J Korean Med Sci. 2025 Mar 3;40(8):e110
https://doi.org/10.3346/jkms.2025.40.e110
eISSN 1598-6357·pISSN 1011-8934

JKMS

Review Article
Medicine General & Health Policy

Statistical Methods for Baseline Adjustment and Cohort Analysis in Korean National Health Insurance Claims Data: A Review of PSM, IPTW, and Survival Analysis With Future Directions

Dong Wook Kim

Department of Information and Statistics, Department of Bio & Medical Big Data, Research Institute of Natural Science, Gyeongsang National University, Jinju, Korea

Received: Feb 18, 2025
Accepted: Feb 19, 2025
Published online: Feb 27, 2025

Address for Correspondence:
Dong Wook Kim, PhD
Department of Information and Statistics, Research Institute of Natural Science, Gyeongsang National University, 501 Jinju-daero, Jinju 52858, Republic of Korea.
Email: kimdw2269@gmail.com

ORCID iD
Dong Wook Kim
https://orcid.org/0000-0002-4478-3794

Disclosure

The author has no potential conflicts of interest to disclose.

Disclosure of Artificial Intelligence (AI)-Assisted Technology

Artificial intelligence (AI)-assisted technology was utilized in both the writing and editing processes.

ABSTRACT

The utilization of health insurance claims data has expanded significantly, enabling researchers to conduct epidemiological studies on a large scale. This review examines key statistical methods for addressing baseline differences and conducting cohort analyses using Korean National Health Insurance claims data. Propensity score matching and inverse probability of treatment weighting are widely used to mitigate selection bias and enhance causal inference in observational studies. These methods help improve study validity by balancing covariates between treatment and control groups. Additionally, survival analysis techniques, such as the Cox proportional hazards model, are essential for assessing time-to-event outcomes and estimating hazard ratios while accounting for censoring. However, the application of these statistical methods is accompanied by challenges, including unmeasured confounding, instability in weight estimation, and violations of model assumptions.
To address these limitations, emerging approaches, such as Doubly robust estimation, machine learning-based causal inference, and the marginal structural model, have gained prominence. These techniques offer greater flexibility and robustness in real-world data analysis. Future research should focus on refining methodologies for integrating high-dimensional health datasets and leveraging artificial intelligence to enhance predictive modeling and causal inference. Furthermore, the expansion of international collaborations and the adoption of standardized data models will facilitate large-scale multi-center studies. Ethical considerations, including data privacy and algorithmic transparency, should also be prioritized to ensure responsible data use. Maximizing the utility of health insurance claims data requires interdisciplinary collaboration, methodological advancements, and the implementation of rigorous statistical techniques to support evidence-based healthcare policy and improve public health outcomes.

Keywords: Korean National Health Insurance Claims Data; Selection Bias; Propensity Score; Cox Proportional Hazard Model; Inverse Probability of Treatment Weighting

INTRODUCTION

The widespread use of health insurance claims data for research began in 2014 with the release of the National Health Insurance Service (NHIS) sample cohort database. Before this, access to the data was limited to researchers familiar with its structure and variable characteristics, as the acquisition process was complex. Many researchers anticipated that studies utilizing the sample cohort database would lead to high-quality scholarly publications based on prior studies leveraging large-scale health insurance data. However, these expectations were not fully realized. This discrepancy arose primarily due to a limited understanding of the characteristics of claims data and a failure to account for the inherent limitations of big data.

By nature, claims data are not generated explicitly for research purposes. Instead, they constitute an extensive dataset produced automatically through administrative processes, lacking a clearly defined clinical objective. To effectively utilize these data for research, it is essential to comprehend their structure and apply appropriate epidemiologic study designs to establish causal relationships, thereby ensuring meaningful and reliable research outcomes.[1] The growing interest in big data research using claims data stems from its capacity to capture real-time patterns and trends without requiring predefined objectives to investigate public health concerns in response to evolving environmental changes. As a result, researchers have increasingly leveraged these data to examine disease patterns, treatment effects, and broader public health trends.[2]

As the healthcare landscape continues to evolve, big data-driven epidemiologic research is expected to play an increasingly crucial role in shaping healthcare policies and interventions. The integration of diverse health record datasets, including data fusion and biomedical big data initiatives, is anticipated to further enhance the potential of big data in public health research. To fully realize this potential, improvements in data quality, analytical methodologies, epidemiological approaches, and interdisciplinary collaboration will be essential.[3-5]

Objectives

This review aims to identify common challenges faced by novice researchers when utilizing health insurance big data, highlight key considerations for conducting robust research, and provide strategies to prevent premature abandonment of studies before yielding meaningful results. A widely recognized issue in research involving such data is the failure to properly define the study population and establish a control group. However, study populations can often be determined through operational definitions based on disease characteristics within each medical specialty. Furthermore, the increasing number of studies using health insurance big data provides reference materials that can help address methodological challenges.

