

Data Sharing in Epidemiology



Version: 0.052

DOI: [10.15497/rda00049](https://doi.org/10.15497/rda00049)

Authors: RDA-COVID-19-Epidemiology Working Group

Co-chair: Priyanka Pillai

Moderators: Claire Austin, Gabriel Turinici

Published: 15th June 2020

Abstract: An immediate understanding of the COVID-19 disease epidemiology is crucial to slowing infections, minimizing deaths, and making informed decisions about when, and to what extent, to impose mitigation measures, and when and how to reopen society.

Despite our need for evidence based policies and medical decision making, there is no international standard or coordinated system for collecting, documenting, and disseminating COVID-19 related data and metadata, making their use and reuse for timely epidemiological analysis challenging due to issues with documentation, interoperability, completeness, methodological heterogeneity, and data quality.

There is a pressing need for a coordinated global system encompassing preparedness, early detection, and rapid response to newly emergent threats such as SARS-CoV-2 virus and the COVID-19 disease that it causes.

The intended audience for the epidemiology recommendations and guidelines are government and international agencies, policy and decision makers, epidemiologists and public health experts, disaster preparedness and response experts, funders, data providers, teachers, researchers, clinicians, and other potential users.

Keywords: COVID-19; Supporting output; Epidemiology; Recommendations; Data Sharing

Language: English

License: [Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Citation and Download: RDA COVID-19-Epidemiology WG (2020): Data Sharing in Epidemiology, Version 0.052. DOI: [10.15497/rda00049](https://doi.org/10.15497/rda00049)

[Research Data Alliance RDA-COVID19
Epidemiology Work Group](#)

Co-moderators: Claire C. Austin and Gabriel Turinici

DATA SHARING IN EPIDEMIOLOGY

SUPPORTING OUTPUT

Draft v0.052 (2020-06-15)

[CC-BY 4.0 license](#)

CONTENTS

Change Log	2
DATA SHARING IN EPIDEMIOLOGY	3
1 Focus and Description	3
2 Scope	3
3 Policy Recommendations	3
3.1 General.....	3
3.2 Information Technology and Data Management.....	4
3.3 COVID-19 Epidemiological data, analysis, and modeling.....	4
4 Guidelines	5
4.1 COVID-19 Population Level Data Sources	5
4.2 Interoperable COVID-19 epidemiological surveillance: Clinical and Population-based instruments	6
4.3 Preservation of individuals' privacy in shared COVID-19 related data	7
4.4 A Full Spectrum View of the COVID-19 Data Domain: An Epidemiological Data Model.....	8
4.5 Epi-TRACS: Rapid detection and whole system response for emerging pathogens	8
4.6 Putting It All Together: Common Data Model and Epi-Stack.....	8
APPENDIX 1 - COVID19 Population level data sources: Review and analysis	10
Claire C. Austin, Anna Widyastuti; and the RDA-COVID19-WG	
APPENDIX 2 – Interoperable COVID-19 epidemiological surveillance: Clinical & Population-based instruments...21	
Carsten O. Schmidt, Rajini Nagrani, Matthias Löbe, Atinkut Zeleke, Guillaume Fabre, Stefan Sauermann, Anna Widyastuti, Chifundo Kanjala, Jay Greenfield, Claire C Austin; and the RDA-COVID19-WG	
APPENDIX 3 – Preservation of individuals' privacy in shared COVID-19 related data	29
Stefan Sauermann, Chifundo Kanjala, Matthias Templ; and the RDA-COVID19-WG	
APPENDIX 4 – A Full Spectrum View of the COVID-19 data domain: An Epidemiological Data Model	36
Jay Greenfield, Rajini Nagrani, Meg Sears, Claire C. Austin; and the RDA-COVID19-WG	
APPENDIX 5 – Epi-TRACS: Rapid detection and whole system response for emerging pathogens	43
Jay Greenfield, Henri E.Z. Tonnang, Gary Mazzaferro, Claire C. Austin; and the RDA-COVID19-WG	
APPENDIX 6 – Putting It All Together: Common Data Model and Epi-Stack.....	58
Jay Greenfield, Meg Sears, Rajini Nagrani, Gary Mazzaferro, Anna Widyastuti, Claire C. Austin; and the RDA-COVID19-WG	
RDA-COVID19-Epidemiology-WG Bibliography	69-82

Change Log

VERSION	DATE	CHANGE
0.051	2020-06-08	Fixed version number inconsistencies
0.051	2020-06-08	Minor edits
0.051	2020-06-08	Updated Figures
0.051	2020-06-08	Updated Appendices 1-3, and 5-6
0.051	2020-06-08	Updated the Bibliography
0.052	2020-06-15	Updated Data Sharing in Epidemiology (Focus, Description, Scope, Recommendations, Guidelines, and Table 1)
0.052	2020-06-15	Updated Appendix 1
0.052	2020-06-15	Updated the Bibliography

DATA SHARING IN EPIDEMIOLOGY

1 Focus and Description

An immediate understanding of the COVID-19 disease epidemiology is crucial to slowing infections, minimizing deaths, and making informed decisions about when, and to what extent, to impose mitigation measures, and when and how to reopen society.

Despite our need for evidence based policies and medical decision making, there is no international standard or coordinated system for collecting, documenting, and disseminating COVID-19 related data and metadata, making their use and reuse for timely epidemiological analysis challenging due to issues with documentation, interoperability, completeness, methodological heterogeneity, and data quality.

The intended audience for the epidemiology recommendations and guidelines are government and international agencies, policy and decision makers, epidemiologists and public health experts, disaster preparedness and response experts, funders, data providers, teachers, researchers, clinicians, and other potential users.

2 Scope

Epidemiology underpins COVID-19 response strategies and public health measures. The recommendations and guidelines support development of an internationally harmonized specification to enable rapid reporting and integration of epidemiology and related data across domains and between jurisdictions

The guidelines outline a data driven, coordinated global system that encompasses preparedness, early detection, and rapid response to newly emergent threats such as SARS-CoV-2 virus and the COVID-19 disease that it causes.

Supporting Output

The supporting output (doi.org/10.15497/rda00049) provides supplemental resources, and further develops the global data-driven vision described in the guidelines. This includes a proposed computable framework to support system responses for emerging pathogens. It offers compatible and reliable data models, protocols, and action plans for newly identified threats such as COVID-19.

3 Policy Recommendations

3.1 General

1. Urgently update data sharing policies and Memoranda of Understanding (MOUs) across all domains, in government, healthcare systems, and research institutions to support Open Data, Open Science, scientific data modernization, and linked data life cycles that will enable rapid and credible scientific and epidemiologic discovery, and fast-track decision-making.
2. Streamline data flow between sub-national jurisdictions/institutions and their national government, and countries and international organizations.

3. Implement a “data first” publication policy in research by treating publication of data articles in “open” peer-reviewed data journals, including the deposit of data and associated code in a trusted digital repository with tiered access to appropriately credentialed people and machines to preserve data security.
4. Peer-reviewed data articles should be treated as first-class research outputs equal in value to traditional peer-reviewed articles.
5. Call upon the international [Open Government Partnership \(OGP\)](#) to add “Open Science” as one of its Policy Areas to be included in National Action Plans. Hold member countries accountable for developing and implementing Open Science commitments.
6. Publish situational data, analytical models, scientific findings, and reports used in decision-making and justification of decisions ([OGP 2020](#)).

3.2 Information Technology and Data Management

Properly funded state of the art infrastructure is required to support advanced research, as well as the and data management and data sharing required for rapid response and collaboration (See [Infrastructure Investment](#)). For epidemiology in particular:

1. Ensure an appropriate semantic annotation of data to facilitate its comparability across studies and countries, using as much as possible established standards (e.g. LOINC, UMLS).
2. Rapidly develop standardised tools for aggregating microdata to a harmonised format(s) that can be shared and used while minimising the re-identification risk for individual records.
3. Develop machine readable citations and micro-citations for dynamic data. Rapid development of: (a) Resolvable Persistent Identifiers, rather than Uniform Resource Locators (URLs); (b) Machine readable citations; (c) Micro-citations that refer to the specific data used from large datasets; and, (d) Date and Time Access citations for dynamic data (ESIP, 2019).

3.3 COVID-19 Epidemiological data, analysis, and modeling

1. Rapidly develop a consensus standard for COVID-19 surveillance data:
 - a. Definition of and reporting criteria for COVID-19 testing, reporting on testing, and testing turnaround times.
 - b. Policies and definitions: interventions, contact tracing, reporting of cases, deaths, hospitalisations and length of stay, ICU admissions, recoveries, reinfections, time from contact if known, symptoms onset and detection, through clinical course and interventions, to death or recovery, comorbidities, long-term effects in recovered cases, sequelae and immunity, location, demographic, socioeconomic information, and outcome of resolved cases.
 - c. Uniform standard daily reporting cut-off time.
2. Rapidly develop an internationally harmonised specification to enable the export/import/integration of epidemiologic data across different levels of data generation (e.g., clinical systems, population-based surveillance/research data, data from biomarker and omics studies, death certification, health insurance data), and successful record-linkage.
3. Develop systems that support workflows to link and share data between different domains, while protecting privacy and security. Use domain specific, time stamped, encrypted person identifiers for this purpose based on industry-standard encryption and cryptographic constructions.

4. Implement internationally harmonised COVID-19 intervention protocols based on peer-reviewed empirical modelling and epidemiological evidence, considering local conditions.
5. Publish situational data, analytical models, scientific findings, and reports used in decision-making and justification of decisions ([OGP, 2020b](#)).
6. Account for public health decision making demands in COVID-19 studies.
7. Harmonise approaches to comparably assess and quantify side-effects of pandemic containment and mitigation measures.
8. Report underlying assumptions and quantify effects of uncertainties on all reported parameters and conclusions for all model predictions etc.
9. Implement a data-driven approach for early identification of hotspots.

4 Guidelines

These guidelines highlight current system challenges and offer solutions to help support a larger framework designed to coordinate and structure the collection and use of COVID-19 related data. Six focus areas, described in the guidelines and [supporting output](#) (data sources, instruments, privacy, epidemiological data model, computable framework, and an epi-stack architecture), progressively develop a data driven global vision for managing novel biological threats such as COVID-19. We begin with population level data sources that drive the public health strategy and response at all stages of the COVID-19 threat, from emergence through containment, mitigation, and reopening of society. We then survey clinical and population-based instruments that collect data and discuss preservation of individuals' privacy in shared COVID-19 related data. A full spectrum data model is presented encompassing hospital specific surveillance and electronic health records together with field-based demographic and epidemiological surveillance. We propose Epi-TRACS, a computable framework for emerging pathogen action plans, and an epi-stack that uses the Common Data Model with COVID-19 to integrate clinical data and epidemiological data.

4.1 COVID-19 Population Level Data Sources

Although jurisdictions within countries send COVID-19 population level data to the national level, and member countries send data to the WHO, other organisations also collect COVID-19 surveillance data from various sources for a variety of reasons (Table 1). Epidemiologists are thus faced with a situation where it is difficult to assess which datasets are the most up-to-date, complete, and reliable.

See Appendix 1 for further details and discussion.

Table 1. COVID-19 population level data sources

Allen Institute for AI	COVID-19 Open Research Dataset (CORD-19)
Apple Inc.	COVID-19-Mobility Trends Reports
European Centre for Disease Control	Geographic distribution of COVID-19 cases worldwide
European Centre for Disease Control	The European Surveillance System (TESSy)
Institute for Health Metrics and Evaluation (IHME)	Global Health Data Exchange (GHDx)
Google Inc.	COVID-19 Community Mobility Report
Johns Hopkins University	COVID19 dataset
Kieren Healy	Rpackage - COVID19 Case and Mortality Time Series
Oxford University	COVID19 dataset
The Atlantic	COVID Tracking Project
The New York Times	Covid-19 Data in the United States
U.N.	Humanitarian Data Exchange (HDX)
U.S. Centre for Disease Control	Cases of COVID19 in the U.S.
University of Washington	Be Outbreak Prepared
World Bank	Understanding the Coronavirus (COVID-19) pandemic through data
World Health Organization (WHO)	Novel Coronavirus (2019-nCoV) situation reports
Worldometer	COVID19 data

4.2 Interoperable COVID-19 epidemiological surveillance: Clinical and Population-based instruments

International efforts are currently underway to create COVID-19 instruments/questionnaires (Tables 2 and 3). These COVID-specific tools are concentrated at person-level for clinic/hospital surveillance (e.g., Case Report Forms-CRFs), or community surveillance (e.g., questionnaire for general population), and do not necessarily collect the same data. Adherence of new studies to already introduced instruments will strongly enhance the comparability of results.

Table 2. Questionnaire instruments: Reference studies

<u>CLINICAL</u>	
Australia:	NSW Case questionnaire
Germany:	Covid-19 research dataset
Uganda:	Perinatal COVID-19 Uganda
US:	Human Infection with 2019 Novel Coronavirus Person Under Investigation (PUI) and Case Report Form
WORLDWIDE	
(Member states of WHO): Global COVID-19: clinical platform: novel coronavirus (COVID-19): rapid version	
<u>POPULATION-BASED</u>	
Brazil:	Brazil Prevalence of Infection Survey
Europe:	Questionnaire by WHO Europe
Germany:	GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany
Germany:	NAKO COVID-19 Survey tool

Israel	One-minute population wide survey
Low and Middle Income Countries:	LMIC Covid Questionnaire
South Africa:	South African Population Research Infrastructure (SAPRIN) COVID-19 Screening Form
South Asian countries:	National Institute for Health Research (NIHR) Global Health Research Unit
UK	UK COVID-19 Questionnaire
Worldwide (WHO):	Population-based age-stratified sero-epidemiological investigation protocol for COVID-19 virus infection

Table 3. Questionnaire instruments: Resources

NIH	Public Health Emergency and Disaster Research Response (DR2)
NIH	COVID-19OBSSR Research Tools
PhenX	PhenX COVID-19 Toolkit

Some of the questionnaire initiatives shown in Tables 2 and 3 are currently feeding into the construction of a COVID-19 demographic and epidemiological surveillance question bank that can be used to form locality specific surveys with both common and distinct questions by domains and cohorts (Wellcome Trust). Some, such as the [UK COVID-19 Questionnaire](#), or the [Covid-19 research dataset](#) are now being funded. Question banks, once they become operational can be queried and filtered by domain, cohort, question text, etc. Based on such queries, new questionnaire products can be developed that are more or less interoperable, depending on the questions selected and the capture of “localisation” information in the question metadata when questions are reused from one survey to the next.

See Appendix 2 for further details and discussion.

4.3 Preservation of individuals’ privacy in shared COVID-19 related data

Data sharing is essential to improve epidemiological analysis, cross-border pandemic modelling, and coordinated policy development between countries. To ensure privacy, both pseudo-anonymisation of direct identifiers (e.g. patient specific ID’s) and anonymisation of indirect identifiers (e.g. socio-demographic information on individuals) must be applied. In addition, it is necessary to control statistical disclosure risk to prevent identification of individuals and their health status using a combination of indirect identifiers such as education level, sex, age, and clinical condition, among others ([Duncan et al., 2011](#); [Templ et al., 2015](#); [Templ, 2017](#)). Using synthetic data may be an option to lower re-identification risks while retaining properties of the original data sets.

See Appendix 3 for further details and discussion.

4.4 A Full Spectrum View of the COVID-19 Data Domain: An Epidemiological Data Model

The COVID-19 epidemiology that guides public health decisions is dependent on interoperable input data from across a wide variety of domains that include not only clinical, surveillance, research, and modelling data, but also administrative, demographic, socioeconomic, cultural practices and lifestyle, and environmental data, amongst others.

An epidemiological surveillance data model must include the primary data domains that need to be integrated to understand COVID-19, and to improve surveillance and follow-up: (a) clinical event history and disease milestones; (b) epidemiological indicators and reporting data; (c) contact tracing; (d) personal risk factors.

Standardization challenges within each of these domains remain to be solved before data can be effectively integrated across domains for epidemiology studies. For example, on the clinical side, the U.S. Clinical Data Interchange Standards Consortium (CDISC) new specification (Interim User Guide for COVID-19), and the WHO Core and Rapid COVID-19 Case Reporting Forms used in low- and middle-income Countries (LMIC) require additional harmonization.

See Appendix 4 for further details and discussion.

4.5 Epi-TRACS: Rapid detection and whole system response for emerging pathogens

WHO's Global Influenza Surveillance Response System (GISRS) is a well-established network of more than 150 national public health laboratories in 125 countries that monitors the epidemiology and virologic evolution of influenza disease and viruses (WHO, 2020).

Prior to the COVID-19 outbreak, WHO was already engaged in re-examining GISRS's long-term fitness-for-purpose. In line with these short-term considerations, and with GISRS long-term aspirations, we are recommending a real time, adaptable, rapid response system that supports developing countries, and that employs new technology to combat pandemics and other emerging diseases. The RDA-COVID19-Epidemiology group recommends the creation of a WHO-led EPIdemiological Translational Research Action Coordination System (Epi-TRACS) to add an implementation layer to the existing WHO policies, guidelines, partnerships, and information exchange stack adapted to country-specific contexts.

See Appendix 5 for further details and discussion.

4.6 Putting It All Together: Common Data Model and Epi-Stack

Common Data Model (CDM)

Data models may make use of the broad ecosystem of surveillance and clinical data that can also include contact tracing apps, biospecimens, and environmental sample data collected in the community/population or clinic/hospitals.

An emulated trials approach may enable assessment of various risk and prognostic factors ([Hernán et.al., 2016](#)). Application of a Common Data Model (CDM) for COVID-19 would facilitate comparing

clinical burden and patient outcomes in the context of previous environmental and exposures and comorbidities.

Another possible use case is decision support following an early warning system alert of emergence of a novel pathogen such as SARS-CoV-2. The CDM provides a framework for making public health policy decisions, using partial information about the pandemic that leverages population-level population and health information, person-level epidemiological surveillance information collected in the field and, at the same time or alternatively, person-level patient care information collected in a clinic or hospital setting.

Epi-Stack

The WHO has established the Information Network for Epidemics ([Epi-WIN](#)) covering four strategic areas: (a) Identify; (b) Simplify; (c) Amplify; and, (d) Quantify. Evidence is gathered, appraised, and assessed to help form recommendations and policies that have an impact on the health of individuals and population.

The RDA Epi subWG proposes an expanded Epi-Stack feeding into Epi-WIN. This would bring together in a managed system a common data model, the epidemiological surveillance data model, clinical and questionnaire data, population level indicators, and core use cases (Epi-TRACS early warning and response system, decision support research, and patient care research). It is critical that resiliency be built into the system, from the standpoint of how the system functions under stress. When the degree of complexity and interdependencies increase in human made systems, there is always the risk of collapse if not enough balance is built into the system (both the IT infrastructure, data governance, and the "people" part).

See Appendix 6 for further details and discussion.

Acknowledgments

The RDA-COVID19-WG would like to acknowledge the following people who have contributed their time, knowledge, and expertise to generate these recommendations and guidelines. Listed alphabetically by last name and ORCID ID:

Claire C. Austin ([0000-0001-9138-5986](#)); Marlon L Bayot ([0000-0002-5328-150X](#)); Maeve Campman ([0000-0002-6613-7144](#)); David Delmail ([0000-0003-2836-6496](#)); Cyrille Delpierre ([0000-0002-0831-080X](#)); Gayo Diallo ([0000-0002-9799-9484](#)); Jay Greenfield ([0000-0003-2773-5317](#)); Chifundo Kanjala ([0000-0003-0540-8374](#)); Birte Lindstaedt ([0000-0002-8251-1597](#)); Gary Mazzaferro ([0000-0002-3196-7201](#)); Robin Michelet ([0000-0002-5485-607X](#)); Rajini Nagrani ([0000-0002-1708-2319](#)); Carlos Luis Parra-Calderón ([0000-0003-2609-575X](#)); Priyanka Pillai ([0000-0002-3768-8895](#)); Stefan Sauermann ([0000-0003-0824-9989](#)); Carsten Oliver Schmidt ([0000-0001-5266-9396](#)); Meg Sears ([0000-0002-6987-1694](#)); Henri Tonnang ([0000-0002-9424-9186](#)); Gabriel Turinici ([0000-0003-2713-006X](#)); and, Anna Widyastuti ([0000-0003-2149-935X](#)).

Working paper

APPENDIX 1 - COVID19 Population level data sources: Review and analysis

Claire C. Austin^{1,*}, Anna Widyastuti²; and the RDA-COVID19-WG³

¹Environment and Climate Change Canada, ²Hasselt University, Belgium ³This work is linked to the Research Data Alliance [RDA-COVID19-WG, Epidemiology subWG recommendations and guidelines on data sharing](#).

*Corresponding author: claire.austin@mail.mcgill.ca

All views and opinions expressed are those of the co-authors, and do not necessarily reflect the official policy or position of their respective employers, or of any government, agency or organization.

BACKGROUND

COVID-19 data is a Big Data problem, especially with respect to variety, velocity, veracity, and volume (NIST 2019, Volumes 1-9). The WHO receives COVID-19 data from member countries which typically receive it at the national level from provinces/states, counties, municipalities, public health authorities, and/or hospitals. The WHO should then be the central “go to” source to find world-wide, population level COVID-19 data. However, the system is fraught with antiquated data collection and transmission methods which slow down the reporting, introduce errors and incomplete data, and make the data difficult to find and use. As a result, other organizations and individuals have also compiled COVID-19 related population level data in an attempt to provide well organized, analysis ready data for their own use or for the benefit of other end-users. Added to the mix, are a large number of dashboards which also provide data, but whose motivation is primarily to drive traffic to their website to promote the organization or a product, making the completeness and reliability of the underlying data questionable. Epidemiologists and other end users are thus faced with a situation where the data are fragmented, and difficult to find, access, and assess.

The objective of the present paper was the creation and analysis of a dataset to provide:

1. reliable resources for downloading COVID-19 related population level data;
2. a description of the type and format of the available data;
3. links to the actual data; and,
4. information about each of the resources to help guide data users in selecting the resources most suitable for their needs.

METHODS

Dataset

Resources providing compiled COVID-19 population level data were identified from PubMed and Google searches, and from catalogues discovered during those searches.

Inclusion criteria: The resource must provide COVID-19 related population level data in a downloadable, machine-readable format, and must document their data source(s).

Exclusion criteria: Jurisdictional (country, province/state, municipality, hospital, public health authority, etc.) data sources were generally excluded, with some exceptions as described in the discussion. Data embedded in PDF files were not included, an exception being the WHO daily situation reports. Maps and dashboards were not included, unless they also provided links to the underlying code and data used to generate the visualization. Data summaries, summary statistics, and the results of data analyses were not included.

Dataset description

Retained resources were further explored to identify the purpose of the data compilation, types of data available, download links, and especially the sources relied upon to compile the data.

RESULTS

Table 1. COVID-19 related population datasets and sources

Table 2. Resources that provide a corpus of COVID-19 related text

Table 3. Resources that track COVID-19 policy responses

Table 4. Very useful visualizations of COVID-19 data that go beyond the usual “dashboards”

Table 5. Sites that maintain a database or catalogue of resources.

Table 6. COVID-19 related data analysis platforms

Table 1. COVID-19 related population datasets and sources. Resources are listed in alphabetical order.

ID	Website or resource creator	Website dataset/database name	Data or code and file type (available for download from the website)
	Description and sources used to create the dataset/database (as described on the website)		
1	1Point3Acres	1Point3Arces	Need to request data access
	Manually aggregates data from official government websites and latest news reports. Volunteers fact-check and de-duplicate all the updates. This is a data tracker. Its data are used by CDC, JHU COVID19 World Map, UN and FEMA.		
2	Apple	COVID-19-Mobility Trends Reports	CSV file and charts. Downloading data must adhere to the terms of use.
	Data sent from users' devices to the Maps service. Data is associated with random, rotating identifiers so Apple doesn't have a profile of individual movements and searches. Reports are published daily and reflect request for directions.		
3	Chen, Emily	The COVID-19 Tweet IDs dataset	Hydrating using Hydrator (GUI) or Twarc (CLI)
	Data were collected on January 28, 2020, leveraging Twitter's streaming application programming interface (API) and Tweepy to follow certain keywords and accounts that were trending at the time data collection began. Twitter's were used to search API to query for past tweets, resulting in the earliest tweets in the collection dating back to January 21, 2020.		

