAR이 된다. 사용 가능한 자료를 고려한 후에도 결측 확률이 여전히 결측된 값에 의존하는 경우 자료가 무작위가 아닌 결측(missing not at random[MNAR])이다. 자료가 MNAR인 경우 유효한 추론을 위해서는 자료 결측을 초래한 메커니즘에 대한 명시적인 가정이 필요하다.

무작위 자료 결측(MAR)을 처리하는 방법은 크게 세 가지로 분류된다[108,111]: 확률 기반 접근법(likelihood-based approaches)[112], 가중치 추정(weighted estimation)[113], 다중 결측값 대체(multiple imputation)[111,114]. 이 세 가지 접근법 중 다중 결측값 대체는 가장 일반적으로 사용되며 특히 여러 변수에 결측값이 있는 경우 유연성이 뛰어나다[115]. 이러한 접근법을 사용한 결과는 전체 환자 분석의 결과와 비교하고 중요한 차이점을 논의해야 한다. 결측 자료 분석에서 가정한 타당성은 일반적으로 검증할 수 없다. 특히 자료가 MNAR이 아니라 MAR이라는 것을 증명하는 것은 불가능하다. 따라서 이러한 분석은 민감도 분석의 정신으로 보는 것이 가장 좋다(항목 12e 및 17 참조).

12 (d). *코호트연구*—추적관찰 소실을 어떻게 다루었는지 기술한다.

예시

"적극적인 추적관찰을 시행한 치료 프로그램에서 추적관찰을 중단한 환자와 1년간 추적관찰을 시행한 환자의 기저 CD4 세포 수(는 비슷했지만 중앙값 115 세포/ℓ vs 123 세포/ℓ), 적극적인 추적관찰을 시행하지 않은 프로그램에서 추적관찰을 중단한 환자의 CD4 세포 수는 추적관찰을 시행한 환자보다 유의하게 낮았다(중앙값 64 세포/ℓ vs 123 세포/ℓ). (...) 소극적 추적 관찰이 포함된 치료 프로그램은 후속 분석에서 제외되었다"[116].

설명

코호트연구는 개인별 추적 관찰 시간과 관심 질병 발병 시간을 기반으로 하는 생명표 방법이나 기타 접근법을 사용하여 분석한다. 관찰 기간이 종결된 시점에서 질병이 없는 개인 중 추적 관찰 시간은 결과 발생 확률과 관련이 없는 것으로 가정한다. 이는 추적 관찰이 정해진 날짜 또는 특정 연령에 종료되는 경우에 해당된다. 참가자가 해당 날짜 이전에 연구를 철회하면 추적조사 손실이 발생한다. 추적조사 손실이 있는 개인 또는 질병 발병 위험이 높은 사람에게서 선택적으로 발생하는 경우 연구의 유효성을 저해할 수 있다('정보 중도절단', 'informative censoring'). 위의 예에서, 적극적인 추적 관찰이 이루어지지 않은 치료 프로그램에서 추적 관찰이 중단된 환자는 관찰 중인 환자보다 CD4 헬퍼 세포 수가 적었으므로 사망 위험이 더 높았다[116].

연구 종료에 도달한 사람과 추적 관찰에서 탈락한 사람을 구별하는 것이 중요하다. 안타깝게도 통계 소프트웨어는 일반적으로 두 가지 상황을 구분하지 못하며, 두 경우 모두 관찰 기간이 끝나면 추적 관찰 시간이 자동으로 잘린다('중도절단'). 따라서 연구자는 연구 계획 단계에서 추적 조사 손실에 어떻게 대처할지 결정해야 한다.

소수의 환자가 손실된 경우, 연구자는 추적 관찰이 불완전한 개인을 제외하거나 추적 관찰 손실 날짜 또는 연구 종료 시점에 살아있는 상태에서 중도 포기한 것으로 처리할 수 있다. 저자는 추적 관찰에서 손실된 환자 수와 어떤 중도절단 전략을 사용했는지 보고할 것을 권장한다.

12 (d). 환자대조군 연구: 환자군과 대조군의 짝짓기 방법에 대해 설명한다.

예시

"맥네마 검사, 쌍검정, 조건부 로지스틱 회귀 분석을 사용하여 치매 환자와 대조군을 비교하여 심혈관 위험요인, 자발적 뇌색전 발생, 경동맥 질환, 정맥 대 동맥 순환 단락에 대해 비교했다"[117].

설명

개별적으로 짝짓기된 환자대조군연구에서 짝짓기를 무시하고 확률 비율을 조잡하게 분석하면 일반적으로 단일성(unity)에 편향된 추정치가 도출된다(박스 2 참조). 따라서 짝짓기 분석이 필요한 경우가 많다. 이는 직관적으로 계층 분석으로 이해할 수 있는데, 각 환자는 일치하는 대조군 집합을 가진 하나의 계층으로 간주된다. 이 분석은 일치하는 변수가 비슷함에도 불구하고 환자가 대조군보다 더 자주 노출되는지 여부를 고려하는 데 중점을 둔다. 조사자는 '일치하는' 2×2 테이블에서 맨텔-헨첼 방법을 사용하여 이러한 계층 분석을 수행할 수 있다. 가장 간단한 형태로 확률 비율은 노출 변수에 대해 불일치하는 쌍의 비율이 된다. 연령 및 성별과 같이 보편적인 속성을 가진 변수에 대해 짝짓기가 수행된 경우, 분석에서 개별적인 개인 간 짝짓기를 유지할 필요는 없으며 연령 및 성별 범주의 간단한 분석으로 충분하다[50]. 그러나 이웃, 형제 관계 또는 친구 관계와 같은 다른 짝짓기 변수의 경우, 짝짓기된 각 집합을 고유한 계층으로 간주해야 한다. 개별 짝짓기 연구에서 가장 널리 사용되는 분석 방법은 조건부 로지스틱 회귀 분석으로, 각 환자와 그 대조군을 함께 고려한다. 조건부 방법은 환자마다 대조군의 수가 다르거나 일치하는 변수 외에 다른 변수를 보정해야 할 때 필요하다. 독자가 분석에서 짝짓기 설계가 적절하게 고려되었는지 판단할 수 있도록 작성자는 자료 분석에 사용된 통계적 방법을 자세히 설명하는 것이 좋다. 짝짓기를 고려해도 추정치에 거의 영향을 미치지 않는 경우, 저자는 짝짓기되지 않은 분석을 제시할 수 있다.

12 (d). 단면연구: 해당되는 경우, 표본추출 전략을 고려한 분석 방법을 설명한다.

예시

"복잡한 표본 설계에 기반한 추정치의 표본 오차를 추정하기 위해 테일러 확장 방법을 사용하여 표준 오차(SE)를 계산했다. (...) 이완기 혈압에 대한 전체 설계 효과는 남성 1.9, 여성 1.8로 나타났으며 수축기 혈압의 경우 남성 1.9, 여성 2.0으로 나타났다"[118].

설명

대부분의 단면연구는 미리 지정된 추출 전략을 사용하여 모집단에서 참가자를 선택한다. 그러나 추출은 단순한 무작위 표본 추출보다 더 복잡할 수 있다. 여기에는 여러 단계와 참여자의 지역(예: 지역 또는 마을)별 군집 전략이 포함될 수 있다. 비례 계층화는 특정 특성을 가진 하위 그룹이 올바르게 대표되도록 할 수 있다. 불균형 계층화는 특정 관심 하위 그룹을 과도하게 추출하는 데 유용할 수 있다.

복잡한 표본에서 도출된 연관성 추정치는 단순한 무작위 표본에서 도출된 추정치보다 정확도가 다소 떨어질 수 있다. 표준 오차 또는 신뢰 구간과 같은 정밀도 측정은 단순한 무작위 추출 대신 더 복잡한 추출 전략을 사용할 경우 얼마나 많은 정밀도를 얻거나 잃는지를 설명하는 비율 측정치인 설계 효과를 사용하여 보정해야 한다[119]. 대부분의 복잡한 추출 기법은 정밀도를 감소시켜 1보다 큰 설계 효과를 초래한다.

저자는 복잡한 추출 전략을 보정하는 데 사용된 방법을 명확하게 명시하여 독자가 선택한 추출 방법이 얻은 추정치의 정밀도에 어떤 영향을 미치는지 이해할 수 있도록 하는 것이 좋다. 예를 들어 클러스터 추출의 경우, 설계 효과를 보고하면 자료 수집의 용이성과 정밀도 손실 사이의 암묵적인 절충점을 투명하게 알 수 있다. 이 예에서 남성의 설계 효과가 1.9로 계산된 것은 결과 추정치가 동일한 정밀도를 가지려면 실제 표본 크기가 단순 무작위 표본 추출보다 1.9배 더 커야 함을 나타낸다.

12 (e). 모든 민감도 분석에 대해 기술한다.

예시

"자료가 불충분한 사망 환자(38/148 25.7%)의 비율이 생존 환자(15/437 3.4%)에 비해 상대적으로 높았기 때문에(...), 이로 인해 결과가 편향되었을 가능성이 있다. 따라서 민감도 분석을 수행했다. 연구 그룹에서 경구 피임약을 사용하는 여성의 비율이 전체(사망자 19.1%, 생존자 11.4%)에 적용된다고 가정한 다음, 노출된 결측된 환자가 모두 2세대 피임약을 사용하거나 모두 3세대 피임약을 사용한다는 두 가지 극단적인 시나리오를 적용했다"[120].

설명

민감도 분석은 주요 결과가 대체 분석 전략 또는 가정을 통해 얻은 결과와 일치하는지 여부를 조사하는 데 유용하다[121]. 검토할 수 있는 문제에는 분석에 포함할 기준, 노출 또는 결과의 정의[122], 보정이 필요한 교란 변수, 결측 자료 처리[120,123], 노출, 질병 및 기타 변수의 부정확하거나 일관되지 않은 측정으로 인한 선택 바이어스 또는 바이어스 가능성, 양적 변수 처리와 같은 특정 분석 선택(항목 11 참조) 등이 있다. 여러 바이어스나 가정의 영향을 동시에 모델링하기 위해 정교화된 방법이 점점 더 많이 사용되고 있다[124–126].

1959년 CornField 등은 흡연에 따른 폐암의 상대위험도 9가 비흡연자보다 흡연자에서 최소 9배 이상 유병률이 높아야 하기 때문에 생각할 수 있는 교란 요인으로 인한 결과일 가능성은 극히 낮다는 것을 보여주었다[127]. 이 분석은 그러한 요인이 존재할 가능성을 배제하지는 않았지만, 그러한 요인이 존재해야 하는 유병률을 확인했다. 최근 소아 백혈병과 송전선 근처 거주 사이의 연관성을 설명할 수 있는 그럴듯한 교란 요인을 식별하기 위해 동일한 접근 방식이 사용되었다[128]. 보다 일반적으로 민감도 분석은 연관성을 왜곡하는 데 필요한 교란, 선택 바이어스 또는 정보 바이어스의 정도를 식별하는 데 사용할 수 있다. 민감도 분석의 한 가지 중요한 사용법은 연구 결과 노출과 결과 사이의 연관성이 거의 또는 전혀 나타나지 않고 교란 또는 영(0)에 대한 기타 바이어스가 존재할 가능성이 있는 경우이다.

결과

연구 참여자 모집, 연구 집단에 대한 설명, 주요 결과 및 보조 분석에 이르기까지 발견된 내용에 대해 사실적인 설명이 결과에 보고된다. 저자의 견해와 의견을 반영하는 해석이나 담론적 텍스트가 없어야 한다.

13. 참여자:

13 (a). 각 단계별 인원에 대해 밝힌다 – 잠재적 적격자 수, 적격에 대해 조사한 수, 적격 확인 수, 연구 포함된 수, 추적 완료자 수, 분석자 수

예시

"표본으로 추출한 105개의 독립형 술집과 선술집 중 13개 업소는 더 이상 영업을 하지 않고 9개 업소는 식당에 위치하여 83개의 적격 업소가 남았다. 22개의 경우 6회 이상 시도했지만 소유주와 전화 연락이 닿지 않았다. 36개 술집의 주인은 연구 참여를 거부했다. (...) 25개의 참여 술집과 선술집에는 124명의 바텐더가 근무하고 있었으며, 67명의 바텐더가 주당 최소 1일 주간 근무를 하고 있었다. 주간 바텐더 중 54명(81%)이 기본 인터뷰와 폐활량 측정을 완료했으며, 이들 중 53명(98%)이 후속 조치를 완료했다"[129].

설명

연구 참여자 모집 과정에 대한 자세한 정보는 여러 가지 이유로 중요하다. 연구에 포함된 사람들은 대상 인구집단과 여러 면에서 다른 경우가 많다. 이로 인해 대상 인구집단의 경험을 반영하지 않는 유병률 또는 발생률 추정치가 나올 수 있다. 예를 들어, 성행위에 대한 우편 설문조사에 참여하기로 동의한 사람들은 거부한 사람들보다 교회에 덜 자주 출석하고, 덜 보수적인 성적 태도를 가지고 있으며, 첫 성관계 연령이 더 어려웠고, 담배를 피우거나, 술을 마실 가능성이 더 높았다[130]. 이러한 차이로 인해 우편 설문조사 결과가 인구의 성적 자유주의와 활동을 과대평가할 수 있다. 이러한 응답 바이어스(박스 3 참조)는 연구 참여자와 실제 연구에 포함된 사람 사이에 연관성이 다를 경우 노출-질병 연관성을 왜곡할 수 있다. 또 다른 예로, 일부 환자대조군연구에서 관찰된 젊은 산모의 나이와 자녀의 백혈병 사이의 연관성은 환자군과 대조군에 젊은 여성이 차별적으로 참여한 것이 원인일 수 있다[131,132]. 건강한 자녀를 둔 젊은 여성은 건강하지 않은 자녀를 둔 여성보다 참여 가능성이 낮았다[133].

참여율이 낮다고 해서 연구의 타당성이 크게 훼손되는 것은 아니지만, 참여율과 불참 사유에 대한 투명한 정보는 필수적이다. 또한 참여율, 응답률 또는 후속 조치 비율에 대한 보편적으로 합의된 정의가 없기 때문에 독자는 저자가 이러한 비율을 어떻게 계산했는지 이해할 필요가 있다[134].

