https://doi.org/10.4082/kjfm.20.0062 • Korean J Fam Med 2020;41:146-152

KJFM
Korean Journal of Family Medicine

eISSN: 2092-6715

■ Review Article

Conducting and Reporting a Clinical Research Using Korean Healthcare Claims Database

Seonji Kim[1], Myo-Song Kim[2], Seung-Hun You[2], Sun-Young Jung[2,*]

[1]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea
[2]College of Pharmacy, Chung-Ang University, Seoul, Korea

An increasing number of studies are using healthcare claims databases to assess healthcare intervention utilization patterns or outcomes in real-world clinical settings. However, methodological issues affecting study design or data analysis can make conducting and reporting these types of studies difficult. This review presents an overview of the types of information contained in claims data, describes some advantages and limitations of using claims data for research purposes, and outlines steps for utilizing the Korea Health Insurance Review and Assessment and National Health Insurance Service databases. The study also reviews epidemiological approaches utilizing healthcare claims databases (including cross-sectional, case-control, case-crossover, and cohort designs) with respect to protocol development, analysis, and reporting of results, and introduces relevant guidelines and checklists, including the Guidelines for Good Pharmacoepidemiology Practices, the Strengthening the Reporting of Observational Studies in Epidemiology checklist, and the Risk of Bias in Nonrandomized Studies of Interventions tool.

Keywords: Healthcare Administrative Claims; Research Design; Epidemiology; Bias

Received: April 1, 2020, Revised: April 27, 2020, Accepted: April 30, 2020
*Corresponding Author: **Sun-Young Jung** https://orcid.org/0000-0003-2032-112X
Tel: +82-2-820-5678, Fax: +82-2-816-7338, E-mail: jsyoung@cau.ac.kr

Copyright © 2020 The Korean Academy of Family Medicine
This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

In recent years, a rapidly increasing number of studies have begun to use healthcare claims database to assess healthcare intervention utilization patterns or outcomes.[1] Because observational studies using nationwide claims databases offer a large sample size with less strict inclusion and exclusion criteria than randomized controlled trials (RCTs), researchers may generate results more generalizable to real-world clinical settings.

The United States passed the 21st Century Cures Act in December 2016, with the goal of accelerating drug and medical device approval and promoting increased use of real-world data (RWD), including electronic health records, claims databases, registries, and healthcare applications, to generate real-world evidence (RWE) for potential risk and benefit assessments derived from sources other than RCTs.[2] In South Korea, revisions to the Personal Information Protection Act, the Act on Promotion of Information and Communications Network Utilization and Information Protection, and the Credit Information Use and Promotion Act were enacted in January 2020, and the Act on Safety and Support for Advanced Regenerative Medicine and Advanced Biopharmaceuticals will come into effect in August 2020. Based on growing needs to broaden access to healthcare information and generate RWE for the effectiveness and safety of clinical therapeutics, studies using RWD are expected to continue to increase in South Korea. However, methodological issues affecting study design or data analysis can make studies using healthcare claims databases challenging.

This review provides an overview of claims databases, describes some advantages and limitations of using claims data for research purposes, and presents steps for utilizing the Korean Health Insurance Review and Assessment (HIRA) and National Health Insurance Service (NHIS) databases. The study also reviews epidemiological approaches using healthcare claims databases in terms of protocol development, analysis, and reporting of results, and introduces guidelines and checklists including the Guidelines for Good Pharmacoepidemiology Practices (GPP), the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist, and the Risk of Bias in Nonrandomized Studies of Interventions (ROBINS-I) tool.

NATURE OF HEALTHCARE CLAIMS DATABASES IN KOREA

The South Korean health insurance system is a public, single-payer system. All citizens living in South Korea receive healthcare services as a fundamental right. Three major organizations are involved with the health insurance system: the Ministry of Health and Welfare (MoHW), the HIRA, and the NHIS. The MoHW operates and oversees the overall national health insurance system. Each individual (the insured) may receive a variety of medical services from service providers (healthcare institutions), which send reimbursement claims for medical expenses incurred to the HIRA. The HIRA reviews claims, assesses the quality of care provided, and evaluates healthcare services' adequacy. Based on the results of the HIRA's review, the NHIS reimburses services providers for medical care services provided. Throughout the process, all data related to medical services are accumulated in both HIRA and NHIS databases (Figure 1).