4	eCDC-European Centre for Disease Control	COVID-19 surveillance report	Intensity (html) Epidemic curves (html) Severity (html) Risk groups most affected Other epidemiologic characteristics Global COVID-19 data XLSX , CSV , JSON , XML Data quality (html) Read into R: <pre>library(utils) data <- read.csv("https://opendata.ecdc.europa.eu/covid19/casedistribution/csv", na.strings = "", fileEncoding = "UTF-8-BOM")</pre> Need to request access to infectious diseases data. Aggregated data infectious diseases (CSV).
			Daily counts of reported cases and reported deaths collected by ECDC epidemic intelligence COVID-19. Reports from the Early Warning and Response System (EWRS), EU/EEA countries through the European Surveillance System (TESSy), the WHO, and email exchanges with other international stakeholders. This information is complemented by screening up to 500 sources every day to collect COVID-19 figures from 196 countries. This includes websites of ministries of health (43% of the total number of sources), websites of public health institutes (9%), websites from other national authorities (ministries of social services and welfare, governments, prime minister cabinets, cabinets of ministries, websites on health statistics and official response teams) (6%), WHO websites and WHO situation reports (2%), and official dashboards and interactive maps from national and international institutions (10%). In addition, ECDC screens social media accounts maintained by national authorities, for example Twitter, Facebook, YouTube or Telegram accounts run by ministries of health (28%) and other official sources (e.g. official media outlets) (2%). The European Surveillance System (TESSy) is a web application and a database used by European Member States to report surveillance data. eCDC collect and process COVID-19 data. There are three types of access to TESSy data: direct access to data (restricted), access to subsets of data, and access to aggregated published data (unrestricted.). Direct access to data refers to validated data (i.e. data validated by the data providers according to the metadata set validation rules) within TESSy. Access to query the TESSy database is restricted to qualified users. The only unrestricted dissemination is aggregated data via the eCDC, not via TESSy. TESSy metadata report
5	Google	COVID-19 Community Mobility Report	CSV file. Use of data must adhere to Google Terms of service
			Aggregated, anonymized sets of data from users who have turned on the location history setting, which is off by default. The report will be available for a limited time, so long as public health officials find them useful in their work to stop the spread of COVID-19.
6	HGH-Harvard Global Health Institute	COVID-19 Hospital Capacity Estimates 2020	HRR scorecard (Gsheets) Data guide
			Launched in collaboration with ProPublica, and in sync with news reports in The New York Times and CBS Morning News. Sources: COVID-19 model ;
7	Healy, Kieran	Rpackage - COVID19 Case and Mortality Time Series	European Centers for Disease Control; COVID Tracking Project ; New York Times ; COVID-NET Coronavirus Disease 2019 (COVID-19)-Associated Hospitalization Surveillance Network ; Human Mortality Database ; New York Times . Apple mobility trends ; Google mobility trends . CoronaNet Project

8	JHU-Johns Hopkins University	COVID19 dataset	csse covid 19 daily reports ; csse covid 19 daily reports us ; csse covid 19 time series ; who covid 19 sit rep time series ; Data modification records COVID-19 Case tracker updated daily.
GLOBAL DATA: WHO ; European CDC ; China global report (DXY.cn) ; WorldoMeter ; 1Point3Arcs ; BNO News . US DATA: US CDC; The Atlantic COVID Tracker ; Colorado ; Florida and Florida maps ; Maryland ; New York State (Dept Health) ; NYC HMM and NYC Health ; Washington State . NON-US DATA: Australia and Australia COVID Live ; Brazil Canada ; Chile ; China NHC and China CDC ; Colombia and Instituto Nacional de Salud ; France and France key data ; Germany ; Hong Kong ; Israel ; Italy and Italy regions ; Japan ; Kosovo (and amazon); Macau ; Mexico ; OpenCOVID19 France ; Peru ; Russia ; Serbia ; Singapore ; Spain ; Sweden ; Taiwan CDC ; Ukraine ; West Bank and Gaza .			
9	NYT-New York Times	COVID-19 data in the USA	us.csv (raw CSV) . states.csv (raw CSV) counties.csv (Raw CSV file) Geographic exceptions Excess deaths tracker
Journalists working across several time zones to monitor news conferences, analyze data releases and seek clarification from public officials on how they categorize cases. In most instances, the process of recording cases has been straightforward. But because of the patchwork of reporting methods for this data across more than 50 state and territorial governments and hundreds of local health departments, our journalists sometimes had to make difficult interpretations about how to count and record cases. For those reasons, our data will in some cases not exactly match with the information reported by states and counties. For example, when a resident of Florida died in Los Angeles, we recorded her death as having occurred in California rather than Florida, though officials in Florida counted her case in their own records. And when officials in some states reported new cases without immediately identifying where the patients were being treated, we attempted to add information about their locations once it became available.			
10	Panchadsaram et al.	COVID exit strategy	Data (Google sheet)
Criteria . Gating criteria provided by the White House in their <i>Reopening America Again</i> guidelines. Sourced publicly available data that best represents where a state is at. Some sources are more "real-time" like case data , but others can lag a week like influenza-like illness (ILI) data . Data sources: COVID Tracking Project (cases, deaths, tests); rt.live (effective reproduction number); CDC (ILI-Influenza-Like Illness activity level indicator determined by data reported to ILINet); CDC (ICU & Hospital Utilization)			
11	The Atlantic	COVID Tracking Project	COVID-19 Google sheets COVID-19 API (csv and JSON) Grading State data quality (Google sheets) Github
Collects the data directly from the public health authority in each US state and territory (and the District of Columbia). Each of these authorities reports its data in its own way, including online dashboards, data tables, PDFs, press conferences, tweets, and Facebook posts. The data team uses website-scrappers and trackers to alert them to changes, but the actual updates to their datasets are done manually by volunteers who watch press conferences, follow social media tips about changes in data definitions, double-check each change, and extensively annotate areas of ambiguity.			
12	The Atlantic & Antiracist Research & Policy Center (ARPC)	COVID racial data tracker	Racial/ethnicity CSV
Compiles COVID-19 race and ethnicity data for states that are reporting it. This is a challenging dataset to compile and code. While an increasing number of states and territories are providing race and ethnicity data, the data remain incomplete and inconsistent.			

13	UN-United Nations	Humanitarian Data Exchange (HDX)	COVID-19 Pandemic collection Contributors are encouraged to upload data in any common data format. For tabular data, CSV is preferable, and for geographic data, zipped shapefile is preferred.
	HDX is a platform where organizations can upload their data for sharing. Registered users can “follow” the data they are interested in. Updates to datasets appear as a running list in the user dashboard. Users can contact data contributors to ask for more information about their data, and to request access to the underlying data for metadata only entries. HDX uses open-source software (CKAN) for the technical back-end. All of the code is available on GitHub .		
14	UoO-University of Oxford	Our World in Data	COVID19 dataset (CSV, XLSX, JSON) Testing Change log (html) Codebook
	Confirmed cases and deaths from the ECDC . Testing for COVID-19 from official reports, updated twice a week.		
15	UW-University of Washington, Institute for Health Metrics and Evaluation (IHME)	Global Health Data Exchange (GHDx)	Projections Download the most recent data on COVID-19 estimates (CSV, pdf)
	GHDx is a data catalog created and supported by the Institute for Health Metrics and Evaluation (IHME). Data made available for download by IHME can be used, shared, modified, or built upon by non-commercial users via the Open Data Commons Attribution License. Most information about datasets and series comes from the data providers, and also many of the resources noted on Data Sites We Love. The IPUMS, ICPSR, and the World Bank, in particular, provide data and extensive documentation about data. For older data, WorldCat is an invaluable resource. Data providers used to populate the information in GHDx can be found here: • UN Population Division, and the WHO Human Mortality Database: Population figures estimated based on World Population Prospects, 2015 Revision.. • CIA World Factbook: Country start and end dates, notes, and sovereignty. • Organizations’ websites: Organization start/end dates and acronyms, city and country of their location.		
16	UW-University of Washington	Be Outbreak Prepared	Tar.gz Current sources: sources_list.txt Metadata.txt and datasets derived from GBD are used for co-morbidity estimates . GBD provides query tool to download data in CSV format .
	Methods in: Epidemiological data from the COVID-19 outbreak, real-time case information Source list A range of different sources. First, official government sources and peer-reviewed scientific papers that report primary data as the gold standard for data inclusion. Government sources include press releases on the official websites of Ministries of Health or Provincial Public Health Commissions, as well as updates provided by the official social media accounts of governmental or public health institutions. Second, to find additional details for each case or patient, these data are augmented with online reports, mainly captured through news websites (e.g., https://www.163.com) or via news aggregators (e.g., https://bnonews.com/). All data sources are recorded in the database. Third, in some instances more detailed data are available, typically through peer-reviewed research articles, which were subsequently used to modify existing records in the database.		

17	UW-University of Washington – Humanistic GIS Laboratory	Novel Coronavirus (COVID-19) Infection Map	Map, Source code (GitHub), SQLite, CSV
WHO ; U.S. CDC ; State Officials; Canada (PHAC) ; 1point3acres ; NBC News , Wikipedia , China (National Health Commission) , Provincial & Municipal Health Commission, Provincial & Municipal government database, Public data published from Hongkong, Macau and Taiwan official channels); Baidu ; Mapmiao			
18	CDC-U.S. Centre for Disease Control	Cases of COVID19 in the U.S	
Data reported voluntarily by each jurisdiction’s health department. Data confirmed at 4:00pm ET the day before.			
19	World Bank	Understanding the Coronavirus (COVID-19) pandemic through data	.xlsx
JHU , WHO , UNHCR , IHME , GovLab , UNESCO , Indicators (health, water & sanitation, age & population). Datasets were initially identified from World Bank Group research and publications on Coronavirus (COVID-19) as well as through metadata analysis using keywords related to epidemiology and health care (eg. Hospital, doctor, heart disease).			
20	WHO-World Health Organization	Coronavirus disease (COVID-19) pandemic	Novel Coronavirus (2019-nCoV) daily situation reports (.pdf) WHO COVID-19 global data (.csv) downloadable from ArcGIS dashboard map . The data are also available at UN-HDX: WHO COVID-19 global data (.csv download from Gsheet)
Data are voluntarily reported to the UN by member countries. WHO asks that the designated national authorities to provide data directly to the self-reporting platform, which is made publicly available without editing or filtering by WHO. Aggregate data are made available to all Member States and the wider general public through the WHO website, may be pooled with other data to inform international response operations, and periodically published in WHO situation updates and other formats for the benefit of all Member States. Member states can self-report their data in two ways: (a) upload an Excel file directly into the system; or, (b) manually enter data using the submission platform provided. The WHO provides member states with the following files for Global Surveillance for human infection with COVID-19 : Guidance (.pdf) Revised case reporting form for COVID-19 for confirmed cases and their outcome (.pdf) Template for line listing (.xlsx) Revised data dictionary (.xlsx) Weekly reporting of new cases of COVID-19 (.xlsx)			
21	Worldometer	COVID19 data	
Manually analyzes, validates, and aggregates data in real time from a growing list of over 5,000 sources , including official Websites of Ministries of Health or other Government Institutions and Government authorities' social media accounts. Because national aggregates often lag behind regional and local health departments' data, part of the work consists in monitoring thousands of daily reports released by local authorities. The multilingual team also monitors press briefings' live streams throughout the day. Occasionally, news wires with a proven history of accuracy in communicating data reported by Governments in live press conferences before it is published on the Official Websites are used. The source of each data update is provided in the "Latest Updates" (News) section.			

Table 2. Resources that provide a corpus of COVID-19 related text

ID	Website or resource creator	Website name	Data or code and file type (available for download from the website)
	Description and sources used to create the dataset/database (as described on the website)		
1	Allen Institute for AI	COVID-19 Open Research Dataset (CORD-19)	Current and historical data releases (tar.gz) The CORD dataset is fully incorporated into Semantic Scholar search . Explore the dataset in: Quilt ; TIM Updated weekly as new research is published in scientific publications and as Semantic Scholar discovers new related research.
	A corpus of full-text and metadata dataset of COVID-19 and coronavirus-related scholarly articles optimized for machine readability for the purpose of natural language processing. Papers and preprints are collected by Semantic Scholar. Metadata are harmonized and deduplicated, and document files are processed to extract full text. The corpus is updated regularly. See Wang et al. (2020). A usage example is the White House COVID-19 Open Research Dataset Challenge (CORD-19). The dataset contains all COVID-19 and coronavirus-related research (e.g. SARS, MERS, etc.) from the following sources: PubMed's PMC open access corpus using this query (COVID-19 and coronavirus research); Additional COVID-19 research articles from a corpus maintained by the WHO; and, bioRxiv and medRxiv pre-prints using the same query as PMC (COVID-19 and coronavirus research).		

Table 3. Resources that track COVID-19 policy responses

ID	Website or resource creator	Website name	Data or code and file type (available for download from the website)
	Description and sources used to create the dataset/database (as described on the website)		
1	NYU-Abu Dhabi	CoronaNet project on government responses to COVID-19	GitHub repo Core dataset (.csv); code book (pdf) Extended dataset (.csv);
	Excellent visualization. Tracking government responses towards COVID-19. Based on a policy activity index (PAX) developed to summarize the different indicators in the data for each country for each day since Dec 31st 2019. The estimation is done with Stan and is updated daily. Working papers: CoronaNet: A Dyadic Dataset of Government Responses to the COVID-19 Pandemic. A Retrospective Bayesian Model for Measuring Covariate Effects on Observed COVID-19 Test and Case Counts Data sources: Description Sources of news articles - Directory-covid-19-world-policy-tracker UN HDX - ACAPS COVID-19: Government Measures Dataset - Factiva - Nexis Uni CoronaNet may use the following, but they caution that these sources are not quality/controlled and data must be carefully checked before using: OGP Open Government Partnership - Crowd-sourced list of COVID-19 government actions Crowd-sourced coronavirus tech handbook (Note that this site is asking for money donations: "We (@nwspl) are a tiny tech-for-good organisation in the UK that has dropped everything to work on this handbook. We will be bankrupt in 7 weeks. Please donate so we can stay focused on improving it.")		

2	UoO-University of Oxford	OxCGRT - COVID19 government response tracker	GitHub repo Code book Data (.csv)
	Working paper: Lockdown rollback checklist Index methodology		
3	UN-HDX	Global Database of COVID19 PHSM (Public Health and Social Measures)	Clean data 2020-04-29 (.csv) 4PHSM data flow and dataset descriptions (.pdf)
	A number of organizations have begun tracking implementation of PHSMs around the world, using different data collection methods, database designs and classification schemes. A unique collaboration between WHO, the London School of Hygiene and Tropical Medicine, ACAPS, University of Oxford, Global Public Health Intelligence Network, US Centers for Disease Control and Prevention and the Complexity Science Hub Vienna has brought these datasets together, using a common taxonomy and structure, into a single, open-content dataset for public use.		

Table 4. Sites that have stunning and very useful visualizations of COVID-19 data that go beyond the usual “dashboards”

ID	Website or resource creator	Website visualization name	Visualization description
	Sources used to create the dataset/database (as described on the website)		
1	McCandless, David et al.	Information is beautiful	COVID-19 visualizations data: COVID-19 data pack (Gsheet) All infectious diseases visualization data: MicrobeScope - all infectious diseases (Gsheet) Response to Swine Flu Timeline of media inflamed fears (general) All datasets
	COVID-19 data sources: COVID-19 data pack (Gsheet) All information is beautiful data sources		
2	Pro-Publica	Hospital regions would fill their beds: Nine scenarios	
	Data source: Harvard Global Health Institute .		

Table 5. Sites that maintain a database or catalogue of resources from which COVID-19 population data can be downloaded. These sites do not provide the actual data.

ID	Resource	Title	Link to listing page
Description			
1	Exaptive/Wellcome Trust	Cognitive City: COVID-19 Resources	
	A large listing of data sources.		
2	IHME	GHDX	
3	Opendatasoft	COVID-19 Data catalog	
	Free access to consult the datasets published by clients in the public space, without having to open an account.		
4	UN-United Nations	Humanitarian Data Exchange (HDX)	COVID-19 Pandemic collection Contributors are encouraged to upload data in any common data format. For tabular data, CSV is preferable, and for geographic data, zipped shapefile is preferred.
	HDX is a platform where organizations can upload their data for sharing. Registered users can “follow” the data they are interested in. Updates to datasets appear as a running list in the user dashboard. Users can contact data contributors to ask for more information about their data, and to request access to the underlying data for metadata only entries. HDX uses open-source software (CKAN) for the technical back-end. All of the code is available on GitHub .		

Table 6. COVID-19 related data analysis platforms

Resource	Title	Description
Galaxy Project	COVID-19 analysis on usegalaxy	Provides publicly accessible infrastructure and workflows for SARS-CoV-2 data analyses for three types of data: genomics, evolution, and cheminformatics. Each analysis section is continuously updated as new data becomes available
COVID-19 living NMA Initiative	Covid-19 living Data	Data are updated every week. Data were analyzed by using network meta-analysis. Primary sources: The World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP), electronic bibliographic databases (PubMed , MedRxiv , Chinaxiv , China National Knowledge Infrastructure database). Secondary sources: LitCOVID , WHO database of publications on coronavirus disease (COVID-19) , ClinicalTrials.gov , The EU Clinical Trial Register , The Cochrane COVID-19 Study register ; Other sources: EPPI-centre living map of evidence , Meta-evidence .

DISCUSSION

There is no single source of truth for COVID-19 population level data. There is dependency between sources which allows propagation of errors, and there is a lack of transparency in whether the data are updated after the initial acquisition. Although most of the resources identify the sources of their data, it is generally not possible to know the provenance at the record level. Even with the co-dependency exhibited between resources, there are still some discrepancies observed when comparing total global counts.

A Health Data Collaborative (HDC) [progress report 2016-2018](#) found that only 50% of [countries](#) reported cause-of-death data to WHO. More than 100 countries did not accurately count births and deaths. Donors requested reporting on hundreds of health indicators that fell outside of the priorities set by health ministries. Global health partners had developed at least 8 separate health facility survey tools that collected overlapping information. The HDC report concluded that fragmented health data systems hampered effective use of data during disease outbreaks.

Future work: Analysis of the dataset in R to identify and visualize codependences and discrepancies between sources. A machine-friendly data matrix and code used to analyze the dataset will be made available in GitHub.

REFERENCES

- 1Point3Acres. (2019). *COVID-19/Coronavirus Real Time Updates With Credible Sources in US and Canada* | 1Point3Acres. <https://coronavirus.1point3acres.com/en>
- AI challenge with AI2, CZI, MSR, Georgetown, NIH, & The White House. (2020). *COVID-19 Open Research Dataset Challenge (CORD-19)*. <https://kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>
- Brickley, D., et al., & et al. (2020). *Organization of Schemas*. <https://schema.org/docs/schemas.html>
- CDC. (2020a). *Cases of COVID19 in the U.S.* [Dataset]. Centers for Disease Control, USA. <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html>
- CDC. (2020b). *Weekly provisional death counts by select demographic and geographic characteristics* [Dataset]. Center for Disease Control, USA. https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm
- CDC. (2020c). *Weekly provisional death counts from death certificate data: COVID19, pneumonia, flu* [Dataset]. Center for Disease Control, USA. <https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm>
- COVID19India. (2020). *Coronavirus in India: Latest Map and Case Count*. <https://www.covid19india.org>
- ECDC. (2020). *COVID-19 situation update worldwide, as of 9 June 2020*. European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>
- Exaptive. (2020). *Cognitive City: COVID-19 Resources*. Exaptive, and the Bill and Melinda Gates Foundation. <https://covid-19.cognitive.city/cognitive/community/resource-gallery>
- GESIS Panel Team. (2020). *GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany* GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany (Version 1.1.0) [Data set]. GESIS Data Archive. <https://doi.org/10.4232/1.13520>
- Healy, K. (2020). *Rpackage (covdata)—COVID19 Case and Mortality Time Series*. <https://kihealy.github.io/covdata>
- IHME. (2020). *Global Health Data Exchange | GHDx*. Institute for Health Metrics and Evaluation (IHME), University of Washington. <http://ghdx.healthdata.org/>
- International Severe Acute Respiratory and Emerging Infection Consortium, & World Health Organization. (2020a, March 24). *ISARIC_COVID-19_RAPID_CRF_24MAR20_EN.pdf*. https://media.tghn.org/medialibrary/2020/04/ISARIC_COVID-19_RAPID_CRF_24MAR20_EN.pdf

- International Severe Acute Respiratory and Emerging Infection Consortium, & World Health Organization. (2020b, April 23). *ISARIC_WHO_nCoV_CORE_CRF_23APR20.pdf*.
https://media.tghn.org/medialibrary/2020/05/ISARIC_WHO_nCoV_CORE_CRF_23APR20.pdf
- NSW Health. (2020, March 23). *Novel-coronavirus-case-questionnaire.pdf*.
<https://www.health.nsw.gov.au/Infectious/Forms/novel-coronavirus-case-questionnaire.pdf>
- NYT. (2020, April 21). *Coronavirus (Covid-19) Data in the United States*. New York Times.
<https://github.com/nytimes/covid-19-data> (Original work published 2020)
- OpenAIRE. (2020). *COVID-19 Open Research Gateway*. OpenAIRE - Connect. <https://beta.covid-19.openaire.eu/content>
- Oxford University. (2020). *COVID19 government response tracker* [Dataset]. University of Oxford.
<https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker>
- Panchadsaram, R., Davis, N., Wikelius, K., Shahpar, C., Cobb, L., Wosińska, M., Anderson, D., Brown, L. M., Hunt, D., & Goligorsky, D. (2020). *COVID exit strategy: How we reopen safely*. <https://www.covidexitstrategy.org/>
- Park, H. W., Park, S., & Chong, M. (2020). Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea. *JOURNAL OF MEDICAL INTERNET RESEARCH*, 22(5).
<https://doi.org/10.2196/18897>
- Team, G. P. (2020). *GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany*.
<https://doi.org/10.4232/1.13520>
- The Atlantic. (2020). *COVID Tracking Project* [Datasets]. <https://covidtracking.com/>
- The European Health Data & Evidence Network. (2020, April 15). *COVID-19-Data-Partner-Call-Description-v1.4.pdf*.
<https://www.ehden.eu/wp-content/uploads/2020/04/COVID-19-Data-Partner-Call-Description-v1.4.pdf>
- The National Institute of Environmental Health Science. (2020, May 14). *COVID-19_BSSR_Research_Tools.pdf*.
https://www.nlm.nih.gov/dr2/COVID-19_BSSR_Research_Tools.pdf
- UN. (2020). *The Humanitarian Data Exchange (HDX)*. United Nations, Office for the Coordination of Humanitarian Affairs (OCHA), Centre for Humanitarian Data. <https://data.humdata.org/>
- University of Maryland. (2020, April 24). *COVID-19 Impact Analysis Platform*. COVID-19 Impact Analysis Platform.
<https://data.covid.umd.edu/>
- University of Washington - HGIS Lab. (2020). *Novel Coronavirus Infection Map*. University of Washington - Humanistic GIS Laboratory. <https://github.com/jakobzhao/virus>
- U.S. Department of Health and Human Services. (2017, December). *Pan-flu-report-2017v2.pdf*.
<https://www.cdc.gov/flu/pandemic-resources/pdf/pan-flu-report-2017v2.pdf>
- Wang, L. L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., Funk, K., Kinney, R., Liu, Z., Merrill, W., Mooney, P., Murdick, D., Rishi, D., Sheehan, J., Shen, Z., Stilson, B., Wade, A. D., Wang, K., Wilhelm, C., ... Kohlmeier, S. (2020). *CORD-19: The Covid-19 Open Research Dataset*. *ArXiv:2004.10706 [Cs]*.
<http://arxiv.org/abs/2004.10706>
- WHO. (2020a). *COVID-19 situation reports*. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
- WHO. (2020b, May 29). *C-TAP: COVID-19 technology access pool*.
<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/covid-19-technology-access-pool>
- Worldometer. (2020). *COVID19 data* [Dataset]. <https://www.worldometers.info/coronavirus/>
- Xu, B., Gutierrez, B., Mekaru, S., Sewalk, K., Goodwin, L., Loskill, A., Cohn, E. L., Hswen, Y., Hill, S. C., Cobo, M. M., Zarebski, A. E., Li, S., Wu, C.-H., Hulland, E., Morgan, J. D., Wang, L., O'Brien, K., Scarpino, S. V., Brownstein, J. S., ... Kraemer, M. U. G. (2020). Epidemiological data from the COVID-19 outbreak, real-time case information. *Scientific Data*, 7(1), 106. <https://doi.org/10.1038/s41597-020-0448-0>
- Zhang, L., Ghader, S., Pack, M. L., Xiong, C., Darzi, A., Yang, M., Sun, Q., Kabiri, A., & Hu, S. (2020). An interactive COVID-19 mobility impact and social distancing analysis platform. *MedRxiv*, 2020.04.29.20085472.
<https://doi.org/10.1101/2020.04.29.20085472>

Working paper

APPENDIX 2 – Interoperable COVID-19 epidemiological surveillance: Clinical and Population-based instruments

Carsten O. Schmidt^{1,*}, Rajini Nagrani², Matthias Löbe³, Atinkut Zeleke⁴, Guillaume Fabre⁵, Stefan Sauermann⁶, Anna Widyastuti⁷, Chifundo Kanjala⁸, Jay Greenfield⁹, Claire C Austin¹⁰; and the RDA-COVID19-WG¹¹

¹University Medicine Greifswald, ²University of Leipzig, ³University of Greifswald - Medicine, ⁴McGill University - Maelstroem group, ⁵UAS Technikum Wien, ⁶Leibniz Institute for Prevention Research and Epidemiology-BIPS, ⁷Hasselt University, Belgium, ⁸London School of Hygiene and Tropical Medicine, ⁹Data Documentation Initiative, ¹⁰Environment and Climate Change Canada; ¹¹This work is linked to the Research Data Alliance [RDA-COVID19-WG, Epidemiology subWG recommendations and guidelines on data sharing](#).

BACKGROUND

The COVID-19 pandemic has raised a huge demand for patient and population-based data which cover not only the onset and course of the disease, and related treatments, but also side effects due to pandemic mitigation measures. For this purpose, all around the world, new instruments are being developed in a very short time. Heterogeneity between these instruments may limit comparability of results across studies and countries. Unawareness of already existing options for surveying COVID-19 related factors contributes to this heterogeneity. The present study provides an overview of some preselected instruments from different countries to guide the creation of novel tools by reusing existing instruments and items in relation to COVID-19.

METHODS

We searched for publicly available online resources from major organizations in different countries to identify instruments or item banks of potential interest. Inclusion criteria for an instrument or item bank were: (1) directly COVID-19 related by targeting health-related behaviours, symptoms, treatments, clinical outcomes, pandemic mitigation measures, or COVID-19 related attitudes; (2) issued by an initiative or authority of enough weight to impact COVID-19 research, at least at the national level; (3) the instruments or item appears to be methodologically sound; (4) major parts of the instrument are of sufficient general applicability to be used beyond a local context; (5) the instrument is applicable for patient, or for general population surveys; and, (6) the instrument may be used for future research without any necessity for negotiating a license agreement. The focus was on instruments issued in English. However, other major languages were also considered.

We described the selected instruments and compared them at the module-level based on the targeted topic. We also referenced resources that collect instruments.

RESULTS

Table 1a provides an overview of the instruments selected for review. Two broad categories were distinguished: (1) symptom and treatment oriented instruments for patient populations in clinic/hospital surveillance; and, (2) instruments addressing the psychosocial impact of COVID-19 and pandemic mitigation measures, and health related behaviour and attitudes in population-based surveys. Many of the latter instruments also make brief reference to symptoms or treatments. Most instruments were available only as a pdf file, while only one resource conducted an in-depth semantic annotation of all items (1). There was a lack of machine-readable formats to facilitate the incorporation of items and instruments into new data dictionaries. Three instrument resource pages were identified that provide access to a wider range of instruments (Table 1b).

Table 1. Questionnaire instruments: Reference studies

Country	Initiative	Target population	Development stage	Language	Provenance (Influenced by...)	Comments
1 Australia	NSW Case questionnaire	Patients		English		
2 Brazil	Brazil Prevalence of Infection Survey	Rapid tested, Tested positive	In development	English		
3 Europe	Questionnaire by WHO Europe	General population		German, Russian		Single Instrument
France	Barometer Covid19					The "barometer Covid19" is a citizen science initiative driven by the DataCovid association. It aims to provide open-access data from a weekly survey to illuminate the struggle against the epidemic Covid19 from observations on its dynamics, its

Country	Initiative	Target population	Development stage	Language	Provenance (Influenced by...)	Comments
						determinants and its impacts.
France	French COVID-19	Patients	In use	English		Lead by REACTing consortium in collaboration with ISARIC consortium (International Severe Acute Respiratory and emerging Infection Consortium)
4 Germany	Covid-19 research dataset	Patients	In development	German		National Network of German University Clinics to study COVID19.
6 Germany	GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany	General population	In use	German		As the largest European infrastructure institute for the social sciences GESIS provides essential and internationally relevant research-based services
Germany	NAKO COVID-19 Survey tool	General population	In use	German		The German National Cohort (GNC) has been inviting adults aged between 20 and 69 to 18 study centers throughout Germany since 2014 with more than 200.000 participants. The COVID-19 questionnaire may be requested at the NAKO study office (Bohn /Panreck).