Since defining a study population is a prerequisite for causal inference through comparison with a control group, ensuring similarity between groups at the outset of a study is critical. As previously mentioned, claims data are not collected for clinical research purposes and do not have a predefined index time for data collection. Consequently, a designated washout period is necessary to establish a clear study initiation point.[6] At this stage, controlling for covariates beyond the exposure risk factor is essential to improve research quality. This can be achieved through methods such as matching or weighting techniques.[7,8]

CHARACTERISTICS AND UTILIZATION OF KOREAN HEALTH INSURANCE CLAIMS DATA

Health insurance big data encompass information on the health of the entire population, individual health status, and medical institutions. Health information is categorized into medical treatment data—recorded using Korean Standard Classification of Diseases codes, which are localized versions of the International Classification of Diseases—and examination and questionnaire data obtained through national health screening programs. Further details on these datasets can be accessed through the websites of the NHIS and the Health Insurance Review & Assessment Service, where comprehensive manuals are available.[9]

One of the primary advantages of utilizing large claims datasets in research is the significantly larger sample size compared to clinical trials, along with the ability to track subjects longitudinally. However, a key statistical consideration is that large sample sizes substantially increase the likelihood of rejecting the null hypothesis, even when effect sizes are small. Consequently, researchers should prioritize visualizing data trends, reporting measures of variation such as confidence intervals instead of solely relying on *P* values, and employing cross-validation techniques to mitigate overfitting and enhance the generalizability of findings.[10,11]

Since claims data are not collected for clinical research purposes, medical treatments do not always adhere to standardized clinical guidelines and are often influenced by government policies. Changes in government reimbursement policies can impact patient behavior; for instance, some patients who previously visited hospitals regularly may discontinue care when financial support is reduced. Similarly, treatments that receive reimbursement only up to a predefined limit may not be continued beyond that threshold. Additionally, diagnosis codes alone do not provide direct information on treatment outcomes, necessitating the integration of procedure codes to accurately define diseases. Understanding reimbursement histories for medications and medical devices is also critical for designing robust studies.

Moreover, awareness of changes in reimbursement policies is essential, as claims records may sometimes include prescriptions issued before official reimbursement approval or cases where treatments were administered without an associated diagnostic code. These situations may arise due to exceptional circumstances, such as retrospective reimbursement approvals or pilot program applications, making it necessary to analyze the flow of insurance claims comprehensively.[9]

Another key limitation of claims data is the lack of standardized quality control for study subjects, which can introduce biases due to unequal baseline characteristics. Common biases encountered in claims-based research include selection bias, immortal time bias, length-time bias, visit bias, and misclassification bias. Most of these biases stem from disparities in baseline characteristics, underscoring the importance of applying appropriate statistical methods to address them effectively.[12]

STATISTICAL APPROACHES TO CONTROLLING FOR BASELINE DIFFERENCES

Traditional covariate adjustment (regression adjustment)

Statistical analysis seeks to determine the causes of variation in a dependent variable by accounting for multiple influencing factors. To accurately estimate the direct effect of an independent variable on a dependent variable, it is necessary to measure the variation in the independent variable while controlling for potential confounding variables. The process of quantifying the effect of a variable of interest while accounting for the influence of confounders is commonly referred to as multiple analysis. However, this approach does not entirely eliminate confounding effects, as confounders may still influence the independent variable. To mitigate this limitation, researchers must ensure that the distribution of confounding variables is as similar as possible between groups.

PS-based methods

PS-based methods are widely used in observational studies to reduce selection bias and enhance the accuracy of causal inference. These methods can be broadly categorized into matching and adjustment techniques. PSM is a technique that matches individuals with similar PSs to balance covariates between treatment and control groups, minimizing baseline differences and improving treatment effect estimation. The process involves estimating the PS using logistic regression, matching individuals based on their PS, and assessing balance using metrics like the standardized mean difference.[13] An alternative approach is PS Stratification and Adjustment. In stratification, the sample is divided into strata (typically five) based on PS quintiles, and treatment effects are estimated within each stratum before being combined into an overall effect. In adjustment, the PS is included as a covariate in a regression model to estimate the treatment effect. This method has the advantage of utilizing the entire sample without data loss.