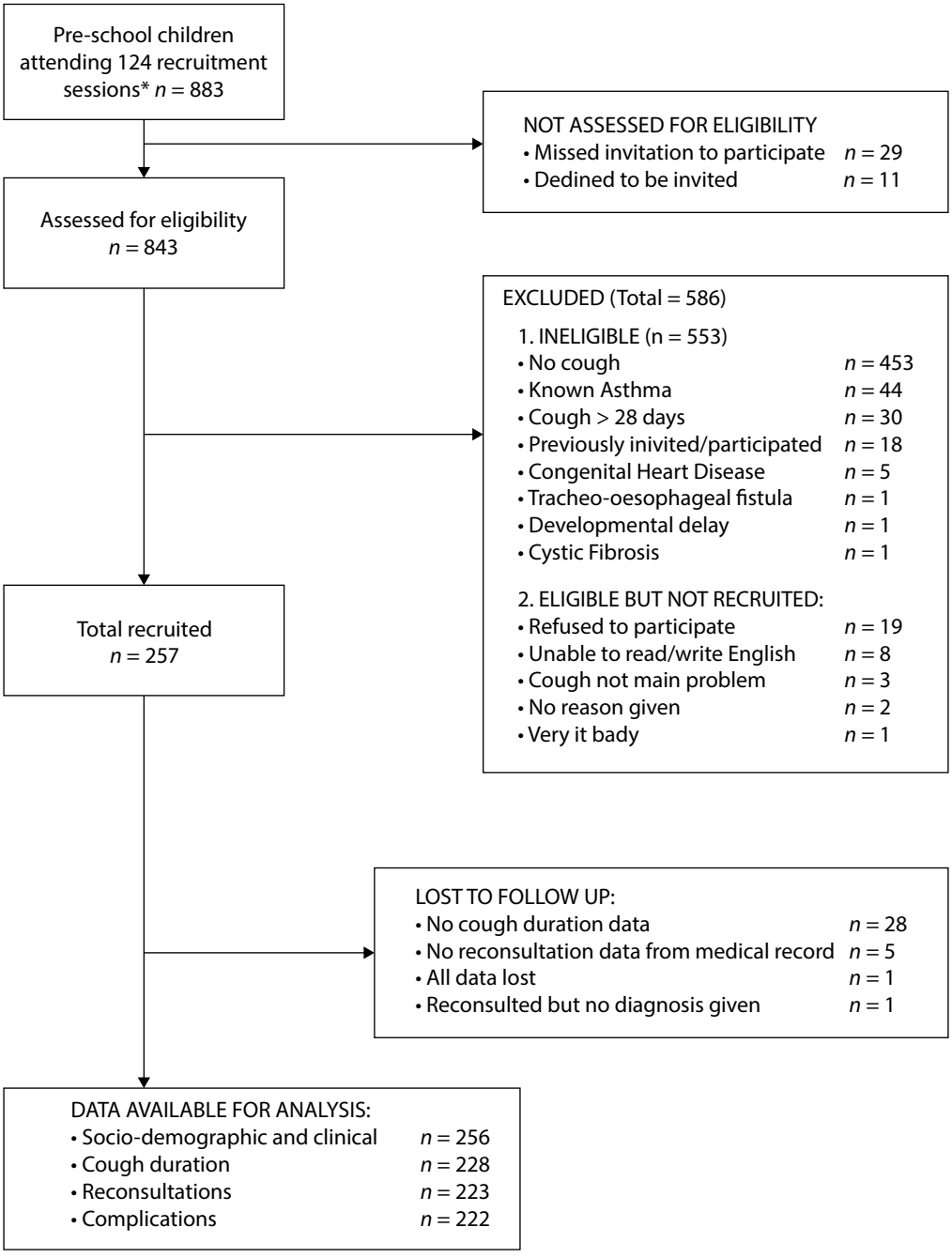
이상적으로 연구자는 대상 모집단 선정부터 분석에 참여자 자료를 포함시키는 것까지 연구 참여자 모집의 각 단계에서 고려한 개인 수에 대한 설명을 제공해야 한다. 연구 유형에 따라 여기에는 잠재적 적격자로 간주되는 사람의 수, 적격성 평가 인원, 적격자로 판명된 인원, 연구에 포함된 인원, 조사된 인원, 추적 관찰된 인원, 분석에 포함된 인원이 포함될 수 있다. 위의 예에서와 같이 연구 참여자의 추출이 두 단계 이상으로 수행되는 경우(다단계 추출) 다른 추출 단위에 대한 정보가 필요할 수 있다. 환자대조군연구의 경우, 저자는 환자 그룹과 대조군에 대해 참여자의 흐름을 별도로 설명하는 것이 좋다[135]. 대조군은 입원 환자 및 지역사회 거주자 등 여러 출처에서 선택할 수 있다. 이 경우 각 유형의 대조군에 대한 참가자 수를 별도로 설명하는 것이 좋다. 올슨과 동료들은 무작위 전화걸기 및 기타 방법을 통해 모집한 대조군에 대한 유용한 보고 지침을 제안했다[136].

최근 10년에 발표된 역학 연구에 대한 설문조사에 따르면 일반 역학, 공중 보건 및 의학 저널에 발표된 107건의 환자 대조군 연구 중 47건(59%), 코호트연구 154건 중 49건(32%), 단면연구 86건 중 51건(59%)에서 참여자에 대한 일부 정보가 제공되었다[137]. 역학 연구에서 참여 및 불참에 대한 보고가 불완전하거나 없는 것이 다른 두 건의 문헌 조사에서도 확인되었다[4,5]. 마지막으로, 최근 수십 년 동안 역학 연구 참여가 감소했을 수 있다는 근거가 있으며[137,138], 이는 투명한 보고의 필요성을 강조한다[139].

13 (b). 각 단계에서 불참 이유를 제시한다.

예시

"불참의 주요 이유는 참여자가 너무 아프거나 인터뷰 전에 사망한 경우(환자 30%, 대조군 1%), 무응답(환



*Denominator data missing from one session at which at least 3 attended with cough, 2 recruaited

자 2%, 대조군 21%), 거부(환자 10%, 대조군 29%), 기타 이유(전문의 또는 일반의의 거부, 비영어권, 정신 장애; 환자 7%, 대조군 5%)였다."[140].

설명

사람들이 더 이상 연구에 참여하지 않은 이유 또는 통계 분석에서 제외된 이유를 설명하면 독자가 연구 집단이 대상 집단을 대표하는지, 바이어스가 있을 가능성이 있는지 판단하는 데 도움이 된다. 예를 들어, 단면 건강 설문조사에서 건강 상태와 관련이 없을 것 같은 이유(예: 잘못된 주소로 인해 초대장이 전달되지 않은 경

우)로 인한 불참은 추정치의 정확성에 영향을 미치지만 바이어스를 유발하지는 않는다. 반대로, 질병이나 건강 상태가 양호하다고 인식되어 설문조사에 불참하는 사람이 많으면 조사 결과가 인구의 질병 유병률을 과소평가하거나 과대평가할 수 있다.

앞의 예에서와 같이 유익하고 잘 구조화된 순서도는 긴 설명이 필요할 수 있는 정보를 쉽고 투명하게 전달할 수 있다[142]. 이 다이어그램에는 사건 수와 같은 주요 결과가 유용하게 포함될 수 있다. 특히 복잡한 관찰 연구의 경우 흐름도을 사용하는 것이 좋지만, 특정 형식을 제안하지는 않는다.

2. 설명 자료:

14 (a). 연구 참여자의 특성(예: 인구통계학적, 임상적, 사회적)과 노출 및 잠재적 교란 요인에 대한 정보를 제공한다.

Example

Table. Characteristics of the Study Base at Enrolment, Castellana G (Italy), 1985-1986

	HCV-Negative n-1458	HCV-Positive n-511	Unknown n-513
Sex (%)			
Male	936 (64%)	296 (58%)	197 (39%)
Female	522 (36%)	215 (42%)	306 (61%)
Mean age at enrolment (SD)	45.7 (10.5)	52.0 (9.7)	52.5 (9.8)
Daily alcohol intake (%)			
None	250 (17%)	129 (25%)	119 (24%)
Moderate[a]	853 (59%)	272 (53%)	293 (58%)
Excessive[b]	355 (24%)	110 (22%)	91 (18%)

HCV, Hepatitis C virus.

[a] Males <60 g ethanol/day, fernales <30 g ethanol/day.

[b] Males >60 g ethanol/day, females >30 g ethanol/day.

Table adapted from Osella et al. [143].

설명

독자는 연구 결과의 일반화 가능성을 판단하기 위해 연구 참여자 및 노출에 대한 설명이 필요하다. 잠재적 교란 요인에 대한 정보(측정 여부 및 방법 포함)는 연구 유효성에 대한 판단에 영향을 미친다. 각 연구 그룹의 연속형 변수는 평균과 표준 편차를 제공하거나 자료가 비대칭 분포인 경우 중앙값과 백분위수 범위(예: 25번째 및 75번째 백분위수)를 제시하여 요약하는 것이 좋다. 소수의 범주(예: 질병의 1단계에서 4단계까지)를 연속형 변수로 제시해서는 안 되며, 각 범주에 대한 숫자와 비율을 제시하는 것이 바람직하다(박스 4 참조). 그룹을 비교하는 연구에서는 위의 예와 같이 그룹별로 기술적 특성과 숫자를 제시해야 한다.

표준 오차 및 신뢰 구간과 같은 추론적 측정값은 특성의 변동성을 설명하는 데 사용해서는 안 되며, 기술적(descriptive) 표에서 유의성 검정은 피해야 한다. 또한 P값은 분석에서 보정할 교란변수를 선택하는 데 적절한 기준이 아니며, 결과에 큰 영향을 미치는 교란변수의 작은 차이도 중요할 수 있다[144,145].

코호트연구에서는 노출이 다른 특성 및 잠재적 교란 요인과 어떻게 관련되는지 문서화하는 것이 유용할 수 있다. 환자대조군연구에서는 환자군와 대조군을 비교하여 잠재적 교란 요인을 판단할 수 없다. 대조군은 원천(source) 집단을 대표하며 일반적으로 여러 측면에서 환자군과 다를 수 있다. 예를 들어 경구 피임약과 심근경색 관련성에 대한 연구에서 심근경색이 있는 젊은 여성 표본은 대조군보다 높은 혈청 콜레스테롤, 흡연, 양성 가족력 등 해당 질환의 위험 요인을 더 자주 가지고 있었다[146]. 경구 피임약 처방이 이러한 위험 요인의 존재에 따라 이루어지지 않는 한(예: 위험 요인이 사건 이후에야 확인되었기 때문에) 경구 피임약의 효과 평가에 영향을 미치지 않는다(박스 5 참조). 환자대조군연구에서 잠재적 교란 요인의 존재 여부에 대해 노출된 그룹과 노출되지 않은 그룹을 동등하게 비교하려면, 대조군이 충분히 크고, 원천 모집단을 대표해야 한다[121,147].

14 (b). 관심 있는 각 변수에 대해 결측된 자료가 있는 참가자 수를 제시한다.

예시

Table. Symptom End Points Used in Survival Analysis

	Cough	Short of Breath	Sleeplessness
Symptom resolved	201 (79%)	138 (54%)	171 (67%)
Censored	27 (10%)	21 (8%)	24 (9%)
Never symptomatic	0	46 (18%)	11 (4%)
Data missing	28 (11%)	51 (20%)	50 (20%)
Total	256 (100%)	256 (100%)	256 (100%)

Table adapted from Hay et al. [141].

설명

결측된 자료는 결과의 일반화 가능성에 바이어스를 초래하거나, 영향을 줄 수 있으므로 저자는 노출, 잠재적 교란 요인 및 기타 환자의 중요한 특성에 대한 결측된 자료의 양을 독자에게 알려야 한다(항목 12c 및 박스 6 참조). 코호트연구의 경우, 불완전한 추적조사는 연구 결과에 바이어스를 초래할 수 있으므로, 저자는 추적조사 손실 정도를 이유와 함께 보고해야 한다(항목 12d 및 13 참조)[148]. 표와 그림을 사용하여 결측된 자료의 양을 열거할 것을 권장한다.

14 (c). 코호트연구: 추적 관찰 기간(예: 평균 및 총기간)을 요약한다.

예시

"4366년(중앙값 5.4년, 최대 8.3년)의 추적 관찰 기간 동안 알츠하이머병 202명을 포함하여 265명의 피험자가 치매 진단을 받았다."[149].

설명

독자는 사용 가능한 결과 자료에 대한 추적 관찰 기간과 범위를 알아야 한다. 저자는 평균 또는 중앙값 또는 둘 모두를 사용하여 평균 추적 관찰 기간에 대한 요약을 제시할 수 있다. 독자는 평균에 연구 참여자 수를 곱하여 총 인년 수를 계산할 수 있다. 또한 저자는 독자에게 추적 관찰 시간의 분포를 보여주기 위해 최소 및 최대 시간 또는 분포의 백분위수를 제시할 수 있다. 총 추적 관찰 기간(인년)을 보고하거나 수집된 잠재적 자료의 비율을 제시할 수도 있다[148]. 이러한 모든 정보는 두 개 이상의 노출 범주에 속한 참가자에 대해 별도로 제시될 수 있다. 암 저널에 실린 132편의 논문(대부분 코호트연구) 중 거의 절반이 추적 관찰 기간에 대한 요약을 제공하지 않았다[37].

15. 결과 자료:

코호트연구: 시간 경과에 따른 사건 발생 건수 또는 요약 측정값을 보고한다.

예시

Table. Rates of HIV-1 Seroconversion by Selected Sociodemographic Variables: 1990-1993

Variable	Person-Years	No. Seroconverted	Rate/1000 Person-Years (95% CI)
Calendar year			
1990	2197.5	18	8.2(4.4-12.0)
1991	3210.7	22	6.9(4.0-9.7)
1992	3162.6	18	5.7(3.1-8.3)
1993	2912.9	26	8.9(5.5-12.4)
1994	1104.5	5	4.5(0.6-8.5)

Table. Continued

Variable	Person-Years	No. Seroconverted	Rate/1000 Person-Years (95% CI)
Tribe			
Bagandan	8433.1	48	5.7(4.1-7.3)
Other Ugandan	578.4	9	15.6(5.4-25.7)
Rwandese	2318.6	16	6.9(3.5-10.3)
Other tribe	866.0	12	13.9(6.0-21.7)
Religion			
Muslim	3313.5	9	2.7(0.9-4.5)
Other	8882.7	76	8.6(6.6-10.5)

CI, confidence interval.
Table adapted from Kengeya-Kayondo et al. [150].

환자 대조 연구: 각 노출 범주의 보고 번호 또는 노출 요약 측정값을 제시한다.

Example
Table. Exposure among Liver Cirrhosis Cases and Controls

	Cases (*n*-40)	Controls (*n*-139)
Vinyl chloride monomer (cumulative exposure: ppm × years)		
<160	7 (18%)	38 (27%)
160-500	7 (18%)	40 (29%)
500-2500	9 (23%)	37 (27%)
>2500	17 (43%)	24 (17%)
Alcohol consumption (g/day)		
<30	1 (3%)	82 (59%)
30-60	7 (18%)	46 (33%)
>60	32 (80%)	11 (8%)
HBsAG/HCV		
Negative	33 (83%)	136 (98%)
Positive	7 (18%)	3 (2%)

HBsAG, hepatitis B surface antigen; HCV, hepatitis C virus.
Table adapted from Mastrangelo et al. [151].