Country	Initiative	Target population	Development stage	Language	Provenance (Influenced by...)	Comments
7 Israel	One-minute population wide survey	Israeli population	In use	Hebrew, Arabic, Russian, Spanish, French, English		Participants asked to fill it out on a daily basis and separately for each family member, including members who are unable to fill it out independently (e.g., children and older people).
8 Low and Middle Income Countries	LMIC Covid Questionnaire		In development		UK COVID-19 questionnaire, SAPRIN COVID-19 screening form	
9 South Africa	South African Population Research Infrastructure (SAPRIN) COVID-19 Screening Form		In development	English, Afrikaans		
10 Uganda	Perinatal COVID-19 Uganda	Women pre-/perinatal		English		

Country	Initiative	Target population	Development stage	Language	Provenance (Influenced by...)	Comments
11 UK	UK COVID-19 Questionnaire	Adult Respondent, Children, Key worker, Partner.	In development	English	NIHR Global Health Research Unit	- World Bank code book for metadata. - Becoming a model for some African countries
12 UK	National Institute for Health Research (NIHR) Global Health Research Unit	Telephone sample		English		
13 US	Human Infection with 2019 Novel Coronavirus Person Under Investigation (PUI) and Case Report Form	Patients	In use	English	CDC	
14 US/WHO	Population-based age-stratified seroepidemiological investigation protocol for COVID-19	Patients	In use	English	WHO	This protocol has been designed to investigate the extent of infection, as determined by seropositivity in the general population, in any country in which COVID-

Country	Initiative	Target population	Development stage	Language	Provenance (Influenced by...)	Comments
	virus infection					19 virus infection has been reported.

Footnotes (Domains)

1. Clinical symptoms, Disease outcome, Exposure sites, Pre-existing conditions, Risk history, Sociodemographics.
2. Demographics, Home life, Test results, Transportation
3. Affect, Behaviour, Conspiracies (perceptions), COVID-19 risk perception: probability and severity, Fairness (perceptions), Frequency of Information, Influenza risk perception: probability and severity,, interventions (perceptions), Knowledge and self-assessed adherence to prevention measures, Knowledge incubation, Knowledge symptoms/treatment, Lifting restrictions (pandemic transition phase), Policies, Preparedness and perceived self-efficacy, Prevention – own behaviours, Resilience (perceptions), Risk group, Rumors (open-ended), Self-assessed knowledge, Socio-demography, Trust in institutions (perceptions), Trust in sources of information, Use of sources of information, Worry.
4. Clinical symptoms, Complications, Imaging, Laboratory markers, Medical treatments, Medication, Sociodemographics.
5. Adherence to risk minimization measures, Changes in lifestyle factors, COVID infections and testing, COVID tracing, General health status, Physical symptoms, Respiratory infections, Workplace/changed employment situation.
6. Changed employment situation, Childcare obligations, Risk perception, Evaluation of political measures & their compliance, Media consumption, Risk minimization measures, Trust in politics and institutions.
- 6b. NAKO: General health status, respiratory diseases, testing for COVID, social contacts, adherence to pandemic mitigation measures, work place, economic and social impact, depression, psychosocial complaints, physical activity, ,alcohol drinking, smokin
7. Age, Geographic location (city and street), Isolation status, Sex, Smoking habits
9. Actions in response to COVID-19, Bounded structure, Eligibility for testing, Epidemiological risk, Household enumeration, Household impact, Quarantine and hygiene directions,, Symptom screen, Travel and movement (mobility), Travel history, Visit attempts.
10. Disease outcome, Sociodemographics, Symptoms

11. Accommodation type, Away from home environment, Behavior changes, Change in benefits, Digital access, Economic activity before and after lockdown, Environmental attitudes, Environmental impact, Family relations, Financial impact, Food security, Impact on employment, Knowledge, Medication, Mental health, Mental health, Physical health, Pre-existing conditions, Social impact, Symptoms, Volunteering.
12. COVID-19 Interventions, Current living conditions, Displacement and mobility, Economic impacts, Impact of COVID-19 on health-related behaviors, Mental health, Precautions, Pre-existing conditions, Social aid, Social impact, Symptoms, Treatments for pre-existing conditions.
13. Diagnostic testing procedures, Clinical course, Medical history, Pre-existing conditions, Risk exposure, Sociodemographics, Symptoms, Treatments.
14. Laboratory results, Sociodemographics, Symptoms.

Table 2. Questionnaire instruments - Resources

Provider	Initiative	Language
US	NIH Public Health Emergency and Disaster Research Response (DR2)	Diverse
NIH	COVID-19OBSSR Research Tools	Diverse
PhenX is funded by the National Institutes of Health (NIH) Genomic Resource Grant	PhenX COVID-19 Toolkit	Diverse

Instruments used in a clinical context

Table 2 Covid19 CRFs Content Overview ([working draft Gsheet](#))

DISCUSSION

A large number of COVID-19 related instruments have been developed in very a short period of short time. We did not attempt to conduct a complete survey. Rather, selected instruments, item banks and resources of estimated high value in the field were identified and compared. This comparison revealed a very large scope of areas targeted by these instruments in relation to COVID-19, ranging from proximal measures such as symptoms of the disease to distal measures such as political attitudes. Amongst population-based instruments, some questionnaires were an add-on module to already existing cohort studies, while others were standalone surveillance questionnaires. Researchers must consider the scope and feasibility of their surveys before implementing the questionnaires in specific situations/communities.

Although there have been some efforts taken to make clinical data interoperable, population-based data still require considerable attention. Finally, if applicable in a specific research context, our recommendation is to adhere as much as possible to existing items and instruments to facilitate the later comparison of results. New instruments should be made publicly available and published in a findable, reusable, and machine-readable format with appropriate semantic annotation.

Working paper

APPENDIX 3 – Preservation of individuals' privacy in shared COVID-19 related data

Stefan Sauermann¹, Chifundo Kanjala², Matthias Templ^{3,*}; and the RDA-COVID19-WG⁴

¹UAS Technikum Wien, ²London School of Hygiene and Tropical Medicine, ³Zurich University of Applied Sciences, ⁴This work is linked to the Research Data Alliance [RDA-COVID19-WG Recommendations and guidelines on data sharing](#)

*Corresponding author: matthias.templ@gmail.com

All views and opinions expressed are those of the co-authors, and do not necessarily reflect the official policy or position of their respective employers, or of any government, agency or organization.

Abstract

This paper gives insight into the pseudo-anonymization and anonymization of COVID-19 data sets. First, methods for the pseudo-anonymization of direct identification variables are discussed. We also discuss different pseudo-IDs of the same person for multi-domain and multi-organization. Essentially, pseudo-anonymization and its encrypted IDs are used to successfully match data later if required and permitted, as well as to restore the true ID (and authenticity) in individual cases of a patient's clarification.

To make the re-identification of individual persons of COVID-19 (that are often enriched with other covariates like age, gender, nationality, etc.) impossible, the successful re-identification by a combination of attribute values must be prevented. This is done with methods of statistical disclosure control for anonymization of data.

BACKGROUND

It is imperative that the response efforts to COVID-19 be coordinated at a global scale to ensure that no part of the world is left behind which would serve as reservoirs of the virus when other regions have put in place control measures. Critical to this globally coordinated approach are the requirements that the data relating to the pandemic are rapidly collected, processed and shared. Data sharing reduces duplication of efforts, facilitates transparency, efficiency and thus the maximization of benefit from the funding and efforts applied in responding to COVID-19.

However, the rapid collection, processing, analysis and sharing of the data should not come at the expense of ethical practices. The privacy and confidentiality of the individuals supplying these much-needed data whether from a healthcare facility, a community-based study or some other surveillance mechanism in place need to be preserved throughout the entire process.

In the case of a pandemic such as COVID-19, the task of preserving data privacy is complicated by the need to link datasets on the same population from disparate sources. The RDA recommendation includes processes that involve anonymization and pseudonymization of data. Technology is in place to address these issues. This paper will briefly summarise these use cases of the RDA COVID-19 recommendation, and illustrate how existing methods and technology can be harnessed and

implemented within the context of the other RDA recommendations. In particular, we consider how the privacy preservation methods can work within the context of proposed Common Data Model in paper #7. Paper # 7 provides guidelines on a model for integrating silos of COVID-19 related data from hospital/clinic surveillance (Electronic Health Records [EHR]), field-based longitudinal demographic and epidemiological surveillance, population level indicators, and monitoring programs (e.g., air quality monitoring). In this paper we describe how the existing privacy preservation technologies and methods can be customised for implementation in the context of the model proposed in paper 7 to ensure pseudonymization and anonymization as needed.

METHODS

To preserve privacy of the individuals there is a need to iteratively assess and control disclosure at each stage of the data integration. First, when the data are harvested from their sources and then second when they are integrated into harmonised databases. The key is to control for disclosure while minimising data loss.

RESULTS

Pseudonymization and anonymization

To ensure privacy, both pseudonymization of direct identifiers (e.g. patient specific ID's) and anonymization of indirect identifiers (e.g. socio-demographic information on individuals) must be applied. Pseudonymization can be done e.g. by first salting the ID's using e.g. the SHA-3 hash function [1, p. 3] and 256 bit salts) and then applying a hash function to the salted ID's for encryption using salted and hashed values (using e.g. the SHA-3 hash function [1, p. 3] and 256 bit salts) and different domains-specific salts for the same patient and a central agency that stores the salts and is able to link different data sets from the same patients. An appropriate IT-infrastructure is relevant here for ID management and linking data. This article introduces basic concepts that can be used to sustain privacy of patient related data in epidemiologic, research and administration contexts.

Pseudo-anonymization

Patient related data are recorded and used in different domains, for example medicine, communities, research, administration, and statistics. Patient specific electronic identifiers (IDs) link specific information to the individual patient records in each domain. Each domain assigns a domain-specific ID to each individual patient. In order to satisfy privacy requirements on direct identifiers, data that carry a domain ID remain within the data holder's secure server, and are not shared.

If data are re-used, for example in research or public health purposes, the domain-specific ID is removed (anonymization) or replaced by a different ID (pseudonym). Pseudonyms can later be traced back by an authorized person to the original domain ID. This must only occur under well-defined conditions, e.g. if a researcher needs to clear ambiguities in incoming information together with the organisation that generated the data.

In multi-domain and multi-organisation scenarios, consistent management of IDs and pseudonyms on a secure authorized server is needed to enable cross-linking of data from different sources using the ID's.

The following requirements apply:

- Patient IDs exist for each domain;
- The local domain ID must not leave the local domain in clear text, to prevent unintended record linkage between domains;
- When providing data from a source domain to a target domain, the target domain patient ID (pseudonym) must become available to users in the target domain; and,
- Domain IDs must enable domains to cooperate e.g. for clearing ambiguities, while preserving privacy.

Example of a multi-domain and multi-organization pseudo-anonymization strategy

In Austria, for example, eGovernment legislation and IT infrastructures are in place to handle domain specific identifiers e.g. for health care, traffic, taxes, and statistics [2], [3]. This is implemented and in operation for example in the Austrian electronic health care record ELGA [4]. Figure 1A describes how data from a health domain can be linked to records in a research domain in this way. Figure 1B introduces the needed IT infrastructure. Figure 1C shows how IDs are mapped between domains while preserving pseudonymization.

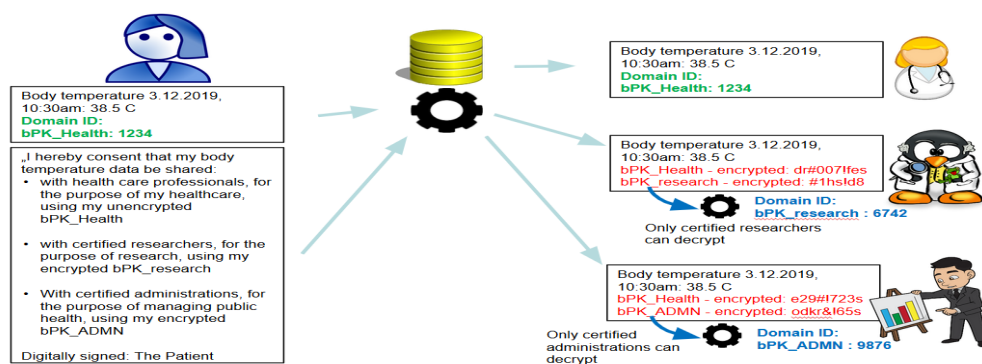


Figure 1A. Sharing or linking a body temperature observation from the healthcare domain with a research and administration domain. In the healthcare domain ID (bPK_Health, green), a patient is identified with a specific ID, (1234, colour green, denoting that it is unencrypted). A doctor will receive this data together with the original ID, as the law allows doctors to share IDs unencrypted. The doctor can attach an encrypted ID (bPK_research - #1hsld8, red, denoting it is encrypted) to the data. The doctor never has access to the unencrypted bPK research ID. A researcher who receives the data, decrypts the encrypted ID. This decrypted ID (bPK_research - 6742, blue, denoting that it is a pseudonym) is specific to the research domain. The same method applies as data are provided to administrations, e.g. for public health purposes. *Contributed by Dr. Carsten Schmidt. LICENCE: CC BY-NC-SA 3.0*

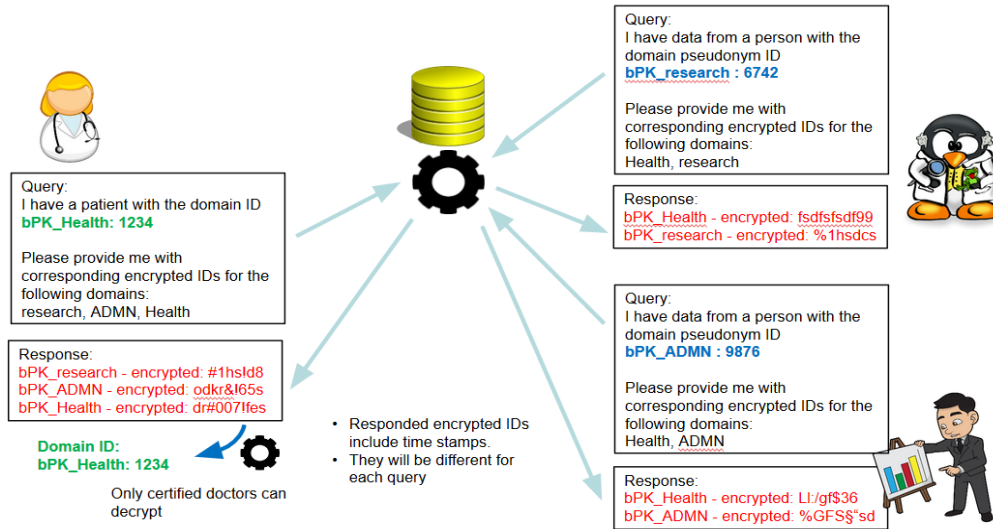


Figure 1B. IT infrastructure for cross-domain IDs management. A user in the medical domain queries the IT service using the ID of the health domain, asking for encrypted IDs of other domains. The service responds with the encrypted IDs. Users in other domains can use the same mechanism. This enables users in different domains to co-operate: For example the researcher can attach the encrypted bPK_Health ID to a message to the doctor, asking for details to support clearing ambiguities in the data the doctor provided earlier. The doctor can then decrypt the patient ID, access the patient related information in the health IT system, and finalise the clearing with the researcher. Contributed by Dr. Carsten Schmidt. [LICENCE: CC BY-NC-SA 3.0](#)

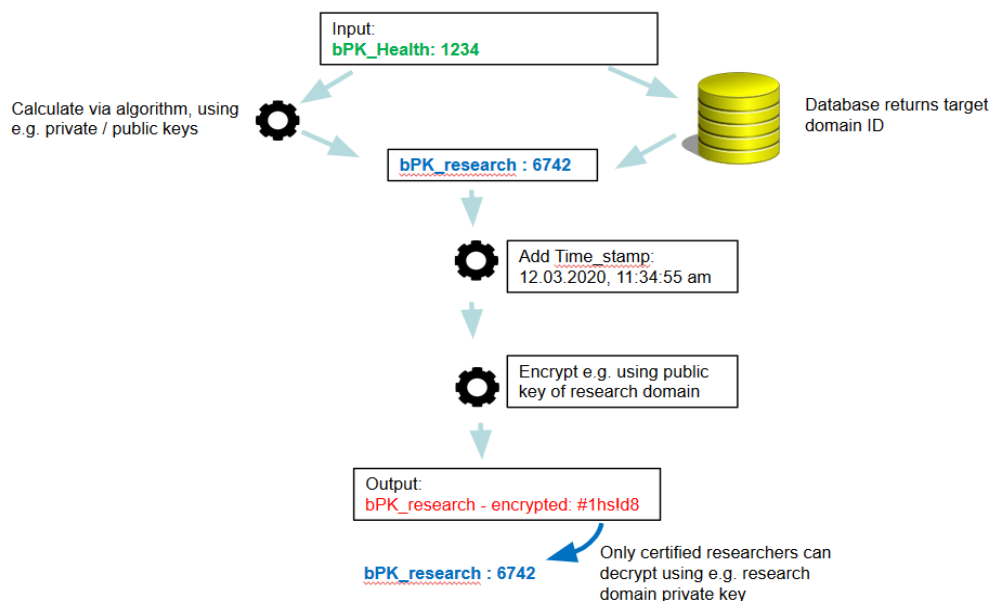


Figure 1C. Data flow for deriving an encrypted ID (#1hsld8) for a target domain (research). The source ID (1234) can be mapped to the target identifier for example in two ways. A mathematical algorithm is used (left branch) to calculate the target domain ID (6742), or a database query returns the ID. A time

stamp (as kind of salting) is then attached to this ID. ID and timestamp together are then encrypted, e.g. using the public key of the target domain. This assures that no two encrypted IDs for the same patient and the same domain are identical, in this way preventing the unintended linking of records.

Contributed by Dr. Carsten Schmidt. [LICENCE: CC BY-NC-SA 3.0](#)

Anonymization: Controlling statistical disclosure risk by anonymization

A pseudonymized data set does not imply that individuals cannot be re-identified and, additionally, the data set still needs to be anonymized using different methods than those regarding pseudonymization. Intruders can reveal patients' identities using a combination of indirect identifiers such as gender, age, education, location, race/ethnicity, etc.

Since the first data protection scandals in the 1990s [5], we have learned that removing or pseudonymizing directly identifying attributes (ID's) such as names, addresses and social insurance numbers is generally not sufficient to prevent data protection violations. Beyond masking these direct identifiers, we need to control statistical disclosure risk. This is the risk of intruders using a combination of indirect identifiers such as race/ethnicity, education level, sex, age, among others, to identify individuals and their health status. The levels of such risks need to be quantified, assessed and controlled in the shared datasets to ensure privacy and trust of the individuals contributing their data and to meet scientific integrity requirements. In order to measure the risk, frequencies of persons with certain attributes are generally counted and evaluated. The lower the frequency, the higher the risk. For example, an 83 year old man with COVID-19 in a community. If he is the 83-year-old and the only 83 year old guy in the community, then we know that he has had COVID-19, as well as the values of all other attributes given for that man. To make this person anonymous, various methods can be used. For example, post randomization of the place of residence, or adding a noise to the actual age, or regrouping age in age groups. After the anonymization the data utility is checked. While the risk for re-identification should be very low, the data utility should be as high as possible, i.e. the data quality of the anonymized data should be very good. If queries are published that results in reporting aggregated information in form of tabular data (aggregates) instead of microdata (each row of the data set corresponds to an individual), other methods of tabular data protection are used. Again, the frequency of cells in a table can be checked (how many persons contribute to the cell value) and if the number of cases is too small, blocking can be applied (primary suppression). These primary locked cells are protected by a secondary suppression to such an extent that the primary suppressed information can no longer be estimated accurately enough. See an introduction to statistical disclosure control [6].

In the LMIC, the World Bank Group and the International Household Survey Network supported the development of the statistical disclosure control software *sdcmicro* [7] and they recommend it (<http://surveys.worldbank.org/sdcmicro>). There is literature and recommended best practices in this regard [8], [9].

DISCUSSION

For COVID-19 data, it must be possible for authorized persons to know the true ID of a patient, i.e. to reverse the encryption of an ID, either in the case that several data sets are merged with this ID or because clarifications are necessary for the patient. To prevent the merging of patient data, the true ID's are salted and encrypted with a hash function. In a multi-domain and multi-organization framework, a different salt for the same person is used for each domain and organization. The management of these

salts is an IT-technical challenge, but it is well established in practice. As an example, a complex pseudo anonymization based on the Austrian health data was shown.

However, pseudo anonymization only prevents the merging of persons across different data sets. However, it will still be possible to clearly identify persons by a combination of attribute values, and further anonymization is necessary, especially if the data is made available to an extended circle of users.

The choice of indirect identifiers is very important. This means asking "what is the knowledge of an attacker"? In other words, what possible databases with intersectional populations may be available to someone who receives the data?

For indirect identifying variables, prior of using SDC techniques, the re-identification risks of the data should be modeled and estimated. This allows the identification of records with high re-identification risks, e.g. those having too less frequency and/or assess high-risk observations.

For categorical key variables apply global recoding and local suppression to establish k anonymity - that at least k observations/persons in the data set shares the same values for the indirect identifiers (e.g. age 65, female, lives in region XY, ...), or alternatively, swapping techniques could be used to change, for example, the place of residence with another person.

Each time an SDC technique is used the risk of re-identification and the increase of information loss should be measured. The anonymization process is typically iterative. Different anonymization methods are tried until the risk of re-identification has been decisively reduced and at the same time the loss of information is considerably low.

The acceptable level of risk depends on many factors. If COVID-19 data are shared as public-use files, they should have much less disclosure risk than scientific-use files for selected researchers or research institutions. The latter are restricted to certain users under certain conditions.

COVID-19 data enriched with information on patients' other comorbidities containing sensitive information such as HIV / AIDS information may require greater intervention for anonymization, compared to data with less sensitive information.

Note that the risk is never 0, but this is not required. De-facto anonymity is reached whenever the burden and efforts of re-identification is higher than the value for the data intruder.

Data protection must also be guaranteed when delivering data internally in the data holders' organization. But it is clear that, in general, the risk for internal data delivery is smaller than externally, since (and only if) the planned processing tasks are described internally, the purposes of the processing may be better known, the risks to data storage can be minimized (the data may be less likely) can be stolen), and it can also be logged who works when with the data, i.e. to be able to control this better than with external data delivery.

Author roles

All authors accept responsibility for the content of the article.

The same author may have multiple roles, and multiple authors may share a single role.

Conceptualization: SS, CK, MT. Methodology: SS, CK, MT. Investigation: SS, CK, MT. Formal analysis: SS, CK, MT. Validation: SS, CK, MT, RDA-COVID19 WG. Writing: SS, CK, MT. Bibliographic review and analysis: SS, CK, MT. Visualization: SS. Supervision and project administration: Claire Austin.

Acknowledgments

The authors would like to thank the following people for their support and contributions to this work: We like to thank Claire Austin for useful comments and support.

REFERENCES

- [1] I. T. L. Computer Security Division, “SHA-3 Project - Hash Functions | CSRC,” *CSRC / NIST*, Jan. 04, 2017. <https://csrc.nist.gov/projects/hash-functions/sha-3-project> (accessed May 30, 2020).
- [2] Austria, “Austrian RIS - E-Government-Gesetz - Bundesrecht konsolidiert, Fassung vom 22.04.2020,” 2018. <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20003230> (accessed Apr. 22, 2020).
- [3] Austria, “Austrian RIS - E-Government-Bereichsabgrenzungsverordnung - Bundesrecht konsolidiert, Fassung vom 22.04.2020,” *Austrian Parliament*, 2013. <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20003476> (accessed Apr. 22, 2020).
- [4] ELGA GmbH, Vienna, Austria, “ELGA: Technischer Aufbau im Überblick,” Apr. 22, 2020. <https://www.elga.gv.at/technischer-hintergrund/technischer-aufbau-im-ueberblick/> (accessed Apr. 22, 2020).
- [5] D. Barth-Jones, “The’re-identification’of Governor William Weld’s medical information: a critical re-examination of health data identification risks and privacy protections, then and now,” *Then July 2012*, 2012.
- [6] M. Templ, B. Meindl, A. Kowarik, and S. Chen, “Introduction to Statistical Disclosure Control (SDC),” *IHSN Working Paper No 007*. p. 25, 2014.
- [7] M. Templ, B. Meindl, and A. Kowarik, *sdcmicro: Statistical Disclosure Control Methods for Anonymization of Data and Risk Estimation*. R package manual, version 5.5.1, 2020.
- [8] G. T. Duncan, M. Elliot, and G. J. J. Salazar, *Statistical Confidentiality: Principles and Practice*. New York: Springer-Verlag, 2011.
- [9] M. Templ, *Statistical disclosure control for microdata: Methods and applications in R*. Springer International Publishing, 2017, p. 287.

Working paper

APPENDIX 4 – A Full Spectrum View of the COVID-19 data domain: An Epidemiological Data Model

Jay Greenfield^{1,*}, Rajini Nagrani², Meg Sears³, Claire C. Austin⁴; and the RDA-COVID19-WG⁵

¹Data Documentation Initiative, ²Leibniz Institute for Prevention Research and Epidemiology-BIPS, ³Ottawa Hospital Research Institute, ⁴Environment and Climate Change Canada, ⁵This work is linked to the Research Data Alliance [RDA-COVID19-WG, Epidemiology subWG recommendations and guidelines on data sharing](#).

*Corresponding author: nightcleaner@gmail.com

All views and opinions expressed are those of the co-authors, and do not necessarily reflect the official policy or position of their respective employers, or of any government, agency or organization.

ABSTRACT

An epidemiological surveillance data model is proposed that is able to guide the construction of a common data model and other database schemas used in COVID-19 epidemiological research. The data model encompasses both hospital specific surveillance in line with the WHO COVID-19 Core and Rapid CRFs and electronic health records together with field-based demographic and epidemiological surveillance now in progress for COVID-19.