In recent years, various studies using data from the NHIS and HIRA have become possible under the Act on Promotion of the Provision and Use of Public Data. However, because these databases are intended for administrative and not research purposes, the data must be pro-

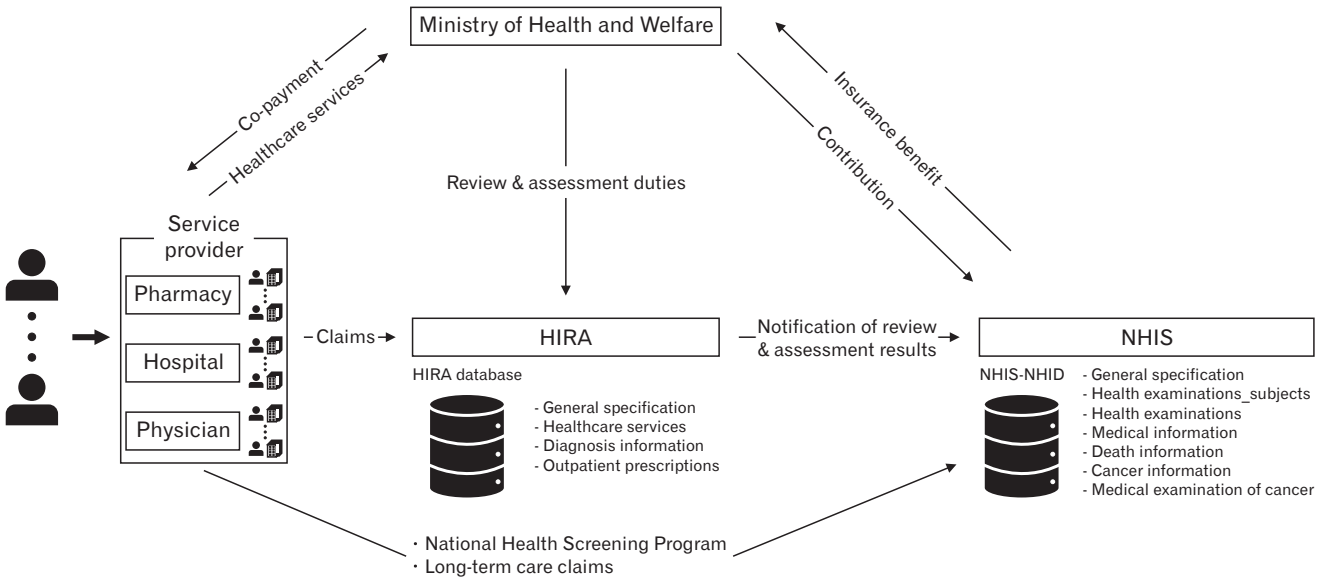

Figure 1. Governance of the healthcare system organization and healthcare claims databases in South Korea. HIRA, Health Insurance Review and Assessment Service; NHIS, National Health Insurance Service; NHID, National Health Information Database.

https://doi.org/10.4082/kjfm.20.0062

cessed before they can be used for research. Therefore, it is necessary for clinical researchers to fully understand the structure of each database.

Both databases are multi-layer in structure. If a patient is provided with medical services multiple times, multiple claims are generated, each of which contains information such as procedures performed, medications taken, and so on. Additionally, single claims are divided into several tables: specifications, treatment details, disease details, and prescription. Each table can be conjoined through a claim's key sequence number. Specifications (designated "Table 20") includes general information regarding the treatment, such as primary/secondary diagnosis, date of visit, and length of treatment in days. Treatment details (designated "Table 30") contains procedure codes, treatment codes, and prescription drugs for inpatients. Disease details (designated "Table 40") include all diagnosis codes pertaining to the patient. Finally, prescription (designated "Table 60" in the NHIS database and "Table 53" in the HIRA database) contains information on medications, such as generic medication codes, daily doses, unit doses, and days of supply for outpatients. Both the NHIS and HIRA databases include their own specific tables in addition to these general medical treatment-related.[3-7]

CURRENT STATUS OF HEALTHCARE CLAIMS DATABASES AND OTHER HEALTHCARE BIG DATA IN SOUTH KOREA

In South Korea, health insurance is a single-payer system managed by the HIRA and NHIS.[8] The government-run national healthcare claims databases cover approximately 98% of the total population and are available to researchers for public research purposes (Table 1).