Despite their advantages, PS methods have certain limitations. The most significant issue is that they cannot control for unobserved confounders. If important covariates are omitted, the estimated treatment effect may remain biased. Various approaches, such as sensitivity analysis, instrumental variable methods, difference-in-differences analysis, and machine learning techniques, have been proposed to address this issue.

Another limitation is data loss during the matching process, particularly when certain subgroups have insufficient sample sizes. This can occur when certain subgroups have low or insufficient sample sizes, reducing the matched dataset. Researchers can mitigate this by using continuous variables for matching, increasing the number of covariates, or employing weighting methods if the number of subjects in the treatment group significantly decreases after matching.

Weighting-based method: IPTW

Similar to PSM, IPTW is a statistical method used to control bias and estimate causal effects.[8,14,15] This method assigns inverse probability weights based on PSs to create a pseudo-randomized controlled trial (RCT) setting, allowing for better adjustment of baseline differences between groups. The key steps involve estimating the PS using logistic regression, calculating the treatment (1/PS) and control group (1/[1–PS]) weights, and applying the weights in analysis.

Table 1. Comparison of statistical approaches to controlling for baseline differences

Methods	Advantages	Disadvantages	Application areas
Regression adjustment	- Efficiently controls for confounding variables. - Facilitates the interpretation of study results.	- Does not fully account for baseline differences in the treatment groups. - Requires assessment of multicollinearity. - Dependent on the assumptions of the regression model.	- Suitable for well-designed studies (RCTs) where treatment group are initially balanced. - Commonly used in clinical settings to adjust for patient characteristics when evaluating treatment effects.
Propensity score matching	- Effectively balances treatment and control groups. - Results are intuitive and easy to interpret.	- Some data loss occurs during the matching process. - Reduced statistical power due to missing data.	- Suitable for large datasets, especially when there are many covariates for matching. - Useful in experiments where a slight reduction in the treatment group does not significantly affect the results.
Propensity score stratification/Adjustment	- Simple and straightforward method. - Retains the entire sample for analysis.	- Performance is suboptimal when the number of events is low. - Results may vary depending on the number of strata. - Assumes a linear relationship between propensity score transformations and outcomes. - Does not fully eliminate confounding compared to propensity score matching.	- Suitable for large datasets, especially when adjusting for multiple confounding variables and propensity scores simultaneously. - Preferred when minimizing reduction in the treatment group is crucial. - Useful when key matching variables are highly imbalanced.
Inverse probability of treatment weighting	- Utilizes the entire sample for analysis. - Can estimate various treatment effects.	- Extreme weights can excessively influence results. - Interpretation is less intuitive due to weighting.	- Suitable when the treatment group size is small. - Effective for comparisons involving multiple treatment groups rather than just two. - Useful for handling missing data. - Ideal for studies where minimizing data loss is crucial.

RCT = randomized controlled trial.

IPTW has the advantage of retaining data from all individuals, thereby minimizing information loss and maintaining statistical power. However, if the PSs are too close to 0 or 1, extremely large weights can be generated, which can lead to estimation instability. Additionally, if important confounders are not accounted for, bias may still occur. The interpretation of IPTW results can also be less intuitive compared to PSM. To address instability in weight estimation, the stabilized weights method is recommended. Stabilized weights help reduce variance and improve estimation robustness by assuming a normal distribution of weights. The methods discussed above are compared in **Table 1**, highlighting their respective advantages, disadvantages, and application areas.

OUTCOME ASSESSMENT IN COHORT STUDIES: SURVIVAL ANALYSIS

Overview of survival analysis

Survival analysis is a fundamental statistical method for evaluating time-to-event outcomes in cohort studies, particularly those utilizing national health insurance big data. The Cox proportional hazards regression model is widely used due to its ability to assess the impact of multiple predictor variables on survival time. A proper understanding of event occurrence time and censoring is essential. Event occurrence time refers to the first diagnosis of the disease or outcome of interest. Censoring arises when an event has not been observed by the end of the study period or when follow-up is no longer possible. However, national health insurance big data are primarily collected for administrative and billing purposes. As a result, the start and end points of follow-up may not be clearly defined. Therefore, careful data preprocessing is required to ensure the validity of survival analyses. Key preprocessing considerations include defining the study period, identifying pre-existing conditions, and excluding individuals who experienced the event before the study period to avoid bias. Additionally, individuals lost to follow-up due to death, emigration, or loss of health insurance eligibility should be identified. Event occurrence criteria must balance accuracy and feasibility, as overly strict definitions can introduce bias while overly lenient definitions may increase false positives. Ensuring consistent application across treatment and control groups is crucial to maintaining comparability.