BACKGROUND

In a full spectrum COVID-19 domain model¹ it is necessary to account for the entire individual experience from susceptibility to exposure, to infection, and through treatment, to death or recovery with and without sequela [Pesquita C, Ferreira JD, Couto FM, Silva MJ, 2014] [World Health Organization, 2020 Apr 23] [World Health Organization, 2020 Mar 24]. Susceptibility includes person-level risk factors, the toolbox of public health countermeasures in play and the use a person is able to make of these countermeasures. In the full spectrum model susceptibility interfaces with exposure through contacts, and this information may be captured prospectively and/or retrospectively. Diagnosis may follow, or not, and can take many forms. Treatment may or may not follow diagnosis. Treatment may or may not include hospitalization. Death may occur without hospitalization. Finally, recovery may

¹ In ontology engineering, a domain model is a formal representation of a knowledge domain with concepts, roles, datatypes, individuals, and rules, typically grounded in a description logic.

be fraught with sequela. The full spectrum domain model presented here accounts for all of these pathways.

METHODS

The full spectrum model is based on the SEIR (Susceptibility, Exposure, Infection, Recovery) domain model [Sun, P and Kang L, 2020]. We augmented the SEIR model to take into account public health countermeasures [Giordano G, Blanchini F, Bruno R *et al*, 2020] and differences in the availability of resources between High Income Countries (HIC) and Low and Middle Income Countries (LMIC). The augmented SEIR domain model might be more informed by the hospital experience in HICs, and by community-based demographic and epidemiological surveillance in LMICs [Wang et al., 2020]. Some LMICs also benefit more directly from lessons learned from experiences with Ebola and HIV AIDS. The augmented SEIR model gives equal weight to both the hospital experience and to the community-based demographic and epidemiological surveillance.

RESULTS

The COVID-19 Epidemiological Surveillance Data Model came about as a result of these augmentations. Figure 1 breaks out the model into three components -- disease milestones, contacts and person risk factors. In Figure 1 component details are hidden; however, the source material that influenced the composition of each component is indicated. Component details are outlined in Figure 2 (disease milestones), Figure 3 (contacts) and Figure 4 (personal risk factors).

Note that the COVID-19 Epidemiological Surveillance Data Model depicted here is an “interim” model; it is very much a work in progress. That is because the COVID-19 pathogen is a novel coronavirus, so we are still learning about COVID-19 disease management, person risk factors and best practices for early case recognition and contact tracing. Furthermore, when it comes to questionnaires for community surveillance, there is a plethora of ongoing overlapping initiatives. A separate effort has been undertaken to contrast and compare some of these initiatives [Schmidt CO, Nagrani R, Löbe M, et al. 2020]

In Figure 1 we trace the provenance of the several components that compose the COVID-19 Epidemiological Data Model. Many standards groups, governmental agencies and non-governmental organizations have developed guidance, financially support initiatives and instruments that have informed and will continue to inform this COVID-19 Epidemiological Surveillance Data Model. Two in particular are outstanding. They pre-date COVID-19 and are part of a longstanding commitment to understand and assess emerging pathogens. They are the WHO Global Influenza Surveillance and Response System (GISRS) which now includes COVID-19 sentinel surveillance [WHO. 2020a] and the Wellcome Trust Longitudinal Populations Strategy which was launched in 2017 [Wellcome Trust. 2017]. The former has spawned the WHO COVID-19 Core Version and Rapid Version CRFs, while the latter has produced the Wellcome Trust COVID-19 LPS HIC and LMIC Questionnaires. Together they form the provenance for both the COVID-19 Epidemiological Surveillance Model disease milestones component and the person risk factors component.

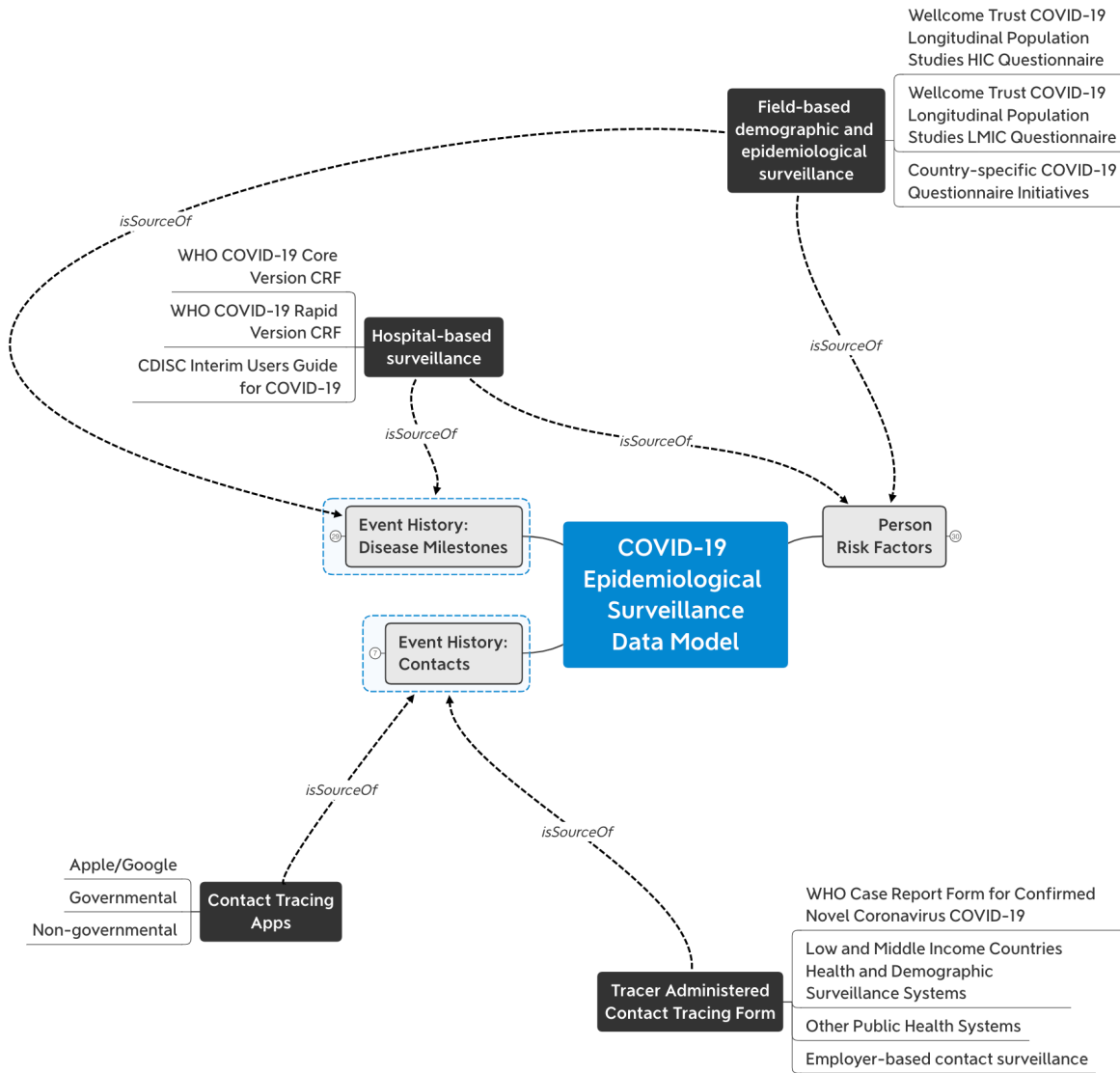


Figure 1: Overview of the COVID-19 Epidemiological Surveillance Data Model Components and their Provenance

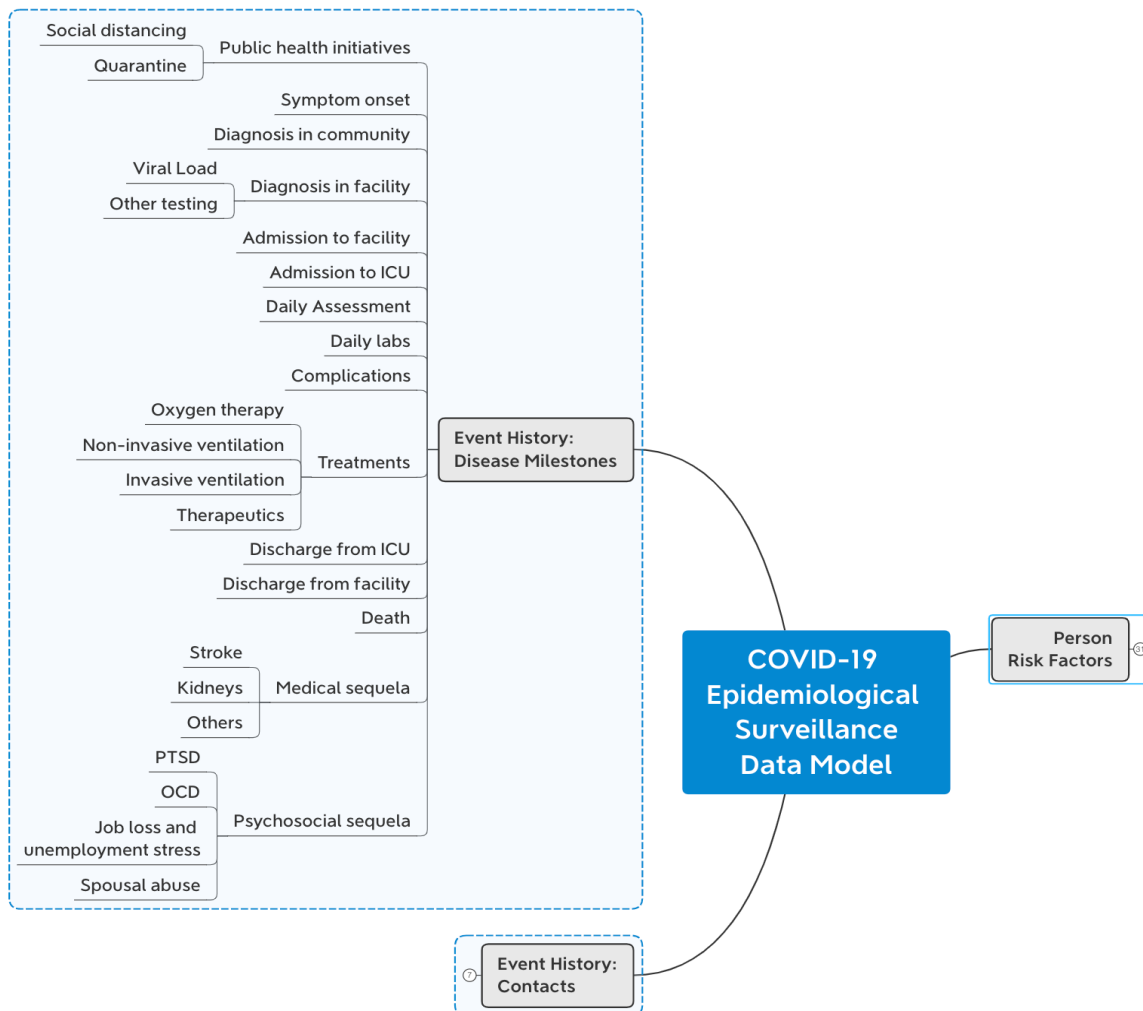


Figure 2: The Disease Milestones Component

Note that the Disease Milestone Component is an “event history”. It needs to support:

- multiple diagnosis encounters in the community and in a facility
- repeat visits to the same or different facilities including clinics, hospitals and nursing homes and
- repeat visits to a person in the community

Repeat visits to a person in the community are needed with a novel coronavirus because we are still learning the sequela and don’t know the observation period over which the appearance of sequela might extend.

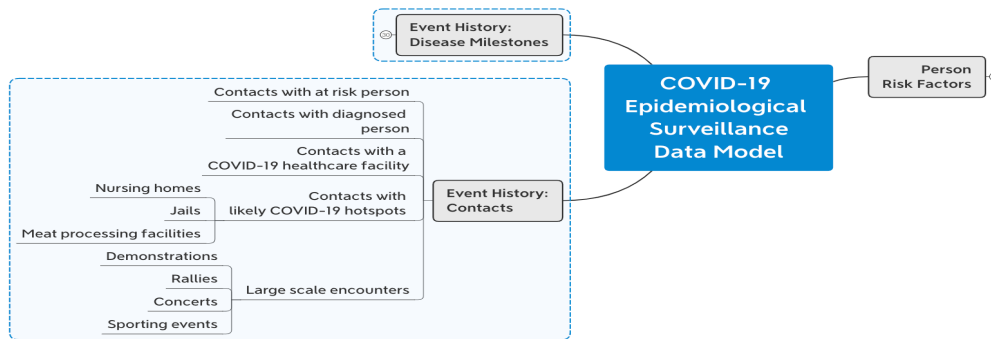


Figure 3: The Contacts Component

Note that with “large scale encounters” it isn’t possible by and large to record individual contacts. Increasingly, demonstrations are happening during the COVID-19 pandemic because of civil unrest. It is likely that “large scale encounters” will lead to new forms of contact tracing.

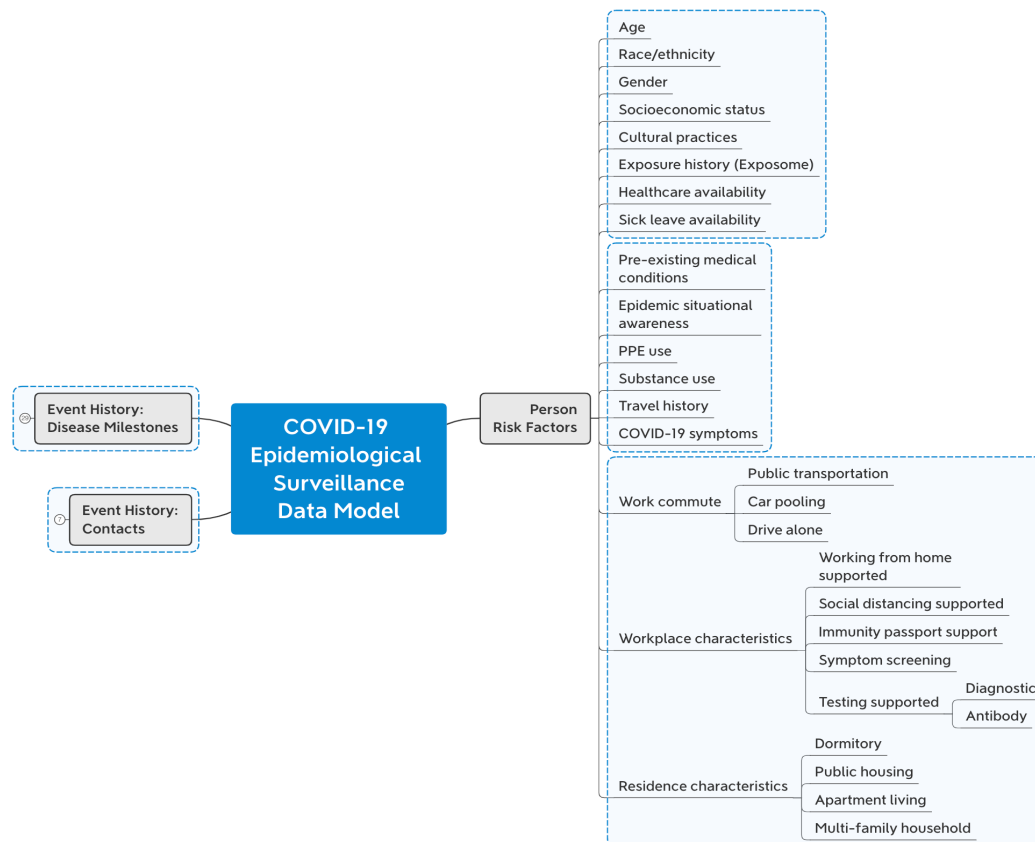


Figure 4: The Person Risk Factors Component

Again with COVID-19, as we learn more about salient pre-conditions, the mechanics of spread and the value of different initiatives in the public health toolbox, the enumeration of person risk factors remains a work in progress.

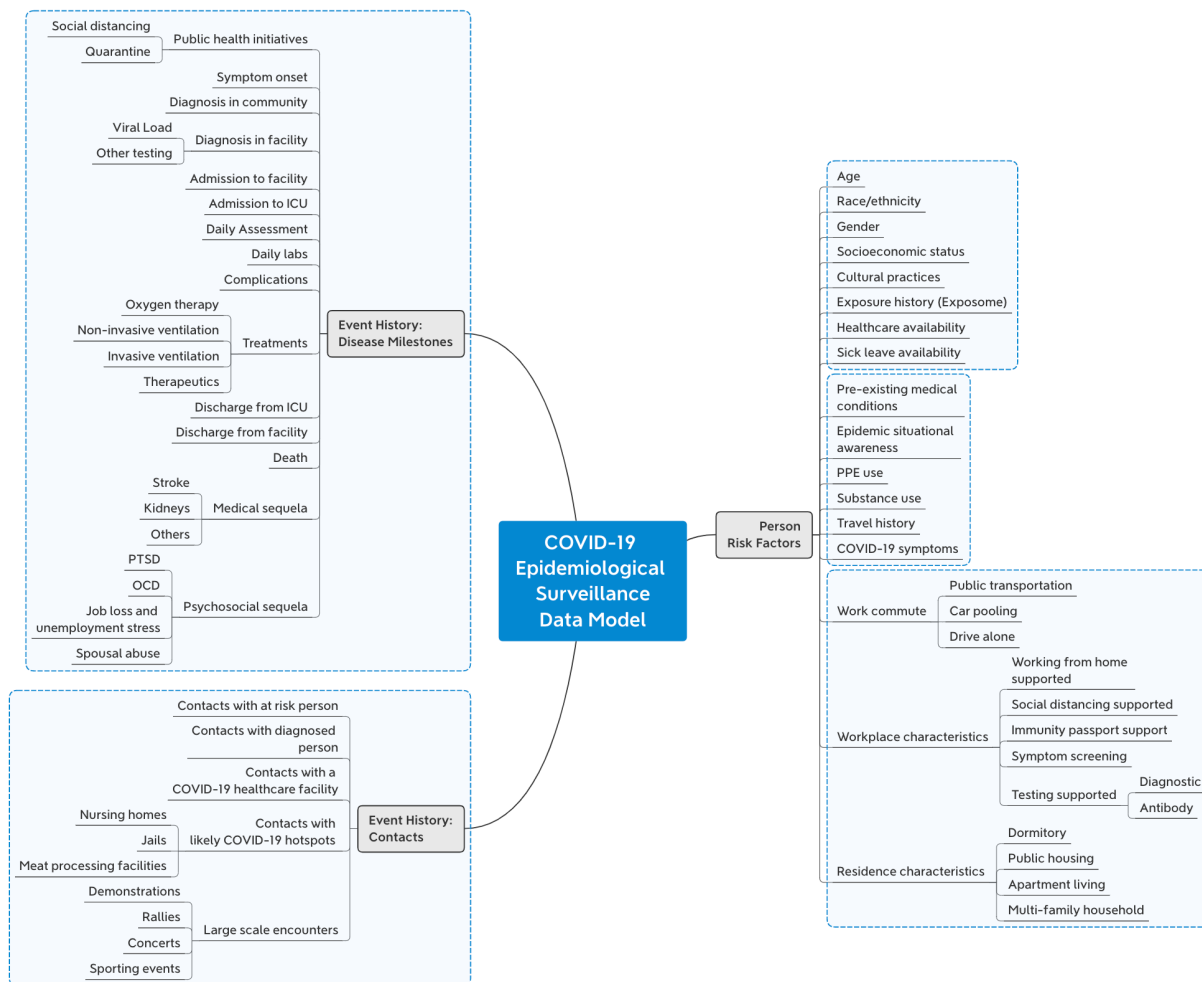


Figure 5: Full View of COVID-19 Epidemiological Surveillance Data Model (less Provenance)

DISCUSSION

The Epidemiological Surveillance Data Model is meant to inform the development of an implementation model composed of extensible database schema. For examples of extensible database schema, see the treatment of schemas in schema.org [schema.org. 2020] and the role of resources in HL7 FHIR [HL7. 2020]. Also, recall that the model is longitudinal in scope so an implementation model of extensible schema needs to be able to record multiple encounters for repeat visits to clinics, hospitals and nursing homes, repeated and changed diagnoses and diagnostic tests, and repeat visits with field workers.

AUTHOR ROLES

All authors accept responsibility for the content of the article

The same author may have multiple roles, and multiple authors may share a single role.

Conceptualization: JG, MS, CCA. Methodology: JG. Investigation: JG, RN, MS, RN. Validation: RN, MS.

Writing: JG, MS, CCA. Visualization: JG

REFERENCES

CDISC. [Website] CDISC Interim User Guide for COVID-19.

<https://wiki.cdisc.org/display/COVID19/CDISC+Interim+User+Guide+for+COVID-19>. 2020

Giordano G, Blanchini F, Bruno R et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med* (2020). <https://doi.org/10.1038/s41591-020-0883-7>

HL7. [Website] Introduction to HL7 FHIR. <https://hl7.org/FHIR/summary.html>. 2020

Pesquita C, Ferreira JD, Couto FM, Silva MJ. The epidemiology ontology: an ontology for the semantic annotation of epidemiological resources. *J Biomed Semantics*. 2014;5(1):4. Published 2014 Jan 17. doi:10.1186/2041-1480-5-4

Schema.org. [Website] Organization of Schemas. <https://schema.org/docs/schemas.html>, 2020

Schmidt CO, Nagrani R, Löbe M, et al. Interoperable COVID-19 epidemiological surveillance: Clinical and Population-based instruments [published online ahead of print, 2020 June 8]

Sun, P and Kang L. *An SEIR Model for Assessment of Current COVID-19 Pandemic Situation in the UK*, 2020. <https://doi.org/10.1101/2020.04.12.20062588>.

Wang C, Pan R, Wan X, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China [published online ahead of print, 2020 Apr 13]. *Brain Behav Immun*. 2020;S0889-1591(20)30511-0. doi:10.1016/j.bbi.2020.04.028

Wellcome Trust. [Webpage] https://wellcome.ac.uk/sites/default/files/longitudinal-population-studies-strategy_0.pdf. 2017

World Health Organization. [Webpage] Global Influenza Surveillance and Response System (GISRS). https://www.who.int/influenza/gisrs_laboratory/en/. 2020a

World Health Organization. COVID-19 Core Version CRF. [Core COVID-19 CRF - English](#). 2020b

World Health Organization. COVID-19 Rapid Version CRF. [RAPID COVID-19 CRF – English](#). 2020c

Working paper

APPENDIX 5 – Epi-TRACS: Rapid detection and whole system response for emerging pathogens

Jay Greenfield^{1,*}, Henri E.Z. Tonnang², Gary Mazzaferro³, Claire C. Austin⁴; and the RDA-COVID19-WG*

¹Data Documentation Initiative, ²International Center of Insect Physiology and Ecology (ICIPE), Kenya, ³Independent Researcher, ⁴Environment and Climate Change Canada, ⁴This work is linked to the Research Data Alliance [RDA-COVID19-WG, Epidemiology subWG recommendations and guidelines on data sharing](#).

*Corresponding author: nightcleaner@gmail.com

All views and opinions expressed are those of the co-authors, and do not necessarily reflect the official policy or position of their respective employers, or of any government, agency or organization.

ABSTRACT

Epi-TRACS proposes a computable framework for emerging pathogen action plans. Epi-TRACS can assess population health and economic outcomes based on sentinel surveillance, public health disease management tools and inclusive growth and recovery tools using a form of causal analysis called causal loop diagrams.

BACKGROUND

COVID-19 threat detection has been slow relative to the speed with which the epidemics and subsequent pandemic spread. As a result, countries have implemented a series of severe public health measures that have differed from country to country depending on a variety of factors.

For both Low and Middle Income Countries (LMIC) that depended on the WHO Global Influenza Surveillance and Response System (GISRS) [WHO, 2020a] and High Income Countries (HIC) that, besides the GISRS, have some form of early response system separate from the WHO; the tempo, stealth, and spread of this novel coronavirus left most countries with little time to act leading to widespread lockdowns to suppress or mitigate the spread [Gates B, 2020; Bai Z, et al. 2020].

Lockdowns in turn have economic consequences for both HICs and LMICs.

These facts on the ground have engendered deep concerns at the WHO and in member countries for the current state of COVID-19 sentinel surveillance and for ongoing influenza sentinel surveillance programs as well where the WHO is reporting that vigilance has begun to suffer:

Since the emergence of the SARS-CoV-2 virus and the disease it causes, COVID-19, early in 2020, and in particular since the designation of the COVID-19 outbreak as a pandemic by WHO, laboratories of the Global Influenza Surveillance and Response System (GISRS) have become COVID-19 testing centres in

many countries. GISRS has also become the main global platform for COVID-19 sentinel surveillance, an essential component of WHO COVID-19 pandemic surveillance. Influenza-like illness (ILI), acute respiratory infection (ARI), and severe acute respiratory infection (SARI) syndromic sentinel surveillance systems have served to monitor community transmission and geographic spread of COVID-19.

However, numbers of specimens tested for influenza and shipments of viruses to WHO Collaborating Centres of GISRS have significantly decreased in the past months compared with the same time period in previous years. Reporting directly or indirectly to FluNet by some countries has also been delayed or ceased altogether. Although influenza activity in the Northern Hemisphere has decreased and remains at inter-seasonal levels, the Southern Hemisphere influenza season is imminent. Moreover, continued vigilance is needed for the emergence of zoonotic and non-seasonal influenza viruses of pandemic potential, as has been seen in the recent past in outbreaks caused by A(H5N1) and A(H7N9) viruses, for example [WHO. 2020b].

Now, seemingly, we are in uncharted territory where we have to:

- Prepare for the co-circulation of influenza and COVID-19 in the upcoming Southern Hemisphere influenza season, both of which are of public health importance;
- Utilize and strengthen existing national influenza surveillance systems for both influenza and COVID-19 responses; integrate COVID-19 sentinel surveillance into ongoing sentinel surveillance systems as much as possible under strategies tailored to the needs and capacity specific to the country [WHO. 2020b].

In this context we are having to learn and relearn how to phase and stage the public health and economic health toolboxes within and between regions and countries as part of a single global response system.

METHODS

In light of the COVID-19 pandemic experience, both the early warning and response parts of our systems need to be rethought. This is necessarily occurring in the midst of the pandemic. Varying responses around the world have created a number of natural experiments: South Korea, Sweden, Kenya, Brazil, Vietnam, California [Zwald ML, et al. 2020], South Dakota, and New York State [Cummings et al., 2020]. Each of these jurisdictions have implemented emergency public health and economic measures in unique ways.