The HIRA maintains a claims database for all patients, known as the HIRA database, along with four types of sampling databases with information from 2009 to 2018: the HIRA-National Patient Sample, HIRA-National Inpatient Sample, HIRA-Aged Population Sample (HIRA-APS), and HIRA-Pediatric Patient Sample.[9] The samples are updated annually and extracted using demographic stratification of age and gender.[10] Researchers can apply to use these claims data online (https://opendata.hira.or.kr/home.do).

The NHIS also maintains a database for the whole population of South Korea, the NHIS-National Health Information Database, and several sampling cohort databases: the NHIS-National Sample Cohort (NHIS-NSC), NHIS-National Health Screening Cohort (NHIS-HEALS), NHIS-senior cohort, NHIS-Female Employees (NHIS-FEM), and NHIS-Infants and Children's Health Screening (NHIS-INCHS). The NHIS-NSC includes a stratified random sample for age, gender, participant's eligibility status, region, and income level based on Korean population in 2006.[5] The NHIS-HEALS, NHIS-senior cohort, and NHIS-FEM are simple random samples of individuals.[11,12] The NHIS-INCHS was extracted from 2008–2012 births and samples 5% of the population by birth year. Researchers can access the NHIS databases and their information online (https://nhiss.nhis.or.kr/bd/ay/bdaya001iv.do).

The two claims databases appear similar, but have several important differences. First, the two institutions include slightly different variables in their datasets. The HIRA research database's main sections include patients' general specifications, healthcare utilization, diagnoses, and outpatient prescriptions (Table 2).[9,13] The NHIS database's main sections include healthcare utilization, sociodemographic variables, health screening, and mortality.[14] Second, the HIRA sample databases include separate cohorts for each year, whereas the NHIS sample databases include longitudinal cohorts.[5,9] Because patients are stratified and resampled annually in the HIRA sample databases, patient information in cannot be linked across years within HIRA sample databases. Therefore, the HIRA sample database is useful for conducting cross-sectional study or short-term follow-up (less than 1 year) studies. In contrast, participants in the NHIS sample cohort databases

Table 1. Types and contents of South Korean healthcare claims databases

Database type	Data period	Sampling description	Size
HIRA database	Depends on data size	Total eligible Korean patients	Over 50 million people
HIRA-NPS	2009–2018	Stratified proportional sample of patients (3% of population)	700,000 inpatients per year; approximately 400,000 outpatients per year
HIRA-NIS	2009–2018	Stratified proportional sample of patients who used inpatient services (13% of inpatients and 1% of outpatients)	1.4 million patients overall per year
HIRA-APS	2009–2018	Annual stratified proportional sample of patients over 65 years (20%)	Approximately 1 million patients per year
HIRA-PPS	2009–2018	Annual stratified proportional sample of patients under 20 years (10%)	Approximately 1.1 million patients per year
NHIS-NHID	Depends on data size	Total eligible Korean population	Over 50 million people
NHIS-NSC	2002–2015	Stratified proportional sample of total eligible Korean population (2%)	Approximately 1 million people
NHIS-HEALS	2002–2015	Simple random sample of population 40 years and over (5%)	Approximately 0.51 million people
NHIS-senior cohort	2002–2015	Simple random sample of population 60 years and over (10%)	Approximately 0.55 million people
NHIS-FEM	2007–2015	Simple random sample of employed women aged 26–64 years (5%)	Approximately 0.18 million people
NHIS-INCHS	2008–2015	5% sample of newborns by birth year between 2008 and 2012	Approximately 0.08 million people