Cox proportional hazards regression model

The Cox proportional hazards regression model is widely used in survival analysis to estimate the effect of predictor variables on survival time. A key assumption of this model is the proportional hazards assumption, which states that the hazard ratio (HR) remains constant over time. The HR quantifies the risk of an event occurring in a particular group compared to reference group. For example, an HR of 1.5 indicates group of interest has a 50% higher risk of experiencing the event compared to reference group. HR estimates are typically reported with confidence intervals, and if the confidence interval includes 1, the association is not statistically significant.

Validation of the proportional hazards assumption can be conducted using non-parametric methods, such as log(−log) plots and observed vs. expected value plots, which provide graphical validation. If the curves for each group are parallel or similar in shape, the assumption of proportional hazards is satisfied. Schoenfeld residual analysis enables both visual and statistical assessment of the assumption.

When the proportional hazards assumption is violated, alternative modeling approaches should be considered. The time-varying Cox model allows HRs to change over time, making it useful for exposures with dynamic effects. The stratified Cox model divides data based on variables that do not satisfy the assumption, improving the reliability of coefficient estimates for other covariates, but it does not provide an HR for the stratified variable. Competing risks analysis is another alternative, particularly when multiple events can occur and some may preclude others, providing a more comprehensive risk evaluation.

Commonly utilized methods and applications in Korean National Health Insurance claims data

An analysis of studies using NHIS big data revealed that research in this field has grown significantly since 2014, peaking between 2015 and 2018 before stabilizing in subsequent years.[16] A review of 678 studies utilizing customized datasets found that retrospective cohort studies were the most frequently used design, accounting for 56.93% of all studies. Other commonly used designs included incidence and prevalence studies (15.34%), exposure-matched cohort studies (12.68%), and cross-sectional studies (8.26%), while nested case-control studies (2.95%), case-control studies (1.47%), case-crossover studies (0.88%), cohort profile reports (0.88%), and review articles (0.59%) were employed less frequently. Following the coronavirus disease 2019 (COVID-19) pandemic in 2020, there was a sharp increase in pandemic-related research, leading to an atypical surge in publications. To avoid potential bias, studies conducted after 2020 were excluded from this trend analysis. The most frequently investigated conditions included diabetes mellitus, hypertension, dyslipidemia, cancer, stroke, chronic kidney disease, myocardial infarction, ischemic heart disease, chronic pulmonary disease, and depression. However, these ten conditions represented only 11.00% of all studies, highlighting the diverse range of diseases examined using NHIS claims data.

Various statistical methodologies have been applied, from basic descriptive analyses to advanced models. Survival analysis (34.51%) was the most commonly used method, followed by descriptive statistics (21.98%) and logistic regression (18.14%). Other frequently employed techniques included Poisson regression, multiple linear regression, time-series analysis, mixed effects model, and machine learning approaches. To minimize confounding bias, matching techniques were used in 18.85% of studies, with PSM (11.65%) being the most common, followed by simple matching methods (6.05%).

NEW STATISTICAL APPROACHES

Since the onset of the COVID-19 pandemic, study designs utilizing claims data have become increasingly diverse and sophisticated. With widespread vaccine administration across all age groups, the complexity of establishing appropriate control groups has increased, and ethical constraints on vaccine and therapeutic trials have led to the evolution of traditional randomized controlled trials (RCTs) models into targeted trial emulation (TTE) and adaptive trial designs to better address the challenges posed by the pandemic. At the same time, statistical analysis methods have evolved to accommodate this complexity. In particular, propensity score matching (PSM) methods have been improved by incorporating the double robustness property, thereby mitigating bias inherent in model specification. In addition, advanced survival analysis techniques such as multistate models have increasingly been utilized to effectively characterize the various health states of COVID-19 patients over time.

This section provides a comprehensive review of these methodological advances, highlighting their theoretical underpinnings, practical applications, and implications for epidemiological research in the post-pandemic era.

