Datensatzdokumentation

Observatory of serological studies on SARS-CoV-2 in Germany

Robert Koch Institute | RKI

Nordufer 20

13353 Berlin

Unit 25 | Physical Health

Hannelore Neuhauser (Project management)

Christina Poethko-Müller

Giselle Sarganas Margolis

Roma Thamm

Unit 34 | HIV/AIDS and other sexually or blood-borne infections

Marlow Zimmermann

Unit 24 | Health Reporting

Franziska Prütz

ZfKD | Centre for Cancer Registry Data

Nina Buttmann-Schweiger

Julia Fiebig

Content and structure of the research project

The infectious disease COVID-19, caused by the novel SARS-CoV-2 virus, which has emerged since 2019, led to public health and societal challenges. In order to be able to take appropriate measures to contain the pandemic and to gain new insights into the pandemic, there is an increased need for research on COVID-19. One starting point for this is using blood samples obtained from both infected and uninfected individuals, which are tested for antibodies to the SARS-CoV-2 virus. They provide information on the proportion of the population that has already been infected with SARS-CoV-2, including unrecognized infections (under-reporting).

The project "Observatory of serological studies on SARS-CoV-2 in Germany" (SERO-OBS Corona) provides an overview of antibody studies (so-called seroepidemiological studies) in Germany. The seroepidemiological studies are based on blood samples from citizens tested for antibodies against the SARS-CoV-2 virus at different times during the pandemic. This will answer questions such as: What is the frequency of SARS-CoV-2 infection in different populations? What is the under-reporting factor, which shows how many times more infections have occurred compared to the previously known (reported) cases? In the present project, seroepidemiological studies on SARS-CoV-2 conducted in Germany since spring 2020 are continuously identified via systematic searches in study registries, literature databases including previous

publications as well as media reports, and study information as well as overviews of results are made available.

The results of the SERO-OBS-Corona project are available and regularly updated on the website www.rki.de/covid-19-ak-studien, in German, and the website www.rki.de/covid-19-serostudies-germany, in English.

Administrative and organizational details

In the project "Observatory of serological studies on SARS-CoV-2 in Germany" (SERO-OBS Corona), different Units of the Department 2 | Epidemiology and Health Monitoring and Department 3 | Infectious Disease Epidemiology of the Robert Koch Institute are involved. We thank the SeroTracker (serotracker.com) team at the University of Calgary for helpful discussions and contributions to the operationalization of the study database and for initial programming of the interactive graphics. Questions regarding the content of the data extraction can be addressed directly to the SERO-OBS-Corona project team at COVID19_AK@rki.de. The publication of the data as well as the quality management of the (meta-)data is done by the department MF 4 | Research Data and Infromation Management. Questions regarding data management and the publication infrastructure can be directed to the Open Data Team of the Department MF4 at OpenData@rki.de.

The project "Observatory of serological studies on SARS-CoV-2 in Germany" (SERO-OBS Corona) is funded by the German Ministry of Health.

Content and structure of the research project

Seroepidemiologic studies have been of high interest since the onset of the COVID-19 pandemic. They provide information on the proportion of the population that has already been infected with SARS-CoV-2, including unrecognized infections (under-reporting).
These studies have been continuously identified since spring 2020 via systematic searches in study registers, literature databases, preprints and media reports. The RKI's overarching search strategy for the novel coronavirus has been published in the journal GMS of the Arbeitsgemeinschaft für Medizinisches Bibliothekswesen (Erling J, Heldt K, Sturm J, 2021).

> Erling J, Heldt K, Sturm J. Lessons learned from the pandemic – ein Praxisbericht aus der Bibliothek des Robert Koch-Instituts. GMS Med Bibl Inf. 2021;21(1-2):Doc05. DOI: 10.3205/mbi000494

The studies are critically assessed, and aggregated data from already available studies from Germany are extracted and processed for analyses, for an interactive map, for a dashboard and for the regular update of a factsheet on the results of seroepidemiological SARS-CoV-2 studies in the general population in Germany. Furthermore, we carry out searches and analyses on under-reporting and infection fatality rates. The results are provided in German on the website www.rki.de/covid-19-ak-studien and in English on the website www.rki.de/covid-19-serostudies-germany. They are regularly updated.

