

COVID-19 DATA in HARVARD DATAVERSE

NIH Webinar on Sharing, Discovering, and Citing COVID-19 Data and Code
in Generalist Repositories

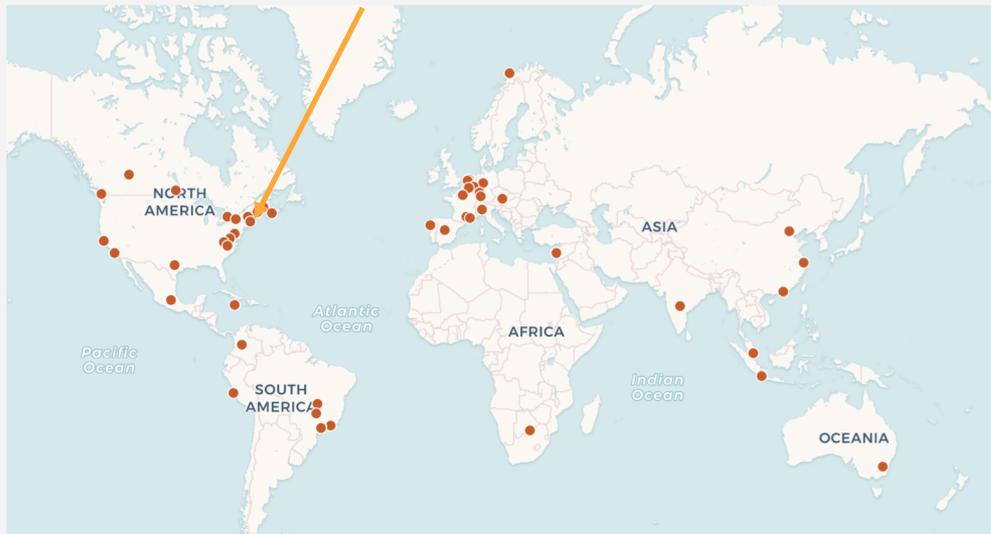
April 24, 2020

Mercè Crosas, Ph.D.

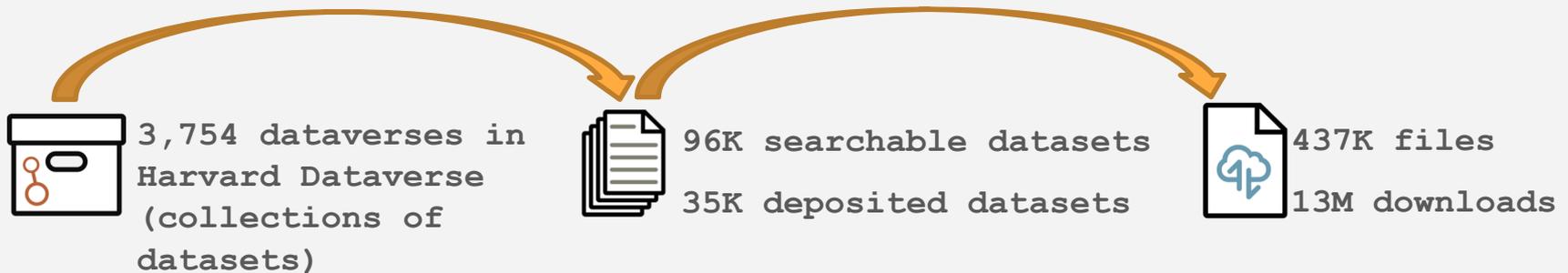
University Research Data Management Officer

Chief Data Science and Technology Officer, IQSS

Harvard University



- **Dataverse** is an open-source software with a growing, active community
- **Harvard Dataverse** is one of the 56 Dataverse repositories world-wide
- Open to **all** research domains
- Researchers can deposit for **free**
- **New** data curation services (fee-based)



COVID-19 Datasets in Harvard Dataverse

- **50 COVID-19 datasets** with **2,043 data files** deposited since February 7
- Total of **48,471 downloads** by April 23
- Include datasets on: COVID-19 statistics, social science studies to evaluate the effectiveness of government measures, survey data, Twitter data, gubernatorial responses, replication data for preprints

Example of COVID-19 datasets in Harvard Dataverse

HARVARD
Dataverse

COVID-19 Research & Evaluations (Population Council)

Harvard Dataverse > Population Council Dataverse > COVID-19 Research & Evaluations

Partnering with national health ministries and other government agencies in sub-Saharan Africa, South Asia, and Latin America, we are conducting COVID-19 public health and social science research to produce relevant and timely evidence to support the evaluation of the effectiveness of prevention and mitigation measures, and assessing longer-term health, social and economic impacts.