단면연구: 결과 발생 건수 또는 요약 측정값을 보고한다.

Cross-sectional study: Report numbers of outcome events or summary measures.

Example
Table. Prevalence of Current Asthma and Diagnosed Hay Fever by Average *Alternaria alternata* Antigen Level in the Household

Categorized *Alternaria* Level*	Current Asthma		Diagnosed Hay Fever	
	N	Prevalence† (95% CI)	*N*	Prevalence† (95% CI)
1st tertile	40	4.8 (3.3-6.9)	93	16.4 (13.0-20.5)
2nd tertile	61	7.5 (5.2-10.6)	122	17.1 (12.8-22.5)
3rd tertile	73	8.7 (6.7-11.3)	93	15.2 (12.1-18.9)

*1st tertile <3.90 μg/g; 2nd tertile 3.90-6.27 μg/g; 3rd tertile ≥6.28 μg/g.
†Percentage (95%CI) weighted for the multistage sampling design of the National Survey of Lead and Allergens in Housing.
Table adapted from Salo et al. [152].

설명

노출(위험 요인)과 결과 사이의 가능한 연관성을 다루기 전에 저자는 관련된 기술자료를 보고해야 한다. 기술 자료를 나타내는 동일한 표에 연관성 측정치를 제시하는 것이 가능하고 의미 있을 수 있다(항목 14a 참조). 사건 발생을 결과로 하는 코호트연구의 경우, 관심 있는 각 결과에 대한 총 사건 수 혹은 1인당 사건 발생률을 보고하는 것을 고려한다. 추적 관찰 시간에 따라 사건의 위험이 변하는 경우, 사건의 수와 비율을 적절한 추적 관찰 간격으로 제시하거나 카플란-마이어 수명 표 또는 도표로 제시한다. 시간에 따른 측정값(예: 평균 및 표준 편차)은 표나 그림으로 제시하는 것이 더 나을 수 있다. 단면적 연구의 경우 가장 흔한 결과 사건이나 요약 측정값에 대해 동일한 유형의 정보를 제시하는 것이 좋다. 환자대조군연구의 경우 환자군과 대조군의 빈도 또는 정량적 요약을 별도로 보고하는 데 중점을 둔다[154]. 모든 설계에 대해 자료가 분석되지 않더라도 지속적인 결과 또는 노출을 범주별로 표로 작성하는 것도 도움이 될 수 있다.

16. 주요 결과

16 a). 비보정 추정치를 제시한다, 가능하면 교란변수 보정 추정치와 정밀도를 제시한다(95% 신뢰구간). 보정 변수를 제시하고 보정변수 선택 이유를 밝힌다.

예시 1

"우리는 처음에 다음과 같은 변수를 만텔-헨첼 계층화 분석에 의한 잠재적 교란변수로 고려했다. (...) 최종 로지스틱 회귀 모델에 포함된 변수는 만텔-헨첼 보정 후 확률 비율에 10%의 변화를 가져온 (...) 변수였다"[155].

Example 2

Table. Relative Rates of Rehospitalisation by Treatment in Patients in Community Care after First Hospitalisation due to Schizophrenia and Schizoaffective Disorder

Treatment	No. of Relapses	Person-Years	Crude Relative Rate (95% CI)	Adjusted Relative Rate (95% CI)	Fully Adjusted Relative Rate (95% CI)
Perphenazine	53	187	0.41 (0.29 to 0.59)	0.45 (0.32 to 0.65)	0.32 (0.22 to 0.49)
Olanzapine	329	822	0.59 (0.45 to 0.75)	0.55 (0.43 to 0.72)	0.54 (0.41 to 0.71)
Clozapine	336	804	0.61 (0.47 to 0.79)	0.53 (0.41 to 0.69)	0.64 (0.48 to 0.85)
Chlorprothixene	79	146	0.79 (0.58 to 1.09)	0.83 (0.61 to 1.15)	0.64 (0.45 to 0.91)
Thioridazine	115	201	0.84 (0.63 to 1.12)	0.82 (0.61 to 1.10)	0.70 (0.51 to 0.96)
Perphenazine	155	327	0.69 (0.58 to 0.82)	0.78 (0.59 to 1.03)	0.85 (0.63 to 1.13)
Risperidone	343	651	0.77 (0.60 to 0.99)	0.80 (0.62 to 1.03)	0.89 (0.69 to 1.16)
Haloperid	73	107	1.00	1.00	1.00
Chlorpromazine	82	127	0.94 (0.69 to 1.29)	0.97 (0.71 to 1.33)	1.06 (0.76 to 1.47)
Levomepromazine	52	63	1.21 (0.84 to 1.73)	0.82 (0.58 to 1.18)	1.09 (0.76 to 1.57)
No antipsychotic drugs	2248	3362	0.98 (0.77 to 1.23)	1.01 (0.80 to 1.27)	1.16 (0.91 to 1.47)

설명

많은 경우 비보정 혹은 최소 보정 분석 결과와 완전 보정 분석 결과를 함께 제시할 수 있다. 보정되지 않은

분석 결과를 주요 자료와 함께 제공하는 것이 좋다(예: 노출된 환자 및 대조군 수). 이를 통해 독자는 연관성 측정의 이면에 있는 자료를 이해할 수 있다(15번 항목 참조). 보정 분석의 경우, 공변량에서 결측된 값으로 인해 이 숫자가 달라질 수 있으므로 분석 참여자 수를 보고한다(항목 12c 참조). 추정치는 신뢰 구간과 함께 제공해야 한다. 독자는 비보정연관성 측정값과 잠재적 교란 요인에 대해 보정된 연관성 측정값을 비교하여 얼마나 많이, 어떤 방향으로 변화했는지 판단할 수 있다. 독자는 '보정' 결과가 연관성 주제 측정의 인과적 부분과 동일하다고 생각할 수 있지만, 보정 결과도 추출 오류, 선택 바이어스, 정보 바이어스 또는 잔존 교란이 없다고 할 수는 없다(박스 5 참조). 따라서 보정 결과를 해석할 때 세심한 주의를 기울여야 하며, 결과의 타당성은 종종 중요한 교란변수에 대한 완전한 지식, 정확한 측정, 통계 모델의 적절한 사양(항목 20 참조)이 중요하기 때문이다[157,158]. 저자는 고려된 모든 잠재적 교란변수와 통계 모델에서 변수를 제외하거나 포함시키는 기준을 설명해야 한다. 변수를 제외할지 아니면 포함할지에 대한 결정은 인과 관계에 대한 지식 또는 명시적 가정에 따라 이루어져야 한다. 부적절한 결정, 예를 들어 노출과 질병 사이의 인과 경로에 있는 변수를 포함함으로써 바이어스를 유발할 수 있다(매개 변수에 의한 영향의 정도를 평가하는 것이 목적이 아닌 경우). 모델에 변수를 포함하기로 한 결정이 추정치의 변화에 근거한 것이라면, 어떤 변화가 그 포함을 정당화할 만큼 충분히 중요하다고 간주되었는지 보고하는 것이 중요하다. 교란변수를 선택하기 위해 '후방 삭제' 또는 '전방 포함' 전략을 사용했다면, 그 과정을 설명하고 귀무가설을 거부할 수 있는 유의 수준을 제시해야 한다. 참고로, 통계적 유의성 테스트에만 근거하여 교란변수를 선택하는 것을 권장하지 않는다[147,159,160]. 최근 역학 연구 보고의 질에 대한 연구에 따르면 대부분의 논문에서 신뢰 구간이 보고된 것으로 나타났다[4]. 그러나 교란 변수의 선택에 대해 설명한 저자는 거의 없었다[4,5].

16 (b). 연속변수를 범주화 했으면 범주와 범위를 제시한다.

Example

Table. Polychlorinated Biphenyls in Cord Serum

Quartile	Range (ng/g)	Number
1	0.07-0.24	180
2	0.24-0.38	181
3	0.38-0.60	181
4	0.61-18.14	180

Table adapted from Sagiv et al. [161].

설명

연속 자료를 분류하는 것은 분석에서 중요한 부분이며(박스 4 참조), 결과 제시에도 영향을 미친다. 표에서는 각 그룹(예: 환자 및 대조군)에 대해 개별적으로 관련성이 있는 경우 위험에 노출된 사람, 시간 등 각 노출 범주에 대한 결과를 제공해야 한다. 사용된 범주의 세부 사항은 연구 비교 및 메타 분석에 도움이 될 수 있다. 체질량지수 임계값[162]과 같은 기존의 임계점을 사용하여 자료를 그룹화한 경우, 최고 및 최저 범주를 제외하고 그룹 경계(즉, 값의 범위)를 쉽게 도출할 수 있다. 사분위수 범주를 사용하는 경우, 자료에서 범주 경계를 유추할 수 없다. 최소한 범주 경계는 보고해야 하며, 자료의 범위와 범주 내 평균 또는 중앙값을 보고하는 것도 도움이 된다.

16 (c). 적절하다면, 상대위험도를 절대위험도로 변환한다.

예시

"10년간 HRT[호르몬 대체 요법]를 사용하면 에스트로겐 단독 제제 사용자 1,000명당 5건(95% CI 3–7)의 추가 유방암이 발생하고 에스트로겐-프로게스타겐 복합제 사용자 1,000명당 19건(15–23)의 추가 암이 발생하는 것으로 추정된다"[163].

설명

노출과 질병 사이의 연관성을 조사한 연구 결과는 일반적으로 위험도, 비율 또는 확률의 비율 같은 상대적인 수치로 보고된다(박스 8 참조). 상대 측정치는 노출과 질병 사이의 연관성의 강도를 파악한다. 상대위험도가 1에서 멀리 떨어져 있으면 교란으로 인한 연관성일 가능성이 적다[164,165]. 상대 효과 또는 연관성은 절대적 측정치보다 연구와 인구집단 간에 일관성이 있는 경향이 있지만, 종종 특정 환자에서는 연관성이 없을 수도 있다. 예를 들어, 관상동맥 심장 질환의 고전적 위험도는 국가마다 크게 다르지만 북아일랜드, 프랑스, 미국 및 독일에 거주하는 남성의 상대 위험도는 유사하였다[166,167]. 대조적으로, 심혈관 질환 사망률의 위험요인으로서 고혈압에 대한 연구에서 자료는 일정한 비의 비율보다 위험도 차이가 더 적합했다[168]. 로지스틱[169] 및 비례 위험(Cox) 회귀[170]를 포함하여 널리 사용되는 통계 모델은 비율 측정에 기반한다. 이러한 모델에서는 비율 효과 측정값의 정합성 이탈만 쉽게 식별할 수 있지만, 상호 작용으로 인한 상대적 초과 위험(RERI, 항목 12b 및 박스 8 참조)과 같이 위험 차이의 추가성 이탈을 평가하는 측정값은 비율 측정값에 기반한 모델에서 추정할 수 있으며, 많은 상황에서 노출과 관련된 절대 위험은 상대 위험보다 더 중요한 관심사이다. 예를 들어, 약물의 부작용에 초점을 맞추는 경우 단위 사용 시간(예: 일, 주 또는 년) 당 추가 환자 수를 알고 싶을 것이다. 위의 예시는 10년간 호르몬 대체 요법을 사용한 여성 1,000명당 유방암 추가 발생 건수를 보여준다[163]. 기여 위험 또는 인구 기여 비율과 같은 측정은 노출을 제거할 경우 질병을 얼마나 예방할 수 있는지 측정하는 데 유용할 수 있다. 이러한 측정치는 통계적 불확실성 측정치(예: 예시에서와 같은 신뢰 구간)와 함께 제시하는 것이 바람직하다. 저자는 위험 요인과 질병 사이의 인과 관계를 포함하여 이러한 맥락에 서 만들어진 강력한 가정을 알고 있어야 한다(박스 7 참조)[171]. 의미적 모호성과 복잡성 때문에 저자는 기여성 위험을 계산하는 데 어떤 방법을 사용했는지 자세히 보고해야 하며, 이상적으로는 사용된 공식을 제공해야 한다[172]. 최근 주요 의학 저널에 게재된 222개 논문의 초록을 조사한 결과, 비율을 포함한 무작위 임상시험 초록의 62%에서 절대 위험도가 제공되었지만 코호트연구 초록의 21%에서만 제공된 것으로 나타났다[173]. 1966년부터 1997년까지 Medline의 무료 텍스트 검색에 따르면 제목 또는 초록에 기여 위험을 언급한 항목은 619개였으며, 상대 위험 또는 오즈비를 사용한 항목은 18,955개로 1에서 31의 비율을 사용했다[174].

17. 다른 분석: 다른 분석에 대해 보고한다 – 하부집단분석, 교호작용(interactions), 민감도분석

Example 1

Table. Analysis of Oral Contraceptive Use, Presence of Factor V Leiden Allele, and Risk for Venous Thromboembolism

Factor V Leiden	Oral Contraceptives	No. of Patients	No. of Controls	Odds Ratio
Yes	Yes	25	2	34.7
Yes	No	10	4	6.9
No	Yes	84	63	3.7
No	No	36	100	1 (Reference)

Table modified from Vandenbroucke et al. [182] by Botto et al. [183].

Example 2

Table. Sensitivity of the Rate Ratio for Cardiovascular Outcome to an Unmeasured Confounder

Prevalence of Unmeasured Binary Confounder in the Exposed Group, %	Prevalence of Unmeasured Binary Confounder in the Comparator Group, %	Unmeasured Binary Confounder Rate Ratio	High Exposure Rate Ratio (95% CI)*
90	10	1.5	1.20 (1.01-1.42)
90	50	1.5	1.43 (1.22-1.67)
50	10	1.5	1.39 (1.18-1.63)
90	10	2	0.96 (0.81-1.13)
90	50	2	1.27 (1.11-1.45)
50	10	2	1.21 (1.03-1.42)

Table. Continued

Prevalence of Unmeasured Binary Confounder in the Exposed Group, %	Prevalence of Unmeasured Binary Confounder in the Comparator Group, %	Unmeasured Binary Confounder Rate Ratio	High Exposure Rate Ratio (95% CI)*
90	50	3	1.18 (1.01-1.38)
50	10	3	0.99 (0.85-1.16)
90	50	5	1.08 (0.85-1.26)

CI, confidence interval.
*Adjusted for age, sex, cardiovascular drug use, and unmeasured binary confounder.
Table adapted from Wei et al. [184].

박스 7. 연관성, 효과 및 영향 측정

관찰 연구는 인구집단에서 건강 문제의 규모와 분포를 설명하기 위해서 수행될 수 있다. 특정 시점에 질병에 걸린 사람의 수(유병률) 또는 정해진 기간 동안 질병에 걸린 사람의 수(발병률)를 조사할 수 있으며, 발병률은 질병에 걸린 사람의 비율(누적 발병률) 또는 추적 관찰 기간 동안의 1인당 비율(발병률)로 표현할 수 있다. 다양한 발생률을 설명하기 위해 사망률, 출생률, 발병률, 환자 사망률 등의 특정 용어가 사용된다. 마찬가지로, 점 유병률 및 기간, 연간 또는 평생 유병률과 같은 용어는 다양한 유형의 유병률을 설명하는 데 사용된다[30].