The project SERO-OBS Corona thus provides an overview of seroepidemiological studies on SARS-CoV-2 conducted in Germany. The following publications resulted from the project:

> Neuhauser H, et. al : Seroepidemiologische Studien zu SARS-CoV-2 in Stichproben der Allgemeinbevölkerung und bei Blutspenderinnen und Blutspendern in Deutschland – Ergebnisse bis August 2021 Epid Bull 2021;37:3 -12 | DOI: 10.25646/8999

Thamm, R., et al. SARS-CoV-2-Seroprävalenz bei Kindern und Jugendlichen in Deutschland – ein Überblick. Bundesgesundheitsblatt 64, 1483–1491 (2021). DOI: 10.1007/s00103-021-03448-0

Poethko-Müller C, et al.: Studien zur Seroprävalenz von SARS-CoV-2 in Deutschland und international. JoHM 2020;5(S(4)):2-16. DOI: 10.25646/7023

Neuhauser H, et al.: Ergebnisse seroepidemiologischer Studien zu SARS-CoV-2 in Stichproben der Allgemeinbevölkerung und bei Blutspenderinnen und Blutspendern in Deutschland (Stand 3.12.2020). Epid Bull 2020;50:3-6. DOI: 10.25646/7728

Robert Koch-Institut (2021) SARS-CoV-2-Seroprävalenz in der Allgemeinbevölkerung in Deutschland. Stand: 23. November 2021. DOI:10.25646/9288

Robert Koch-Institut (2022) SARS-CoV-2-Seroprävalenz in der Allgemeinbevölkerung in Deutschland – Aktualisierung Juni 2022. https://edoc.rki.de/handle/176904/9849.2

Robert Koch-Institut (2022) SARS-CoV-2-Seroprävalenz in der Allgemeinbevölkerung in Deutschland – Aktualisierung September 2022. DOI: 10.25646/9693.3

Selection and description of the object of study

Information and results of epidemiological studies on the seroprevalence of SARS-CoV-2 in random samples of the general population, in blood donors and in specific population subgroups in Germany have been systematically searched for and key data are provided.

Methods, instruments and course of data collection

Search and search strategy: (We searched for...) Seroepidemiological studies on SARS-CoV-2 in Germany that can be found as peer-reviewed publication in literature databases (PubMed and Embase). We also searched for manuscripts published on preprint servers prior to peer review (medRxiv, bioRxiv, arXiv, ChemRxiv, preprints.org, ResearchSquare, Social Science Research Network (SSRN)).

After an initial search for SARS-CoV-2 and related terms, the defined search terms in the titles of the publications are "sero OR antibod OR immune OR immunity OR immunology OR fatality rate OR population-based OR cohort study OR dried blood OR test strategy".

We also browsed reports and notifications from WHO, the European Centre for Disease Prevention and Control (ECDS) and the Center for Disease Control and Prevention (CDC) and trial registries (Deutsches Register Klinischer Studien, ClinicalTrials.gov) and followed media coverage as well

Content and structure of the data set

The data set contains results and information from seroepidemiologial studies on SARS-CoV-2 in Germany. In addition, basic metadata, information on the usage license as well as data set documentation are provided. Included in the data set are:

- Table with information on the studies and results summaries
- Data set documentation in German
- License file with the usage license of the data set in German and English

- Metadata file with contextual information

Data from the Observatory of Serological Studies on SARS-CoV-2 in Germany

The data of the project "Observatory of Serological Studies on SARS-CoV-2 in Germany" (SERO-OBS Corona) are available in the root directory, in German and English language, in the data formats .csv, .json and .xlsx:

> [SERO-OBS_SARS-CoV-2_in_Deutschland.csv](#)
> [SERO-OBS_SARS-CoV-2_in_Deutschland.json](#)
> [SERO-OBS_SARS-CoV-2_in_Deutschland.xlsx](#)
>
> [SERO-OBS_SARS-CoV-2_in_Germany.csv](#)
> [SERO-OBS_SARS-CoV-2_in_Germany.json](#)
> [SERO-OBS_SARS-CoV-2_in_Germany.xlsx](#)

The data provided differentiates according to overarching characteristics:

- Information on the study and the study design
- Information on results
    - concerning seroprevalence (value, confidence interval, adjustment, etc.)
    - concerning the under-reporting factor

Variables and variable characteristics

A study always has a unique name, but can include multiple publications, have multiple study locations, etc. When changing from preprint to full print, the related data is updated if the specific results (e.g. by age, region, time period) have been recalculated and differ from each other. The entries are created by one person, quality-checked by a second person and discussed in the working group in case of disagreement.