HIRA, Health Insurance Review and Assessment Service; HIRA-NPS, HIRA-National Patient Sample; HIRA-NIS, HIRA-National Inpatient Sample; HIRA-APS, HIRA-Aged Population Sample; HIRA-PPS, HIRA-Pediatric Patient Sample; NHIS, National Health Insurance Service; NHIS-NHID, NHIS-National Health Information Database; NHIS-NSC, NHIS-National Sample Cohort; NHIS-HEALS, NHIS-National Health Screening Cohort; NHIS-FEM, NHIS-Female Employees; NHIS-INCHS, NHIS-Infants and Children's Health Screening.

Table 2. Databases and information available for linkage in South Korea

Source	Database	Data period	Contents and variables*
HIRA	HIRA database	2007–2018	- General specifications (billing statement identification key, age, gender, type of insurance, date of treatment, primary diagnosis, secondary diagnosis, surgery, etc.) - Healthcare services (billing statement identification key, inpatient prescriptions, treatments, diagnostic tests, unit price, days of supply, etc.) - Diagnosis (billing statement identification key, diagnostic code, department, etc.) - Outpatient prescriptions (billing statement identification key, drug codes, unit price, days of supply, etc.)
NHIS	NHIS-NHID	2007–2018	- General specifications (year, age, gender, region, grade of disability, contribution amount, etc.) - Health examinations - subjects (year, working type) - Health examinations (disease history, physical activity, current medications, smoking, drinking, height, weight, blood pressure, laboratory tests, etc.) - Medical institution (year, location, number of doctors, number of nurses, number of pharmacists, number of beds, etc.) - Death information (death year and month) - Cancer information (breast/colorectal/cervical/liver/gastric cancer) - Medical examination of cancer (medical examination experience, medical history, year, family history, etc.)
NCC	KNCI DB	2002–2016	- Age, gender, date of diagnosis, Surveillance Epidemiology and End Results code, diagnosis code, primary cancer site, treatment, histological type, etc.
KCDC	KNHANES	2007–2017	- Age, gender, socioeconomic status, educational status, chronic disease, health status, cancer examination, cost, quality of life information, injury, height, weight, blood pressure, laboratory tests, nutritional intake, dietary supplements, nutritional knowledge, etc.)
KCDC	Quarantine database	2013–2018	- Date of quarantine, type of quarantine, site of quarantine, country of departure, transportation, number of crew, number of passengers, number of suspicious entrants, pollution, major freight
KCDC	KTBS system database	2013–2018	- Year, age, age group, gender, region, nationality, reporting public health center, reporting medical institution, date of reporting, type of tuberculosis, disease code, patient type, smear screening
KCDC	KoGES	2001–2013	- Cohort name, age, gender, chronic disease, smoking, drinking, exercise, blood pressure, height, weight, laboratory tests, etc.
KCDC	Immunization registry data	2012–2018 (only NIP)	- Vaccination name, date of vaccination, medical institution, region of medical institution

HIRA, Health Insurance Review and Assessment Service; NHIS, National Health Insurance Service; NHIS-NHID, NHIS-National Health Information Database; NCC, National Cancer Center; KNCI DB, Korea National Cancer Incidence database; KCDC, Korea Centers for Disease Control and Prevention; KNHANES, Korea National Health and Nutrition Examination Survey; KTBS, Korean Tuberculosis Surveillance; KoGES, Korean Genome and Epidemiology Study; NIP, National Immunization Program.
*Information based on the Healthcare Big Data platform (https://hcdl.mohw.go.kr/BD/Portal/Enterprise/DefaultPage.bzr).

can be followed for up to 13 years. For example, researchers can assess exposure status during 2002 and follow up until the incidence of the study outcome or the end of the study period in 2015. Therefore, the NHIS sample cohort database is appropriate for study hypotheses requiring long-term follow-up.