Search this dataverse... **Q Find** [Advanced Search](#)

Dataverses (0)

Datasets (3)

Files (14)

Publication Year
2020 (3)

Subject
[Social Sciences \(3\)](#)

Author Name
[Ngo, Thoai D. \(3\)](#)
[White, Corinne \(3\)](#)
[Pinchoff, Jessie \(2\)](#)
[Abuya, Timothy \(1\)](#)
[Archarya, Rajib \(1\)](#)

[More...](#)

Keyword Term
[COVID-19 \(3\)](#)
[attitudes \(3\)](#)
[coronavirus \(3\)](#)
[knowledge \(3\)](#)
[indigenous \(1\)](#)

[More...](#)

Geographic Coverage Country / Nation
[Guatemala \(1\)](#)
[India \(1\)](#)

POPULATION COUNCIL
Ideas. Evidence. Impact.

Search

[About](#) [Our Research](#) [Events](#) [News & Views](#) [DONATE](#)

Responding to the COVID-19 Pandemic

The Population Council is producing relevant and timely evidence to help control the spread of coronavirus, evaluate the effectiveness of prevention measures, and assess immediate and longer-term effects of the pandemic.

[READ MORE](#)

Example of COVID-19 replication data

medRxiv THE PREPRINT SERVER FOR HEALTH SCIENCES

CSH Cold Spring Harbor Laboratory **BMJ** Yale

HOME | ABOUT | SUBMIT | ALERTS / RSS

Search [] Advanced Search

Comments on this paper

Population-Level Mortality Rates from Novel Coronavirus (COVID-19) in South Korea

Samir Soneji, Hiram Beltrán-Sánchez, JaeWon Yang, Caroline Mann
doi: <https://doi.org/10.1101/2020.03.23.20041814>

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Abstract Info/History Metrics Preview PDF

Abstract

Background: South Korea was among the first countries to report a case of the novel coronavirus (COVID-19) outside of China. As of 22 March, 2020, South Korea reported 8897 confirmed cases of and 104 deaths from COVID-19. Methods: We collected the number of laboratory-confirmed cases and deaths in South Korea from the World Health Organization (as of 21 March, 2020) and case distribution and fatality rates by age from the Korean Center for Disease Control and Prevention (as of 22 March, 2020). We estimated population-level mortality rates by fitting a negative binomial regression model with the number of deaths as the outcome and population by age as an offset. We then calculated the age-standardized death rate (ASDR) based on the current COVID-19 figures and for alternative scenarios of increased prevalence. Findings: The COVID-19 population-level mortality rate (per 100,000 person-years) increased with age: from 0.1 deaths among 30-39 year olds to 9.5 deaths among ≥80 year olds. The ASDR (per 100,000 person-years) was 0.8 deaths. The ASDR would increase to 5.9

Preprint

Replication Data + Code

HARVARD Dataverse

Add Data Search About User Guide Support Sign Up Log In

Replication Data for: Population-Level Mortality Rates from Novel Coronavirus (COVID-19) in South Korea

Version 2.0

Soneji, Samir, 2020, "Replication Data for: Population-Level Mortality Rates from Novel Coronavirus (COVID-19) in South Korea", <https://doi.org/10.7910/DVN/SLJ10C>, Harvard Dataverse, V2, UNF:6:5i89rp8FJEdIANpkWzsUQ== [fileUNF]

Dataset Metrics 45 Downloads

Files Metadata Terms Versions

Search this dataset... Find

Filter by File Type: All Access: All

1 to 7 of 7 Files

File Name	Size	Created	Downloads	Actions
asdr.pdf	63.6 KB	Apr 21, 2020	0 Downloads	Explore Download
cod_kor.tab	584 B	Apr 21, 2020	15 Downloads	Explore Download
covid19_kor.R	2.9 KB	Apr 21, 2020	0 Downloads	Explore Download
covid19_mortality_rates_south_korea.pdf	18.8 KB	Apr 21, 2020	0 Downloads	Explore Download
functions.R	11.2 KB	Apr 21, 2020	0 Downloads	Explore Download