Doubly robust estimation (DRE)

DRE has emerged as a powerful methodology for estimating the average treatment effect in causal inference. It combines an outcome model and a PS model, providing a dual safeguard: consistent estimates are ensured if at least one of the two models is correctly specified.[17] However, if both models are misspecified, biased estimates may result. Additionally, extreme weights can be generated when PSs are close to 0 or 1, leading to instability in estimation.[18]

Machine learning/deep learning-based causal inference

Recent advances in computational environments have enabled the application of machine learning and deep learning techniques in causal inference. These methods excel in handling nonlinear relationships and high-dimensional data, often outperforming traditional statistical methods in identifying causal relationships. Despite these advantages, limited interpretability remains a significant barrier to widespread adoption. Explainable artificial intelligence (AI) models have been developed to enhance interpretability, but challenges persist in ensuring transparency and applying these models to real-world health insurance data, limiting their practical implementation.

Marginal structural model (MSM)

The MSM, introduced by James Robins and colleagues in 2000, is a sophisticated statistical approach for causal inference that effectively controls for time-varying confounders. Unlike traditional multivariate regression analysis, which may struggle with complex causal relationships, MSM is well-suited for adjusting for time-dependent confounding. It employs the IPTW technique, assigning weights to observations to reflect the extent to which certain characteristics are over- or underrepresented in the sample.

FUTURE PROSPECTS

Increasing demand for sophisticated causal inference

The growing availability of big data has enabled researchers to analyze causal relationships using large observational datasets, such as electronic health records, genomic data, and biometric information from wearable devices. These data sources are critical for refining causal inference methodologies and facilitating more robust analyses.

While RCTs have long been considered the gold standard for establishing causal relationships, real-world research primarily relies on observational data. Therefore, the development of advanced causal inference methodologies for observational data is crucial. Quasi-experimental approaches—such as difference-in-differences analysis and regression discontinuity design—have emerged as powerful tools in non-experimental settings, contributing to improved research quality.

The emergence of precision medicine emphasizes the need to accurately assess treatment effects within specific patient groups. Causal inference is expected to play a pivotal role in advancing precision medicine by facilitating the development of optimized, personalized treatment strategies based on individual patient characteristics.

Application of machine learning and AI techniques for high-dimensional and nonlinear data

Given the complexity and scale of the National Health Insurance Database, the adoption of advanced machine learning techniques—such as deep learning, reinforcement learning, and natural language processing—is expected to expand.[19] These techniques will enable the analysis of unstructured electronic medical record data, the integration of multi-omics datasets, and the development of personalized predictive models for precision medicine.[20]

Hybrid designs for incorporating real-world evidence

Hybrid research designs that combine big data analytics with prospective clinical studies are expected to become increasingly important for generating real-world evidence.[21] The integration of pragmatic clinical trials with big data analysis is anticipated to enhance the external validity of clinical research and facilitate the assessment of the real-world effectiveness of medical interventions.[22]

International collaboration and multi-center studies

The widespread adoption of the Observational Medical Outcomes Partnership Common Data Model is likely to promote international and multi-institutional comparative studies.[23] This development will enable large-scale cohort studies on topics such as health disparities across ethnic and national groups, rare disease research, and drug safety evaluations. Additionally, distributed network analysis methods will facilitate collaborative research while addressing legal and ethical challenges related to data sharing.[24]

Policy and ethical considerations

As the use of big data expands, privacy-preserving technologies—such as differential privacy and federated learning—will become increasingly necessary to protect sensitive information.[25] Methodological advancements will also be needed to mitigate algorithmic bias and improve the transparency and interpretability of AI models.[26] Institutional measures, including the establishment of robust data governance frameworks and the development of research ethics guidelines, will play a crucial role in ensuring responsible data use.

CONCLUSION

This study reviewed the utilization of Korean National Health Insurance claims data and various statistical methodologies for managing baseline differences in observational studies. As large-scale real-world data, health insurance claims data are valuable resources for medical research and health policy development. However, inherent biases in non-experimental data necessitate the application of appropriate study designs and robust statistical techniques.

PS-based methods—such as PSM and IPTW—along with traditional models like the Cox proportional hazards regression model, offer practical solutions for mitigating confounding. However, each method has inherent limitations. For example, PS methods are ineffective in controlling for unobserved confounders, while the Cox model relies on the proportional hazard assumption.

