Guidelines for COVID-19 International Research Co-hosted by MoHW and HIRA of Korea(ver.1)

보건복지부

Ministry of Health and Welfare

건강보험심사평가원

Health Insurance Review & Assessment Service

1. Background

As of March 26, 2020, 446,185 people have been confirmed with COVID-19 from 200 countries, and the total death toll reached 18,805 globally. The COVID-19 pandemic is expected to continue to spread further for the time being.

The novel coronavirus is threatening the livelihood and safety of global citizens, and social and economic activities have been hit hard in many countries. Yet, there is little evidence of real-world clinical data available for physicians or policy makers.

Against this backdrop, the government of the Republic of Korea decided to share the world's first de-identified COVID-19 nationwide patient data with domestic and international researchers. The data sets are digitally collected and processed promptly, thanks to the Korean National Health Insurance System, covering the entire population across the nation. We hope that researchers from home and abroad join us and actively engage in producing effective measures to fight the global spread of the deadly disease.

2. Objective

COVID-19 International Research will share the nationwide patient Big Data of Health Insurance Review and Assessment Service of Korea to help humanity as a whole overcome the disease and produce evidence for effective policy enforcement.

3. Data Attribute

The data source of this research is insurance benefit claims submitted to the Health Insurance Review and Assessment Service of Korea. In nature, the claims data are both real-world data (hereafter RWD) and administrative data, that include clinical contents. The data have accumulated history of all medical service use of the entire population in sequence (patient unit panel data), based on fee-for-service payment system. Provided, as the data collection was not intended for clinical use, information nonessential for benefit reimbursement (uncovered services and tests results for covered services) is not included. Please be reminded of this in your research.

4. Data Extraction Criteria

4-1. Data Scope

1) The list of patients of COVID-19 (using submitted claims data) is connected to their history of medical service use for the past 5 years (using finalized claims data, from January 2015 to February 2020), and the entire data set is de-identified.
2) All benefit claims relevant to COVID-19 as of March 25, 2020
   - The total number of claims is 34,131, and the total number of patients is 32,083

4-2. Extraction Criteria

1) Insurer type: all (National Health Insurance, Medical Aid, Korea Veterans Service)
2) Claim statement of COVID-19: claim statements with code MT043 (national disaster fundholding type) 3/02
3) A claim statement that includes following codes in the table below are considered a confirmed case of COVID-19 under Korean Standard Classification of Diseases (hereafter KCD)

[Note] COVID-19 Disease Code

Code	Name	Note
B34.2	Coronavirus infection in unspecified area	One of the five codes shall be added to diagnosis of confirmed patients
B97.2	Coronavirus as the cause of other diseases in other chapters	
U18	Tentative designation or emergency use of a new disease in Korea	
U18.1	Novel coronavirus infection	
U07.1	Coronavirus disease 2019 [COVID-19]	

4-3. Data Schema

1) 8 Data Tables
   - Four reception tables to check COVID-19 infection of the patient, and four tables to check history of medical service use of the patient
   - In total, there are 69 variables (including duplication)
2) For more information on tables and variables, please refer to the "Data Schema" file on the website.

5. Research Method

5-1. Online platform

1) Domestic and international academic societies, state institutions and researchers can participate through online platform.
   - Website address: https://covid19data.hira.or.kr
2) Register your name, organization, email account and other basic information to join and log in. Submit your research proposal and analysis code.
   - The email account should be given by the researcher's organization. Commercial email accounts (gmail.com, etc.) are not allowed.
3) HIRA will keep the data set for analysis within its closed network, and share data schema only (which explains table structure, variable configuration, etc.).
   - All data sets will be shared in the form of SAS, R, or OMOP-CDM
4) **The researcher is responsible for producing analysis code** using the data schema and sample data. The analysis code should be uploaded on the website**.**
5) HIRA will run the analysis code on the applicable data, and deliver the result values (statistics) only.
   - In case of error, the error log will be shared.
   - The researchers should submit documents demonstrating approval or exemption of the study from IRB review for export or publication of the results.
6) All researchers must sign "Data Use Agreement" (online signature).
   - Please sign the agreement when you register on the website.
7) Research period is scheduled from late March to early April, but is subject to change if needed.

5-2. Analysis Tools (statistical package)

1) Data set analysis can be done with SAS, R, or CDM.
   - Refer to a separate manual for CDM.
2) Researchers are required to produce analysis code using SAS or R.
   - If the researcher did not have an access to SAS, he or she could visit HIRA Healthcare Big Data Hub (https://opendata.hira.or.kr) to use SAS Studio
     ※Healthcare Big Data Hub>Healthcare Big Data>Big Data Analysis Drill>Apply to use SAS Studio
   - For researchers who use R for analysis, he or she could upload the analysis code and R package that was used
     ※To avoid conflicts from difference between versions

5-3. Sample Data Set (SAS, R)

1) As a reference of analysis code production, a sample data set is uploaded on the website with the same structure and variables
   - Identification Key: De-identify claim statement key, patient key, and healthcare provider code
   - History of medical service use: processed data by HIRA
   - Variable coding: refer to coding book (digital file) for coding table

5-4. Sample Data Set (OMOP-CDM, please contact us for inquiry)

1) HIRA's data set in OMOP-CDM 5.3.1 version is prepared in the same manner. For more detailed data structure of OMOP-CDM 5.3.1 version, visit the following link.
   https://github.com/OHDSI/CommonDataModel/blob/v5.3.1/OMOP_CDM_v5_3_1.pdf

2) Analysis code can be produced through virtual OMOP-CDM data and analysis package on cloud service. (Please refer to the link below. Cost may be incurred for use of cloud service.)

3) Inquiries related to OMOP-CDM should be directed to OHDSI Korea Forum.(link below)
   https://forums.ohdsi.org/c/For-collaborators-wishing-to-communicate-in-Korean

6. Use of Outcome

6-1. Open access

All research outcomes and the following publication using HIRA-provided data (through academic journal or any other form of formal release) should be open access.

6-2. In-advance Sharing of Research Outcome with Data Provider

Prior to any form of formal release of research outcome using HIRA-provided data, the content must be shared with the Korean government.

6-3. Acknowledgement

Researchers are encouraged to add following acknowledgement to their research outcome using HIRA-provided data.

> (Example) 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.