Most popular dataset in Harvard Dataverse in 2020

The screenshot shows the Harvard Dataverse interface for the dataset "World COVID-19 Daily Cases with Basemap". The page includes a navigation bar with "Add Data", "Search", "About", "User Guide", "Support", "Sign Up", and "Log In". The dataset title is "World COVID-19 Daily Cases with Basemap" (Version 15.0). A "Dataset Metrics" box shows "5,466 Downloads". The description states it was updated to April 18, 2020, and covers Earth and Environmental Sciences, Medicine, Health and Life Sciences, and Social Sciences. The keyword is "2019-nCoV, statistics". Below the description are tabs for "Files", "Metadata", "Terms", and "Versions". A search bar and filter options are present. The file list shows four items: "Countries-Confirmed.tab", "Countries-Deaths.tab", "Countries-Recovered.tab", and "World_Map_0302.zip". A dropdown menu is open over the "Download" button for "Countries-Confirmed.tab", showing options: "Original File Format (MS Excel Spreadsheet)", "Tab-Delimited", "RData Format", "Variable Metadata", and "Data File Citation".

Published on February 20

5,466 Downloads

8,293 Page Views

From China Data Lab

Resources for Novel Coronavirus Global Research

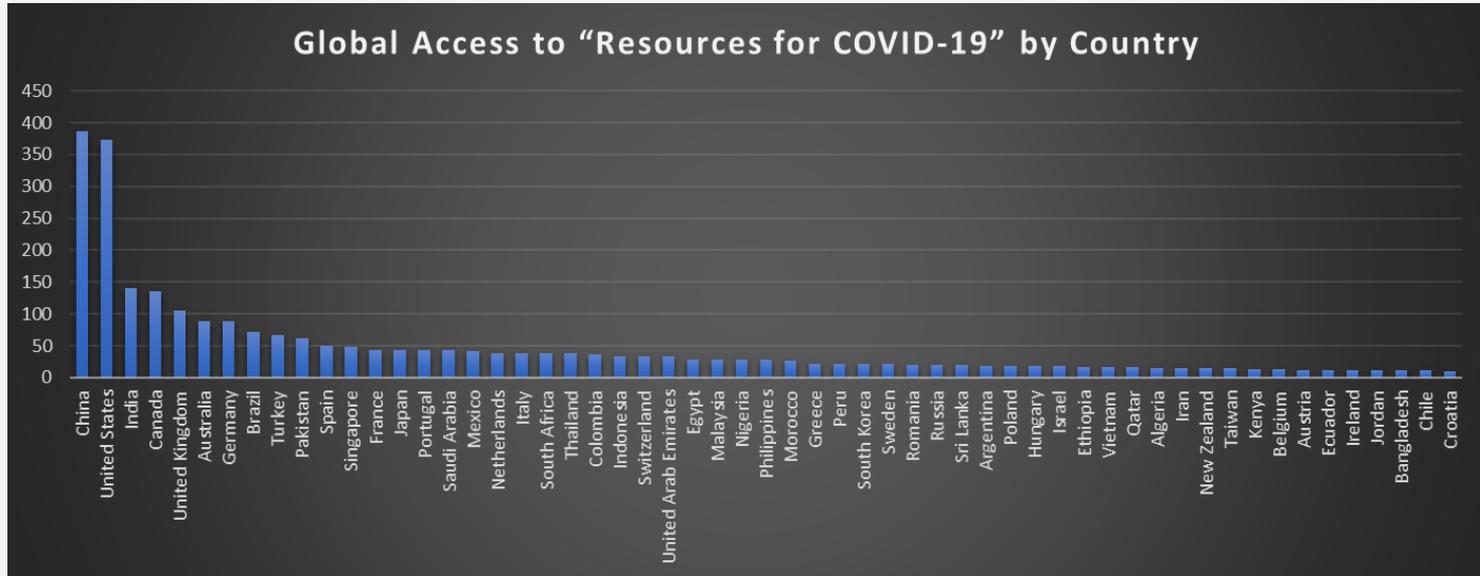
The screenshot shows the Harvard University China Data Lab website. At the top, there is a black navigation bar with the Harvard University logo and 'HARVARD UNIVERSITY' on the left, and 'HARVARD.EDU' on the right. Below this is a white header area with the 'GDL China Data Lab' logo and the text '@Center for Geographic Analysis, IQSS'. A search bar is located to the right of the logo. A red navigation bar contains the following menu items: HOME, People, Resources, Partners, Events, About. On the left side, there is a vertical sidebar with a red header 'RESOURCES' and a list of links: Enter the Lab, Resources for COVID-19, Publications, Presentations, and Data Case Studies. The main content area is titled 'Resources for COVID-19' and is divided into four quadrants. The top-left quadrant features the Harvard Dataverse logo and lists two links: 'Resources for COVID-19 at Dataverse' and 'User Guide for "Resources for COVID-19"'. The top-right quadrant features the 'Global Research on Novel Coronavirus' logo and lists two links: 'Cloud for Global Research on COVID-19' and 'User Guide for "Cloud Platform for COVID-19"'. The bottom-left quadrant is titled 'CALL FOR PAPERS COVID-19' and lists three links: 'China Economic Review: Special Issue on Virtual Pandemic Disease', 'Journal of Data Science: Special Issue on Data Science in Action in Response to the Outbreak of COVID-19', and 'Transportation Research Interdisciplinary Perspectives: Special Issue on Transportation Impacts and Responses to the Coronavirus (COVID-19)'. The bottom-right quadrant is titled 'Free User Account on Cloud Platform for Global Research on COVID-19' and lists three links: 'Guideline for Applications', 'Apply online:' (with sub-links for 'The Online Application in Chinese' and 'The Online Application in English'), and 'Download the application form:' (with sub-links for 'The application form in Chinese' and 'The application form in English').