Data collected from each of these natural experiments will help us to identify sentinels (*"canaries in the coal mine"*), and response measures that prove useful or not under various conditions. An early warning and response system called Epidemiology Translational Research Action Coordinated System (Epi-TRACS) has been proposed in the light of these natural experiments. It is more descriptive than prescriptive. The saliency of sentinels and responses proposed in Epi-TRACS need to be placed in a computable framework where they can be evaluated.

One candidate for this computable proposed here is a type of causal analysis called system thinking. In system thinking, circles of causality are identified together with the patterns these circles fall into -- called system archetypes [Bradley DT, Mansouri MA, Key F, et al. 2020].

Here the transmission of COVID-19 is represented as a dynamic process conditioned by multiple variables which can interact and escalate across various system sectors including the contagion subsystem, healthcare and the economy. These influences occur through a series of so-called “reinforcing” and “balancing” loops that together form a circle of causality in which variables associated with the contagion, healthcare and the economy are manipulated or not in a succession of action plans on a temporal and spatial scale across a plethora of social environments.

RESULTS

Epi-TRACS

We propose an Epidemiology Translational Research Action Coordinated System (Epi-TRACS) led by the WHO which even now is rethinking and reimagining itself based on lessons learned from COVID-19 (**Figure 1**).

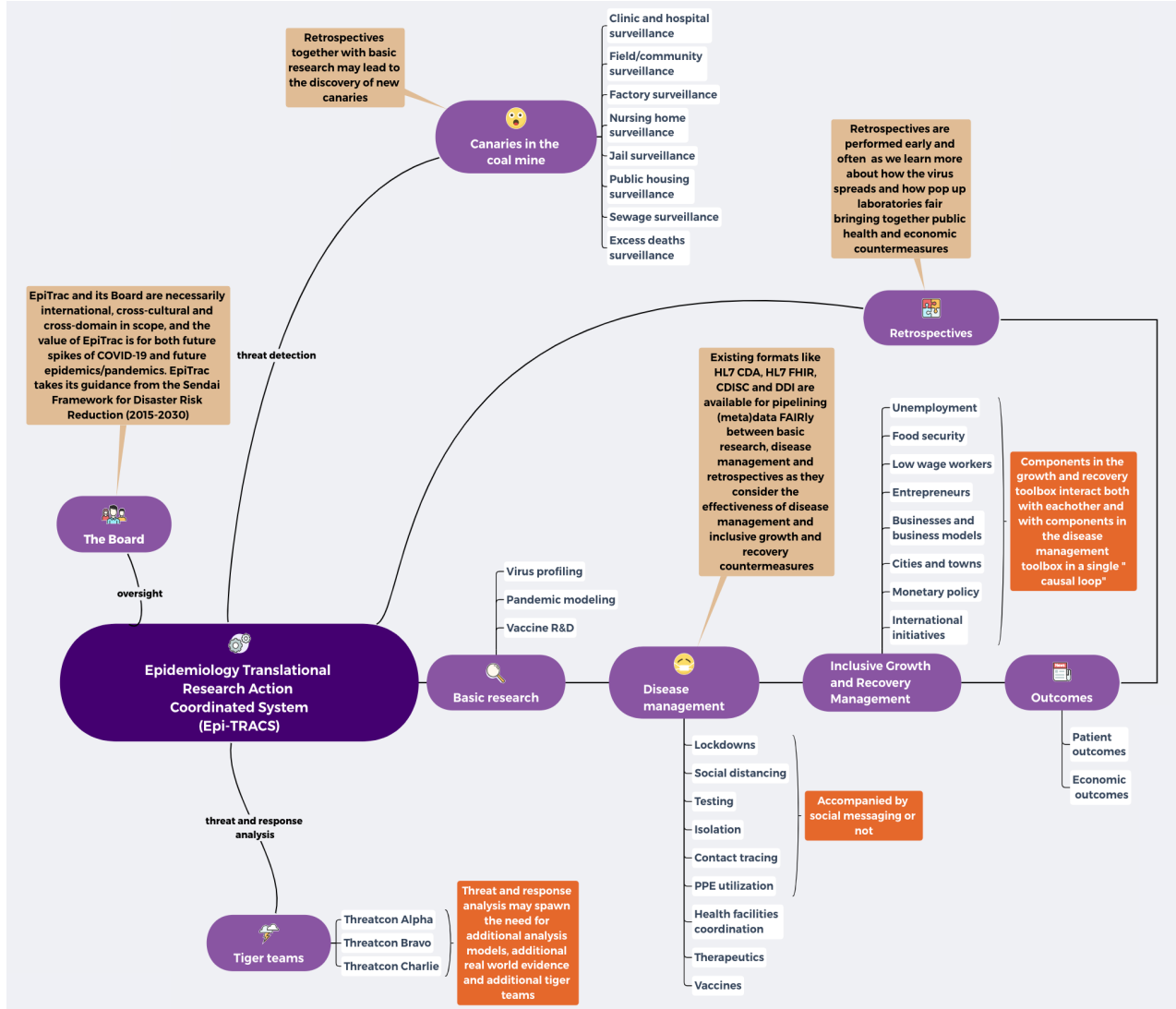


Figure 1. Epi-TRACS. Proposed WHO-led COVID-19 EPIdemiological Translational Research Action Coordinated System.

Epi-TRACS is a translational research workbench in a model derived from activity-based intelligence [Attwood, 2015]. The components of this model that are not colored in Figure 2, including a **common data model** and a complete description of **full spectrum epidemiology**, are discussed elsewhere [Full spectrum epidemiology and the common data model, work in progress].

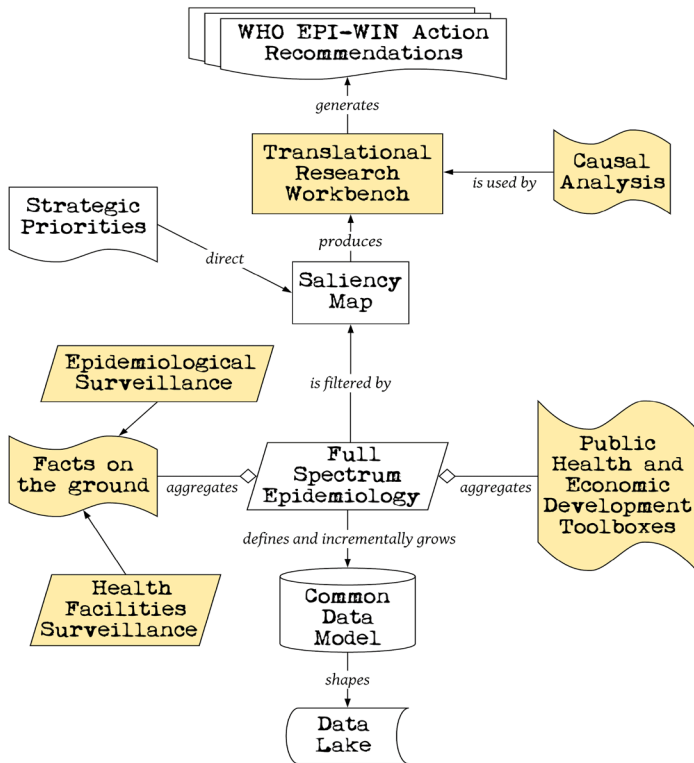


Figure 2. Underlying Epi-TRACS Intelligence Model: A model adapted from the activity-based intelligence model to support full spectrum COVID-19 epidemiology.

Epi-TRACS puts sentinel surveillance facts on the ground (the *canaries in the coal mine*), together with public health disease management tools, and inclusive growth and recovery tools in an action plan that posits a set of patient health and economic outcomes (Figure 3).

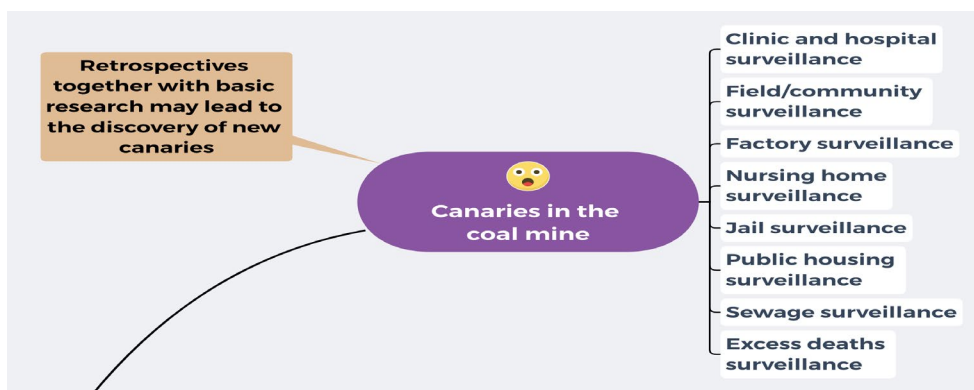


Figure 3. Epi-TRACS Sentinel Surveillance. The canaries in the coal mine.

The public health toolbox includes a set of tools that can be used singly or in combination depending the availability of countermeasures and the current state of disease spread in a population (Figure 4).

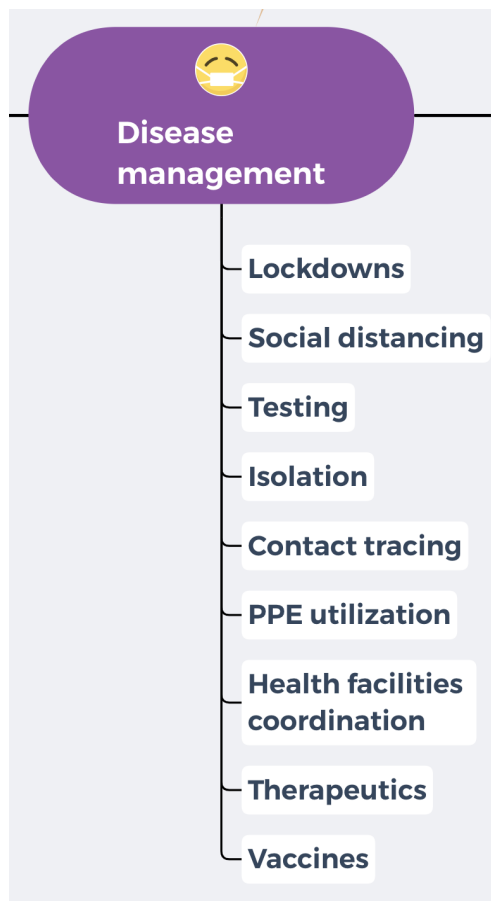


Figure 4. Public Health Disease Management Tools: The public health toolbox.

The economic development toolbox includes monetary and fiscal policy. However, growth and recovery management also include reimagining business models, cities and towns, and support for the disadvantaged (Figure 5).

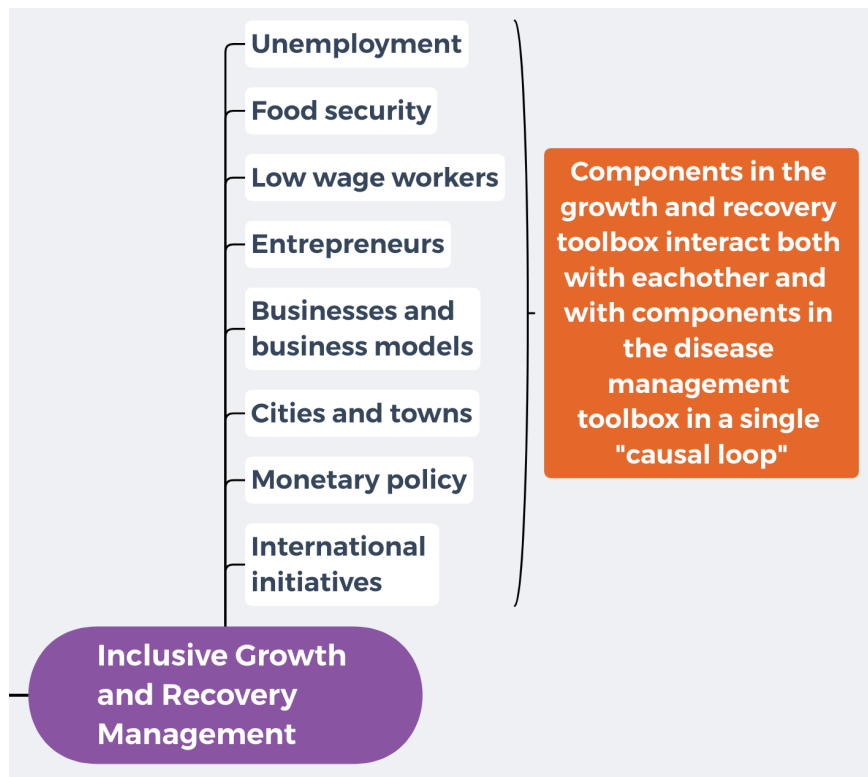


Figure 5. Inclusive Growth and Recovery Management Tools: The economic development toolbox.

Both retrospectively and prospectively, Epi-TRACS uses data science *tiger teams* and *causal analysis* to assess the results these action plans produce. Epi-TRACS is governed by an *international, cross-domain and cross-cultural board* that gives direction to the data scientists and the early warning and response system as a whole.

The Emergency Public Health and Economic Measures Causal Loop

Alongside the Epi-TRACS proposal for an early warning and response system, a computable framework is proposed in which the actions taken in response to COVID-19 sentinel surveillance can be simulated and assessed both retrospectively and prospectively. It has been shown that causal loop modeling may be valuable in assessing *system sustainability* and *system resiliency* [Ricciardi F, De Bernardi P, Cantino V. 2020].

The basic causal loop diagram (CLD) characterizing the dynamics and interactions between the COVID-19 contagion, the healthcare subsystem and the economy subsystem are shown in **Figure 6**.

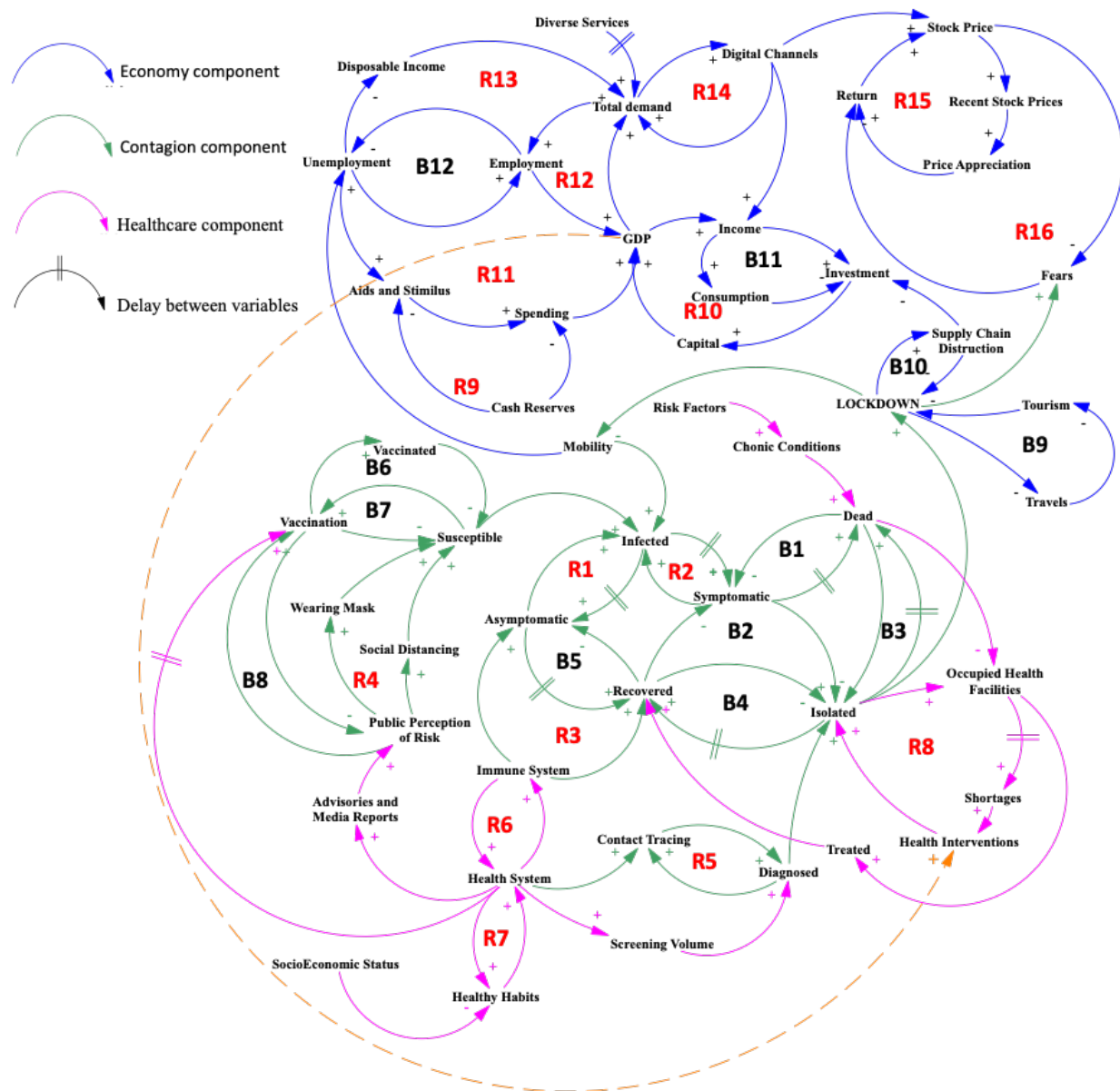


Figure 6. A notional COVID-19 emergency public health and economic measures causal loop. The colours green, violet and blue indicate pandemic effects on the population, healthcare system, and economic sector, respectively.

The whole interacting system consists of twelve directly recognizable balancing loops (B1-B12) and sixteen obvious reinforcing loops (R1-R16) distributed as follows:

- The contagion system component in green has eight balancing loops (B1-B8) and five reinforcing loops (R1-R5).
- The healthcare component has three reinforcing loops (R6-R8).

- And the economy component encompasses four balancing loops (B9-B12) and eight reinforcing loops (R9-R16).

The contagion subsystem is based on the SEIR (Susceptible, Exposed, Infectious, Recovered) model. Table 1 breaks out the loops from Figure 6 into system components. Each loop is identified, its endpoints enumerated, and its implication described.

Table 1. COVID-19 causal loops by system component

Loop	Description	Implication
Contagion on the human population system component		
R1 - Reinforcing	Asymptomatic-infected-asymptotic	Exponential growth of infected individuals among human population
R2 - Reinforcing	Infected – symptomatic - infected	Upsurge in number of dead and isolated individuals
R3 - Reinforcing	Immune system – asymptomatic – recovered - immune system	Strong immune system upsurges the number of asymptomatic individuals
R4 - Reinforcing	Susceptible - social distance - public perception -wearing mask - susceptible	Wearing mask and social distancing growth the number of susceptible individuals
R5 - Reinforcing	Contact tracing – diagnosed - contact tracing	Contact tracing growth number of diagnosed individuals
B1 - Balancing	Symptomatic – dead - symptomatic	Symptomatic individual after a time delay could die
B2 - Balancing	Recovered – symptomatic – isolated - recovered	Recovered individual reduce the number of symptomatic
B3 - Balancing	Isolated - dead - isolated	After delay, section of isolated individuals dies contributing to increase the number of dead individuals

B4 - Balancing	Recovered – isolated - recovered	After delay section of isolated individuals recovered
B5 - Balancing	Asymptomatic – recovered - asymptomatic	After delay section of asymptomatic individuals recovered
Potential contribution of vaccine		
B6 - Balancing	Vaccination-vaccinated-susceptible-vaccination	Vaccine contribute to decline in the number of the susceptible individuals
B7 - Balancing	Vaccination – susceptible - vaccination	Number of susceptible individuals reduced due to vaccine
B8 - Balancing	Vaccination - public perception of risk - vaccination	Public perception influences vaccine intake
Healthcare component		
R6 - Reinforcing	Health system - immune system - health system	Upright health system enhances immune system
R7 - Reinforcing	Healthy habit - health system - healthy habits	Healthy habits upsurge number of individuals with strong immune system
R8 - Reinforcing	Health intervention – isolated - occupied health facilities -shortage - health interventions	Increase health interventions upsurge the number of isolated and occupied health facilities contributing to shortage
Economy component		
R9 - Reinforcing	Aid and stimulus – cash reserve – spending	Increase aid and stimulus boots spending which may reduce GDP
R10 - Reinforcing	Capital - GDP – income – investment - Capital	Injection of capital boosts the GDP which increases income and investment

R11 - Reinforcing	Unemployment - aids and stimulus – spending – GDP -employment - unemployment	Unemployment triggers aid and stimulus for spending that contribute to GDP loss and reduced employment
R12 - Reinforcing	Employment – GDP – demand - employment	Job creation growth the GDP which contribute to total demand
R13 - Reinforcing	Employment – unemployment - disposable income -demand - employment	Reduces in employment surges the number of unemployed individuals reducing disposable income
R14 - Reinforcing	Total demand - digital channels - total demand	Exponential growth in the use of digital channels
R15 - Reinforcing	Return - stock price - recent stock price - price appreciation - return	Volatility increases in recent stock prices provoking price appreciation and return
R16 - Reinforcing	Stock price – fears – returns – stock price	Stock volatility create fears, leading to low returns
B9 - Balancing	Lockdown – travel – tourism - lockdown	Lockdown reduces travel with negative impact on tourism
B10 - Balancing	Lockdown - supplier chain - lockdown	Lockdown reduces supply chain causing distortion
B11 - Balancing	Income – consumption - investment – capital – GDP - income	Income contributes to consumption limiting investment
B12 - Balancing	Unemployment – employment – unemployment	Aids and stimulus boost unemployment reducing employment

Reviewing the CLD, potential ways to reduce the number of infected individuals is to boost the population immunity and reduce human movements. Effective social distancing associated with wearing a mask seems to also reduce the number of infected individuals. However, to reduce the number of dead, special care should be given to people with high risk factors that lead to chronic conditions. Vaccine, when available will contribute to decline in the number of the susceptible individuals, but adequate advocacy is necessary as public perception influences vaccine intake.

Figure 7. Interaction channels between the contagion, healthcare and economic system components

Table 2 breaks out the interacting loops from Figure 7 across the contagion, healthcare, and economic components. Each loop is identified, its endpoints enumerated, and its implication described.

Table 2. COVID-19 causal loop diagram by system.

Loop	Description	Implication
RCHE1 - Reinforcing	GDP – health interventions – shortages – occupied health facilities – treated – recovered – isolated – lockdown – fears – return– stock price – digital channels – income – investment - GDP	Overcoming fears and upsurge interest in investing in stock market may help increase the GDP and revive national economy to exponential growth
RCHE2 - Reinforcing	GDP – health interventions – isolated – lockdown – fears – return– stock price – digital channels – income – investment - GDP	Overcoming fears and upsurge interest in investing in stock market may help increase the GDP and revive national economy to exponential growth
BCHE1 - Balancing	GDP – health interventions – shortages – occupied health facilities – treated – recovered – isolated – lockdown – mobility – unemployment – aids and stimulus – spending - GDP	Continuing depletion of the GDP through aids, stimulus and interventions in health system component will lead national economy to recession and loss of employment
BCHE2 - Balancing	GDP – health interventions – shortages – occupied health facilities – treated – recovered – isolated – lockdown – mobility – unemployment – employment - GDP	Continuing depletion of the GDP through aids, stimulus and interventions in health system component will lead national economy to recession and loss of employment

BCHE3 - Balancing	GDP – health interventions – isolated – lockdown – mobility – unemployment – aids and stimulus – spending - GDP	Continuing depletion of the GDP through aids, stimulus and interventions in health system component will lead national economy to recession and loss of employment
BCHE4 - Balancing	GDP – health interventions – isolated – lockdown – mobility – unemployment – employment - GDP	Continuing depletion of the GDP through aids, stimulus and interventions in health system component will lead national economy to recession and loss of employment

DISCUSSION

“Inclusive Growth and Recovery Management” in **Figure 1** is also the name of a challenge, with a call for entries being supported by the Rockefeller Foundation [The Rockefeller Foundation, 2020]. In this challenge and elsewhere “data science breakthroughs” are being counted upon to inform the development of early warning and response systems for COVID-19, future novel viruses and other potential threats characterized by breathtaking tempos.

In this context causal loop diagramming (CLD) holds promise and may be fit to purpose. More specifically, it may be possible to use CLD to determine the sustainability and resiliency of the COVID-19 contagion, healthcare and economic growth and recovery system under different action plans.

As governments and non-governmental organizations review their research portfolios, we expect more challenges like the Rockefeller Foundation “Inclusive Growth and Recovery Management” challenge to emerge. And, in response to these challenges, over the next several months we expect researchers will be undertaking assessments of the sustainability and resiliency of various pandemic action plans for growth and recovery using CLD.

AUTHOR ROLES

All authors accept responsibility for the content of the article

The same author may have multiple roles, and multiple authors may share a single role.