다른 종류의 관찰 연구는 인과 관계를 다룬다. 이러한 연구들은 조사 대상 위험 요인에 노출된 사람과 노출되지 않은 사람 사이의 관심 대상 사건의 위험, 비율 또는 유병률을 비교하는 데 중점을 둔다. 이러한 연구는 종종 '상대위험도'을 추정하는데, 이는 위험비(누적 발생률의 비율)와 비율(발생률의 비율)을 의미할 수 있다. 환자대조군연구에서는 원인 집단(대조군)의 일부만 포함된다. 결과는 환자군와 대조군 간의 노출 확률의 비율로 표현된다. 이 오즈비는 환자와 대조군의 추출에 따른 위험 또는 비율 비율의 추정치를 제공한다(박스 1 참조)[175,176]. 단면연구의 유병률 또는 유병률 오즈비는 일부 상황에서 유용할 수 있으며[177], 결과를 상대적 및 절대적 용어로 표현하는 것이 종종 도움이 될 수 있다.

예를 들어, 영국 남성 의사를 대상으로 한 연구에서 50년 추적 관찰 기간 동안 폐암으로 인한 사망률은 흡연자의 경우 연간 10만 명당 249명인 반면 비흡연자의 경우 연간 10만 명당 17명으로 14.6(249/17)의 배율을 보였다[178]. 관상동맥심장질환(CHD)의 경우, 해당 비율은 연간 10만 명당 1,001명과 619명으로, 비율비는 1.61(1,001/619)이었다. 흡연이 사망에 미치는 영향은 CHD보다 폐암에서 훨씬 더 강하게 나타난다. 흡연의 절대적인 영향을 고려하면 그림이 달라진다. 발생률의 차이는 폐암의 경우 연간 10만 명당 232명(24,917명), 만성 폐쇄성 폐질환의 경우 382명(1,001,619명)이었다. 따라서 흡연을 하는 의사들 사이에서 흡연은 폐암보다 만성 폐쇄성 폐질환으로 인한 사망을 유발할 가능성이 더 높았다. 노출을 제거함으로써 인구의 질병 부담을 얼마나 예방할 수 있을까? 한 연구에 따르면 전체 폐암의 91%, 만성 폐쇄성 폐질환의 40%, 2000년 남성 전체 사망의 33%가 흡연에 기인한 것으로 나타났다[179]. 인구기여 비율은 일반적으로 특정 노출로 인한 환자의 비율로 정의되지만 여러 개념(통일된 용어 없음)이 존재하며 다른 요인을 보정하기 위한 잘못된 접근 방식이 때때로 사용된다[172,180].

보고의 의미는 무엇인가? 상대적 측정치는 연관성의 강도를 강조하며 병인 연구에 가장 유용하다. 노출과의 인과관계가 제시되고 연관성이 효과로 해석되는 경우, 공중보건 정책의 가능한 영향을 측정하기 위해 상대위험도 추정치를 절대적 위험 측정치로 변환할 수 있다(항목 16c 참조)[181]. 그러나 저자는 이러한 맥락에서 만들어진 강력한 가정을 인식해야 한다[171]. 개별 상황에 적합한 개념과 방법을 결정할 때 주의가 필요하다.

설명

관찰 연구에서는 주 분석 외에도 다른 분석이 수행되는 경우가 많다. 이러한 분석은 특정 하위 그룹, 위험 요인 간의 잠재적 상호 작용, 기여 위험 계산을 다루거나 민감도 분석에서 연구 변수에 대한 다른 정의 사용 등이 있다.

하위 그룹 분석과 관련된 위험성과 분석 다양성에 대한 논쟁이 있다[4,104]. 우리의 의견으로는, 전체 결과에서 효과가 거의 또는 전혀 없을 때 하위 그룹별 연관성을 보는 근거를 찾는 경향이 너무 크다. 반면에 전체 연관성이 여러 하위 그룹, 특히 각 하위 그룹에서 충분한 자료를 확보할 수 있을 만큼 연구 규모가 큰 경우 미리 지정된 여러 하위 그룹에서 일관되게 나타나는지 살펴보는 것도 가치가 있다. 두 번째 쟁점 사항으로 자료 분석 중에 발생한 관심 하위 그룹에 관한 것이다. 이는 중요한 결과일 수도 있지만 우연히 발생할 수도 있다. 향후 다른 자료에 대한 분석을 통해 초기의 흥미로운 결과가 시간의 경과에 따라 어느 정도 지속될 수 있는지 알 수 없기 때문에 수행된 모든 하위 그룹 분석에 대해 알리는 것은 가능하지도 않고 필요하지도 않다고 주장하기도 한다[9]. 어떤 분석이 계획된 분석이고 어떤 분석이 계획되지 않은 분석인지 보고할 것이 필요하다(항

목 4, 12b 및 20 참조). 이를 통해 독자는 발견에서 검증 또는 반박에 이르는 연속체에서 연구의 위치를 고려하여 다중성의 의미를 판단할 수 있다.

세 번째 논쟁 영역은 위험 요인 간의 공동 효과와 교호성을 평가하는 방법, 즉 덧셈 모형 혹은 곱셈 모형으로 평가해야 하는지, 그것도 아니면 가장 적합한 통계 모델로 척도를 결정해야 하는지이다(항목 12b 및 박스 8 참조). 가능한 경우 위의 첫 번째 예[183] 또는 Martinelli 등의 연구에서와 같이 각 노출의 개별 효과와 공동 효과를 표로 작성하여 보고하는 것이 좋다[185]. 이러한 표는 독자에게 덧셈 상호 작용뿐만 아니라 곱셈 상호 작용을 평가할 수 있는 충분한 정보를 제공한다(이러한 계산 방법은 박스 8에 나와 있다). 개별 및 공동 효과에 대한 신뢰 구간은 독자가 자료의 강도를 판단하는 데 도움이 될 수 있다. 또한 상호 작용으로 인한 초과 상대위험도(RERI)와 같은 상호 작용 측정치에 대한 신뢰 구간은 상호 작용 테스트 또는 동질성 테스트와 관련이 있다. 한 가지 반복되는 문제는 저자가 하위 그룹 간의 P값 비교를 사용하여 효과 변경자(effect modifiers)에 대한 잘못된 주장을 하는 것이다. 예를 들어, 한 범주(예: 남성)에서는 통계적으로 유의미한 연관성이 있지만 다른 범주(예: 여성)에서는 그렇지 않다고 해서 그 자체로 효과 변경의 근거가 되는 것은 아니다. 마찬가지로, 각 점 추정치에 대한 신뢰 구간 간극이 겹친다는 것을 근거로 교호성이 없다고 추론하는 경우가 있는데 이는 부적절하다. 유효한 추론을 하려면 연관의 크기가 하위 그룹 간에 차이가 있는지 직접 평가해야 한다.

민감도 분석은 통계 분석에서 선택한 항목의 영향을 조사하거나 결측된 자료 또는 가능한 바이어스에 대한 결과의 강건성을 조사하는 데 유용하다(항목 12b 참조). 이러한 분석의 보고 수준과 관련하여 판단이 필요하다. 민감도 분석이 다수 수행된 경우, 모든 민감도 분석에 대해 상세한 결과를 제시하는 것은 비현실적일 수 있다. 민감도 분석이 수행되었으며 제시된 주요 결과와 일치한다고 보고하는 것으로 충분할 수 있다. 조사된 문제가 주요 관심사이거나 효과 추정치가 상당히 다른 경우 상세히 제시해야 한다[59,186].

Pocock 등은 관찰 연구를 보고한 73편의 논문 중 43편에 하위 그룹 분석이 포함되어 있으며, 대다수가 그룹 간 차이를 주장했지만 교호성에 대한 공식적인 평가를 보고한 논문은 8편에 불과했다고 하였다(항목12b 참조)[4].

고찰

고찰에서는 연구의 타당성과 의미에 대해 다룬다[191]. 설문 조사에 따르면 고찰은 연구 결과와 그 의미에 대한 불완전하거나 편향된 평가와 저자의 결과를 뒷받침하는 내용이 많이 나타나고 있다[192,193]. 고찰을 구조화하면 저자가 독자에게 결과 내용을 안내하면서 결과에 대한 부당한 추측과 과도한 해석을 피하는 데 도

박스 8. 교호성(효과 변경): 관절효과 분석

노출과 질병 위험의 연관성이 다른 노출에 따라 달라질 때 교호성(interaction)이 존재한다. 교호성을 평가하고 보고할 때의 문제 중 하나는 노출 효과가 상대 위험도(또는 비율) 또는 위험도 차이(또는 비율 차이)의 두 가지 방법으로 측정될 수 있다는 것이다. 상대 위험도를 사용하면 곱셈(multiplicative) 모델이 되고, 위험도 차이를 사용하면 덧셈(additive) 모델이 된다[187,188]. 곱셈 또는 덧셈 모델 모두 아닌 '통계적 상호 작용'과 덧셈 모델에서 벗어나 측정되는 '생물학적 상호 작용'이 때때로 구별된다[189]. 그러나 덧셈 모델이나 곱셈 모델 모두 특정 생물학적 메커니즘을 나타내지 않는다.

모델 선택에 관계없이 주요 목표는 두 노출의 결합 효과가 (다른 노출이 없는 경우) 개별 효과와 어떻게 다른지 이해하는 것이다. Human Genomic Epidemiology Network(HuGENet)는 다양한 유형의 교호성을 평가할 수 있는 개별 및 공동 효과를 투명하게 제시하기 위한 레이아웃을 제안했다[183]. 제안을 설명하기 위해 경구 피임약 및 인자 V Leiden 돌연변이에 관한 연구 자료가 사용되었으며 이 예는 항목 17에도 사용되었다. 경구 피임약과 인자 V Leiden 돌연변이 모두 정맥 혈전증의 위험을 증가시킨다. 이들의 개별 효과와 공동 효과는 2×4 표(항목 17의 예 1 참조)에서 계산할 수 있다. 여기서 OR 1은 경구 피임약을 사용하지 않는 인자 V Leiden이 없는 여성의 기준선을 나타낸다.

어려운 점은 사례 관리 연구와 같은 일부 연구 설계와 로지스틱 또는 Cox 회귀 모델과 같은 여러 통계 모델이 상대 위험도를 추정하고 본질적으로 곱셈 모델로 이어진다는 것이다. 이 경우 상대 위험도는 추가 규모로 변환될 수 있다. 항목 17의 예 1에서 개별 OR은 3.7과 6.9이다. 결합 승산비는 34.7이다. 이러한 자료를 승법 모델로 분석하면 결합 승산비는 25.7(3.73–6.9)이 될 것으로 예상된다. 관찰된 결합 효과 34.7은 곱셈 척도(34.7/25.7)에서 예상되는 것보다 1.4배 더 크다. 이는 곱셈 상호 작용의 승산비이다. 이는 로지스틱 회귀 모델에서 추정된 상호 작용 계수의 역로그와 같다. 가산 모델에서 결합 승산비는 9.6(3.7+6.9–1)이 될 것으로 예상된다. 관찰된 결합 효과는 가산성과 크게 다르다. 차이는 25.1(34.7–9.6)이다. 승산비가 상대적 위험(또는 비율)으로 해석될 때 후자의 양(25.1)은 RERI(교호성으로 인한 상대적 초과 위험)이다[190]. 이는 기준 값(OR=1과 동일)이 정맥 혈전증의 기준 발생률, 예를 들어 1/10,000 여성년을 나타낸다고 상상하면 더 쉽게 이해될 수 있다. 이는 개별노출과 공동노출이 있을 때 증가한다.

움이 될 수 있다[194,195]. 예를 들어, Annals of Internal Medicine[196]에서는 저자가 다음을 제시하여 고찰 파트을 구성할 것을 권장한다 : (1) 주요 결과에 대한 간략한 개요; (2) 가능한 메커니즘과 설명; (3) 다른 출판 연구의 관련 결과와의 비교; (4) 연구의 한계; (5) 진료와 연구에 미치는 영향에 대한 내용. 다른 사람들도 비슷한 제안을 했다[191,194]. 연구 권고 사항에 대한 부분과 연구의 한계에 대한 부분은 서로 밀접하게 연결되어야 한다. 연구자들은 '더 많은 연구가 필요하다'고 단호하게 말하기보다는 후속 연구가 자신의 연구를 개선할 수 있는 부분을 제시해야 한다[197,198]. 우리는 적절한 부제목을 사용하여 고찰 파트을 구성할 것을 권장한다.

18. 주요 결과: 연구 목적에 비추어 주요 결과를 요약한다.

예시

'우리는 소수 민족이 더 높은 수준의 심혈관 질환(CVD) 위험 요소와 관련이 있고, 그 연관성은 사회경제적 상태에 의해 실질적으로 설명될 것이라는 가설을 세웠다. 우리의 가설은 확인되지 않았다. 연령과 SES를 보정한 후에도 백인 여성과 흑인 및 멕시코계 미국인 여성 사이에는 체질량 지수, 혈압, 당뇨병, 신체 활동 부족에서 매우 유의미한 차이가 남아 있었다. 또한 우리는 SES에 의한 CVD 위험 요소에 큰 차이가 있음을 발견했는데, 이는 소수 민족 여성과 SES가 낮은 백인 여성 모두의 고위험 상태를 보여주는 결과이다[199].

설명

연구의 주요 결과에 대한 간략한 요약으로 고찰을 시작하는 것이 좋다. 짧은 요약은 독자에게 주요 결과를 상기시키고 저자가 제공한 해석과 의미가 결과에 의해 뒷받침되는지 여부를 평가하는 데 도움이 될 수 있다.

19. 제한점: 잠재적 바이어스나 비정밀을 고려하면서 연구의 제한점에 대해 고찰한다. 잠재적 바이어스의 방향성과 크기를 고찰한다.

예시

"비만 수준이 증가함에 따라 상담 시행률도 증가하기 때문에 우리의 추정치는 실제 시행률을 과대평가할 수 있다. 전화 설문조사 역시 상담의 실제 시행률을 과대평가할 수 있다. 전화가 없는 사람은 전화가 있는 사람과 비슷한 수준의 과체중을 가지고 있지만, 전화가 없는 사람은 교육 수준이 낮은 경향이 있으며 이는 우리 연구에서 상담 수준이 낮은 것과 관련된 요소이다. 또한, 체중에 대한 질문에 응답을 거부한 사람들뿐만 아니라 참여를 거부한 사람들로 인해 발생할 수 있는 바이어스도 우려된다. 더욱이, 자료는 단면적으로 수집되었기 때문에 환자의 체중 감량 시도에 앞서 상담이 선행되었다고 추론할 수 없다."[200].