The data set contains the following variables and their characteristics:

Variable	Type	Values	Description
title	String	`$Name`	Name from own information, DRKS, study title.

Variable	Type	Values	Description
type	String	`Cross sectional` , `Cross sectional and selective longitudinal` , `Longitudinal/Prospectice cohort study` , `Repeated cross-sectional`	Cross-sectional if sampled once; longitudinal if same participants are sampled repeatedly over time; cross-sectional and selective longitudinal if participants are selected at baseline and are selected selectively from the pool of those already included for further sampling. Repeated sampling at one site/over one recruitment method at multiple time points.
start_date	Date	`YYYY-MM-DD`	Most definite date of the start of sampling. If only month specification was available, the 1st of the month was taken. If a time span was specified, the 15th of the month was taken.
end_date	Date	`YYYY-MM-DD`	Most definite date of the end of sampling. If only month specification was available, the 1st of the month was taken. If a time span was specified, the 15th of the month was taken.
organisation_$	String	`$Name`	Name of the conducting Organisation.
organisation_raw	String	`$Name,$Name,....`	Names of the conducting Organisations.
location_$	String	`$Name`	Finest spatial resolution available (city, or county, municipality) of the study site (single values)

Variable	Type	Values	Description
location_raw	String	`$Name,$Name,...`	Finest spatial resolution available (city, or county, municipality) of the study site
region	String	`Nationwide`, `Baden-Wuerttemberg`, `Bavaria`, `...`, `Thuringia`, `Multilocal`	Name of the federal state as well as a value for the entire federal territory and cross-state surveys (Mulitlocal)
registry_url	String	`https://www.NAME.de`	Link to the study registry entry (e.g. DRKS).
information_url	String	`https://www.NAME.de`	Link to study webpage.
resource_url_$	String	`https://www.Name1.de`	Link to publication/report published within the framework of the study. (single values)
scope_regionality	String	`Local/Regional` `Nationwide`	The described or implied scope of the estimate (i.e., what region is this estimate intended to cover) s see Serotracker. `Local/Regional`: e.g. hospital, workplace, city, region, state, several cities/regions in one state, `Nationwide`: several states.

Variable	Type	Values	Description
scope_population	String	`Allgemeinbevölkerung`, `Besondere Bevölkerungsgruppen`, `Allgemeinbevölkerung und besondere Bevölkerungsgruppen`	The sample represents the general population or gives an idea of seroprevalence in populations that represent a sample of the general population. For example, the target population of a study of school-age students in a state is intended to reflect the seroprevalence of the general population of that age; teaching staff, on the other hand, represent a special population as the target population.
scope_age	String	`Adults`, `Children/adolescents`, `Multiple`	Targeted age group. The selection is exclusive. Children/adolescents are below 18 years.
subject_frame	String		Selection of the sample frame in which the sample was drawn for which the seroprevalence result is mapped. The selection reflects the group that was most specifically described in the study (e.g., residents of nursing homes instead of residential population).
subject_information	String		Combination of information from the sample frame and study location and other relevant characteristics of the sample in terms of a brief description.
subject_vaccination	String		Vaccination status of participants for results shown in the graphs.

Variable	Type	Values	Description

sample_design	String	`Blood donation`
		`Cohort-sample - Sampling from existing cohort study in the general population`
		`Convenience sample`
		`Entire sample`
		`Multistage, partial random sampling`
		`Random sampling (central registration-based)`
		`Random sampling (not central registration-based)`
			Convenience: Sampling based on pragmatism and convenient access.
			Blood donation: Anonymised residual samples from blood donations are used for the study
			Entire sample: Every person within a population is included in the study
			Cohort sample: Sample from an existing general population-cohort study
			Random sampling (not central registration-based): random selection of participants not based on a register of residents
			Random sampling (central registration-based): random sampling based on the register of residents
			Multistage partial random sampling: Multi-stage sampling, not random at all levels.

sample_analysis_unit	String	`Households with participants`
		`Medical practices with participants`
		`Participants`
		`Samples of participants`
			Unit of samplin.

age_max	Integer		Maximum age of participants included in the study.
age_min	Integer		Minimum age of participants included in the study.
sample_size_obtained	Integer		Number of participants included in the study.