In response to recent emphasis on the importance of big data, the Healthcare Big Data Platform has been established, which can link to claims databases. Linkable databases include the Korea National Cancer Incidence database provided by the National Cancer Center, the Korea National Health and Nutrition Examination Survey database, the Quarantine database, the Korean Tuberculosis Surveillance System database, the Korean Genome and Epidemiology Study database, and immunization registry data provided by the Korea Centers for Disease Control and Prevention.[15-17] All databases can be linked to each other, accessed online via the Healthcare Big Data Platform (https://hcdl.mohw.go.kr/BD/Portal/Enterprise/DefaultPage.bzr).

APPLICABILITY OF HEALTHCARE CLAIMS DATABASES

Healthcare claims databases are useful for clinical epidemiological research, particularly medication research on prescribing patterns, medication adherence, and adverse drug events.[18] Among observational research studies of clinical outcomes, analytical study designs can be roughly divided into cross-sectional studies, case-control studies, and cohort studies. A cross-sectional study measures both exposure and outcome at the same time; a case-control study first measures outcome, then determines any previous exposure; and a cohort study classifies groups according to exposure and follows up to confirm the outcome.[19] Recently, a number of observational studies using healthcare claims databases have been reported in Korea. This section considers examples of such studies by design.

An example cross-sectional study used a HIRA-APS dataset (stratified proportional sample of patients over the age of 65 years) to assess medication use among elderly patients in intensive care units.[20] Using this dataset, the researchers analyzed patterns of medication use in real-world settings according to duration of mechanical ventilation, patient age, and annual trends, and assessed patient factors related to the use of sedatives and analgesics in elderly patients.

An example nested case-control study examined the risk of esophageal or gastric cancer after exposure to oral bisphosphonates in the Korean population using the NHIS-NSC database.[21] From a cohort of over 160,000 patients with osteoporosis, 1,708 cases were selected (patients aged 40 years and above with initial esophageal or gastric cancer). For each case, four controls were matched for age, gender, and income level. The study did not confirm a significant association be-

Table 3. Strengths and limitations of healthcare claims databases

Strengths	Limitations
- High generalizability for the Korean population - Minimized risk of recall bias - Thorough cover of disease conditions - Sufficiently large sample size to retain statistical power - Various information on healthcare utilization, diagnoses, procedures, treatment, and payments - Relatively inexpensive to use - Linkable to other databases	- Risk of confounding bias such as confounding by indication and healthy user effect - Often no measurement of potential confounders such as laboratory data, disease severity, and health behaviors - Risk of misclassification bias (may affect internal validity) - Not applicable to research on healthcare services not covered by insurance - Insufficient information on patient adherence to treatment - Time gap between actual provision of health services and availability of the claim data for research

tween bisphosphonates and upper gastrointestinal cancer in real-world settings.

An example cohort study was conducted using the NHIS-HEALS, a database constructed using the NHIS claims database and the national health screening databases.[22] The study estimated the association between various risk factors (e.g., body mass index and health-related behaviors such as smoking and alcohol consumption) and dementia using a Cox proportional-hazards model. Because this dataset provided health screening data biennially for each individual, weight change could be identified.[11] The study found that both weight gain and weight loss are potential risk factors for dementia, and therefore that weight changes should be carefully monitored.

ADVANTAGES AND LIMITATIONS OF USING HEALTHCARE CLAIMS DATABASES FOR RESEARCH

Healthcare claims databases offer several important advantages for research (Table 3). First, because almost all Korean populations are covered by national insurance, research results are highly generalizable.[23] Second, because claims databases are constructed during the course of medical services, and are thus not dependent on the memory of patients or healthcare professionals, recall bias is minimized. Third, they cover disease conditions thoroughly utilizing international disease code classifications. Fourth, the databases have sufficiently large sample sizes to retain statistical power, and contain various information on healthcare utilization, diagnoses, procedures, treatment, and payments. Fifth, use of a healthcare database is relatively quick and inexpensive compared to implementation of a clinical trial. Finally, these databases can be linked to various others, including the Korea National Cancer Incidence database and information on mortality (date and cause of death) from Statistics Korea. For example, a study has assessed the association between fatal motor vehicle collisions and zolpidem prescription by linking the database of the Korea Road Traffic Authority with health insurance data from the NHIS.[24]