Conceptualization: GM, JG. Methodology: JG, HT, GM. Investigation: CCA. Formal analysis: GM, HT. Writing: JG. Bibliographic review and analysis: JG, MC. Visualization: JG, HT, CCA

REFERENCES

- Atwood CP. Activity-Based Intelligence: Revolutionizing Military Intelligence Analysis. *Joint Force Quarterly*. 2015; 77 (2nd Quarter, April 2015)
- Bai Z, Gong Y, Tian X, Cao Y, Liu W, Li J. The Rapid Assessment and Early Warning Models for COVID-19 [published online ahead of print, 2020 Apr 1]. *Virus Sin*. 2020;1-8. doi:[10.1007/s12250-020-00219-0](https://doi.org/10.1007/s12250-020-00219-0)
- Bradley DT, Mansouri MA, Key F, et al. A systems approach to preventing and responding to COVID-19 [published online ahead of print, 2020 Apr 1]. *EClinicalMedicine* 2020;21. doi:[10.1016/j.eclinm.2020.100325](https://doi.org/10.1016/j.eclinm.2020.100325)
- Cummings MJ, Baldwin MR, Abrams D, et al. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study [published online ahead of print, 2020 May 19]. *The Lancet* 2020. doi:[10.1016/S0140-6736\(20\)31189-2](https://doi.org/10.1016/S0140-6736(20)31189-2)
- Gates B. Responding to Covid-19 — A Once-in-a-Century Pandemic? *N Engl J Med* 2020; 382:1677-1679. doi:[10.1056/NEJMp2003762](https://doi.org/10.1056/NEJMp2003762)
- Ricciardi F, De Bernardi P, Cantino V, 2020. "System dynamics modeling as a circular process: The smart commons approach to impact management," *Technological Forecasting and Social Change*, Elsevier, vol. 151(C). doi:[10.1016/j.techfore.2019.119799](https://doi.org/10.1016/j.techfore.2019.119799)
- Rockefeller Foundation. [Webpage] Call for entries: Data Science Breakthroughs for an Inclusive Recovery. https://www.rockefellerfoundation.org/blog/call-for-entries-data-science-breakthroughs-for-an-inclusive-recovery/?utm_source=Rockefeller+Foundation+eAlerts&utm_campaign=3e2ba36552-EMAIL_CAMPAIGN_2020_MAY_TSG_LAUNCH&utm_medium=email&utm_term=0_6138ee88b7-3e2ba36552-215466361&goal=0_6138ee88b7-3e2ba36552-215466361&mc_cid=3e2ba36552&mc_eid=20f10e51cc. 2020
- WHO. [Webpage] Global Influenza Surveillance and Response System (GISRS). https://www.who.int/influenza/gisrs_laboratory/en/. 2020a
- WHO. [Webpage] Preparing GISRS for the upcoming influenza seasons during the COVID-19 pandemic – practical considerations. https://apps.who.int/iris/bitstream/handle/10665/332198/WHO-2019-nCoV-Preparing_GISRS-2020.1-eng.pdf. 2020b
- Zwald ML, Lin W, Sondermeyer Cooksey GL, et al. Rapid Sentinel Surveillance for COVID-19 - Santa Clara County, California, March 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:419-421. doi:[10.15585/mmwr.mm6914e3](https://doi.org/10.15585/mmwr.mm6914e3)

Working paper

APPENDIX 6 – Putting It All Together: Common Data Model and Epi-Stack

Jay Greenfield¹, Meg Sears², Rajini Nagrani³, Gary Mazzaferro⁴, Anna Widyastuti⁵, Claire C. Austin⁶; and the RDA-COVID19-WG⁷

¹Data Documentation Initiative, ²Ottawa Hospital Research Institute, ³Leibniz Institute for Prevention Research and Epidemiology-BIPS, ⁴Independent Researcher, ⁵Hasselt University, ⁶Environment and Climate Change Canada, ⁷This work is linked to the Research Data Alliance [RDA-COVID19-WG, Epidemiology subWG recommendations and guidelines on data sharing](#).

*Corresponding author: nightcleaner@gmail.com

All views and opinions expressed are those of the co-authors, and do not necessarily reflect the official policy or position of their respective employers, or of any government, agency or organization.

ABSTRACT

An architecture for hosting COVID-19 datasets is proposed. The architecture supports full spectrum COVID-19 research. Full spectrum includes schemas for clinical care, demographic surveillance and epidemiological surveillance data based on the Common Data Model and the extensions it supports for questionnaire data. An emulated clinical trial framework based on the Common Data Model that employs causal analysis is described.

BACKGROUND

COVID-19 epidemiology observes and identifies factors that influence the spread of COVID-19 within individuals, in their communities, in a country as a whole and across borders. In addition to resources and successes in disease detection and treatment, important determinants of health include individual and population risk factors, environmental context and exposures, and policy and public health responses.

COVID-19 presents very differently among individuals, from asymptomatic (albeit contagious) infections, to lethal. A consistent observation in the USA [Garg, Kim, Whitaker, et al. 2020], Italy [EpiCentro 2020] and China [Liu, Fang, Deng et al. 2020], however, is that individuals with pre-existing chronic conditions such as hypertension, diabetes, asthma, chronic obstructive pulmonary disease (COPD) and/or other comorbid risk factors are much more likely to be hospitalized, admitted to the Intensive Care Unit (ICU) and to die. Indeed, virtually all of the deaths are in individuals with pre-existing comorbidities, across all age ranges - not merely in the elderly [Garg, Kim, Whitaker, et al. 2020].

Risk factors being studied include socioeconomic status, housing, crowding, occupation, environmental exposures, lifestyle-related factors, health-related behaviours and pre-existing conditions [Greenfield J, Sears M, Nagrani R, et al. 2020] The disease progression, severity and sequelae may be further impacted

by patient care procedures, emergency public health and economic measures [Greenfield J, Sears M, Nagrani R, et al. 2020]. Much of this information is available from hospital/clinic surveillance (Electronic Health Records [EHR]), field-based longitudinal demographic and epidemiological surveillance, population level indicators, and monitoring programs (e.g., air quality monitoring).

To date, clinical and observational research is largely conducted as stand-alone studies with little integration, even when conducted in the same population. This leads to duplication of efforts by investigators and data collection personnel to extract similar information for a given population. A further shortcoming is that even a small change in construction of a question by two different investigators may magnify the challenges to integration [Schmidt CO, Nagrani R, Löbe M, et al. 2020]. For example, age at event requires manipulation to translate into the year of event or numbers of years since the event. Further, the purpose of information collected in clinics is different from data collected in the field (observational studies). Medical staff in clinics are often overburdened by the sheer number of patients, compromising the quality of information collected which may not be relevant to treatment (e.g., age at menstruation, family history of disease, occupational history, etc.) but may be important in epidemiological association studies. This poses challenges in consolidating information between clinics and observational studies. In the recent years, there have been considerable efforts to streamline Case Report Forms (CRFs) and population-based questionnaires using the approach called Common Data Model [Garza M, Fiol GD, Tenenbaum J, et al. 2018].

The Common Data Model

Common data models (CDMs) were first developed to host and support twenty-first century patient care research using EHR data:

...a CDM is used to standardize and facilitate the exchange, pooling, sharing, or storing of data from multiple sources... Within the last decade, several CDMs have been collaboratively developed and have risen to the level of **de facto standards for clinical research data** [emphasis added] [Garza M, Fiol GD, Tenenbaum J, et al. 2018].

Now these CDMs host:

...the increased exploration of observational, quasi-experimental, and experimental clinical study designs all using healthcare data... [Garza M, Fiol GD, Tenenbaum J, et al. 2018]

where “quasi-experimental” designs include the use of Electronic Health Record (EHR) data to support emulated clinical trials when clinical trials are impractical [Hernán and Robins, 2016]. Similarly, and in line with their support for multiple data sources, CDMs have also evolved to include questionnaire data from surveys conducted in the field [Blacketer M, Voss EA, Ryan PB, 2015].

Together this information forms an epidemiology intelligence model in which there is person-level information from both clinics and community on the ground, and countermeasures or action candidates. Action candidates fall into at least two categories -- a public health toolbox and an economic development toolbox. This intelligence model supports the exercise of cross-domain full spectrum epidemiology [Atwood, 2015].

We are proposing to use the CDM with COVID-19 to integrate clinical data and epi data for COVID-19.

METHODS

An activity-based intelligence model was adapted to create Epi-STACK - a COVID-19 epidemiology model that supports cross-domain, full spectrum COVID-19 epidemiology. Several COVID-19 specific components emerged.

RESULTS

Figure 1 illustrates how Epi-STACK may yield many workbenches or, again, analytical models. Models may be explored as emulated trials to discern associations and causal relationships, as illustrated in **Figure 2**. Finally, one of these models, the COVID-19 Risk Assessment Workbench is exemplified in **Figure 3**.

The activity-based intelligence model yields a series of translational research workbenches to develop WHO EPI-WIN action recommendations. EPI-WIN “*covers common myths; information for health workers; effects on travel and tourism; and tailored advice for the general public, businesses and employers, and WHO member states*” [WHO, 2020].

For full spectrum COVID-19 epidemiology we considered multiple disease surveillance sources together with the public health and economic development toolboxes that mitigate spread and severity of disease in a population. For more details about the intelligence that underlies this full spectrum view, see *A Full Spectrum View of the COVID-19 data domain: An Epidemiological Data Model* [Greenfield J, Sears M, Nagrani R, Austin CC, the RDA-COVID19-WG, 2020].

Depending on the characteristics of the pathogen, its history in a population and the resources and practices of a population, different facts on the ground and different components of public health and economic development toolboxes are salient. Full spectrum COVID-19 epidemiology is hosted by a Common Data Model which is grounded in a Data Lake. One of the many research workbenches that Epi-STACK can yield is a Risk Assessment Research Workbench for COVID-19 Health Care, that uses causal analysis in the course of producing WHO EPI-WIN action recommendations.

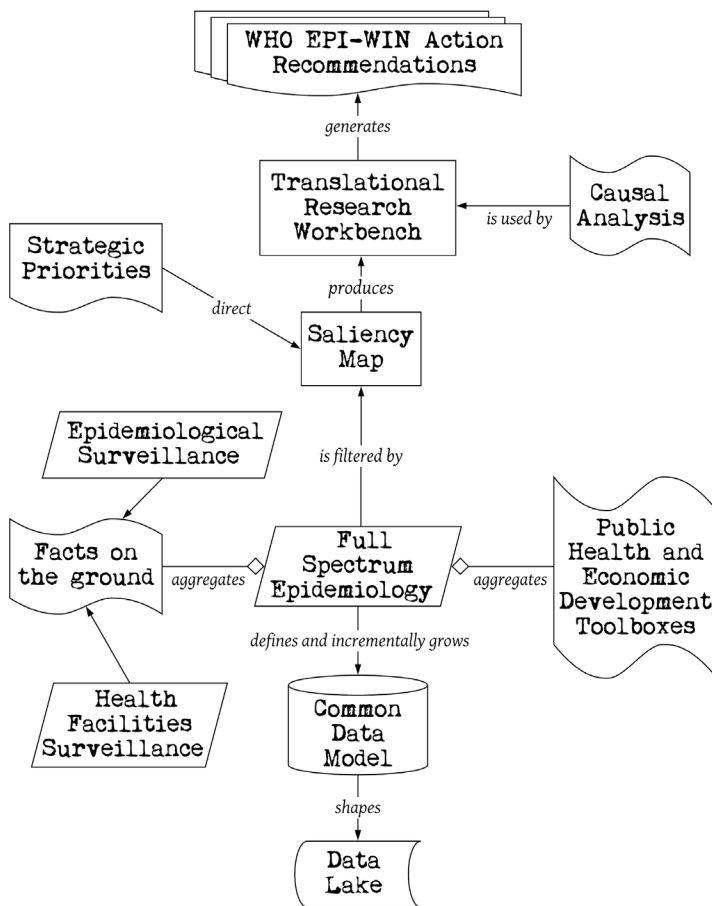


Figure 2. Epi-Stack. Proposed evidence support system as input to the WHO’s Epi-WIN communication channels targeting various audiences.

Quasi-experimental design and causal analysis are integral to the Risk Assessment Research Workbench for COVID-19 Health Care specific to this pathogen. One feature of SARS-CoV-2, the COVID-19 pathogen, that is especially remarkable and challenging, is its pace. Rapid pace is evident in rapid spread and short window for action, which in turn impacts the usefulness of various countermeasures. The pace of novel viruses can also contribute to social unrest [Nii-Trebi, 2017]:

Infectious diseases are a significant burden on public health and economic stability of societies all over the world. They have for centuries been among the leading causes of death and disability and presented growing challenges to health security and human progress. The threat posed by infectious diseases is further deepened by the continued emergence of new, unrecognized, and old infectious disease epidemics of global impact.

Risk Assessment Research Workbench for COVID-19 Health Outcomes

COVID-19 pace motivates us to consider quasi-experimental designs with observational EHR data for clinical research, modelling random assignment to treatment as time and ethical considerations allow. The risk assessment workbench uses observational data and causal analysis to produce interim results and to reduce uncertainty. This was demonstrated with several meta analyses that compare the results of certain observational designs with randomized clinical trials [Benson K, Hartz AJ. 2000]. In separate papers, we develop proposals for two additional workbenches that take into account the rapid pace of the COVID-19 pathogen: (a) an early warning and response system [Epi-TRACS, work in progress]; and (b) a decision support workbench for the selection of public health and economic response measures [Novel pathogens decision support, work in progress].

CDM may be considered suitable for integration of clinical and observational data in support of translational research, such as evaluation of the effect of comorbidities, patient care or both on the clinical course of COVID-19 patient cohorts.

The concept behind “different environmental exposures” is the “exposome,” defined as “the totality of exposures to which an individual is subjected from conception to death, including those resulting from environmental agents, socioeconomic conditions, lifestyle, diet, and endogenous processes” [Porta, 2016; Dagnino, Macherone eds. 2019]. Development, health and chronic diseases are strongly affected by the exposome. Comorbidities are in turn strongly associated with poorer clinical course across all ages [Garg S, Kim L, Whitaker M, et al. 2020], suggesting a potential for population resiliency (milder disease) and lower health care burden, with lower levels of adverse environmental exposures. Environmental exposure history questionnaires are used to assess, at least in part, the exposome [Marshall, Weir E, et al. 2002; Dagnino, Macherone, eds. 2019], and how environmental exposures are developmental origins of disease [Heindel, Balbus et al. 2015].

The same or other questionnaires are useful in comparing and contrasting populations in areas that are more or less polluted or disadvantaged, noting also potential roles of nutrition, housing (including congregate housing), occupations and activities [Liu, 2020; EpiCentro, 2020].

In support of the health care risk assessment research workbench the Common Data Model takes as input both health facilities intelligence and sensor intelligence that takes the form of epidemiological surveillance questionnaires administered in the field repeatedly over time.

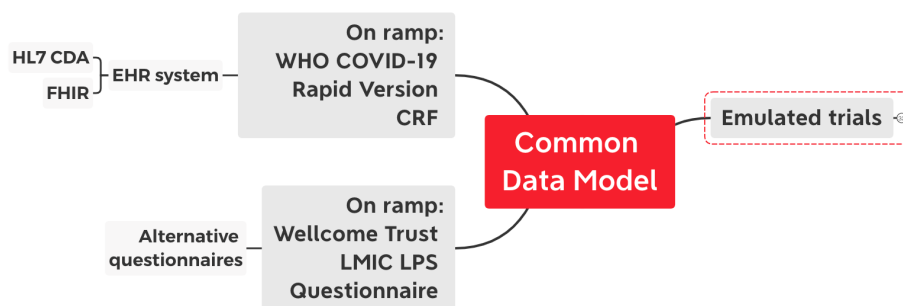


Figure 2. Full spectrum translational research enabled by the Common Data Model. The Common Data Model is a full spectrum provider for emulated trials conducted to understand the risk factors that influence outcome

Next the CDM places all this intelligence in a framework called the emulated trial. In an emulated trial the impact of various treatments taken separately or together on the treatment course is assessed using “messy” EHR and/or claims data alongside exposure and/or other “confounding” data [Gershman B, Guo DP, Dahabreh IJ, 2018].

An emulated trial is observational research that models a hypothetical randomized clinical trial. In the emulated trial, risk factors that could have an impact on the disease course apart from and or in conjunction with treatments may be included as stratification variables.

Causal Analysis

Risk factors may operate by themselves and/or with each other to influence treatment effectiveness and the clinical course. Causal analysis is a way of considering a chain of factors. Causal analysis is conducted using various methodological constructs including causal loops [Bradley et al. 2020], Markov chains [Poole and Crowley. 2013] and linear structural equations [Bullock et al. 1994]. Hernan and Robins, who pioneered emulated trials using CDMs and big data, recently published a primer on causal analysis in observational research called Causal Inference: What If [Hernán and Robins. 2020]: *“The book provides a cohesive presentation of concepts of, and methods for, causal inference that are currently scattered across journals in several disciplines. We expect that the book will be of interest to anyone interested in causal inference, e.g., epidemiologists, statisticians, psychologists, economists, sociologists, political scientists, computer scientists.”*

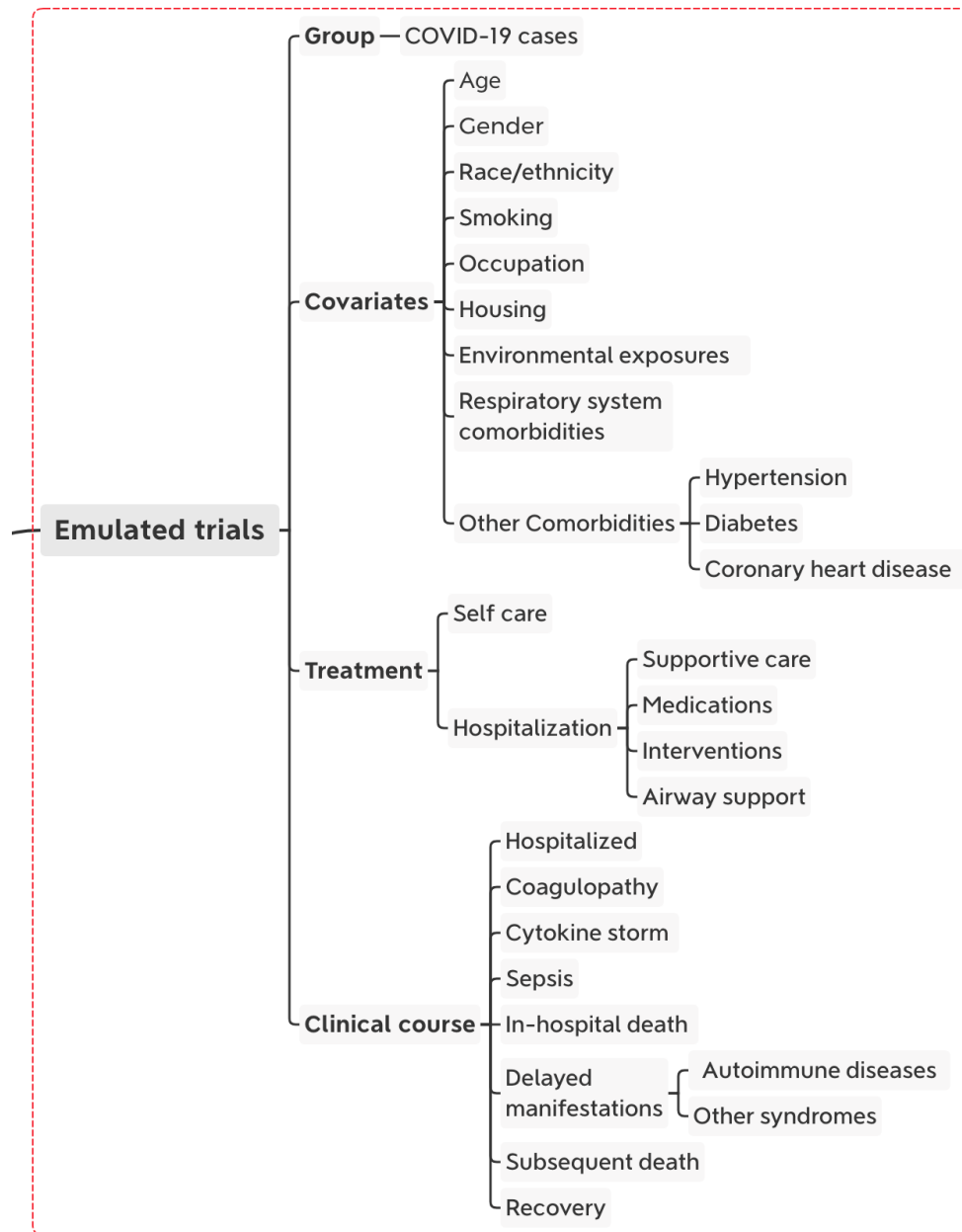


Figure 3. COVID-19 Risk Assessment Workbench for Health Care and Population Health. Details are provided as examples; lists are not comprehensive.

Epi-STACK and EPI-WIN

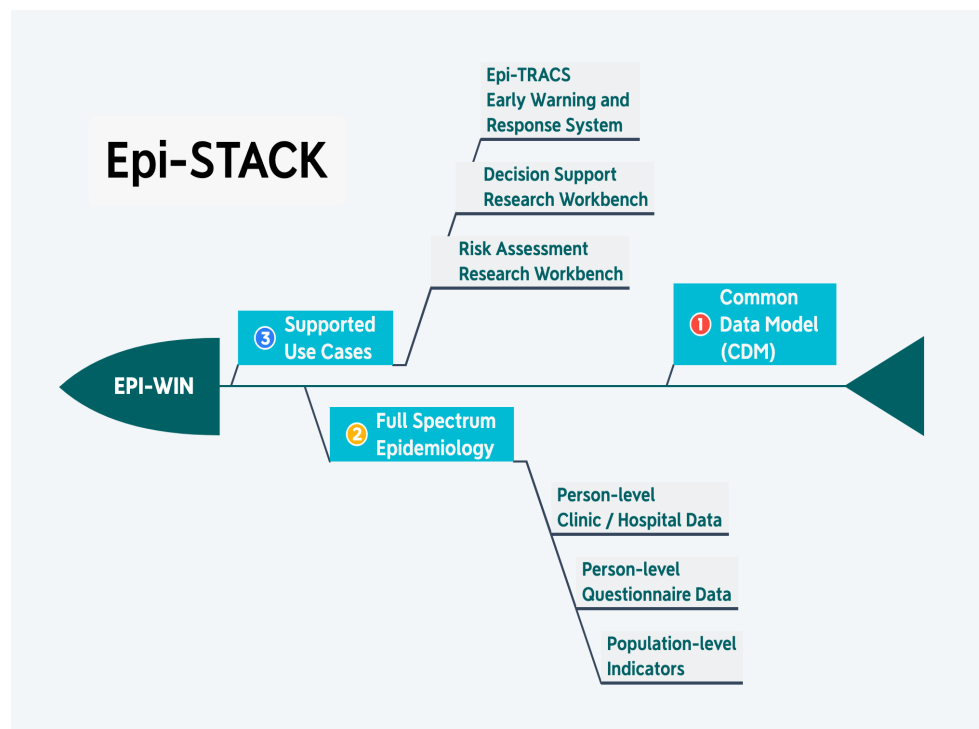


Figure 4. Another View of Epi-STACK. In this view the three workbench use cases are enumerated.

Figure 3 illustrates that integration of environmental history and comorbidities may be useful to derive causal relationships in the course of COVID-19, with information from cohorts that have regular follow-up since birth, combined with data collected in CRFs.

Other use cases include an early warning and response system [Epi-TRACS, work in progress], and a decision support system for public health and economic interventions [Novel pathogens decision support, work in progress]. At present these three use cases and the intelligence that they necessitate define Full Spectrum Epidemiology which in turn gives shape to the CDM.

All of these use cases require a series of clinical datasets, survey datasets and population indicators over time, and this is what makes Epi-STACK fit for purpose with the WHO EPI-WIN communications facility: Epi-STACK has the ability over time to produce a series of interim results that the WHO is able to channel to a wide range of target audiences.

DISCUSSION

Both the increasing frequency with which novel pathogens are occurring and now the pace of these pathogens has overwhelmed the current public health intelligence apparatus, which includes the WHO

and the CDC. This has led to recriminations between countries and within regions. These recriminations together with the failure of an early public health response are affecting trade, supply and demand with the result that the world is experiencing a perhaps unprecedented economic contraction and serious social unrest.

Under the circumstances, across the world people are pinning their hopes on epidemiology, clinical research, and omics, laboratory and trials to deliver insights, protocols, therapeutics and a vaccine that will mitigate at least some of the effects of COVID-19.

So far the speed of science has not been able to match the speed of COVID-19; the virus has always been ahead of us. As a consequence, the pressure for research methods to reach reliable, useful results is growing. In response, medical science is trying to expand its methodologies such as clinical trials and disease modeling, while doing no harm (or at least minimizing harm related to delays).

It is in this context that the workbenches being designed here need to be viewed. For example, emulated trials performed over time may be able to produce early insights and be one basis for so-called sentinel medical science. At the same time, privacy may be protected with anonymization of data [COVID-19 data privacy, work in progress] including choice of tracking apps with minimal and time-limited centralized data retention [WHO (2) 2020].

With appropriate data sets, CDMs are able to narrowly focus on specific concerns such as congregate care facilities, and/or zoom out to discern differences among communities with different demographics, sources of employment, housing, environmental quality, etc. Recognizing and addressing differences in viral transmission and disease severity, often among more vulnerable populations, is key to protecting public health regionally, nationally and ultimately globally.

Finally, along the way to sentinel medical science that does no harm, it remains to be seen whether innovations like full spectrum epidemiology are FAIRER (Findable, Accessible, Interoperable, Reusable, Ethical and Reproducible). Besides doing no harm, FAIRER is the second test sentinel medical science innovations like emulated trials need to pass. Indeed, there have been several FAIR assessments of a CDM called the Observational Medical Outcomes Partnership (OMOP) with EHRs. A European Health Data and Evidence Network (EHDEN) initiative based on the FAIR assessment attempts to harmonize 100 million health records [van Bochove K, Vos E, van Winzu A, Kurps J, Moinat M. 2020]. A recent EHDEN initiative is earmarked to use the OMOP CDM for COVID-19 research for registry and cohort data alongside EHRs and hospital information [The EHDEN Consortium. 2020].

AUTHOR ROLES

All authors accept responsibility for the content of the article

The same author may have multiple roles, and multiple authors may share a single role.

Conceptualization: GM MS RN CCA. Methodology: GM HT. Investigation: MC. Formal analysis: CCA. Validation: CCA MC. Writing: JG MS. Bibliographic review and analysis: MC AW. Visualization: JG

REFERENCES

- Atwood CP. Activity-Based Intelligence: Revolutionizing Military Intelligence Analysis. *Joint Force Quarterly*. 2015; 77 (2nd Quarter, April 2015)
- Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000;342(25):1878-1886. doi:10.1056/NEJM200006223422506
- Blacketer M, Voss EA, Ryan PB. Applying the OMOP Common Data Model to Survey Data [Webpage]. 2015 https://www.ohdsi.org/web/wiki/lib/execute/fetch.php?media=resources:using_the_omop_cdm_with_survey_and_registry_data_v6.0.pdf.
- Bradley DT, Mansouri MA, Key F, et al. A systems approach to preventing and responding to COVID-19 [published online ahead of print, 2020 Apr 1]. *EClinicalMedicine* 2020;21. doi:10.1016/j.eclinm.2020.100325
- Bullock HE, Harlow LL, Mulaik SA. Causation issues in structural equation modeling research, *Structural Equation Modeling: A Multidisciplinary Journal*, 1994;1:3, 253-267, DOI: 10.1080/10705519409539977
- EpiCentro. Characteristics of COVID-19 patients dying in Italy. Accessed May 31, 2020. <https://www.epicentro.iss.it/en/coronavirus/sars-cov-2-analysis-of-deaths>
- Dagnino, S ;Macherone A. eds. *Unraveling the Exposome: A Practical View*. Cham: Springer International Publishing, 2019. <https://doi.org/10.1007/978-3-319-89321-1>.
- Garg S, Kim L, Whitaker M, et al. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States, March 1–30, 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:458–464. DOI: <http://dx.doi.org/10.15585/mmwr.mm6915e3>
- Garza M, Fiol GD, Tenenbaum J, et al. Evaluating common data models for use with a longitudinal community registry. *Journal of Biomedical Informatics*. 2018;64(12):333-341. DOI: <https://doi.org/10.1016/j.jbi.2016.10.016>
- Gershman B, Guo DP, Dahabreh IJ. Using observational data for personalized medicine when clinical trial evidence is limited. *Fertil Steril*. 2018;109(6):946-951. doi:10.1016/j.fertnstert.2018.04.005
- Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016;183(8):758-764. doi:10.1093/aje/kwv254
- Hernán MA, Robins JM (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.
- Heindel JJ, Balbus J, Birnbaum L, Brune-Drisse MN, Grandjean P, Gray K, Landrigan PJ, et al. Developmental Origins of Health and Disease: Integrating Environmental Influences. *Endocrinol*. 2015;156(10): 3416–21. <https://doi.org/10.122015-139410/EN>.
- Kurt OK, Zhang J, Pinkerton KE. Pulmonary health effects of air pollution. *Curr Opin Pulm Med*. 2016;22(2):138-143. doi:10.1097/MCP.0000000000000248
- Liu, K; Fang, Y-Y; Deng, Y; Liu, Wei; Wang, M-F; Ma, J-P; Xiao, W; Wang, Y-N; Zhong, M-H; Li, C-H; Li, G-C; Liu, H-G. Clinical characteristics of novel coronavirus cases in tertiary hospitals in Hubei Province, Chinese Medical Journal: May 5, 2020 - Volume 133 - Issue 9 - p 1025-1031 doi:10.1097/CM9.0000000000000744

Marshall L, Weir E, Abelsohn A, Sanborn MD. Identifying and Managing Adverse Environmental Health Effects: 1. Taking an Exposure History. *CMAJ* 2002;166(8): 1049–55.

