Corona 19 International Cooperation Research Manual

(ver.1) Guidelines for COVID-19

Backgro
und

2020 As of March 26, 446,185 cases of corona19 and 18,805 deaths occurred in 200 countries around the world, and for the time being, the global pandemic of corona19 infectious diseases is expected to continue to spread.

This is threatening the lives and safety of the people of the world and shrinking socio-economic activities, but it is based on actual clinical data-based evidence and government policies that can help clinicians diagnose and treat patients at the treatment site. The provision of necessary information is not enough.

Accordingly, the government of the Republic of Korea was the first to release the world's first dataset of corona19 patients that were quickly collected, purified, and de-identified digitally through the national health insurance system. We look forward to proactively participating in domestic and foreign researchers to effectively respond to serious diseases and to draw meaningful results.

2. The goal

to overcome the Corona 19 international cooperation, the entire human disease by opening the Health Insurance Review and Assessment Service of the Republic of Korea Big Data holds the medical records of all citizens, through the study and use in the production of evidence for the effective implementation of government policies It's possible.

3. Dataproperty

Sourceof this research is the medical expenses as a medical claims billing data of the Health Insurance Review and Assessment Service, and clinical information, clinical data including (real-world data, since the RWD) and administrative data. Information that is not essential for reimbursement of medical expenses because the medical use history of the entire nation is accumulated in detailed and time-series (patient-level panel data) based on fee-for-service. (Non-paid medical treatment and test results, etc.) are not included. Please refer to this point when performing analysis.

4. The data extraction criteria

4-1. Data range

1)medical use data (of trial datapast 5 years to patient list (receipt data utilization) received through Corona 19)

Non-identification after linkingutilization, '15 .1 ~ '20 .2.) For the2) 2020.3.25. Currently, the total number of billed statements related to Corona19-34,131 total bill statements and 32,083 patients

4-2. Data Extraction Criteria

1) Insurer Category: All (Health Insurance, Medical Benefits, Veterans Affairs) 2) Corona 19 Billing Statement: Myeongserial Unit Specific History Category Code MT043 (National Disaster Medical Expenses Supported Type) “3/02” Specification 3) If the billing statement is claimed with the code in the table below, it is judged as a confirmed patient (applied with the Korean Standard Disease /Classification (KCD) code)

[Table] Coronavirus Infection-19Code

Code KoreanRemarks

SignalCorporativeNameB34.2 Coronavirus in detail Infection

B97.2 As a cause of the diseases classified in other chapters,coronavirus newthrough thetest

U18 the Korean temporary designation ofnew diseaseoremergency confirmed asthean U18.1 new coronavirus infectioncoronavirus infection should be included. U07.1 Coronavirus disease 2019 [Corona-19] Coronavirus disease 2019 [COVID-19]

4-3. Data Schema

1) 8 data tables-data tables 4 reception tables for corona patient identification and 4for the patient's past medical use-69 total variables (including duplicates) 2) For detailed descriptions of the tables and variables, see the “ Data Schema ”file

-

Research Method

5-5.5-1. Conduct research using online platformResearch

1)institutes, government agencies and researchers at home and abroad participate in the research online

-Website address: covid19data.hira.or.kr 2) Research questions after registering basic information such as name, affiliation and email (login) Analysis code upload

-Commercial email account registration is not possible (e.g. gmail.com) and registered by affiliate email 3) The dataset for analysis is kept in the closed network of Simpyeongwon,

and only the data schema (table structure, variable configuration, etc.) is

disclosed- Dataset is prepared in SAS, R and OMOP-CDM format 4)using data schema and sample data published on **Researcher writes analysis codeit**

and uploadsto analysis boardbulletin board 5) Health Insurance Review & Assessment Service puts analysis code on data After execution, only the result (statistics) value is exported.

-When the analysis code is executed, the log contents are exported.-Use of this data related to corona19. In order to do so, IRB's research

approval or deliberation exemption documents are required. 6) Researchers participating in this study must sign (online) the “Data Utilization Agreement”

-Check when registering the website 7) Research schedule is from late March to April 2020 Elderly, can be adjusted as needed (research schedule mayason the website

varydiscussed)

5-2. Analysis tool
(statistics package)

1) Data set analysis can be performed in SAS or R. 2)
Researcher writes analysis code using SAS or R.

-If the researcher has difficulty accessing SAS in the local environment, the Health Insurance Review & Assessment Service Health Care Big You can connect to the data opening system (https://opendata.hira.or.kr) and apply for SAS Studio.

※ Health and medical big data opening system> Medical big data> Big data analysis practice> SAS Studio application -Analysis with R If desired, upload the R package used by the researcher along with the analysis code

※ Prevent the possibility of errors due to the latest version mismatch -OMOP-CDM-based research is based on R analysis, and the available packages and versions are shown below

.

5-3 . Sample dataset (SAS, R
version)

1) Samples with the same structure and variable name of the research dataset for reference when writing the analysis code

Publish the dataset on the bulletin board-Identification key: statement key, patient key, nursing home sign -Non-identifying measures-Medical use details: Details processed by the Health Insurance Review and Assessment Service-Variable coding details: For coding table, refer to bulletin board data coding book (file)

5-4. Sample data set (OMOP-CDM version, please inquire separately) 1) Simpyeongwon data set extracted as above is composed of

OMOP-CDM 5.3.1 version, and detailed data structure of OMOP-CDM 5.3.1 version is as follows. See link

https://github.com/OHDSI/CommonDataModel/blob/v5.3.1/OMOP_CDM_v5_3_1.pdf

2) Analysis code can be written using the virtual OMOP-CDM data and analysis package of the

cloud (see link below, cloud possible

costs) 3)OMOP-CDM-related inquiries, contact us at OHDSI Korea forum (link)

https://forums.ohdsi.org/c/For-collaborators-wishing-to-communicate-in-Korean

6 Use of achievements

6-1. Principle of free access (open access)

The results of research using this Corona19 data (such as the publication of journal papers) should be published free of charge.

6-2. Principle of pre-sharing

**research results In principle,
prior**

to the official announcement of the research results using this Corona19 data, the principle should be shared in advance with the Government of the Republic of Korea

6-3. Statement of Acknowledgment
(Acknowledgement) It

is recommended to include the following acknowledgment in the research results using this Corona19 data.

(Korean) This study, including Korean medical staffs who are doing their best to treat patients For this, we would like to thank the Ministry of Health and Welfare and the Health Insurance Review and Assessment Service for promptly providing the corona19 patient's medical use information.
The authors appreciate healthcare professionals dedicated to treating COVID-19 patients in Korea, and the Ministry of Health and Welfare and the Health Insurance Review & Assessment Service of Korea for sharing invaluable national health insurance claims data in a prompt manner.