- Part of a larger NSF IUCRC “Spatial Data Lab” project
- Focuses on COVID-19 data and analytical tools (GIS, Statistics)
- Enables access to cloud computing resources
- Facilitates creating, sharing analytical workflows
- All datasets published in Harvard Dataverse
- Datasets include: China census, socio-economic data, China, US, and Global daily COVID-19 data, health facilities, policies and regulations, climate, population mobilities.
- Datasets updated weekly

Credit to Wendy Guan, Tao Hu (IQSS, CGA, Harvard University)

Resources for Novel Coronavirus Global Research

- “Resource for COVID-19” dataverse has had more than 30,000 downloads
- Over 130 countries have accessed the datasets



Credit to Wendy Guan, Tao Hu (IQSS, CGA, Harvard University)

Making European Union COVID-19 Datasets FAIR

- In the European Union coronavirus datasets mostly **not shared as FAIR data**
- Some countries (Italy, Austria) publish their data on GitHub, without persistent identifiers
- Most countries share the official statistics as PDFs, sometimes with incorrect or missing data points
- In Spain and the Netherlands all data collected and processed by volunteers and shared on GitHub
- It's a big challenge to force official institutions to translate all their variables from the national language to English and use the same codebook for all countries to make data interoperable
- Slava Tykhonov from DANS-KNAW created an **European COVID-19 data hub and archive with FAIR datasets in the Harvard Dataverse**, updated daily:
<https://dataverse.harvard.edu/dataverse/covid-19-eu>

Another interesting initiative that started collecting and processing coronavirus data through Harvard Dataverse is [CoronaWhy](#), a global distributed organization, mobilizing volunteers and community partners to address the current challenge through data science, AI, and knowledge sharing.

Thanks and Upcoming Events

OpenDP Community Meeting, May 13-15, 2020: Session on Differential Privacy to Explore and Analyze COVID-19 sensitive data

 HARVARD UNIVERSITY HARVARD.EDU



building an open-source suite of tools for deploying differential privacy

[Home](#) [About](#) [People](#) [Events](#)

OpenDP Community Meeting 2020

On May 13th - 15th from 11 AM - 3 PM EDT each day, we will hold an online workshop to share detailed plans for OpenDP and obtain community feedback on them. We will cover topics such as the programming framework, governance, system integrations, use cases, statistical functionality, and collaborations.

The detailed agenda is below. We hope you will be able to participate in the workshop, and we look forward to your input and future engagement with OpenDP!

[Registration](#)

Detailed Agenda

▼ Wednesday, May 13th

11 AM - 3 PM EDT: Plenary Presentations
Eastern Daylight Time (UTC-04:00)

[11-11:30 Overview of OpenDP and Goals for Meeting](#)

Dataverse Community Meeting, June 17-19, 2020: Session on Global COVID-19 Data Sharing

 HARVARD UNIVERSITY HARVARD.EDU

Dataverse Community Meeting 2020

Held Remotely on June 17, 18, 19



[Home](#) [Plenary Session](#) [Breakout Sessions](#)



#dataverse2020

#dataverse2020 will be an online event. [Read more...](#)

The annual Dataverse Community Meeting is an opportunity to build, grow, and enrich the global community. Like the open-source Dataverse product itself, the activities of the Dataverse Community Meetings are community-driven. Over three days of presentations, workshops, and working group meetings we aim to promote and learn about behavioral and technical solutions and standards for curating, sharing, and preserving data that can be discovered and reused across disciplines to reproduce and advance research.

Tweets about #dataverse2020

 [shlakeuva](#) Tim says "Thanks!" #dataverse2020 t.co/GxOvzEGVjm t.co/TJOGMCMDDy