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC100881/>

Nii-Trebi NI. Emerging and Neglected Infectious Diseases: Insights, Advances, and Challenges. *Biomed Res Int*. 2017;2017:5245021. doi:10.1155/2017/5245021

Observational Health Data Sciences and Informatics (OHDSI). OMOP Common Data Model. [Webpage]. 2015; <http://www.ohdsi.org/data-standardization/the-common-data-model/>, 20 Jul 2015.

Porta, M. A Dictionary of Epidemiology (6 ed). Oxford University Press. 2014. DOI: 10.1093/acref/9780199976720.001.0001

Poole D, Crowley M. Cyclic Causal Models with Discrete Variables : Markov Chain Equilibrium Semantics and Sample Ordering. *IJCAI International Joint Conference on Artificial Intelligence*. 2013.

Schmidt CO, Nagrani R, Löbe M, et al. Interoperable COVID-19 epidemiological surveillance: Clinical and Population-based instruments [published online ahead of print, 2020 June 8]

The EHDEN Consortium. Data Partner Rapid Collaboration Call on COVID-19. V 1.4, April 15 2020

van Bochove K, Vos E, van Winzu A, Kurps J, Moinat M. Implementing FAIR in OHDSI and EHDEN. <https://www.ohdsi.org/wp-content/uploads/2020/05/Implementing-FAIR-in-OHDSI.pdf>. The Hyve, Utrecht, The Netherlands. 2020

WHO. EPI-WIN: WHO information network for epidemics [Webpage]. 2020; <https://www.who.int/teams/risk-communication>

RDA-COVID19-Epidemiology-WG Bibliography

- 1Point3Acres. (2019). *COVID-19/Coronavirus Real Time Updates With Credible Sources in US and Canada* | 1Point3Acres. <https://coronavirus.1point3acres.com/en>
- Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *JOURNAL OF MEDICAL INTERNET RESEARCH*, 22(4). <https://doi.org/10.2196/19016>
- Access Now. (2020). *Recommendations on privacy and data protection in the fight against COVID-19*. <https://www.accessnow.org/cms/assets/uploads/2020/03/Access-Now-recommendations-on-Covid-and-data-protection-and-privacy.pdf>
- Aguilar-Gallegos, N., Romero-García, L. E., Martínez-González, E. G., García-Sánchez, E. I., & Aguilar-Ávila, J. (2020). Dataset on dynamics of Coronavirus on Twitter. *Data in Brief*, 105684. <https://doi.org/10.1016/j.dib.2020.105684>
- AI challenge with AI2, CZI, MSR, Georgetown, NIH, & The White House. (2020). *COVID-19 Open Research Dataset Challenge (CORD-19)*. <https://kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>
- Aitsi-Selmi, A., & Murray, V. (2016). Protecting the Health and Well-being of Populations from Disasters: Health and Health Care in The Sendai Framework for Disaster Risk Reduction 2015-2030. *Prehospital and Disaster Medicine*, 31(1), 74–78. Cambridge Core. <https://doi.org/10.1017/S1049023X15005531>
- Allocati, N., Petrucci, A. G., Di Giovanni, P., Masulli, M., Di Ilio, C., & De Laurenzi, V. (2016). Bat–man disease transmission: Zoonotic pathogens from wildlife reservoirs to human populations. *Cell Death Discovery*, 2(1), 1–8. <https://doi.org/10.1038/cddiscovery.2016.48>
- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine*, 26(4), 450–452. <https://doi.org/10.1038/s41591-020-0820-9>
- Apple. (2020). *COVID-19—Mobility Trends Reports*. AppleMaps. <https://www.apple.com/covid19/mobility>
- Bai, Z., Gong, Y., Tian, X., Cao, Y., Liu, W., & Li, J. (2020). The Rapid Assessment and Early Warning Models for COVID-19. *Virologica Sinica*, 1–8. <https://doi.org/10.1007/s12250-020-00219-0>
- Bajwa, S. J. S., Sarna, R., Bawa, C., & Mehdiratta, L. (2020). Peri-operative and critical care concerns in coronavirus pandemic. *INDIAN JOURNAL OF ANAESTHESIA*, 64(4), 267–274. https://doi.org/10.4103/ija.IJA_272_20
- Barton, C. M., Alberti, M., Ames, D., Atkinson, J.-A., Bales, J., Burke, E., Chen, M., Diallo, S. Y., Earn, D. J. D., Fath, B., Feng, Z., Gibbons, C., Hammond, R., Heffernan, J., Houser, H., Hovmand, P. S., Kopainsky, B., Mabry, P. L., Mair, C., ... Tucker, G. (2020). Call for transparency of COVID-19 models. *Science*, 368(6490), 482.2-483. <https://doi.org/10.1126/science.abb8637>
- Battegay, M., Kuehl, R., Tschudin-Sutter, S., Hirsch, H. H., Widmer, A. F., & Neher, R. A. (2020). 2019-novel Coronavirus (2019-nCoV): Estimating the case fatality rate – a word of caution. *Swiss Medical Weekly*, 150(0506). <https://doi.org/10.4414/smw.2020.20203>
- Berry, I., Soucy, J.-P. R., Tuite, A., & Fisman, D. (2020). Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. *Canadian Medical Association Journal*, 192(15), E420–E420. <https://doi.org/10.1503/cmaj.75262>
- Blacketer, M. S., Voss, E. A., & Ryan, P. B. (2013). Applying the OMOP Common Data Model to Survey Data. *Medical Care*, S45–S52. https://www.ohdsi.org/web/wiki/lib/exe/fetch.php?media=resources:using_the_omop_cdm_with_survey_and_registry_data_v6.0.pdf

- Bradley, D. T., Mansouri, M. A., Kee, F., & Garcia, L. M. T. (2020). A systems approach to preventing and responding to COVID-19. *EClinicalMedicine*, 21. <https://doi.org/10.1016/j.eclinm.2020.100325>
- Brickley, D., et al., & et al. (2020). *Organization of Schemas*. <https://schema.org/docs/schemas.html>
- Bullock, H. E., Harlow, L. L., & Mulaik, S. A. (2009). Causation Issues in Structural Equation Modeling Research. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(3), 253–267. Scopus. <https://doi.org/10.1080/10705519409539977>
- Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V., & Schountz, T. (2006). Bats: Important Reservoir Hosts of Emerging Viruses. *Clinical Microbiology Reviews*, 19(3), 531–545. <https://doi.org/10.1128/CMR.00017-06>
- CDC. (2018, December 4). *National Pandemic Strategy*. <https://www.cdc.gov/flu/pandemic-resources/national-strategy/index.html>
- CDC. (2020a). *Cases of COVID19 in the U.S.* [Dataset]. Centers for Disease Control, USA. <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html>
- CDC. (2020b). *Weekly provisional death counts by select demographic and geographic characteristics* [Dataset]. Center for Disease Control, USA. https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm
- CDC. (2020c). *Weekly provisional death counts from death certificate data: COVID19, pneumonia, flu* [Dataset]. Center for Disease Control, USA. <https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm>
- CDC. (2020d, March 24). *Public Health and Promoting Interoperability Programs (formerly, known as Electronic Health Records Meaningful Use)*. <https://www.cdc.gov/ehrmeaningfuluse/introduction.html>
- CDC. (2020e). *Person Under Investigation (PUI) and Case Report F.pdf*. 2. https://www.phenxtoolkit.org/toolkit_content/PDF/CDC_PUI.pdf
- CDISC. (2020). *CDISC Standards in the Clinical Research Process* [Text]. CDISC. <https://www.cdisc.org/standards>
- Chan, A. T., Drew, D. A., Nguyen, L. H., Joshi, A. D., Ma, W., Guo, C.-G., Lo, C.-H., Mehta, R. S., Kwon, S., Sikavi, D. R., Magicheva-Gupta, M. V., Fatehi, Z. S., Flynn, J. J., Leonardo, B. M., Albert, C. M., Andreotti, G., Beane Freeman, L. E., Balasubramanian, B. A., Brownstein, J. S., ... Spector, T. (2020). The Coronavirus Pandemic Epidemiology (COPE) Consortium: A Call to Action. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*. <https://doi.org/10.1158/1055-9965.EPI-20-0606>
- Chan, J. F.-W., To, K. K.-W., Tse, H., Jin, D.-Y., & Yuen, K.-Y. (2013). Interspecies transmission and emergence of novel viruses: Lessons from bats and birds. *Trends in Microbiology*, 21(10), 544–555. <https://doi.org/10.1016/j.tim.2013.05.005>
- Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X., & Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *The Lancet*, 395(10223), 507–513. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7)
- Cheng, V. C. C., Wong, S.-C., Chen, J. H. K., Yip, C. C. Y., Chuang, V. W. M., Tsang, O. T. Y., Sridhar, S., Chan, J. F. W., Ho, P.-L., & Yuen, K.-Y. (2020). Escalating infection control response to the rapidly evolving epidemiology of the coronavirus disease 2019 (COVID-19) due to SARS-CoV-2 in Hong Kong. *INFECTION CONTROL AND HOSPITAL EPIDEMIOLOGY*, 41(5), 493–498. <https://doi.org/10.1017/ice.2020.58>

- Chowdhury, R., Heng, K., Shawon, M. S. R., Goh, G., Okonofua, D., Ochoa-Rosales, C., Gonzalez-Jaramillo, V., Bhuiya, A., Reidpath, D., Prathapan, S., Shahzad, S., Althaus, C. L., Gonzalez-Jaramillo, N., Franco, O. H., & The Global Dynamic Interventions Strategies for COVID-19 Collaborative Group. (2020). Dynamic interventions to control COVID-19 pandemic: A multivariate prediction modelling study comparing 16 worldwide countries. *European Journal of Epidemiology*.
<https://doi.org/10.1007/s10654-020-00649-w>
- COAR. (2020). *Recommendations for COVID-19 resources in repositories*. Confederation of Open Access Repositories. <https://www.coar-repositories.org/news-updates/covid19-recommendations/>
- Coccia, M. (2020). *Two mechanisms for accelerated diffusion of COVID-19 outbreaks in regions with high intensity of population and polluting industrialization: The air pollution-to-human and human-to-human transmission dynamics*. Cold Spring Harbor Laboratory Press.
<https://www.medrxiv.org/content/10.1101/2020.04.06.20055657v1>
- CODATA, C. on D. of the I. S. C., CODATA International Data Policy Committee, CODATA and CODATA China High-level International Meeting on Open Research Data Policy and Practice, Hodson, S., Mons, B., Uhler, P., & Zhang, L. (2019). *The Beijing Declaration on Research Data*.
<https://doi.org/10.5281/zenodo.3552330>
- CODATA, R. (2017). *International Research Data Management glossary (IRiDium)*.
<https://codata.org/initiatives/working-groups/standard-glossary-for-research-data-management-iridium/>
- COVID19India. (2020). *Coronavirus in India: Latest Map and Case Count*. <https://www.covid19india.org>
- Datacovid. (2020). *Barometer Covid19*. <https://datacovid.org/>
- Davis, L. (2020). *Corona Data Scraper* [HTML]. COVID Atlas.
<https://github.com/covidatlas/coronadatascraper> (Original work published 2020)
- DDI Alliance. (2020). *Data Documentation Initiative*. <https://ddialliance.org/>
- DeBord, D. G., Carreon, T., & Lentz, T. (2016). Use of the “Exposome” in the Practice of Epidemiology: A Primer on -Omic Technologies. *Am J Epidemiology*, 184(4), 312–314.
<https://doi.org/10.1093/aje/kwv325>
- Djalante, R., Shaw, R., & DeWit, A. (2020). Building resilience against biological hazards and pandemics: COVID-19 and its implications for the Sendai Framework. *Progress in Disaster Science*, 6, 100080.
<https://doi.org/10.1016/j.pdisas.2020.100080>
- Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet. Infectious Diseases*, 20(5), 533–534. PubMed. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
- Drew, D. A., Nguyen, L. H., Steves, C. J., Menni, C., Freydin, M., Varsavsky, T., Sudre, C. H., Cardoso, M. J., Ourselin, S., Wolf, J., Spector, T. D., Chan, A. T., & COPE Consortium. (2020). Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science (New York, N.Y.)*.
<https://doi.org/10.1126/science.abc0473>
- DSI, D. S., CEF eHealth. (2020, March 24). *EHDSI INTEROPERABILITY SPECIFICATIONS, Requirements and Frameworks (normative artefacts)—EHealth DSI Operations—CEF Digital*.
<https://ec.europa.eu/cefdigital/wiki/pages/viewpage.action?pageId=35210463>
- Duncan, G. T., Elliot, M., & Salazar, G. J. J. (2011). *Statistical Confidentiality: Principles and Practice*. Springer-Verlag. <https://doi.org/10.1007/978-1-4419-7802-8>
- Duncan, M. A., Drociuk, D., Belflower-Thomas, A., Van Sickle, D., Gibson, J. J., Youngblood, C., & Daley, W. R. (2011). Follow-Up Assessment of Health Consequences after a Chlorine Release from a Train

- Derailment-Graniteville, SC, 2005. *Journal of Medical Toxicology*, 7(1), 85–91. Scopus. <https://doi.org/10.1007/s13181-010-0130-6>
- ECDC. (2019). *The European Surveillance System (TESSy)*. European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/publications-data/european-surveillance-system-tesy>
- ECDC. (2020a). *COVID-19 situation update worldwide, as of 9 June 2020*. European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>
- ECDC. (2020b, April 15). *Geographic distribution of COVID-19 cases worldwide*. <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>
- ESIP. (2019). *Data Citation Guidelines for Earth Science Data , Version 2*. Earth Science Information Partners. <https://doi.org/10.6084/m9.figshare.8441816.v1>
- EU. (2019). *EHealth Network Guidelines to the EU Member States and the European Commission on an interoperable eco-system for digital health and investment programmes for a new/updated generation of digital infrastructure in Europe*, ev_20190611_co922_en.pdf. EU EHealth Network. https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20190611_co922_en.pdf
- European Commission. (2020). *Pseudonymisation tool*. EUPID - European Platform on Rare Disease Registration. https://eu-rd-platform.jrc.ec.europa.eu/node/2_en
- COMMISSION RECOMMENDATION (EU) 2020/518 of 8 April 2020 on a common Union toolbox for the use of technology and data to combat and exit from the COVID-19 crisis, in particular concerning mobile applications and the use of anonymised mobility data, (2020). https://ec.europa.eu/info/sites/info/files/recommendation_on_apps_for_contact_tracing_4.pdf
- Exaptive. (2020). *Cognitive City: COVID-19 Resources*. Exaptive, and the Bill and Melinda Gates Foundation. <https://covid-19.cognitive.city/cognitive/community/resource-gallery>
- FAIR4Health. (2020). *FAIR4Health at RDA Germany Conference 2020—Resources*. FAIR4Health Consortium. <https://www.fair4health.eu/en/resources>
- Fauci, A. S., & Morens, D. M. (2012). The Perpetual Challenge of Infectious Diseases. *New England Journal of Medicine*, 366(5), 454–461. <https://doi.org/10.1056/NEJMr1108296>
- FDA. (2019). *Sentinel Common Data Model | Sentinel Initiative*. FDA Sentinel Initiative. <https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model>
- Finnie, T., South, A., & Bento, A. (2016). EpiJSON: A unified data-format for epidemiology. *Epidemics*, 15(June, 2016), 20–26. <https://doi.org/10.1016/j.epidem.2015.12.002>
- FitzHenry, F., Resnic, F. S., Robbins, S. L., Denton, J., Nookala, L., Meeker, D., Ohno-Machado, L., & Matheny, M. E. (2015). Creating a Common Data Model for Comparative Effectiveness with the Observational Medical Outcomes Partnership. *Applied Clinical Informatics*, 6(3), 536–547. <https://doi.org/10.4338/ACI-2014-12-CR-0121>
- French COVID-19. (2020, March 12). COVID-19 funding opportunities. *Reacting*. <https://reacting.inserm.fr/covid-19-funding-opportunities/>
- Frontera, A., Martin, C., Vlachos, K., & Sgubin, G. (2020). Regional air pollution persistence links to COVID-19 infection zoning. *The Journal of Infection*. <https://doi.org/10.1016/j.jinf.2020.03.045>
- Gates, B. (2020). Responding to Covid-19—A once-in-a-century pandemic? *New England Journal of Medicine*, 382(18), 1677–1679. Scopus. <https://doi.org/10.1056/NEJMp2003762>
- German Data Forum (RatSWD). (2020). Remote Access to data from official statistics agencies and social security agencies. *RatSWD Output Paper Series*. <https://www.ratswd.de/en/publication/output-series/2855>

- German National Cohort. (2020, March 31). *NAKO Gesundheitsstudie—Kontakt*.
<https://nako.de/allgemeines/kontakt/>
- Gershman, B., Guo, D. P., & Dahabreh, I. J. (2018). Using observational data for personalized medicine when clinical trial evidence is limited. *Fertility and Sterility*, 109(6), 946–951.
<https://doi.org/10.1016/j.fertnstert.2018.04.005>
- GESIS Panel Team. (2020). *GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany*GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany (Version 1.1.0) [Data set]. GESIS Data Archive. <https://doi.org/10.4232/1.13520>
- Ghani, A. C., Donnelly, C. A., Cox, D. R., Griffin, J. T., Fraser, C., Lam, T. H., Ho, L. M., Chan, W. S., Anderson, R. M., Hedley, A. J., & Leung, G. M. (2005). Methods for Estimating the Case Fatality Ratio for a Novel, Emerging Infectious Disease. *American Journal of Epidemiology*, 162(5), 479–486.
<https://doi.org/10.1093/aje/kwi230>
- GHRU. (2020, March 1). *GHRU - COVID Questionnaire—V6.docx*. Dropbox.
<https://www.dropbox.com/s/auvvey4utibd85s/GHRU%20-%20COVID%20Questionnaire%20-%20v6.docx?dl=0>
- Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Filippo, A. D., Matteo, A. D., & Colaneri, M. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 1–6. <https://doi.org/10.1038/s41591-020-0883-7>
- GLOPID, Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2018). *Principles of data sharing in public health emergencies*. <https://www.glopid-r.org/wp-content/uploads/2018/06/glopid-r-principles-of-data-sharing-in-public-health-emergencies.pdf>
- Goldberg, M., & Villeneuve, P. (2020). *Air pollution, COVID-19 and death: The perils of bypassing peer review*. The Conversation. <http://theconversation.com/air-pollution-covid-19-and-death-the-perils-of-bypassing-peer-review-136376>
- Google. (2020a). *Google Dataset Search*. <https://datasetsearch.research.google.com/>
- Google. (2020b). *Google LLC ‘COVID-19 Community Mobility Report’*. COVID-19 Community Mobility Report. <https://www.google.com/covid19/mobility?hl=en>
- Government of the Democratic Republic of the Congo, & World Health Organization. (2019). *Strategic_response_plan.pdf*.
https://www.un.org/ebolaresponsedrc/sites/www.un.org.ebolaresponsedrc/files/strategic_response_plan.pdf
- Grange, E. S., Neil, E. J., Stoffel, M., Singh, A. P., Tseng, E., Resco-Summers, K., Fellner, B. J., Lynch, J. B., Mathias, P. C., Mauritz-Miller, K., Sutton, P. R., & Leu, M. G. (2020). Responding to COVID-19: The UW Medicine Information Technology Services Experience. *APPLIED CLINICAL INFORMATICS*, 11(2), 265–275. <https://doi.org/10.1055/s-0040-1709715>
- Hare, S. S., Rodrigues, J. C. L., Jacob, J., Edey, A., Devaraj, A., Johnstone, A., McStay, R., Nair, A., & Robinson, G. (2020). A UK-wide British Society of Thoracic Imaging COVID-19 imaging repository and database: Design, rationale and implications for education and research. *Clinical Radiology*, 75(5), 326–328. <https://doi.org/10.1016/j.crad.2020.03.005>
- HCSRN. (2019). *VDW Data Model*. Healthcare Systems Research Network.
<http://www.hcsrn.org/en/Tools%20&%20Materials/VDW/>

- Healy, K. (2020). *Rpackage (covdata)—COVID19 Case and Mortality Time Series*.
<https://khealy.github.io/covdata>
- Hernán, M. A., & Robins, J. M. (2016). Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *American Journal of Epidemiology*, 183(8), 758–764.
<https://doi.org/10.1093/aje/kwv254>
- HL7. (2010). *HL7 Standards Product Brief—CDA® Release 2 | HL7 International*.
http://www.hl7.org/implement/standards/product_brief.cfm?product_id=7
- HL7. (2019a, November 1). *Fast Healthcare Interoperability Resources Standard*. HL7®.
<https://www.hl7.org/fhir/>
- HL7. (2019b, November 1). *Summary—FHIR v4.0.1*. <https://hl7.org/FHIR/summary.html>
- Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., Diaz, G., Cohn, A., Fox, L., Patel, A., Gerber, S. I., Kim, L., Tong, S., Lu, X., Lindstrom, S., ... Invest, W. S. C. (2020). First Case of 2019 Novel Coronavirus in the United States. In *NEW ENGLAND JOURNAL OF MEDICINE* (Vol. 382, Issue 10, pp. 929–936). MASSACHUSETTS MEDICAL SOC. <https://doi.org/10.1056/NEJMoa2001191>
- Homeland Security Council. (2006, May). *Pandemic-influenza-implementation.pdf*.
<https://www.cdc.gov/flu/pandemic-resources/pdf/pandemic-influenza-implementation.pdf>
- Hughes, J. M., Wilson, M. E., Pike, B. L., Saylor, K. E., Fair, J. N., LeBreton, M., Tamoufe, U., Djoko, C. F., Rimo, A. W., & Wolfe, N. D. (2010). The Origin and Prevention of Pandemics. *Clinical Infectious Diseases*, 50(12), 1636–1640. <https://doi.org/10.1086/652860>
- IHME. (2020). *Global Health Data Exchange | GHDx*. Institute for Health Metrics and Evaluation (IHME), University of Washington. <http://ghdx.healthdata.org/>
- International Organization for Standardization. (2015). *ISO/TS 17975:2015*. ISO.
<https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/11/61186.html>
- International Severe Acute Respiratory and Emerging Infection Consortium, & World Health Organization. (2020a, March 24). *ISARIC_COVID-19_RAPID_CRF_24MAR20_EN.pdf*.
https://media.tghn.org/medialibrary/2020/04/ISARIC_COVID-19_RAPID_CRF_24MAR20_EN.pdf
- International Severe Acute Respiratory and Emerging Infection Consortium, & World Health Organization. (2020b, April 23). *ISARIC_WHO_nCoV_CORE_CRF_23APR20.pdf*.
https://media.tghn.org/medialibrary/2020/05/ISARIC_WHO_nCoV_CORE_CRF_23APR20.pdf
- ISARIC. (2020). *COVID-19 Clinical research resources*. International Severe Acute Respiratory and Emerging Infection Consortium - The Global Health Network.
<https://infograph.venngage.com/pe/Mxg90X1jTc?border=false>
- Italian Civil Protection Department, Morettini, M., Sbrolini, A., Marcantoni, I., & Burattini, L. (2020). COVID-19 in Italy: Dataset of the Italian Civil Protection Department. *Data in Brief*, 30, 105526.
<https://doi.org/10.1016/j.dib.2020.105526>
- Ivanov, D. (2020). Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *TRANSPORTATION RESEARCH PART E-LOGISTICS AND TRANSPORTATION REVIEW*, 136.
<https://doi.org/10.1016/j.tre.2020.101922>
- Jackson, J. K., Weiss, M. A., Schwarzenberg, A. B., & Nelson, R. M. (2020). Global Economic Effects of COVID-19. *Congressional Research Service*, 2020-05–15.
<https://crsreports.congress.gov/product/pdf/R/R46270>