However, research using healthcare databases is also subject to certain limitations. First, confounding biases may be introduced. Confounding by indication results when the patient's condition for which the drug is prescribed is itself is related to the outcome. For example, a study of the association of suicide and selective serotonin reuptake inhibitors (SSRIs) may be vulnerable to confounding by indication because SSRIs are indicated to treat depression, which may cause suicidal ideation. This could lead to erroneous conclusions or overestimation of the strength of any association;[25] confounding by indication may thus bias the relative risk of adverse events away from the null. A healthy user effect, in which receiving treatment is associated with underlying patient characteristics like high education level and attitude to pursue health,[26,27] may also distort interpretation of the results. For instance, observational studies of hormone replacement therapy (HRT) have shown that women who took HRT tended to demonstrate more healthy behaviors, such as regular exercise and healthy diet, compared to the nontreatment group; the apparent protective effect of HRT against cardiovascular disease appears to reflect these differences in patients' underlying characteristics.[26] Additionally, unmeasurable potential confounders such as laboratory data, disease severity, or patient-reported outcomes prevent complete control of confounding effects.[27] For example, although the databases contain a diagnosis code for cancer, they do not record information on the stage or severity of the disease. Second, misclassification bias can occur when defining both exposure and outcome variables.[28] Due to insurance reimbursement policies and the fee-for-service system, up-coding issues may arise, and discrepancies between diagnosis coding and patients' actual health conditions may exist. A previous study reported only 70% accuracy of diagnoses in claims databases.[29] Third, because the purpose of claims databases is to reimburse healthcare services, they are not applicable to research on healthcare services not covered by insurance or over-the-counter drugs. Fourth, it is impossible to accurately measure medication adherence using claims data; prescription of a drug does not mean that the patient actually took the drug. Fifth, there is a time gap between the time health services are actually provided and the time a claim for the service becomes available for research.[30] Finally, diseases with low prevalence may be difficult to study using HIRA or NHIS sample databases because of small sample sizes and lack of representativeness to the target population.

GUIDELINES FOR CONDUCTING AND REPORTING OBSERVATIONAL STUDIES USING HEALTHCARE CLAIMS DATABASES

Several methodological criteria and checklists for conducting and reporting observational studies using the healthcare claims database have been developed (Table 4). The Guide on Methodological Stan-

Table 4. Guidelines for observational studies using big data

Guidelines	Publication year	Source	Checklist items	Link
Guide on Methodological Standards in Pharmacoepidemiology	2018 (version 7)	ENCePP	Research question, study design, data sources, source and study population, definition and measurement of exposures/outcomes, bias, effect measure modification, data management, data analysis, quality control, ethical/data protection issues, communication of study results	http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml
GPP	2016 (version 4)	Public Policy Committee and International Society of Pharmacoepidemiology	Population, definition of exposures/outcomes/other risk factors, study size, statistical precision, data management, data analysis, quality assurance, quality control	https://doi.org/10.1002/pds.3891
STROBE	2007 (version 4)	STROBE Initiative	Introduction (background, objective), methods (study design, setting, participants, data source, bias, study size, statistical analysis), results (descriptive data, outcome, main results, other analysis), discussion (interpretation, generalizability, limitations), funding information	https://www.strobe-statement.org/index.php?id=available-checklists
ROBINS-I	2016 (version 1)	Researchers, many involved with Cochrane systematic reviews	Bias related to confounding factors, selection of participants, classification of interventions, deviations from the intended interventions, missing data, measurement of outcomes, and selection of reporting results	http://dx.doi.org/10.1136/bmj.i4919

ENCePP, European Network of Centres for Pharmacoepidemiology and Pharmacovigilance; GPP, Guidelines on Good Pharmacoepidemiology Practices; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology; ROBINS-I, Risk of Bias in Nonrandomized Studies of Interventions.

dards in Pharmacoepidemiology version 7, published in 2018 by the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance, addresses the overall steps for conducting a pharmacoepidemiological study, from formulating research questions to addressing ethical issues and communicating study results to ensure scientifically independent and transparent research. Researchers can refer to the related checklist for study protocols, developed based on the criteria in this guideline, to consider and be aware of key epidemiological principles.