설명

연구의 한계를 확인하고 논의하는 것은 과학적 보고의 필수적인 부분이다. 결과에 영향을 미칠 수 있는 바이어스와 교란의 원인을 확인하는 것뿐만 아니라 잠재적인 바이어스의 방향과 크기, 상대적 중요성을 논의하는 것도 중요하다(항목 9 및 박스 3 참조).

저자는 결과의 비정밀성에 대해서도 논의해야 한다. 연구 규모(항목 10), 노출 측정, 교란요인 및 결과(항목 8)를 포함하여 연구의 여러 측면과 관련하여 비정밀성이 발생할 수 있다. 노출의 실제 값을 정확하게 측정할 수 없으면 단일성을 향하는 바이어스가 발생할 수 있다. 위험 요소를 덜 정확하게 측정할수록 바이어스가 커진다. 이 효과는 '감쇠'[201,202] 또는 최근에는 '회귀 희석 바이어스'[203]로 설명되었다. 그러나 서로 연관되어 있는 위험 요인들이 서로 다른 정도로 비정밀하게 측정되면 이와 관련된 보정 상대위험도는 단일성(unity)으로 편향되거나 단일성에서 멀어질 수 있다[204–206].

한계를 논의할 때 저자는 제시된 연구를 타당성, 일반화 가능성 및 정밀도 측면에서 다른 연구와 비교할 수 있다. 이 접근 방식에서 각 연구는 독립되어 있지 않고, 서로 다른 연구에 영향을 미치는 것으로 볼 수 있다[207]. 놀랍게도 연구의 중요한 한계에 대한 논의가 출판된 보고서에서 생략되는 경우가 있다. *The Lancet*에 독창적인 연구 논문을 발표한 저자들을 대상으로 한 설문 조사에 따르면 연구의 중요한 약점이 조사 설문

지에 보고되었지만 출판된 논문에는 보고되지 않은 것으로 나타났다[192].

20. 해석: 연구목적, 제한점, 유사연구 결과, 다양한 분석, 다른 관련 근거을 고려하면서 주의 깊게 결과에 대한 해석을 수행한다.

예시

"심근경색으로 인한 사망과 2세대 경구 피임약 사용 사이의 연관성에 대한 모든 설명은 추측적이어야 한다. 직접적인 생물학적 메커니즘을 제안하는 공개된 근거는 없으며 관련 결과가 있는 다른 역학 연구도 없다. (...) 절대 위험의 증가는 매우 작으며 아마도 주로 흡연자에게 적용될 것이다. 확증적인 증거가 부족하고 분석이 상대적으로 적은 숫자를 기반으로 하기 때문에 해당 주제에 대한 더 많은 증거가 필요하다. 우리는 이러한 결과를 바탕으로 처방을 변경하는 것을 권장하지 않는다."[120].

설명

고찰 파트의 핵심은 연구 결과의 해석이다. 과잉 해석은 흔하고 인간적이다. 우리가 객관적인 평가를 하기 위해 열심히 노력할 때에도, 심사자들은 우리가 어떤 면에서 너무 멀리 나아갔다고 올바르게 지적하는 경우가 많다. 결과를 해석할 때 저자는 발견과 검증의 스텍트럼에서 어디에 속하고, 탈락과 비참여같은 바이어스의 잠재적 원인을 고려해야 한다(항목 9, 12 및 19 참조). 교란(항목 16a), 관련 민감도 분석 결과, 다중분석 및 하위 그룹 분석 문제(항목 17)를 적절히 고려해야 한다. 저자는 또한 측정되지 않은 변수나 정밀하지 않은 교란변수 측정으로 인한 잔여 교란도 고려해야 한다. 예를 들어, 사회경제적 상태는 많은 건강 결과와 연관되어 있으며 종종 비교 대상 그룹마다 다를 수 있다. SES를 측정하는 데 사용되는 변수는 정의되지 않고 측정되지 않은 기타 노출에 대한 대용이며, 실제 교란요인은 정의에 따라 오류로 측정될 수 있다[208]. 저자는 신뢰 구간에 반영된 통계적 불확실성보다 더 큰 추정치의 실제 불확실성 범위를 다루어야 한다. 통계적 불확실성은 연구의 설계, 구현 및 측정 방법에서 발생하는 다른 불확실성을 고려하지 않는다[209].

인과관계에 대한 사고와 결론을 안내하기 위해 일부에서는 1965년 Bradford Hill이 제안한 기준이 도움이 될 수 있다[164]. 노출과의 연관성은 얼마나 강한가? 질병이 발병하기 전에 발생했나? 다양한 연구와 환경에서 연관성이 일관되게 관찰되나? 실험실 및 동물 연구를 포함한 실험 연구에서 뒷받침하는 증거가 있나? 노출의 추정 효과는 얼마나 구체적이며 용량-반응 관계가 있나? 연관성이 생물학적으로 타당한가?

그러나 이러한 기준을 기계적으로 적용해서는 안된다. 예를 들어 일부에서는 2 또는 3 미만의 상대 위험도를 무시해야 한다고 주장했다[210,211]. 이는 Cornfield 등의 주장을 뒤집은 것이다. 큰 상대적 위험의 강도에 대해(항목 12b 참조)[127] 인과적 영향은 상대 위험도가 9일 때 더 가능성이 높지만, 3 미만의 위험도가 반드시 허위라는 뜻은 아니다. 예를 들어, 자궁 내 방사선 조사 후 소아 백혈병 위험이 약간 증가하는 것은 다른 설명이 분명하지 않은 의료 절차의 부작용과 관련되기 때문에 신뢰할 수 있다[212]. 게다가 방사선의 발암성 영향도 잘 알려져 있다. 일주일에 2–4개의 계란을 섭취하는 것과 관련된 난소암 위험이 두 배로 증가한다는 것은 즉시 신뢰할 수 없다. 왜냐하면 식습관은 SES뿐만 아니라 많은 생활 방식 요인과 연관되어 있기 때문이다[213]. 대조적으로, 다양한 유형의 경구 피임약 사이의 혈전증 위험 차이에 대해 많이 논의된 역학적 지식의 신뢰성은 무작위 교차 시험에서 발견된 응고의 차이로 인해 크게 향상되었다[214]. 다양한 유형의 연구에서 나온 기존 외부 증거에 대한 논의가 항상 포함되어야 하지만, 위험의 작은 증가를 보고하는 연구에서는 특히 중요할 수 있다. 또한 저자는 자신의 결과를 유사한 연구와 연관시키고 이상적으로는 체계적인 문헌고찰을 참조하여 새로운 연구가 기존 증거에 어떻게 영향을 미치는지 설명해야 한다.

21. 일반화 가능성: 연구 결과의 일반화 가능성(외적타당도)에 대해 고찰한다.

예시

"우리의 추정치가 다른 HIV-1 감염 환자에게 얼마나 적용할 수 있을까? 예후 모델을 개발하는 데 사용된 자료밖의 다른 자료에 적용하면 예측 모델의 정확도가 낮아지는 경향이 있기 때문에, 이는 중요한 질문이다. 우리는 모델 복잡성에 페널티를 적용하고 추정 절차에서 생략된 집단에 가장 잘 일반화된 모델을 선택하여 이

문제를 해결했다. 우리 데이터베이스에는 다양한 환경에서 치료를 받은 유럽과 북미의 여러 국가의 환자가 포함되어 있다. 환자의 범위는 남성과 여성, 청소년부터 노인까지 광범위했으며, 주요 노출 범주가 잘 표현되어 있었다. 기준치에서 면역결핍의 심각도는 측정할 수 없는 것부터 매우 심각한 것까지, 바이러스 수치는 감지할 수 없는 것부터 극도로 높은 것까지 다양했다."[215].

설명

외부 타당성 또는 적용 가능성이라고도 불리는 일반화 가능성은 연구 결과를 다른 상황에 적용할 수 있는 정도를 말한다[216]. 외적타당도는 명확하게 명시된 조건과 관련해서만 의미가 있다[217]. 연령, 성별, 민족, 질병의 중증도 및 동반 질환과 관련하여 연구에 등록한 사람들과 다른 개인, 그룹 또는 인구 집단에 결과를 적용할 수 있을까? 노출의 성격과 수준이 비교 가능하고 결과의 정의가 다른 환경이나 모집단과 관련이 있을까? 수년 전에 추적 연구를 통해 수집된 자료가 오늘날에도 여전히 관련이 있을까? 한 국가의 의료 서비스 연구 결과가 다른 국가의 의료 시스템에 적용 가능할까? 연구 결과가 외적 타당도를 가지고 있는지 여부에 대한 질문은 종종 연구 환경, 참가자의 특성, 조사된 노출 및 평가된 결과에 따라 판단의 문제이다. 따라서 저자가 독자에게 환경과 위치, 자격 기준, 노출 및 측정 방법, 결과 정의, 모집 및 추적 기간에 대한 적절한 정보를 제공하는 것이 중요하다. 비참여 정도와 결과가 나타나는 노출되지 않은 참가자의 비율도 관련이 있다. 모집단에 따라 종종 달라지는 절대 위험과 노출 확산에 대한 지식은 결과를 다른 설정 및 모집단에 적용할 때 도움이 된다 (박스 7 참조).

기타 정보

22. 자금원: 있다면, 현재 논문의 기초가 된 연구의 자금원과 자금지원자의 역할을 제시한다.

설명

일부 저널에서는 저자에게 재정적 및 기타 이해 상충의 유무를 공개하도록 요구한다[100,218]. 여러 조사에서는 자금 출처와 연구 논문의 결론 사이에 강한 연관성이 있음을 보여준다[219–222]. 무작위 임상시험의 결론은 효과 크기를 보정한 후에도 영리 단체의 자금 지원을 받은 연구의 경우 실험 약물을 더 많이 권고하였다 (OR 5.3). 다른 연구에서는 자금을 지원한 연구에 대한 담배 및 통신 산업의 영향이 있었다[224–227]. 스폰서가 정부 또는 비영리 조직인 경우에도 과도한 영향력을 행사하는 사례가 있다. 저자 또는 자금 제공자는 다음 사항에 영향을 미치는 이해 상충을 가질 수 있다: 연구 설계[228]; 노출 선택[228,229], 결과[230], 통계 방법[231], 건강결과[230] 및 연구[232]의 선택적 보고. 결과적으로 자금 제공자의 역할을 자세히 설명해야 한다: 연구의 어떤 부분에서 직접 책임을 졌는지(예: 설계, 자료 수집, 분석, 원고 초안 작성, 출판 결정)[100]. 부당한 영향을 미치는 다른 출처로는 고용주(예: 학계 연구원의 경우 대학 관리자, 정부 감독자, 특히 정부 연구원의 경우 정치적 지명자), 자문 위원회, 소송 당사자 및 특수 이익 단체가 있다.

결론

STROBE Statement는 역학 관찰 연구 보고에 대한 권고안을 제공하는 것이 목표이다. 좋은 보고는 연구의 강점과 약점을 드러내고 연구 결과의 건전한 해석과 적용을 촉진한다. STROBE Statement는 또한 관찰 연구를 계획하는 데 도움이 될 수 있으며 동료 심사자와 편집자가 원고를 평가하는 데 지침이 될 수 있다.

우리는 관찰 연구에서 투명하고 완전하게 하는 보고의 중요성을 논의하고, 체크리스트에 포함된 다양한 항목의 근거를 설명하고, 좋은 보고라고 생각하는 출판된 기사의 예를 제공하기 위해 이 설명 기사를 썼다. 여기에 제시된 자료가 작성자와 편집자가 STROBE를 사용하는 데 도움이 되기를 바란다.

우리는 연구 보고에 대한 STROBE 및 기타 권장 사항[13,233,234]이 지속적인 평가, 개선 및 필요한 경우 변경이 필요한 문서로 보아야 함을 강조한다[235,236]. 예를 들어, 병렬 그룹 무작위 시험 보고를 위한 CONSORT statement은 1990년대 중반에 처음 개발되었다[237]. 그 이후로 그룹 구성원들은 정기적으로 만나 권장 사항 수정의 필요성을 검토했다. 개정판은 2001년에 등장했으며[233] 추가 버전이 개발 중이다.

마찬가지로, 이 기사와 STROBE 체크리스트에 제시된 원칙은 새로운 증거와 비판적 의견이 축적됨에 따라 변경될 수 있다. STROBE 웹 사이트(http://www.strobe-statement.org/)는 체크리스트, 이 설명 문서 및 역학 연구의 올바른 보고에 대한 정보의 개선을 위한 고찰과 제안을 위한 포럼을 제공한다. 몇몇 저널에서는 저자들에게 STROBE Statement를 따르도록 요청한다(현재 목록은 http://www.strobe-statement.org/ 참조). 우리는 다른 저널에도 STROBE 선언문을 채택하도록 요청하고 당사 웹사이트를 통해 연락하여 이를 알려드린다. STROBE 권장 사항을 출판하는 저널은 오픈 액세스로 제공한다. 따라서 STROBE Statement는 생물의학 커뮤니티에서 널리 접근할 수 있다.

Authors' contributions
Translation: Kim SY
Scientific review: Kim HC

Conflict of interest
The authors have declared that no competing interests exist. The translators and reviewers have declared that no competing interests exist.

Funding
The initial STROBE workshop was funded by the European Science Foundation (ESF). Additional funding was received from the Medical Research Council Health Services Research Collaboration and the National Health Services Research & Development Methodology Programme. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability
Not applicable.

Acknowledgments
Not applicable.

Supplementary materials
Not applicable.