To address these limitations, emerging statistical methods—including DRE, Targeted Maximum Likelihood Estimation, and machine learning-based models—have gained attention. These approaches offer greater flexibility and robustness by relaxing restrictive assumptions and accommodating complex data structures.

As more datasets become integrated with health insurance claims data, high-dimensional data analysis techniques will become essential, driving the development of more sophisticated research methodologies. The importance of hybrid study designs for accurate causal inference will increase, and advancements in machine learning and AI techniques are expected to facilitate the development of more precise predictive models and causal inference methods. Furthermore, the expansion of international collaborative research will foster the integration and comparative analysis of multinational datasets, promoting large-scale cohort studies on diverse topics such as health disparities and rare diseases.

Maximizing the utility of Korean health insurance claims data requires a comprehensive understanding of their its characteristics, careful selection of statistical methodologies that align with research objectives, and continuous adaptation to advancements in analytical techniques. These efforts will contribute to more accurate and reliable research outcomes, ultimately supporting evidence-based healthcare policy development and public health improvements. Future research should also address ethical and legal considerations, particularly regarding data privacy and algorithmic transparency, alongside methodological advancements.

REFERENCES

1. Mazzali C, Duca P. Use of administrative data in healthcare research. *Intern Emerg Med* 2015;10(4):517-24.
PUBMED | CROSSREF
2. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* 2013;309(13):1351-2.
PUBMED | CROSSREF
3. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;2:3. PUBMED | CROSSREF

4. Khoury MJ, Ioannidis JPA.Big data meets public health. *Science* 2014;346(6213):1054-5. PUBMED | CROSSREF
5. Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L, et al. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J Am Med Inform Assoc* 2015;22(1):43-50. PUBMED | CROSSREF
6. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003;158(9):915-20. PUBMED | CROSSREF
7. Stuart EA. Matching methods for causal inference: a review and a look forward. *Stat Sci* 2010;25(1):1-21. PUBMED | CROSSREF
8. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011;46(3):399-424. PUBMED | CROSSREF
9. 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):103-10. PUBMED | CROSSREF
10. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58(4):323-37. PUBMED | CROSSREF
11. Cox DR, Kartsonaki C, Keogh RH. Big data: some statistical issues. *Stat Probab Lett* 2018;136:111-5. PUBMED | CROSSREF
12. Ellenberg JH. Selection bias in observational and experimental studies. *Stat Med* 1994;13(5-7):557-67. PUBMED | CROSSREF
13. Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, et al. Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies. *J Am Coll Cardiol* 2017;69(3):345-57. PUBMED | CROSSREF
14. Williamson E, Morley R, Lucas A, Carpenter J. Propensity scores: from naive enthusiasm to intuitive understanding. *Stat Methods Med Res* 2012;21(3):273-93. PUBMED | CROSSREF
15. Heinze G, Jüni P. An overview of the objectives of and the approaches to propensity score analyses. *Eur Heart J* 2011;32(14):1704-8. PUBMED | CROSSREF
16. Lim HS, Oh HC, Jang JH, Yoon S, Lee JK, Park S, et al. Research on the development of an analysis method inspection tool to improve the quality of big data research using the National Health Information DB - Methodology review of the literature on the use of the National Health Information DB. https://repository.nhimc.or.kr/bitstream/2023.oak/185/2/2020-20-015.pdf. Updated 2021. Accessed January 21, 2025.
17. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics* 2005;61(4):962-73. PUBMED | CROSSREF
18. Tsiatis AA, Davidian M. Comment: demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Stat Sci* 2007;22(4):569-73. PUBMED | CROSSREF
19. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436-44. PUBMED | CROSSREF
20. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25(1):44-56. PUBMED | CROSSREF
21. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med Care* 2012;50(3):217-26. PUBMED | CROSSREF
22. Ford I, Norrie J. Pragmatic trials. *N Engl J Med* 2016;375(5):454-63. PUBMED | CROSSREF
23. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015;216:574-8. PUBMED
24. Platt L, Grenfell P, Meiksin R, Elmes J, Sherman SG, Sanders T, et al. Associations between sex work laws and sex workers' health: a systematic review and meta-analysis of quantitative and qualitative studies. *PLoS Med* 2018;15(12):e1002680. PUBMED | CROSSREF
25. Roth A, Dwork C. The algorithmic foundations of differential privacy. *Found Trends Theor Comput Sci* 2013;9(3-4):211-407. CROSSREF
26. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 2019;1(5):206-15. PUBMED | CROSSREF