The GPP version 4, developed by the Public Policy Committee and International Society of Pharmacoepidemiology in 2016,[31] suggests essential principles to consider as check points to ensure methodological quality when conducting and evaluating pharmacoepidemiologic studies. The checklists include definitions of exposures, outcomes, other risk factors, statistical precision, data management and analysis, and quality control.

The STROBE Initiative's established recommendations for conducting observational research,[32] the STROBE Statement, was updated up to revision 4 in 2007 and presents checklists for researchers according to study design. Because the STROBE Statement's aim is to improve the quality of observational research reporting, the checklist items pertain to procedures for reporting research in papers, such as the title and abstract, introduction, methods, results, and discussion sections.

The Cochrane Bias Methods Group developed an evaluation tool, the ROBINS-I, to assess the risk of bias in nonrandomized studies in 2016, using criteria for RCTs.[33] The tool focuses on internal validity and utilizes a hypothetical ideal target trial. It is designed for use in observational studies and assesses seven bias domains: selection of participants, confounding, classification of interventions, missing data, deviations from the interventions, selection of reporting results, and measurement of outcomes.

CONCLUSION

Korean national health insurance claims databases are a useful source of data for generating RWEs with high generalizability in the Korean population. However, these databases also have inherent limitations, including confounding bias, selection bias, and validity of study variables. Therefore, clinical research studies using and reporting results based on Korean healthcare insurance claims databases must be well designed, with rigorous analysis and careful interpretation considering the risks of bias.

CONFLICT OF INTEREST

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

ACKNOWLEDGMENTS

This research was supported by a Korea Health Technology R&D Project grant (HI19C1202) through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare.

ORCID

Seonji Kim: https://orcid.org/0000-0002-4993-7537
Myo-Song Kim: https://orcid.org/0000-0002-9094-8780
Seung-Hun You: https://orcid.org/0000-0002-2430-6907