References

1. Glasziou P, Vandenbroucke J, Chalmers I. Assessing the quality of research. *BMJ* 2004;328(7430):39-41. https://doi.org/10.1136/bmj.328.7430.39
2. Funai EF, Rosenbush EJ, Lee MJ, Del Priore G. Distribution of study designs in four major US journals of obstetrics and gynecology. *Gynecol Obstet Invest* 2001;51(1):8-11. https://doi.org/10.1159/000052882
3. Scales CD, Norris RD, Peterson BL, Preminger GM, Dahm P. Clinical research and statistical methods in the urology literature. *J Urol* 2005;174(4 Pt 1):1374-1379. https://doi.org/10.1097/01.ju.0000173640.91654.b5
4. Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, Kalish LA, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ* 2004;329(7471):883. https://doi.org/10.1136/bmj.38250.571088.55
5. Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. *Am J Epidemiol* 2005;161(3):280-288. https://doi.org/10.1093/aje/kwi042
6. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med* 2007;4(10):e296. https://doi.org/10.1371/journal.pmed.0040296
7. Mihailovic A, Bell CM, Urbach DR. Users' guide to the surgical literature. Case-control studies in surgical journals. *Can J Surg* 2005;48(2):148-151.
8. Rushton L. Reporting of occupational and environmental research: use and misuse of statistical and epidemiological methods. *Occup Environ Med* 2000;57(1):1-9. https://doi.org/10.1136/oem.57.1.1
9. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology* 1990;1(1):43-46. https://doi.org/10.1097/00001648-199001000-00010
10. Moonesinghe R, Khoury MJ, Janssens ACJW. Most published research findings are false—but a little replication goes a long way. *PLoS Med* 2007;4(2):e28. https://doi.org/10.1371/journal.pmed.0040028

11. Jenicek M. Clinical case reporting in evidence-based medicine. Oxford: Butterworth-Heinemann; 1999.
12. Vandenbroucke JP. In defense of case reports and case series. *Ann Intern Med* 2001;134(4):330-334. https://doi.org/10.7326/0003-4819-134-4-200102200-00017
13. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Intern Med* 2003;138(1):41-44.
14. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005;93(4):387-391.
15. Ioannidis JPA, Gwinn M, Little J, Higgins JPT, Bernstein JL, Boffetta P, et al. A road map for efficient and reliable human genome epidemiology. *Nat Genet* 2006;38(1):3-5. https://doi.org/10.1038/ng0106-3
16. Rodrigues L, Kirkwood BR. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. *Int J Epidemiol* 1990;19(1):205-213. https://doi.org/10.1093/ije/19.1.205
17. Rothman KJ, Greenland S. Case-control studies. In: Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed. New York: Lippincott Raven; 1998. p.93-114.
18. Forand SP. Leukaemia incidence among workers in the shoe and boot manufacturing industry: a case-control study. *Environ Health* 2004;3(1):7. https://doi.org/10.1186/1476-069X-3-7
19. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;342:1878-1886. https://doi.org/10.1056/NEJM200006223422506
20. Gøtzsche PC, Harden A. Searching for non-randomised studies [Internet]. London (UK): Cochrane; c2002 [cited 2007 Sep 10]. Available from: http://www.cochrane.dk/nrsmg
21. Lohse N, Hansen ABE, Pedersen G, Kronborg G, Gerstoft J, Sørensen HT, et al. Survival of persons with and without HIV infection in Denmark, 1995–2005. *Ann Intern Med* 2007;146(2):87-95. https://doi.org/10.7326/0003-4819-146-2-200701160-00003
22. American Journal of Epidemiology. Information for authors [Internet]. Oxford (UK): Oxford University Press; c2007 [cited 2007 Sep 10]. Available from: http://www.oxfordjournals.org/aje/for_authors/index.html
23. Brian Haynes R, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med* 1990;113(1):69-76. https://doi.org/10.7326/0003-4819-113-1-69
24. Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR, et al. Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. *CMAJ* 1994;150(10):1611-1615.
25. Hartley J, Sydes M. Which layout do you prefer? An analysis of readers' preferences for different typographic layouts of structured abstracts. *J Inf Sci* 1996;22(1):27-37. https://doi.org/10.1177/016555159602200103
26. Viner RM, Cole TJ. Adult socioeconomic, educational, social, and psychological outcomes of childhood obesity: a national birth cohort study. *BMJ* 2005;330(7504):1354. https://doi.org/10.1136/bmj.38453.422049.E0
27. McCauley J, Kern DE, Kolodner K, Dill L, Schroeder AF, DeChant HK, et al. The "battering syndrome": prevalence and clinical characteristics of domestic violence in primary care internal medicine practices. *Ann Intern Med* 1995;123(10):737-746. https://doi.org/10.7326/0003-4819-123-10-199511150-00001
28. McEvoy SP, Stevenson MR, McCartt AT, Woodward M, Haworth C, Palamar P, et al. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ* 2005;331(7514):428. https://doi.org/10.1136/bmj.38537.397512.55
29. Vandenbroucke JP. Prospective or retrospective: what's in a name? *BMJ* 1991;302(6771):249-250. https://doi.org/10.1136/bmj.302.6771.249
30. Last JM. A dictionary of epidemiology. New York: Oxford University Press; 2000.
31. Miettinen OS. Theoretical epidemiology: principles of occurrence research in medicine. New York: Wiley; 1985.
32. Rothman KJ, Greenland S. Types of epidemiologic studies. In: Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed. New York: Lippincott Raven; 1998. p.74-75.
33. MacMahon B, Trichopoulos D. Epidemiology: principles and methods. 2nd ed. Boston: Little, Brown; 1996.
34. Lilienfeld AM. Foundations of epidemiology. New York: Oxford University Press; 1976.
35. Ridker PM, Hennekens CH, Lindpaintner K, Stampfer MJ, Eisenberg PR, Miletich JP. Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. *N Engl J Med* 1995;332(14):912-917. https://doi.org/10.1056/NEJM199504063321403
36. Goodman KJ, O'Rourke K, Sue Day R, Wang C, Nurgalieva Z, Phillips CV, et al. Dynamics of *Helicobacter pylori* infection in a US–Mexico cohort during the first two years of life. *Int J Epidemiol* 2005;34(6):1348-1355. https://doi.org/10.1093/ije/dyi152
37. Altman DG, De Stavola BL, Love SB, Stepniewska KA. Review of survival analyses published in cancer journals. *Br J Cancer* 1995;72(2):511-518. https://doi.org/10.1038/bjc.1995.364

38. Cerhan JR, Wallace RB, Folsom AR, Potter JD, Munger RG, Prineas RJ, et al. Transfusion history and cancer risk in older women. *Ann Intern Med* 1993;119(1):8-15. https://doi.org/10.7326/0003-4819-119-1-199307010-00002
39. Beane Freeman LE, Dennis LK, Lynch CF, Thorne PS, Just CL. Toenail arsenic content and cutaneous melanoma in Iowa. *Am J Epidemiol* 2004;160(7):679-687. https://doi.org/10.1093/aje/kwh267
40. Canto JG, Allison JJ, Kiefe CI, Fincher C, Farmer R, Sekar P, et al. Relation of race and sex to the use of reperfusion therapy in Medicare beneficiaries with acute myocardial infarction. *N Engl J Med* 2000;342(15):1094-1100. https://doi.org/10.1056/NEJM200004133421505
41. Metzkor-Cotter E, Kletter Y, Avidor B, Varon M, Golan Y, Ephros M, et al. Long-term serological analysis and clinical follow-up of patients with cat scratch disease. *Clin Infect Dis* 2003;37(9):1149-1154. https://doi.org/10.1086/378738
42. Johnson ES. Bias on withdrawing lost subjects from the analysis at the time of loss, in cohort mortality studies, and in follow-up methods. *J Occup Environ Med* 1990;32(3):250-254. https://doi.org/10.1097/00043764-199003000-00013
43. Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Biom Bull* 1946;2(3):47-53. https://doi.org/10.2307/3002000
44. Feinstein AR, Walter SD, Horwitz RI. An analysis of Berkson's bias in case-control studies. *J Chronic Dis* 1986;39(7):495-504. https://doi.org/10.1016/0021-9681(86)90194-3
45. Jick H, Vessey MP. Case-control studies in the evaluation of drug-induced illness. *Am J Epidemiol* 1978;107(1):1-7. https://doi.org/10.1093/oxfordjournals.aje.a112502
46. Hackam DG, Mamdani M, Li P, Redelmeier DA. Statins and sepsis in patients with cardiovascular disease: a population-based cohort analysis. *Lancet* 2006;367(9508):413-418. https://doi.org/10.1016/S0140-6736(06)68041-0
47. Smeeth L, Cook C, Fombonne E, Heavey L, Rodrigues LC, Smith PG, et al. MMR vaccination and pervasive developmental disorders: a case-control study. *Lancet* 2004;364(9438):963-969. https://doi.org/10.1016/S0140-6736(04)17020-7
48. Costanza MC. Matching. *Prev Med* 1995;24(5):425-433. https://doi.org/10.1006/pmed.1995.1069
49. Stürmer T, Brenner H. Flexible matching strategies to increase power and efficiency to detect and estimate gene-environment interactions in case-control studies. *Am J Epidemiol* 2002;155(7):593-602. https://doi.org/10.1093/aje/155.7.593
50. Rothman KJ, Greenland S. Matching. In: Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed. New York: Lippincott Raven; 1998. p.147-161.
51. Szklo MF, Javier Nieto F. Epidemiology, beyond the basics. Sudbury: Jones & Bartlett Learning; 2000. p.40-51.
52. Cole P, MacMahon B. Attributable risk percent in case-control studies. *Br J Prev Soc Med* 1971;25(4):242-244. https://doi.org/10.1136/jech.25.4.242
53. Gissler M, Hemminki E. The danger of overmatching in studies of the perinatal mortality and birthweight of infants born after assisted conception. *Eur J Obstet Gynecol Reprod Biol* 1996;69(2):73-75. https://doi.org/10.1016/0301-2115(95)02517-0
54. Gefeller O, Pfahlberg A, Brenner H, Windeler J. An empirical investigation on matching in published case–control studies. *Eur J Epidemiol* 1998;14(4):321-325. https://doi.org/10.1023/a:1007497104800
55. Artama M, Ritvanen A, Gissler M, Isojärvi J, Auvinen A. Congenital structural anomalies in offspring of women with epilepsy: a population-based cohort study in Finland. *Int J Epidemiol* 2006;35(2):280-287. https://doi.org/10.1093/ije/dyi234
56. Ebrahim S. Cohorts, infants and children. *Int J Epidemiol* 2004;33(6):1165-1166. https://doi.org/10.1093/ije/dyh368
57. Walker M, Whincup PH, Shaper AG. The British Regional Heart Study 1975–2004. *Int J Epidemiol* 2004;33(6):1185-1192. https://doi.org/10.1093/ije/dyh295
58. Wieland S, Dickersin K. Selective exposure reporting and Medline indexing limited the search sensitivity for observational studies of the adverse effects of oral contraceptives. *J Clin Epidemiol* 2005;58(6):560-567. https://doi.org/10.1016/j.jclinepi.2004.11.018
59. Anderson HR, Atkinson RW, Peacock JL, Sweeting MJ, Marston L. Ambient particulate matter and health effects: publication bias in studies of short-term associations. *Epidemiology* 2005;16(2):155-163. https://doi.org/10.1097/01.ede.0000152528.22746.0f
60. Winkelmayer WC, Stampfer MJ, Willett WC, Curhan GC. Habitual caffeine intake and the risk of hypertension in women. *JAMA* 2005;294(18):2330-2335. https://doi.org/10.1001/jama.294.18.2330
61. Lukanova A, Söderberg S, Kaaks R, Jellum E, Stattin P. Serum adiponectin is not associated with risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 2006;15(2):401-402. https://doi.org/10.1158/1055-9965.EPI-05-0836
62. Becher H. The concept of residual confounding in regression models and some applications. *Stat Med* 1992;11(13):1747-1758. https://doi.org/10.1002/sim.4780111308

63. Brenner H, Blettner M. Controlling for continuous confounders in epidemiologic research. *Epidemiology* 1997;8(4):429-434. https://doi.org/10.1097/00001648-199707000-00014
64. Phillips MR, Yang G, Zhang Y, Wang L, Ji H, Zhou M, et al. Risk factors for suicide in China: a national case-control psychological autopsy study. *Lancet* 2002;360(9347):1728-1736. https://doi.org/10.1016/S0140-6736(02)11681-3
65. Pasquale LR, Kang JH, Manson JE, Willett WC, Rosner BA, Hankinson SE, et al. Prospective study of type 2 diabetes mellitus and risk of primary open-angle glaucoma in women. *Ophthalmology* 2006;113(7):1081-1086. https://doi.org/10.1016/j.ophtha.2006.01.066
66. Craig SL, Feinstein AR. Antecedent therapy versus detection bias as causes of neoplastic multimorbidity. *Am J Clin Oncol* 1999;22(1):51-56. https://doi.org/10.1097/00000421-199902000-00013
67. Rogler LH, Mroczek DK, Fellows M, Loftus ST. The neglect of response bias in mental health research. *J Nerv Ment Dis* 2001;189(3):182-187. https://doi.org/10.1097/00005053-200103000-00007
68. Murphy EA. The logic of medicine. Baltimore: Johns Hopkins University Press; 1976.
69. Sackett DL. Bias in analytic research. *J Chronic Dis* 1979;32(1–2):51-63. https://doi.org/10.1016/0021-9681(79)90012-2
70. Johannes CB, Crawford SL, McKinlay JB. Interviewer effects in a cohort study: results from the Massachusetts Women's Health Study. *Am J Epidemiol* 1997;146(5):429-438. https://doi.org/10.1093/oxfordjournals.aje.a009296
71. Bloemenkamp KWM, Rosendaal FR, Büller HR, Helmerhorst FM, Colly LP, Vandenbroucke JP, et al. Risk of venous thrombosis with use of current low-dose oral contraceptives is not explained by diagnostic suspicion and referral bias. *Arch Intern Med* 1999;159(1):65-70. https://doi.org/10.1001/archinte.159.1.65
72. Feinstein AR. Clinical epidemiology: the architecture of clinical research. Philadelphia: W.B. Saunders Company; 1985.
73. Yadon ZE, Rodrigues LC, Davies CR, Quigley MA. Indoor and peridomestic transmission of American cutaneous leishmaniasis in northwestern Argentina: a retrospective case-control study. *Am J Trop Med Hyg* 2003;68(5):519-526. https://doi.org/10.4269/ajtmh.2003.68.519
74. Anoop S, Saravanan B, Joseph A, Cherian A, Jacob KS. Maternal depression and low maternal intelligence as risk factors for malnutrition in children: a community based case-control study from South India. *Arch Dis Child* 2004;89(4):325-329. https://doi.org/10.1136/adc.2002.009738
75. Carlin JB, Doyle LW. Sample size. *J Paediatr Child Health* 2002;38(3):300-304. https://doi.org/10.1046/j.1440-1754.2002.00855.x
76. Rigby AS, Vail A. Statistical methods in epidemiology. II: A commonsense approach to sample size estimation. *Disabil Rehabil* 1998;20(11):405-410. https://doi.org/10.3109/09638289809166102
77. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet* 2005;365(9467):1348-1353. https://doi.org/10.1016/S0140-6736(05)61034-3
78. Drescher K, Timm J, Jockel KH. The design of case-control studies: the effect of confounding on sample size requirements. *Stat Med* 1990;9(7):765-776. https://doi.org/10.1002/sim.4780090706
79. Devine OJ, Smith JM. Estimating sample size for epidemiologic studies: the impact of ignoring exposure measurement uncertainty. *Stat Med* 1998;17(12):1375-1389. https://doi.org/10.1002/(SICI)1097-0258(19980630)17:12<1375::AID-SIM857>3.0.CO;2-D
80. Linn S, Levi L, Grunau PD, Zaidise I, Zarka S. Effect measure modification and confounding of severe head injury mortality by age and multiple organ injury severity. *Ann Epidemiol* 2007;17(2):142-147. https://doi.org/10.1016/j.annepidem.2006.08.004
81. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 1994;86(11):829-835. https://doi.org/10.1093/jnci/86.11.829
82. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006;25(1):127-141. https://doi.org/10.1002/sim.2331
83. Greenland S. Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. *Epidemiology* 1995;6(4):450-454. https://doi.org/10.1097/00001648-199507000-00025
84. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 1999;28(5):964-974. https://doi.org/10.1093/ije/28.5.964
85. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychol Methods* 2002;7(1):19-40. https://doi.org/10.1037/1082-989X.7.1.19
86. Altman DG. Categorizing continuous variables. In: Armitage P, Colton T, editors. Encyclopedia of biostatistics. 2nd ed.