Variable	Type	Values	Description
sample_size_rounded	Integer		Rounded number of participants included in the study.
analysis_data_source	String	`blood plasma` `capillary blood` `no blood, sputum` `whole blood`	Blood sampling method for antibody detection.
analysis_target	String	`anti -RBD domain` `anti -S1 domain` `anti Nucleocapsid` `multiple`	Viral targets of the antibody.
analysis_isotype	String	`IgG` `Multiple antibodies` `Total antibody`	Information on the antibody class to which the reported seroprevalence refers.
analysis_test	String	`ELISA` `Elektrochemolumineszenz (ECL)-Immunassay; LIPS` `Mehrere/andere`	Information on the laboratory analytical test method used to determine the isotypes.
analysis_test_manufacturer_$	String	`$Name`	Manufacturer of the test assay. (single values)
analysis_test_manufacturer_raw	String	`$Name, Name, Name, ...`	Manufacturer of the test assay.
results_publication_url	String	`https://www.NAME.de`	Reference to the source where information on the respective result (e.g.seroprevalence, under-reporting) is given.
results_publication_date	Date	`YYYY-MM-DD`	Date of publication of the reference to the source where information on the respective result (e.g.seroprevalence, under-reporting) is given.

Variable	Type	Values	Description
results_publication_type	String	`Institutional report` `Peer-reviewed` `Publikation` `Preprint` `Presentation` `Press release` `Webseite`	Type of the source where information on the respective result (e.g.seroprevalence, under-reporting) is given (e.g. preprint, full print as peer-reviewed publication).
results_seroprevalence_ adjustment	String	`Adjusted for test-specificity/test-sensitivity` `Adjusted for test-specificity/test-sensitivity and population-weighted` `Not adjusted` `Population-weighted`	Type of the source where information on the respective result (e.g.seroprevalence, under-reporting) is given (e.g. preprint, full print as peer-reviewed publication).
results_seroprevalence_ sensitivity	Float		Test performance features of the antibody test used for the analyses.
results_seroprevalence_ specificity	Float		Test performance features of the antibody test used for the analyses.
results_seroprevalence_ positive	Float		Seroprevalence data with the highest available or from the study data calculable adjustment level.
results_seroprevalence_ positive_95pct_lower	Float		Lower 95% confidence interval for seroprevalence.
results_seroprevalence_ positive_95pct_upper	Float		Upper 95% confidence interval for seroprevalence.

Variable	Type	Values	Description
results_seroprevalence_ information	String		Combination of information on seroprevalence, which includes, for example, the study location and the sample period and, if applicable, other relevant characteristics that describe the seroprevalence determined within a study.
results_underreporting	String		Factor showing how many times more infections occurred compared to the cumulative cases known (reported) up to the study midpoint.
record_created	Date	`YYYY-MM-DD`	Date on which a new result was entered.
record_modified	Date	`YYYY-MM-DD`	Date on which an already existing result was adapted (e.g. peer-reviewed full publication instead of preprint).

Data fromating

The data of the study are contained in the dataset as a comma-separated .csv file. The character set used in the .csv file is UTF-8. The separator of the individual values is a comma ",".

- Character set: UTF-8
- .csv separator: Comma ","

Metadata

To increase findability, the provided data is described with metadata. Metadata is distributed to the corresponding platforms via GitHub Actions. A specific metadata file exists for each platform and is stored in the metadata folder:

> Metadaten/

Versioning and DOI assignment is done via Zenodo.org. The metadata provided for import into Zenodo is stored in zenodo.json. Documentation of the individual metadata variables can be found at https://developers.zenodo.org/#representation.

> Metadaten/zenodo.json

Notes on the subsequent use of the dataen

Open research data of the RKI are made available on GitHub.com, Zenodo.org und Edoc.rki.de:

- https://github.com/robert-koch-institut
- https://zenodo.org/communities/robertkochinstitut
- https://edoc.rki.de

License

The dataset "Observatory of serological studies on SARS-CoV-2 in Germany" is licensed under the Creative Commons Attribution 4.0 International Public License | CC-BY 4.0 International