Sun-Young Jung: https://orcid.org/0000-0003-2032-112X

REFERENCES

1. Singh G, Schulthess D, Hughes N, Vannieuwenhuyse B, Kalra D. Real world big data for clinical research and drug development. Drug Discov Today 2018;23:652-60.
2. The Senate and House of Representatives of the United States of America in Congress. 21st Century Cures Act. H.R.34, 114th Congress [Internet]. Washington (DC): The United States Congress; 2016 [cited 2020 Mar 29]. Available from: https://www.gpo.gov/fdsys/pkg/BILLS-114hr34enr/pdf/BILLS-114hr34enr.pdf.
3. Chun CB, Kim SY, Lee JY, Lee SY. Republic of Korea: health system review. Health Syst Transit 2009;11:1-184.
4. Lee EK, Park JA, Cole A, Mestre-Ferrandiz J. Data governance arrangements for real-world evidence: South Korea. London: Office of Health Economics; 2017.
5. Lee J, Lee JS, Park SH, Shin SA, Kim K. Cohort profile: the National Health Insurance Service-National Sample Cohort (NHIS-NSC), South Korea. Int J Epidemiol 2017;46:e15.
6. Ryu DR. Introduction to the medical research using National Health Insurance Claims Database. Ewha Med J 2017;40:66-70.
7. Health Insurance Review and Assessment Service. Healthcare system in Korea: health security system [Internet]. Wonju: Health Insurance Review and Assessment Service [cited 2020 Mar 29]. Available from: https://www.hira.or.kr/dummy.do?pgmid=HIRAJ010000006002.
8. Seong SC, Kim YY, Khang YH, Park JH, Kang HJ, Lee H, et al. Data resource profile: the National Health Information Database of the National Health Insurance Service in South Korea. Int J Epidemiol 2017;46:799-800.
9. Kim L, Kim JA, Kim S. A guide for the utilization of Health Insurance Review and Assessment Service National Patient Samples. Epidemiol Health 2014;36:e2014008.
10. Kim L, Sakong J, Kim Y, Kim S, Kim S, Tchoe B, et al. Developing the inpatient sample for the National Health Insurance claims data. Health Policy Manag 2013;23:152-61.
11. Seong SC, Kim YY, Park SK, Khang YH, Kim HC, Park JH, et al. Cohort profile: the National Health Insurance Service-National Health Screening Cohort (NHIS-HEALS) in Korea. BMJ Open 2017;7:e016640.
12. Kim YI, Kim YY, Yoon JL, Won CW, Ha S, Cho KD, et al. Cohort profile: National health insurance service-senior (NHIS-senior) cohort in Korea. BMJ Open 2019;9:e024344.
13. Kim JA, Yoon S, Kim LY, Kim DS. Towards actualizing the value potential of Korea Health Insurance Review and Assessment (HIRA) data as a resource for health research: strengths, limitations, applications, and strategies for optimal use of HIRA data. J Korean Med Sci 2017;32:718-28.
14. Chung H, Kim SY, Kim HS. Clinical research from a Health Insurance Database: practice and perspective. Korean J Med 2019;94:463-70.
15. Lew WJ, Lee EG, Bai JY, Kim HJ, Bai GH, Ahn DI, et al. An internet-based surveillance system for tuberculosis in Korea. Int J Tuberc Lung Dis 2006;10:1241-7.
16. Jung KW, Won YJ, Kong HJ, Lee ES. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2016. Cancer Res Treat 2019;51:417-30.
17. Cho HY, Kim CH, Go UY, Lee HJ. Immunization decision-making in the Republic of Korea: the structure and functioning of the Korea Advisory Committee on Immunization Practices. Vaccine 2010;28 Suppl 1:A91-5.
18. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2005;58:323-37.
19. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet 2002;359:57-61.
20. Jung SY, Lee HJ. Utilisation of medications among elderly patients in intensive care units: a cross-sectional study using a nationwide claims database. BMJ Open 2019;9:e026605.
21. Jung SY, Sohn HS, Park EJ, Suh HS, Park JW, Kwon JW. Oral bisphosphonates and upper gastrointestinal cancer risks in Asians with osteoporosis: a nested case-control study using national retrospective cohort sample data from Korea. PLoS One 2016;11:e0150531.
22. Park S, Jeon SM, Jung SY, Hwang J, Kwon JW. Effect of late-life weight change on dementia incidence: a 10-year cohort study using claim data in Korea. BMJ Open 2019;9:e021739.
23. Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet 2002;359:248-52.
24. Yang BR, Kim YJ, Kim MS, Jung SY, Choi NK, Hwang B, et al. Prescription of zolpidem and the risk of fatal motor vehicle collisions: a population-based, case-crossover study from South Korea. CNS Drugs 2018;32:593-600.
25. Didham RC, McConnell DW, Blair HJ, Reith DM. Suicide and self-harm following prescription of SSRIs and other antidepressants: confounding by indication. Br J Clin Pharmacol 2005;60:519-25.
26. Shrank WH, Patrick AR, Brookhart MA. Healthy user and related biases in observational studies of preventive interventions: a primer for physicians. J Gen Intern Med 2011;26:546-50.
27. Prada-Ramallal G, Takkouche B, Figueiras A. Bias in pharmacoepidemiologic studies using secondary health care databases: a scoping review. BMC Med Res Methodol 2019;19:53.
28. Walraven CV. A comparison of methods to correct for misclassification bias from administrative database diagnostic codes. Int J Epidemiol 2018;47:605-16.
29. Park BJ, Sung J, Park K, Seo SW, Kim SH. Studying on diagnosis accuracy for health insurance claims data in Korea. Seoul: Seoul National University; 2003.
30. Strom BL. Overview of automated databases in pharmacoepidemiology. In: Strom BL, Kimmel SE, Hennessy S, editors. Pharmacoepidemiology. 5th ed. Chichester: John Wiley & Sons; 2012. p. 158-62.
31. Public Policy Committee, International Society of Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practice (GPP). Pharmacoepidemiol Drug Saf 2016;25:2-10.
32. Von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147:573-7.
33. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919.