Chichester: John Wiley & Sons; 2005. p.708-711.
87. Cohen J. The cost of dichotomization. *Appl Psychol Meas* 1983;7(3):249-253. https://doi.org/10.1177/014662168300700301
88. Zhao LP, Kolonel LN. Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies. *Am J Epidemiol* 1992;136:464-474. https://doi.org/10.1093/oxfordjournals.aje.a116520
89. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* 1968;24(2):295-313. https://doi.org/10.2307/2528036
90. Clayton D, Hills M. Models for dose-response. In: Clayton D, Hills M, editors. Statistical models in epidemiology. Oxford: Oxford University Press; 1993. p.249-260.
91. Cox DR. Note on grouping. *J Am Stat Assoc* 1957;52(280):543-547. https://doi.org/10.1080/01621459.1957.10501411
92. Il'yasova D, Hertz-Picciotto I, Peters U, Berlin JA, Poole C. Choice of exposure scores for categorical regression in meta-analysis: a case study of a common problem. *Cancer Causes Control* 2005;16(4):383-388. https://doi.org/10.1007/s10552-004-5025-x
93. Berglund A, Alfredsson L, David Cassidy J, Jensen I, Nygren A. The association between exposure to a rear-end collision and future neck or shoulder pain: a cohort study. *J Clin Epidemiol* 2000;53(11):1089-1094. https://doi.org/10.1016/S0895-4356(00)00225-0
94. Slama R, Werwatz A. Controlling for continuous confounding factors: non- and semiparametric approaches. *Rev Epidemiol Sante Publique* 2005;53(2S):65-80. https://doi.org/10.1016/S0398-7620(05)84769-8
95. Greenland S. Introduction to regression modelling. In: Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed. New York: Lippincott Raven; 1998. p.401-432.
96. Douglas Thompson W. Statistical analysis of case-control studies. *Epidemiol Rev* 1994;16(1):33-50. https://doi.org/10.1093/oxfordjournals.epirev.a036143
97. Schlesselman JJ. Logistic regression for case-control studies. In: Schlesselman JJ, editor. Case-control studies: design, conduct, analysis. Oxford: Oxford University Press; 1982. p.235-241.
98. Clayton D, Hills M. Choice and interpretation of models. In: Clayton D, Hills M, editors. Statistical models in epidemiology. Oxford: Oxford University Press; 1993. p.271-281.
99. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *Br Med J* 1983;286(6376):1489-1493. https://doi.org/10.1136/bmj.286.6376.1489
100. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. *N Engl J Med* 1997;336(4):309-315. https://doi.org/10.1056/NEJM199701233360422
101. Müllner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med* 2002;136(2):122-126. https://doi.org/10.7326/0003-4819-136-2-200201150-00009
102. Olsen J, Basso O. Re: residual confounding. *Am J Epidemiol* 1999;149(3):290. https://doi.org/10.1093/oxfordjournals.aje.a009805
103. Hallan S, de Mutsert R, Carlsen S, Dekker FW, Aasarød K, Holmen J. Obesity, smoking, and physical inactivity as risk factors for CKD: are men more vulnerable? *Am J Kidney Dis* 2006;47(3):396-405. https://doi.org/10.1053/j.ajkd.2005.11.027
104. Gøtzsche PC. Believability of relative risks and odds ratios in abstracts: cross sectional study. *BMJ* 2006;333(7561):231-234. https://doi.org/10.1136/bmj.38895.410451.79
105. Szklo MF, Nieto J. Communicating results of epidemiologic studies. In: Szklo MF, Nieto J, editors. Epidemiology, beyond the basics. Sudbury: Jones and Bartlett; 2000. p.408-430.
106. Chandola T, Brunner E, Marmot M. Chronic stress at work and the metabolic syndrome: prospective study. *BMJ* 2006;332(7540):521-525. https://doi.org/10.1136/bmj.38693.435301.80
107. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of *ad hoc* methods of correcting for missing values for confounding variables. *Am J Epidemiol* 1991;134(8):895-907. https://doi.org/10.1093/oxfordjournals.aje.a116164
108. Little RJ, Rubin DB. A taxonomy of missing-data methods. In: Little RJ, Rubin DB, editors. Statistical analysis with missing data. New York: Wiley; 2002. p.19-23.
109. Ware JH. Interpreting incomplete data in studies of diet and weight loss. *N Engl J Med* 2003;348(21):2136-2137. https://doi.org/10.1056/NEJMe030054
110. Rubin DB. Inference and missing data. *Biometrika* 1976;63(3):581-592. https://doi.org/10.1093/biomet/63.3.581
111. Schafer JL. Analysis of incomplete multivariate data. London: Chapman & Hall; 1997.
112. Lipsitz SR, Ibrahim JG, Chen MH, Peterson H. Non-ignorable missing covariates in generalized linear models. *Stat Med* 1999;18(17-18):2435-2448. https://doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2435::AID-SIM267>3.0.CO;2-B

113. Rotnitzky A, Robins J. Analysis of semi-parametric regression models with non-ignorable non-response. *Stat Med* 1998;16(1):81-102.
https://doi.org/10.1002/(SICI)1097-0258(19970115)16:1<81::AID-SIM473>3.0.CO;2-0
114. Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons; 1987.
115. Barnard J, Meng XL. Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat Methods Med Res* 1999;8(1):17-36.
https://doi.org/10.1177/096228029900800103
116. Braitstein P, Brinkhof MWG, Dabis F, Schechter M, Boulle A, Miotti P, et al. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet* 2006;367(9513):817-824.
https://doi.org/10.1016/S0140-6736(06)68337-2
117. Purandare N, Burns A, Daly KJ, Hardicre J, Morris J, Macfarlane G, et al. Cerebral emboli as a potential cause of Alzheimer's disease and vascular dementia: case-control study. *BMJ* 2006;332(7550):1119-1124.
https://doi.org/10.1136/bmj.38814.696493.AE
118. Steyn K, Gaziano TA, Bradshaw D, Laubscher R, Fourie J. Hypertension in South African adults: results from the Demographic and Health Survey, 1998. *J Hypertens* 2001;19(10):1717-1725.
https://doi.org/10.1097/00004872-200110000-00004
119. Lohr SL. Design effects. Sampling: design and analysis. Pacific Grove: Duxbury Press; 1999.
120. Dunn NR, Arscott A, Thorogood M. The relationship between use of oral contraceptives and myocardial infarction in young women with fatal outcome, compared to those who survive: results from the MICA case-control study. *Contraception* 2001;63(2):65-69.
https://doi.org/10.1016/S0010-7824(01)00172-X
121. Rothman KJ, Greenland S. Basic methods for sensitivity analysis and external adjustment. In: Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed. New York: Lippincott Raven; 1998. p.343-358.
122. Custer B, Longstreth WT Jr, Phillips LE, Koepsell TD, Van Belle G. Hormonal exposures and the risk of intracranial meningioma in women: a population-based case-control study. *BMC Cancer* 2006;6:152.
https://doi.org/10.1186/1471-2407-6-152
123. Wakefield MA, Chaloupka FJ, Kaufman NJ, Tracy Orleans C, Barker DC, Ruel EE, et al. Effect of restrictions on smoking at home, at school, and in public places on teenage smoking: cross sectional study. *BMJ* 2000;321(7257):333-337.
https://doi.org/10.1136/bmj.321.7257.333
124. Greenland S. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Stat Assoc* 2003;98(461):47-54.
https://doi.org/10.1198/01621450338861905
125. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003;14(4):451-458.
https://doi.org/10.1097/01.EDE.0000071419.41011.cf
126. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 2003;14(4):459-466.
https://doi.org/10.1097/01.ede.0000072106.65262.ae
127. Cornfield J, Haenszel W, Cuyler Hammond E, Lilienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst* 1959;22(1):173-203.
https://doi.org/10.1093/jnci/22.1.173
128. Langholz B. Factors that explain the power line configuration wiring code–childhood leukemia association: what would they look like? *Bioelectromagnetics* 2001;22(S5):S19-S31.
https://doi.org/10.1002/1521-186X(2001)22:5+<::AID-BEM1021>3.0.CO;2-I
129. Eisner MD, Smith AK, Blanc PD. Bartenders' respiratory health after establishment of smoke-free bars and taverns. *JAMA* 1998;280(22):1909-1914.
https://doi.org/10.1001/jama.280.22.1909
130. Dunne MP, Martin NG, Bailey JM, Heath AC, Bucholz KK, Madden PA, et al. Participation bias in a sexuality survey: psychological and behavioural characteristics of responders and non-responders. *Int J Epidemiol* 1997;26(4):844-854.
https://doi.org/10.1093/ije/26.4.844
131. Schüz J, Kaatsch P, Kaletsch U, Meinert R, Michaelis J. Association of childhood cancer with factors related to pregnancy and birth. *Int J Epidemiol* 1999;28(4):631-639.
https://doi.org/10.1093/ije/28.4.631
132. Cnattingius S, Zack M, Ekbom A, Gunnarskog J, Linet M, Adami HO. Prenatal and neonatal risk factors for childhood myeloid leukemia. *Cancer Epidemiol Biomarkers Prev* 1995;4(5):441-445.
133. Schüz J. Non-response bias as a likely cause of the association between young maternal age at the time of delivery and the risk of cancer in the offspring. *Paediatr Perinat Epidemiol* 2003;17(1):106-112.
https://doi.org/10.1046/j.1365-3016.2003.00460.x
134. Slattery ML, Edwards SL, Caan BJ, Kerber RA, Potter JD. Response rates among control subjects in case-control studies. *Ann Epidemiol* 1995;5(3):245-249.
https://doi.org/10.1016/1047-2797(94)00113-8
135. Schulz KF, Grimes DA. Case-control studies: research in reverse. *Lancet* 2002;359(9304):431-434.
https://doi.org/10.1016/S0140-6736(02)07605-5
136. Olson SH, Voigt LF, Begg CB, Weiss NS. Reporting participation in case-control studies. *Epidemiology* 2002;13(2):123-126.
https://doi.org/10.1097/00001648-200203000-00004

137. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. *Am J Epidemiol* 2006;163(3):197-203.
https://doi.org/10.1093/aje/kwj036
138. Olson SH. Reported participation in case-control studies: changes over time. *Am J Epidemiol* 2001;154(6):574-581.
https://doi.org/10.1093/aje/154.6.574
139. Sandler DP. On revealing what we'd rather hide: the problem of describing study participation. *Epidemiology* 2002;13(2):117.
https://doi.org/10.1097/00001648-200203000-00001
140. Hepworth SJ, Schoemaker MJ, Muir KR, Swerdlow AJ, van Tongeren MJA, McKinney PA. Mobile phone use and risk of glioma in adults: case-control study. *BMJ* 2006;332(7546):883-887.
https://doi.org/10.1136/bmj.38720.687975.55
141. Hay AD, Wilson A, Fahey T, Peters TJ. The duration of acute cough in pre-school children presenting to primary care: a prospective cohort study. *Fam Pract* 2003;20(6):696-705.
https://doi.org/10.1093/fampra/cmg613
142. Egger M, Jüni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001;285(15):1996-1999.
https://doi.org/10.1001/jama.285.15.1996
143. Osella AR, Misciagna G, Guerra VM, Chiloiro M, Cuppone R, Cavallini A, et al. Hepatitis C virus (HCV) infection and liver-related mortality: a population-based cohort study in southern Italy. *Int J Epidemiol* 2000;29(5):922-927.
https://doi.org/10.1093/ije/29.5.922
144. Dales LG, Ury HK. An improper use of statistical significance testing in studying covariables. *Int J Epidemiol* 1978;7(4):373-376.
https://doi.org/10.1093/ije/7.4.373
145. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. *Am J Epidemiol* 1993;138(11):923-936.
https://doi.org/10.1093/oxfordjournals.aje.a116813
146. Tanis BC, van den Bosch MAAJ, Kemmeren JM, Cats VM, Helmerhorst FM, Algra A, et al. Oral contraceptives and the risk of myocardial infarction. *N Engl J Med* 2001;345(25):1787-1793.
https://doi.org/10.1056/NEJMoa003216
147. Rothman KJ, Greenland S. Precision and validity in epidemiologic studies. In: Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed. New York: Lippincott Raven; 1998. p.120-125.
148. Clark TG, Altman DG, De Stavola BL. Quantification of the completeness of follow-up. *Lancet* 2002;359(9314):1309-1310.
https://doi.org/10.1016/S0140-6736(02)08272-7
149. Qiu C, Fratiglioni L, Karp A, Winblad B, Bellander T. Occupational exposure to electromagnetic fields and risk of Alzheimer's disease. *Epidemiology* 2004;15(6):687-694.
https://doi.org/10.1097/01.ede.0000142147.49297.9d
150. Kengeya-Kayondo JF, Kamali A, Nunn AJ, Ruberantwari A, Wagner HUH, Mulder DW. Incidence of HIV-1 infection in adults and socio-demographic characteristics of seroconverters in a rural population in Uganda: 1990–1994. *Int J Epidemiol* 1996;25(5):1077-1082.
https://doi.org/10.1093/ije/25.5.1077
151. Mastrangelo G, Fedeli U, Fadda E, Valentini F, Agnesi R, Magarotto G, et al. Increased risk of hepatocellular carcinoma and liver cirrhosis in vinyl chloride workers: synergistic effect of occupational exposure with alcohol intake. *Environ Health Perspect* 2004;112(11):1188-1192.
https://doi.org/10.1289/ehp.6972
152. Salo PM, Arbes SJ Jr, Sever M, Jaramillo R, Cohn RD, London SJ, et al. Exposure to *Alternaria alternata* in US homes is associated with asthma symptoms. *J Allergy Clin Immunol* 2006;118(4):892-898.
https://doi.org/10.1016/j.jaci.2006.07.037
153. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet* 2002;359(9318):1686-1689.
https://doi.org/10.1016/S0140-6736(02)08594-X
154. Sasieni P. A note on the presentation of matched case-control data. *Stat Med* 1992;11(5):617-620.
https://doi.org/10.1002/sim.4780110506
155. Lee GM, Neutra RR, Hristova L, Yost M, Hiatt RA. A nested case-control study of residential and personal magnetic field measures and miscarriages. *Epidemiology* 2002;13(1):21-31.
https://doi.org/10.1097/00001648-200201000-00005
156. Tiihonen J, Walhbeck K, Lönnqvist J, Klaukka T, Ioannidis JPA, Volavka J, et al. Effectiveness of antipsychotic treatments in a nationwide cohort of patients in community care after first hospitalisation due to schizophrenia and schizoaffective disorder: observational follow-up study. *BMJ* 2006;333(7561):224.
https://doi.org/10.1136/bmj.38881.382755.2F
157. Christenfeld NJ, Sloan RP, Carroll D, Greenland S. Risk factors, confounding, and the illusion of statistical control. *Psychosom Med* 2004;66(6):868-875.
https://doi.org/10.1097/01.psy.0000140008.70959.41
158. Smith GD, Phillips A. Declaring independence: why we should be cautious. *J Epidemiol Community Health* 1990;44(4):257-258.
https://doi.org/10.1136/jech.44.4.257
159. Greenland S, Neutra R. Control of confounding in the assessment of medical technology. *Int J Epidemiol* 1980;9(4):361-367.
https://doi.org/10.1093/ije/9.4.361

160. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology* 2001;12(3):313-320. https://doi.org/10.1097/00001648-200105000-00011
161. Sagiv SK, Tolbert PE, Altshul LM, Korrick SA. Organochlorine exposures during pregnancy and infant size at birth. *Epidemiology* 2007;18(1):120-129. https://doi.org/10.1097/01.ede.0000249769.15001.7c
162. World Health Organization [WHO]. Body mass index (BMI) [Internet]. Geneva (CH): WHO; c2007 [cited 2007 Sep 10]. Available from: http://www.euro.who.int/nutrition/20030507_1
163. Beral V. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet* 2003;362(9382):419-427. https://doi.org/10.1016/S0140-6736(03)14065-2
164. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58(5):295-300. https://doi.org/10.1177/003591576505800503
165. Vineis P. Causality in epidemiology. *Soz Praventiv Med* 2003;48(2):80-87. https://doi.org/10.1007/s00038-003-1029-7
166. Empana JP, Ducimetière P, Arveiler D, Ferrières J, Evans A, Ruidavets JB, et al. Are the Framingham and PROCAM coronary heart disease risk functions applicable to different European populations? The PRIME Study. *Eur Heart J* 2003;24(21):1903-1911. https://doi.org/10.1016/j.ehj.2003.09.002
167. Tunstall-Pedoe H, Kuulasmaa K, Mähönen M, Tolonen H, Ruokokoski E, Amouyel P. Contribution of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA project populations. *Lancet* 1999;353:1547-1557. https://doi.org/10.1016/S0140-6736(99)04021-0
168. Cambien F, Chretien JM, Ducimetiere P, Guize L, Richard JL. Is the relationship between blood pressure and cardiovascular risk dependent on body mass index? *Am J Epidemiol* 1985;122(3):434-442. https://doi.org/10.1093/oxfordjournals.aje.a114124
169. Hosmer DW, Taber S, Lemeshow S. The importance of assessing the fit of logistic regression models: a case study. *Am J Public Health* 1991;81(12):1630-1635. https://doi.org/10.2105/AJPH.81.12.1630
170. Tibshirani R. A plain man's guide to the proportional hazards model. *Clin Invest Med* 1982;5(1):63-68.
171. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am J Public Health* 1998;88(1):15-19. https://doi.org/10.2105/AJPH.88.1.15
172. Uter W, Pfahlberg A. The application of methods to quantify attributable risk in medical practice. *Stat Methods Med Res* 2001;10(3):231-237. https://doi.org/10.1177/096228020101000305
173. Schwartz LM, Woloshin S, Dvorin EL, Gilbert Welch H. Ratio measures in leading medical journals: structured review of accessibility of underlying absolute risks. *BMJ* 2006;333(7581):1248. https://doi.org/10.1136/bmj.38985.564317.7C
174. Nakayama T, Zaman MM, Tanaka H. Reporting of attributable and relative risks, 1966-97. *Lancet* 1998;351(9110):1179. https://doi.org/10.1016/S0140-6736(05)79123-6
175. Cornfield J. A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. *J Natl Cancer Inst* 1951;11(6):1269-1275.
176. Pearce N. What does the odds ratio estimate in a case-control study? *Int J Epidemiol* 1993;22(6):1189-1192. https://doi.org/10.1093/ije/22.6.1189
177. Rothman KJ, Greenland S. Measures of disease frequency. In: Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed. New York: Lippincott Raven; 1998. p.44-45.
178. Doll R, Bradford Hill A. The mortality of doctors in relation to their smoking habits: a preliminary report. 1954. *Br Med J* 2004;328(7455):1529-1533. https://doi.org/10.1136/bmj.328.7455.1529
179. Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. *Lancet* 2003;362(9387):847-852. https://doi.org/10.1016/S0140-6736(03)14338-3
180. Greenland S. Applications of stratified analysis methods. In: Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed. New York: Lippincott Raven; 1998. p.295-297.
181. Rose G. Sick individuals and sick populations. *Int J Epidemiol* 2001;30(3):427-432. https://doi.org/10.1093/ije/30.3.427
182. Vandenbroucke JP, Koster T, Rosendaal FR, Briët E, Reitsma PH, Bertina RM. Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation. *Lancet* 1994;344(8935):1453-1457. https://doi.org/10.1016/S0140-6736(94)90286-0
183. Botto LD, Khoury MJ. Commentary: facing the challenge of gene-environment interaction: the two-by-four table and beyond. *Am J Epidemiol* 2001;153(10):1016-1020. https://doi.org/10.1093/aje/153.10.1016
184. Wei L, MacDonald TM, Walker BR. Taking glucocorticoids by prescription is associated with subsequent cardiovascular disease. *Ann Intern Med* 2004;141(10):764-770. https://doi.org/10.7326/0003-4819-141-10-200411160-00007
185. Martinelli I, Taioli E, Battaglioli T, Podda GM, Passamonti SM, Pedotti P, et al. Risk of venous thromboembolism after air travel: interaction with thrombophilia and oral contraceptives. *Arch Intern Med* 2003;163(22):2771-2774.

https://doi.org/10.1001/archinte.163.22.2771
186. Kyzas PA, Loizou KT, Ioannidis JPA. Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst* 2005;97(14):1043-1055.
https://doi.org/10.1093/jnci/dji184
187. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *Am J Epidemiol* 1980;112(4):467-470.
https://doi.org/10.1093/oxfordjournals.aje.a113015
188. Saracci R. Interaction and synergism. *Am J Epidemiol* 1980;112(4):465-466.
https://doi.org/10.1093/oxfordjournals.aje.a113014
189. Rothman KJ. Epidemiology: an introduction. Oxford: Oxford University Press; 2002.
190. Rothman KJ. Interactions between causes. Modern epidemiology. Boston: Little Brown; 1986.
191. Hess DR. How to write an effective discussion. *Respir Care* 2004;49(10):1238-1241.
192. Horton R. The hidden research paper. *JAMA* 2002;287(21):2775-2778.
https://doi.org/10.1001/jama.287.21.2775
193. Horton R. The rhetoric of research. *BMJ* 1995;310(6985):985-987.
https://doi.org/10.1136/bmj.310.6985.985
194. Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ* 1999;318(7193):1224-1225.
https://doi.org/10.1136/bmj.318.7193.1224
195. Perneger TV, Hudelson PM. Writing a research article: advice to beginners. *Int J Qual Health Care* 2004;16(3):191-192.
https://doi.org/10.1093/intqhc/mzh053
196. Annals of Internal Medicine. Information for authors [Internet]. Philadelphia (PA): American College of Physicians; c2007 [cited 2007 Sep 10]. Available from: http://www.annals.org/shared/author_info.html
197. Maldonado G, Poole C. More research is needed. *Ann Epidemiol* 1999;9(1):17-18.
198. Phillips CV. The economics of 'more research is needed'. *Int J Epidemiol* 2001;30(4):771-776.
https://doi.org/10.1093/ije/30.4.771
199. Winkleby MA, Kraemer HC, Ahn DK, Varady AN. Ethnic and socioeconomic differences in cardiovascular disease risk factors: findings for women from the Third National Health and Nutrition Examination Survey, 1988-1994. *JAMA* 1998;280(4):356-362.
https://doi.org/10.1001/jama.280.4.356
200. Galuska DA, Will JC, Serdula MK, Ford ES. Are health care professionals advising obese patients to lose weight? *JAMA* 1999;282(16):1576-1578.
https://doi.org/10.1001/jama.282.16.1576
201. Spearman C. The proof and measurement of association between two things. *Am J Psychol* 1904;15(1):72-101.
https://doi.org/10.2307/1412159
202. Fuller WA, Hidiroglou MA. Regression estimation after correcting for attenuation. *J Am Stat Assoc* 1978;73(361):99-104.
https://doi.org/10.1080/01621459.1978.10480011
203. MacMahon S, Peto R, Collins R, Godwin J, MacMahon S, Cutler J, et al. Blood pressure, stroke, and coronary heart disease. Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990;335(8692):765-774.
https://doi.org/10.1016/0140-6736(90)90878-9
204. Phillips AN, Smith GD. How independent are "independent" effects? Relative risk estimation when correlated exposures are measured imprecisely. *J Clin Epidemiol* 1991;44(11):1223-1231.
https://doi.org/10.1016/0895-4356(91)90155-3
205. Phillips AN, Smith GD. Bias in relative odds estimation owing to imprecise measurement of correlated exposures. *Stat Med* 1992;11(7):953-961.
https://doi.org/10.1002/sim.4780110712
206. Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980;112(4):564-569.
https://doi.org/10.1093/oxfordjournals.aje.a113025
207. Poole C, Peters U, Il'yasova D, Arab L. Commentary: this study failed? *Int J Epidemiol* 2003;32(4):534-535.
https://doi.org/10.1093/ije/dyg197
208. Kaufman JS, Cooper RS, McGee DL. Socioeconomic status and health in blacks and whites: the problem of residual confounding and the resiliency of race. *Epidemiology* 1997;8(6):621-628.
https://doi.org/10.1097/00001648-199711000-00006
209. Greenland S. Randomization, statistics, and causal inference. *Epidemiology* 1990;1(6):421-429.
https://doi.org/10.1097/00001648-199011000-00003
210. Taubes G. Epidemiology faces its limits: the search for subtle links between diet, lifestyle, or environmental factors and disease is an unending source of fear—but often yields little certainty. *Science* 1995;269(5221):164-169.
https://doi.org/10.1126/science.7618077
211. Temple R. Meta-analysis and epidemiologic studies in drug development and postmarketing surveillance. *JAMA* 1999;281(9):841-844.
https://doi.org/10.1001/jama.281.9.841
212. Greenberg RS, Shuster JL Jr. Epidemiology of cancer in children. *Epidemiol Rev* 1985;7(1):22-48.
https://doi.org/10.1093/oxfordjournals.epirev.a036284
213. Kushi LH, Mink PJ, Folsom AR, Anderson KE, Zheng W, Lazovich D, et al. Prospective study of diet and ovarian cancer. *Am J Epidemiol* 1999;149(1):21-31.
https://doi.org/10.1093/oxfordjournals.aje.a009723

214. Kemmeren JM, Algra A, Meijers JCM, Tans G, Bouma BN, Curvers J, et al. Effect of second- and third-generation oral contraceptives on the protein C system in the absence or presence of the factor V_{Leiden} mutation: a randomized trial. *Blood* 2004;103(3):927-933.
https://doi.org/10.1182/blood-2003-04-1285
215. Egger M, May M, Chêne G, Phillips AN, Ledergerber B, Dabis F, et al. Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *Lancet* 2002;360(9327):119-129.
https://doi.org/10.1016/S0140-6736(02)09411-4
216. Campbell DT. Factors relevant to the validity of experiments in social settings. *Psychol Bull* 1957;54(4):297-312.
https://doi.org/10.1037/h0040950
217. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130(6):515-524.
https://doi.org/10.7326/0003-4819-130-6-199903160-00016
218. Krimsky S, Rothenberg LS. Conflict of interest policies in science and medical journals: editorial practices and author disclosures. *Sci Eng Ethics* 2001;7(2):205-218.
https://doi.org/10.1007/s11948-001-0041-7
219. Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA* 2003;289(4):454-465.
https://doi.org/10.1001/jama.289.4.454
220. Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med* 1986;1(3):155-158.
https://doi.org/10.1007/BF02602327
221. Stelfox HT, Chua G, O'Rourke K, Detsky AS. Conflict of interest in the debate over calcium-channel antagonists. *N Engl J Med* 1998;338(2):101-106.
https://doi.org/10.1056/NEJM199801083380206
222. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326(7400):1167-1170.
https://doi.org/10.1136/bmj.326.7400.1167
223. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290(7):921-928.
https://doi.org/10.1001/jama.290.7.921
224. Barnes DE, Bero LA. Why review articles on the health effects of passive smoking reach different conclusions. *JAMA* 1998;279(19):1566-1570.
https://doi.org/10.1001/jama.279.19.1566
225. Barnes DE, Bero LA. Industry-funded research and conflict of interest: an analysis of research sponsored by the tobacco industry through the Center for Indoor Air Research. *J Health Polit Policy Law* 1996;21(3):515-542.
https://doi.org/10.1215/03616878-21-3-515
226. Glantz SA, Barnes DE, Bero L, Hanauer P, Slade J. Looking through a keyhole at the tobacco industry: the Brown and Williamson documents. *JAMA* 1995;274(3):219-224.
https://doi.org/10.1001/jama.1995.03530030039032
227. Huss A, Egger M, Hug K, Huwiler-Müntener K, Röösli M. Source of funding and results of studies of health effects of mobile phone use: systematic review of experimental studies. *Environ Health Perspect* 2007;115(1):1-4.
https://doi.org/10.1289/ehp.9149
228. Safer DJ. Design and reporting modifications in industry-sponsored comparative psychopharmacology trials. *J Nerv Ment Dis* 2002;190(9):583-592.
https://doi.org/10.1097/00005053-200209000-00002
229. Aspinall RL, Goodman NW. Denial of effective treatment and poor quality of clinical information in placebo controlled trials of ondansetron for postoperative nausea and vomiting: a review of published trials. *BMJ* 1995;311(7009):844-846.
https://doi.org/10.1136/bmj.311.7009.844
230. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291(20):2457-2465.
https://doi.org/10.1001/jama.291.20.2457
231. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326(7400):1171-1173.
https://doi.org/10.1136/bmj.326.7400.1171
232. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2007;(2):MR000005.
https://doi.org/10.1002/14651858.MR000005.pub3
233. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357(9263):1191-1194.
https://doi.org/10.1016/S0140-6736(00)04337-3
234. Stroup DF, Berlin JA, Morton SC, Olkin I, David Williamson G, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000;283(15):2008-2012.
https://doi.org/10.1001/jama.283.15.2008
235. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134(8):663-694.

https://doi.org/10.7326/0003-4819-134-8-200104170-00012
236. Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. *JAMA* 1998;279(18):1489-1491.
https://doi.org/10.1001/jama.279.18.1489
237. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996;276(8):637-639.
https://doi.org/10.1001/jama.1996.03540080059030