- Jee, Y. (2020). WHO International Health Regulations Emergency Committee for the COVID-19 outbreak. *EPIDEMIOLOGY AND HEALTH*, 42. <https://doi.org/10.4178/epih.e2020013>
- JHU. (2020). *COVID19 dataset* [Data repository]. Johns Hopkins University, CSSEGISandData. <https://github.com/CSSEGISandData/COVID-19> (Original work published 2020)
- Jiménez, R. C., Kuzak, M., Alhamdoosh, M., Barker, M., Batut, B., Borg, M., Capella-Gutierrez, S., Chue Hong, N., Cook, M., Corpas, M., Flannery, M., Garcia, L., Gelpí, J. Ll., Gladman, S., Goble, C., González Ferreiro, M., Gonzalez-Beltran, A., Griffin, P. C., Grüning, B., ... Crouch, S. (2017). Four simple recommendations to encourage best practices in research software. *F1000Research*, 6, 876. <https://doi.org/10.12688/f1000research.11407.1>
- Kherbst, & Eehlers. (2020, April 27). *COVID-19 Screening Form 2020.pdf*. Dropbox. <https://www.dropbox.com/s/iuhe4366msdfq5c/COVID-19%20Screening%20Form%202020.pdf?dl=0>
- Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H., & Lipsitch, M. (2020). Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, 368(6493), 860–868. <https://doi.org/10.1126/science.abb5793>
- Knight, G., Dharan, N., & Fox, G. (2016). Bridging the gap between evidence and policy for infectious diseases: How models can aid public health decision-making. *Int J Infect Dis.*, 42, 17–23. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4996966/>
- Kurt, O. K., Zhang, J., & Pinkerton, K. E. (2016). Pulmonary health effects of air pollution. *Current Opinion in Pulmonary Medicine*, 22(2), 138–143. PubMed. <https://doi.org/10.1097/MCP.0000000000000248>
- Kushida, C. A., Nichols, D. A., Jadrnicek, R., Miller, R., Walsh, J. K., & Griffin, K. (2012). Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. *Medical Care*, 50(Suppl), S82-101. <https://doi.org/10.1097/MLR.0b013e3182585355>
- Laine, J. E., & Robinson, O. (2019). Framing Fetal and Early Life Exposome Within Epidemiology. In S. Dagnino & A. Macherone (Eds.), *Unraveling the Exposome: A Practical View* (pp. 87–123). Springer International Publishing. https://doi.org/10.1007/978-3-319-89321-1_4
- Lamprecht, A.-L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., Dominguez Del Angel, V., van de Sandt, S., Ison, J., Martinez, P. A., McQuilton, P., Valencia, A., Harrow, J., Psomopoulos, F., Gelpi, J. L., Chue Hong, N., Goble, C., & Capella-Gutierrez, S. (2019). Towards FAIR principles for research software. *Data Science, Preprint*, 1–23. <https://doi.org/10.3233/DS-190026>
- LSRI. (2020). *LSRI Response to COVID-19*. European Life Science Research Infrastructure. <https://lifescience-ri.eu/ls-ri-response-to-covid-19.html>
- Luis, A. D., Hayman, D. T. S., O’Shea, T. J., Cryan, P. M., Gilbert, A. T., Pulliam, J. R. C., Mills, J. N., Timonin, M. E., Willis, C. K. R., Cunningham, A. A., Fooks, A. R., Rupprecht, C. E., Wood, J. L. N., & Webb, C. T. (2013). A comparison of bats and rodents as reservoirs of zoonotic viruses: Are bats special? *Proceedings of the Royal Society B: Biological Sciences*, 280(1756), 20122753. <https://doi.org/10.1098/rspb.2012.2753>
- Luke, D. A., & Stamatakis, K. A. (2012). *Systems science methods in public health: Dynamics, networks, and agents*. 33, 357–376. <https://doi.org/10.1146/annurev-publhealth-031210-101222>
- Mahmood, S., Hasan, K., Carras, M. C., & Labrique, A. (2020). Global Preparedness Against COVID-19: We Must Leverage the Power of Digital Health. *JMIR PUBLIC HEALTH AND SURVEILLANCE*, 6(2), 226–232. <https://doi.org/10.2196/18980>
- Martelletti, L., & Martelletti, P. (2020). Air Pollution and the Novel Covid-19 Disease: A Putative Disease Risk Factor. *Sn Comprehensive Clinical Medicine*, 1–5. <https://doi.org/10.1007/s42399-020-00274-4>

- Mavragani, A. (2020). Tracking COVID-19 in Europe: Infodemiology Approach. *JMIR PUBLIC HEALTH AND SURVEILLANCE*, 6(2), 233–245. <https://doi.org/10.2196/18941>
- McMichael, T. M., Currie, D. W., Clark, S., Pogojans, S., Kay, M., Schwartz, N. G., Lewis, J., Baer, A., Kawakami, V., Lukoff, M. D., Ferro, J., Brostrom-Smith, C., Rea, T. D., Sayre, M. R., Riedo, F. X., Russell, D., Hiatt, B., Montgomery, P., Rao, A. K., ... Duchin, J. S. (2020). Epidemiology of Covid-19 in a Long-Term Care Facility in King County, Washington. *The New England Journal of Medicine*. Scopus. <https://doi.org/10.1056/NEJMoa2005412>
- Morgan, D., & Sargent, J. F. (2020). *Effects of COVID-19 on the Federal Research and Development Enterprise* (No. R46309; CRS Reports, p. 22). Congressional Research Service (CRS), United States Library of Congress. <https://crsreports.congress.gov/product/pdf/R/R46309>
- Munthali, G. N. C., & Xuelian, W. (2020). Covid-19 Outbreak on Malawi Perspective. *ELECTRONIC JOURNAL OF GENERAL MEDICINE*, 17(4). <https://doi.org/10.29333/ejgm/7871>
- National Institute of Allergy and Infectious Disease (NIAID). (2013). *Data Sharing and Release Guidelines*. <https://www.niaid.nih.gov/research/data-sharing-and-release-guidelines>
- NIH. (2020a). *ClinicalTrials—Listed clinical studies related to the coronavirus disease (COVID-19)*. U.S. National Institutes of Health - Information on Clinical Trials and Human Research Studies - National Library of Medicine. <https://clinicaltrials.gov/ct2/results?cond=COVID-19>
- NIH. (2020b). *NIH Public Health Emergency and Disaster Research Response (DR2) COVID-19 Research Tools—Training Material*. NIH Public Health Emergency and Disaster Research Response (DR2). <https://dr2.nlm.nih.gov/>
- NIH. (2020c). *COVID-19 OBSSR Research Tools*. National Institutes of Health. https://www.nlm.nih.gov/dr2/COVID-19_BSSR_Research_Tools.pdf
- NIST PWG. (2015). *Big Data Interoperability Framework: Volume 5, Architectures White Paper Survey* (NIST SP 1500-5; p. NIST SP 1500-5, version 3 (final)). National Institute of Standards and Technology, Big Data Public Working Group. <https://doi.org/10.6028/NIST.SP.1500-5>
- NIST PWG. (2019a). *Big Data Interoperability Framework: Volume 1, Definitions* (NIST SP 1500-1r2; p. NIST SP 1500-1r2, version 3 (final)). National Institute of Standards and Technology, Big Data Public Working Group. <https://doi.org/10.6028/NIST.SP.1500-1r2>
- NIST PWG. (2019b). *Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements* (NIST SP 1500-3r2; p. NIST SP 1500-3r2, version 3 (final)). National Institute of Standards and Technology, Big Data Public Working Group. <https://doi.org/10.6028/NIST.SP.1500-3r2>
- NIST PWG. (2019c). *Big Data Interoperability Framework: Volume 4, Security and Privacy* (NIST SP 1500-4r2; p. NIST SP 1500-4r2, version 3 (final)). National Institute of Standards and Technology, Big Data Public Working Group. <https://doi.org/10.6028/NIST.SP.1500-4r2>
- NIST PWG. (2019d). *Big Data Interoperability Framework: Volume 6, Reference Architecture* (NIST SP 1500-6r2; p. NIST SP 1500-6r2, version 3 (final)). National Institute of Standards and Technology, Big Data Public Working Group. <https://doi.org/10.6028/NIST.SP.1500-6r2>
- NIST PWG. (2019e). *Big Data Interoperability Framework: Volume 7, Standards Roadmap* (NIST SP 1500-7r2; p. NIST SP 1500-7r2, version 3 (final)). National Institute of Standards and Technology, Big Data Public Working Group. <https://doi.org/10.6028/NIST.SP.1500-7r2>
- NIST PWG. (2019f). *Big Data Interoperability Framework: Volume 8, Reference Architecture Interfaces* (NIST SP 1500-9r1; p. NIST SP 1500-9r1, version 3 (final)). National Institute of Standards and Technology, Big Data Public Working Group. <https://doi.org/10.6028/NIST.SP.1500-9r1>

- NIST PWG. (2019g). *Big Data Interoperability Framework: Volume 9, Adoption and Modernization* (NIST SP 1500-10r1; p. NIST SP 1500-10r1, version 3 (final)). National Institute of Standards and Technology, Big Data Public Working Group. <https://doi.org/10.6028/NIST.SP.1500-10r1>
- NIST PWG. (2019h). *Big Data Interoperability Framework: Volume 2, Big Data Taxonomies* (NIST SP 1500-2r2; p. NIST SP 1500-2r2, version 3 (final)). National Institute of Standards and Technology, Big Data Public Working Group. <https://doi.org/10.6028/NIST.SP.1500-2r2>
- NSW Health. (2020, March 23). *Novel-coronavirus-case-questionnaire.pdf*. <https://www.health.nsw.gov.au/Infectious/Forms/novel-coronavirus-case-questionnaire.pdf>
- NYC Health. (2020). *Nychealth/coronavirus-data*. NYC Department of Health and Mental Hygiene. <https://github.com/nychealth/coronavirus-data> (Original work published 2020)
- NYT. (2020, April 21). *Coronavirus (Covid-19) Data in the United States*. New York Times. <https://github.com/nytimes/covid-19-data> (Original work published 2020)
- OECD. (2019). *Recommendation of the Council on Health Data Governance*. <https://www.oecd.org/health/health-systems/Recommendation-of-OECD-Council-on-Health-Data-Governance-Booklet.pdf>
- Ogunyemi, O. I., Meeker, D., Kim, H.-E., Ashish, N., Farzaneh, S., & Boxwala, A. (2013). Identifying Appropriate Reference Data Models for Comparative Effectiveness Research (CER) Studies Based on Data from Clinical Information Systems. *MEDICAL CARE*, 51(8, 3), S45–S52. <https://doi.org/10.1097/MLR.0b013e31829b1e0b>
- OHDSI. (2019). *OMOP Common Data Model – OHDSI*. Observational Health Data Sciences and Informatics. <https://www.ohdsi.org/data-standardization/the-common-data-model/>
- OpenAIRE. (2020). *COVID-19 Open Research Gateway*. OpenAIRE - Connect. <https://beta.covid-19.openaire.eu/content>
- Oxford University. (2020a). *COVID19 dataset* [Dataset]. <https://github.com/owid/covid-19-data>
- Oxford University. (2020b). *COVID19 government response tracker* [Dataset]. University of Oxford. <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker>
- Panchadsaram, R., Davis, N., Wikelius, K., Shahpar, C., Cobb, L., Wosińska, M., Anderson, D., Brown, L. M., Hunt, D., & Goligorsky, D. (2020). *COVID exit strategy: How we reopen safely*. <https://www.covidexitstrategy.org/>
- Park, H. W., Park, S., & Chong, M. (2020). Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea. *JOURNAL OF MEDICAL INTERNET RESEARCH*, 22(5). <https://doi.org/10.2196/18897>
- Pathak, E. B., Salemi, J. L., Sobers, N., Menard, J., & Hambleton, I. R. (2020). COVID-19 in Children in the United States: Intensive Care Admissions, Estimated Total Infected, and Projected Numbers of Severe Pediatric Cases in 2020. *Journal of Public Health Management and Practice, Publish Ahead of Print*. <https://doi.org/10.1097/PHH.0000000000001190>
- PCORnet. (2020). *Patient-Centered Outcomes Research Institute*. The National Patient-Centered Clinical Research Network. <https://pcornet.org/>
- Peirlinck, M., Linka, K., Sahli Costabal, F., & Kuhl, E. (2020). Outbreak dynamics of COVID-19 in China and the United States. *BIOMECHANICS AND MODELING IN MECHANOBIOLOGY*. <https://doi.org/10.1007/s10237-020-01332-5>
- periCOVID Uganda CRF. (2020, April 10). *PeriCOVID Uganda CRF.docx*. Dropbox. <https://www.dropbox.com/s/1p32oudodv8bm1h/periCOVID%20Uganda%20CRF.docx?dl=0>
- PhenX. (2020). *PhenX Toolkit—COVID-19 Protocols*. <https://www.phenxtoolkit.org/covid19>

- Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L., & Lloyd-Smith, J. O. (2017). Pathways to zoonotic spillover. *Nature Reviews Microbiology*, 15(8), 502–510. <https://doi.org/10.1038/nrmicro.2017.45>
- Priyadarsini, S. L., & Suresh, M. (2020). Factors influencing the epidemiological characteristics of pandemic COVID 19: A TISM approach. *INTERNATIONAL JOURNAL OF HEALTHCARE MANAGEMENT*. <https://doi.org/10.1080/20479700.2020.1755804>
- RDA-COVID19-Epidemiology WG. (2020). *Data sharing in epidemiology—Supporting output*. <https://www.rd-alliance.org/group/rda-covid19-epidemiology/outcomes/data-sharing-epidemiology>
- Ricciardi, F., De Bernardi, P., & Cantino, V. (2020). System dynamics modeling as a circular process: The smart commons approach to impact management. *Technological Forecasting and Social Change*, 151(C). <https://ideas.repec.org/a/eee/tefoso/v151y2020ics0040162519310923.html>
- Rockefeller Foundation. (2020, May 19). *Call for Entries: Data Science Breakthroughs for an Inclusive Recovery*. The Rockefeller Foundation. <https://www.rockefellerfoundation.org/blog/call-for-entries-data-science-breakthroughs-for-an-inclusive-recovery/>
- Sansone, S.-A., Rocca-Serra, P., Field, D., Maguire, E., Taylor, C., Hofmann, O., Fang, H., Neumann, S., Tong, W., Amaral-Zettler, L., Begley, K., Booth, T., Bougueleret, L., Burns, G., Chapman, B., Clark, T., Coleman, L.-A., Copeland, J., Das, S., ... Hide, W. (2012). Toward interoperable bioscience data. *Nature Genetics*, 44(2), 121–126. <https://doi.org/10.1038/ng.1054>
- Saulnier, K. M., Bujold, D., Dyke, S. O. M., Dupras, C., Beck, S., Bourque, G., & Joly, Y. (2019). Benefits and barriers in the design of harmonized access agreements for international data sharing. *Scientific Data*, 6(1), 297. <https://doi.org/10.1038/s41597-019-0310-4>
- Semantic Scholar. (2020). *CORD-19*. <https://pages.semanticscholar.org/coronavirus-research>
- Setti, L., Passarini, F., Gennaro, G. D., Baribieri, P., Perrone, M. G., Borelli, M., Palmisani, J., Gilio, A. D., Torboli, V., Pallavicini, A., Ruscio, M., Piscitelli, P., & Miani, A. (2020). *SARS-Cov-2 RNA Found on Particulate Matter of Bergamo in Northern Italy: First Preliminary Evidence*. Cold Spring Harbor Laboratory Press. <https://www.medrxiv.org/content/10.1101/2020.04.15.20065995v2>
- SNOMED. (2020). *COVID-19 Data Coding using SNOMED CT - COVID-19 Guide—SNOMED Confluence*. <https://confluence.ihtsdotools.org/display/DOCCV19>
- Staunton, C., Slokenberga, S., & Mascalzoni, D. (2019). The GDPR and the research exemption: Considerations on the necessary safeguards for research biobanks. *European Journal of Human Genetics*, 27(8), 1159–1167. <https://doi.org/10.1038/s41431-019-0386-5>
- Sun, P., Lu, X., Xu, C., Sun, W., & Pan, B. (2020). Understanding of COVID-19 based on current evidence. *Journal of Medical Virology*, 92(6), 548–551. <https://doi.org/10.1002/jmv.25722>
- Team, G. P. (2020). *GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany*. <https://doi.org/10.4232/1.13520>
- Templ, M. (2015). Quality indicators for statistical disclosure methods: A case study on the structure of earnings survey. *Journal of Official Statistics*, 31(4), 737–761. Scopus. <https://doi.org/10.1515/JOS-2015-0043>
- Templ, M. (2017). *Statistical disclosure control for microdata: Methods and applications in R* (p. 287). Springer International Publishing; Scopus. <https://doi.org/10.1007/978-3-319-50272-4>
- Templ, Matthias, Meindl, B., & Kowarik, A. (2020). *sdcmicro: Statistical Disclosure Control Methods for Anonymization of Data and Risk Estimation*. <https://cran.r-project.org/web/packages/sdcMicro/index.html>

- Templ, Matthias, Meindl, B., Kowarik, A., & Chen, S. (2014). *Introduction to Statistical Disclosure Control (SDC)*. 25. <https://www.ihsn.org/sites/default/files/resources/ihsn-working-paper-007-Oct27.pdf>
- The Atlantic. (2020). *COVID Tracking Project* [Datasets]. <https://covidtracking.com/>
- The European Data Protection Board. (2020). *Guidelines in the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak*. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202003_healthdatascientificresearchcovid19_en.pdf
- The European Health Data & Evidence Network. (2020, April 15). *COVID-19-Data-Partner-Call-Description-v1.4.pdf*. <https://www.ehden.eu/wp-content/uploads/2020/04/COVID-19-Data-Partner-Call-Description-v1.4.pdf>
- The National Institute of Environmental Health Science. (2020, May 14). *COVID-19_BSSR_Research_Tools.pdf*. https://www.nlm.nih.gov/dr2/COVID-19_BSSR_Research_Tools.pdf
- UN. (2018). *The Humanitarian Exchange Language (HXL)*. United Nations, Office for the Coordination of Humanitarian Affairs (OCHA), Centre for Humanitarian Data. <https://hxlstandard.org/standard/1-1final/>
- UN. (2020). *The Humanitarian Data Exchange (HDX)*. United Nations, Office for the Coordination of Humanitarian Affairs (OCHA), Centre for Humanitarian Data. <https://data.humdata.org/>
- UNDRR. (2020). *Disaster risk management for health: Overview*. <https://www.undrr.org/publication/disaster-risk-management-health-overview>
- Universidade Federal de Pelotas. (2020). *Brazil COVID serological survey questionnaire 20200428.docx*. Dropbox. <https://www.dropbox.com/s/l9hhblb83ybr70u/Brazil%20COVID%20serological%20survey%20questionnaire%2020200428.docx?dl=0>
- University of Maryland. (2020, April 24). *COVID-19 Impact Analysis Platform*. COVID-19 Impact Analysis Platform. <https://data.covid.umd.edu/>
- University of Washington. (2020). *COVID19 data: Beoutbreakprepared* [Data repository]. <https://github.com/beoutbreakprepared>
- University of Washington - HGIS Lab. (2020). *Novel Coronavirus Infection Map*. University of Washington - Humanistic GIS Laboratory. <https://github.com/jakobzhao/virus>
- U.S. Department of Health and Human Services. (2017, December). *Pan-flu-report-2017v2.pdf*. <https://www.cdc.gov/flu/pandemic-resources/pdf/pan-flu-report-2017v2.pdf>
- van Bochove, K., Vos, E., van Winzum, A., Kurps, J., & Moinat, M. (2020). *Implementing FAIR in OHDSI: Challenges and opportunities for EHDEN*. 1.
- Voss, E. A., Makadia, R., Matcho, A., Ma, Q., Knoll, C., Schuemie, M., DeFalco, F. J., Londhe, A., Zhu, V., & Ryan, P. B. (2015). Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION*, 22(3), 553–564. <https://doi.org/10.1093/jamia/ocu023>
- Wang, L. L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., Funk, K., Kinney, R., Liu, Z., Merrill, W., Mooney, P., Murdick, D., Rishi, D., Sheehan, J., Shen, Z., Stilson, B., Wade, A. D., Wang, K., Wilhelm, C., ... Kohlmeier, S. (2020). CORD-19: The Covid-19 Open Research Dataset. *ArXiv:2004.10706 [Cs]*. <http://arxiv.org/abs/2004.10706>
- Warren, M. (2020, April 20). *CDISC Interim User Guide for COVID-19—CDISC Interim User Guide for COVID-19—Wiki*. <https://wiki.cdisc.org/display/COVID19/CDISC+Interim+User+Guide+for+COVID-19>

- Webster, R. G. (2004). Wet markets—A continuing source of severe acute respiratory syndrome and influenza? *The Lancet*, 363(9404), 234–236. [https://doi.org/10.1016/S0140-6736\(03\)15329-9](https://doi.org/10.1016/S0140-6736(03)15329-9)
- Wellcome. (2020, April 23). *Final UK Covid Questionnaire_23 April.pdf*. Dropbox. https://www.dropbox.com/s/hy6jdpvgvkpgxfi/Final%20UK%20Covid%20Questionnaire_23%20April.pdf?dl=0
- Wellcome Trust. (2017). *Longitudinal Population Studies Strategy*. https://wellcome.ac.uk/sites/default/files/longitudinal-population-studies-strategy_0.pdf
- WHO. (2003). *Climate change and infectious diseases*. Climate Change and Human Health; World Health Organization. <https://www.who.int/globalchange/summary/en/index5.html>
- WHO. (2015, September 2). *Developing Global Norms for Sharing Data and Results during Public Health Emergencies*. World Health Organization; World Health Organization. http://www.who.int/medicines/ebola-treatment/data-sharing_phe/en/
- WHO. (2020a). *About EPI-WIN*. <https://www.who.int/teams/risk-communication/about-epi-win>
- WHO. (2020b). *COVID-19 situation reports*. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
- WHO. (2020c). *Ethical considerations to guide the use of digital proximity tracking technologies for COVID-19 contact tracing*. <https://www.who.int/publications-detail/WHO-2019-nCoV-Ethics-Contact-tracing-apps-2020.1WHO>
- WHO. (2020d). *Novel Coronavirus (2019-nCoV) situation reports*. World Health Organization. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
- WHO. (2020e). *Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection*. <https://apps.who.int/iris/bitstream/handle/10665/332188/WHO-2019-nCoV-Seroepidemiology-2020.2-eng.pdf?sequence=1&isAllowed=y>
- WHO. (2020f). *Risk Communication: EPI-WIN*. <https://www.who.int/teams/team-preview/risk-communication---global>
- WHO. (2020g). *Survey tool and guidance: Behavioural insights on COVID-19*. WHO Regional Office for Europe. http://www.euro.who.int/_data/assets/pdf_file/0007/436705/COVID-19-survey-tool-and-guidance.pdf?ua=1
- WHO. (2020h). *WHO | Global Influenza Surveillance and Response System (GISRS)*. WHO; World Health Organization. http://www.who.int/influenza/gisrs_laboratory/en/
- WHO. (2020i, February). *COVID-19 CRF • ISARIC*. <https://isaric.tghn.org/COVID-19-CRF/>
- WHO. (2020j). *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)*. World Health Organization. [https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19))
- WHO. (2020k, March 20). *Global surveillance for COVID-19 caused by human infection with COVID-19 virus: Interim guidance*. <https://apps.who.int/iris/bitstream/handle/10665/331506/WHO-2019-nCoV-SurveillanceGuidance-2020.6-eng.pdf>
- WHO. (2020l, March 20). *Surveillance, rapid response teams, and case investigation*. Coronavirus Disease (COVID-19) Technical Guidance. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/surveillance-and-case-definitions>
- WHO. (2020m, March 23). *WHO Global COVID-19 Clinical Platform Case Record Form (CRF)*. [https://www.who.int/publications-detail/global-covid-19-clinical-platform-novel-coronavirus-\(covid-19\)-rapid-version](https://www.who.int/publications-detail/global-covid-19-clinical-platform-novel-coronavirus-(covid-19)-rapid-version)

- WHO. (2020n, April 29). *Modes of transmission of virus causing COVID-19: Implications for IPC precaution recommendations*. <https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations>
- WHO. (2020o, May 26). *Preparing GISRS for the upcoming influenza seasons during the COVID-19 pandemic – practical considerations*. https://apps.who.int/iris/bitstream/handle/10665/332198/WHO-2019-nCoV-Preparing_GISRS-2020.1-eng.pdf
- WHO. (2020p, May 29). *C-TAP: COVID-19 technology access pool*. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/covid-19-technology-access-pool>
- Woo, P. C., Lau, S. K., & Yuen, K. (2006). Infectious diseases emerging from Chinese wet-markets: Zoonotic origins of severe respiratory viral infections. *Current Opinion in Infectious Diseases*, 19(5), 401–407. <https://doi.org/10.1097/01.qco.0000244043.08264.fc>
- World Bank. (2020). *Understanding the Coronavirus (COVID-19) pandemic through data* [Datasets]. <http://datatopics.worldbank.org/universal-health-coverage/covid19/>
- Worldometer. (2020). *COVID19 data* [Dataset]. <https://www.worldometers.info/coronavirus/>
- Wu, D., Wu, T., Liu, Q., & Yang, Z. (2020). The SARS-CoV-2 outbreak: What we know. *International Journal of Infectious Diseases*, 94, 44–48. Scopus. <https://doi.org/10.1016/j.ijid.2020.03.004>
- Xu, B., Gutierrez, B., Mekaru, S., Sewalk, K., Goodwin, L., Loskill, A., Cohn, E. L., Hswen, Y., Hill, S. C., Cobo, M. M., Zarebski, A. E., Li, S., Wu, C.-H., Hulland, E., Morgan, J. D., Wang, L., O'Brien, K., Scarpino, S. V., Brownstein, J. S., ... Kraemer, M. U. G. (2020). Epidemiological data from the COVID-19 outbreak, real-time case information. *Scientific Data*, 7(1), 106. <https://doi.org/10.1038/s41597-020-0448-0>
- Yang, C., Qiu, X., Fan, H., Jiang, M., Lao, X., Zeng, Y., & Zhang, Z. (2020). Coronavirus disease 2019: Reassembly attack of coronavirus. *INTERNATIONAL JOURNAL OF ENVIRONMENTAL HEALTH RESEARCH*. <https://doi.org/10.1080/09603123.2020.1747602>
- Yasaka, T. M., Lehigh, B. M., & Sahyouni, R. (2020). Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App. *JMIR MHEALTH AND UHEALTH*, 8(4). <https://doi.org/10.2196/18936>
- Zastrow, M. (2020). Open science takes on the coronavirus pandemic. *Nature*, 581(7806), 109–110. <https://doi.org/10.1038/d41586-020-01246-3>
- Zhang, L., Ghader, S., Pack, M. L., Xiong, C., Darzi, A., Yang, M., Sun, Q., Kabiri, A., & Hu, S. (2020). An interactive COVID-19 mobility impact and social distancing analysis platform. *MedRxiv*, 2020.04.29.20085472. <https://doi.org/10.1101/2020.04.29.20085472>
- Zhang, Y. (2020). *The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19)—China, 2020*. Cina CDC Weekly. <http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51>
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., ... Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270–273. <https://doi.org/10.1038/s41586-020-2012-7>
- Zhou, S.-J., Zhang, L.-G., Wang, L.-L., Guo, Z.-C., Wang, J.-Q., Chen, J.-C., Liu, M., Chen, X., & Chen, J.-X. (2020). Prevalence and socio-demographic correlates of psychological health problems in Chinese

adolescents during the outbreak of COVID-19. *EUROPEAN CHILD & ADOLESCENT PSYCHIATRY*.
<https://doi.org/10.1007/s00787-020-01541-4>