

## **Data Access and Use Policy for ClinEpiDB, a VEuPathDB Database**

Updated January 27, 2020

The Clinical Epidemiology Database, ClinEpiDB (<http://ClinEpiDB.org>) is operated under funding from the US National Institute of Allergy and Infectious Diseases (NIAID) and the Bill and Melinda Gates Foundation. ClinEpiDB is charged with ensuring that epidemiologic datasets are conveniently and freely accessible to the global biomedical researcher community. This document summarizes policies associated with the access and use of datasets on ClinEpiDB.

### ***Access and use agreement:***

ClinEpiDB is committed to ensuring the appropriate protection of all study data made available through its platform.

Prior to accessing any data provided through ClinEpiDB, the User acknowledges that study data are interpretable only to the extent that the study design and context are understood and agrees:

- To adhere to the principles presented in the access system description below
- To use the datasets solely for research and statistical analysis and not for investigating specific research participants
- To guard the confidentiality of research participants by making no attempt to identify individual participants
- To inform ClinEpiDB immediately in the case of inadvertently identifying a participant and making no further use of that knowledge
- Not to redistribute data without the approval of ClinEpiDB and in some cases the data providers
- To contact data providers if considering publishing on the dataset
- To acknowledge in all oral or written presentations, disclosures, or publications the contributing investigator(s) who conducted the original study, the funding organization(s) that supported the work, the specific dataset(s) and applicable accession number(s), and ClinEpiDB as the mechanism through which the investigator accessed the data
- To adhere to any additional requirement that may be requested by the data provider

### ***Data access restrictions:***

Each dataset that is submitted to ClinEpiDB is unique and may require implementation of a different level of data access restriction even once the dataset is publicly released. Each dataset will fall into one of the five categories described in the table below: public, controlled, limited, protected, and private. These categories differ in what levels of access are required to see aggregated vs disaggregated data. Aggregated data includes what can be seen in the search wizard and Shiny App analysis tools, while disaggregated data includes what can be seen on the results page, individual record pages, and the downloaded files.

Access level	Description
Public	No access restrictions. Users can view and download the data as “Guests” without logging in.
Controlled	Guest users can view data and access analysis tools and record pages, but must obtain approval from the data providers to download data. The request may be submitted via a form that pops up when a user logs in with a registered account and clicks “Download data”.
Limited	Guest users can view aggregated data via the search wizard and Shiny App analysis tools, but must log in with a registered account to view more than 20 rows of data on the results page or view individual record pages. Users must request and obtain approval from the data providers to download data. The request may be submitted via a form that pops up when a user is logged in and clicks “Download data”.
Protected	Guest users can view aggregated data via the search wizard and Shiny App analysis tools, but must request and obtain approval from the data providers to view more than 20 rows of data on the results page, view individual record pages, or download data. The request may be submitted via a form that pops up when a user is logged in and clicks “Download data” or attempts to access the other information.
Private	A registered user must request and obtain approval to access any aspect of the data. The request may be submitted via a form that pops up when a user is logged in and clicks on the study.

### ***General principles:***

**Quality control** - Data from research studies is not made publicly available until thorough quality control is carried out. Research data are displayed on an internal protected site and examined by the original data providers and ClinEpiDB to ensure accuracy and adequate representation of the data prior to public release.

**Participant confidentiality** - Out of respect to the principles of participant confidentiality, ClinEpiDB does not make data available that allows for the identification of study participants and applies processing algorithms to obfuscate potentially identifying information. ClinEpiDB also does not publicly release data or variables that data providers are not comfortable disseminating.

**Data security** - ClinEpiDB.org websites are served over encrypted connections using secure hypertext transfer protocol (HTTPS) to protect against interception and manipulation of sensitive data. Access to development and other pre-release versions of ClinEpiDB websites is password-restricted and available only to VEUPathDB staff and scientific working group members approved by the original data provider.

**Contacting and acknowledging data providers** - While ClinEpiDB aims to make data broadly accessible and reduce technical barriers to data exploration, we fully endorse proper attribution to original study principal investigators and study teams and acknowledgement of participation of study participants.

ClinEpiDB aims to provide clear access to data provider information, including contact information, so that data users may contact study teams if they have any questions or concerns specific to the study.

**Ethical approval** - The ClinEpiDB platform has received approval from the University of Pennsylvania under IRB#7, Protocol #828806. All studies that are included in ClinEpiDB have undergone ethical approval at applicable institutions prior to protocol execution and ClinEpiDB is generally not involved in this process.