Downloading Data

1. Submitting data access requests
2. Downloading complete datasets
3. Generating a subset of data to download

Data Access Levels

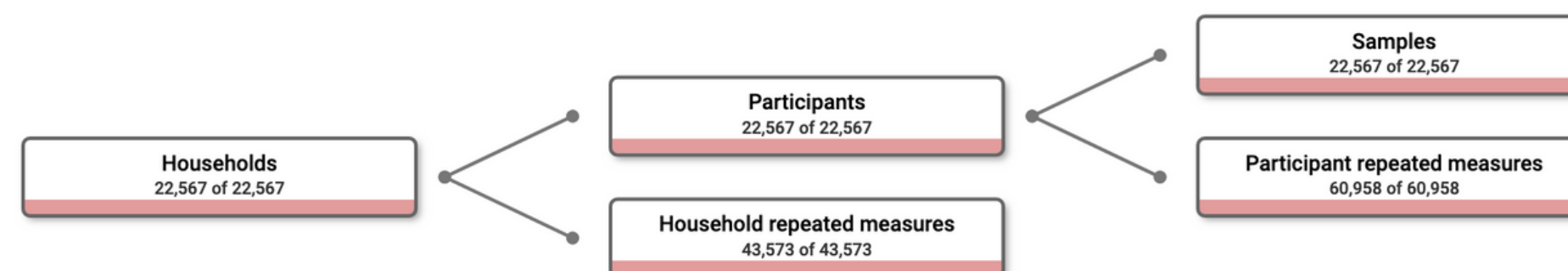| Access level	Description
Public	No access restrictions. Users can view and download the data as "Guests" without logging in.
Controlled	Guest users can view aggregated data via the "Browse and subset" and "Visualize" tabs, but must register and request access to download data. The request may be submitted via a form that pops up when a user logs in with a registered account and clicks on any file in the "Download" tab. Data can be downloaded by the registered user immediately following request submission.
Protected	Guest users can view aggregated data via the "Browse and subset" and "Visualize" tabs, but must register, request, and obtain approval from the data providers to download data. The request may be submitted via a form that pops up when a registered user is logged in and clicks on any file in the "Download" tab.
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 registered user is logged in and clicks on the study.
Prerelease	The study is under active development for future release, and only the study team can access the data. Guest and registered users can view the study page only.

Submitting Data Access Request

ClinEpiDB.org

GEMS1 Case Control

+ New Analysis   ▦ My Analyses

Unnamed Analysis ✎

View Study Details   Browse and Subset   Visualize   Download   Record Notes

1 Data Provenance
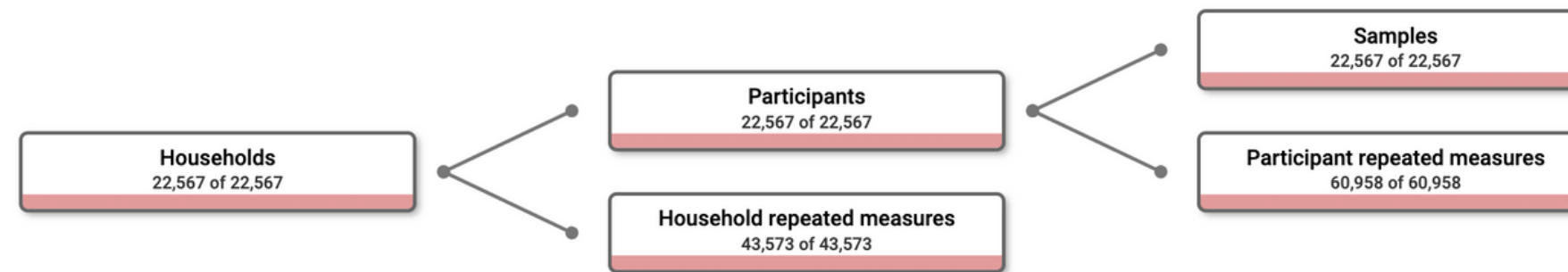
Data accessibility ❓ Protected

Submitting Data Access Request

Submitting Data Access Request

ClinEpiDB.org

GEMS1 Case Control

+ New Analysis   ⊞ My Analyses

Unnamed Analysis ✏️

Households
22,567 of 22,567

Participants
22,567 of 22,567

Household repeated measures
43,573 of 43,573

Samples
22,567 of 22,567

Participant repeated measures
60,958 of 60,958

View Study Details   Browse and Subset   Visualize   Download   Record Notes

Data downloads for this study are protected. You must register or log in and request access to download data; data can be downloaded after the study team reviews your request and grants you access. Click here to request access.

My Subset
Configure and download one or more tabular files of the filtered dataset

⊞⬇ **22,567 of 22,567 Households**
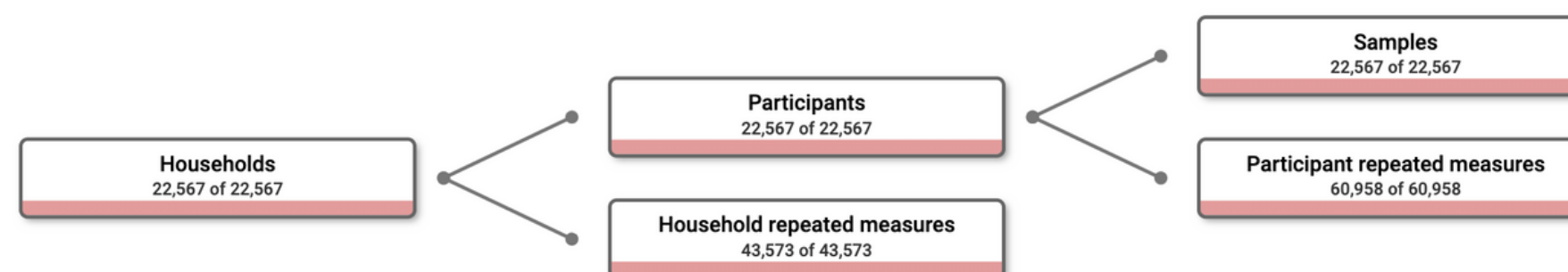
⊞⬇ **43,573 of 43,573 Household repeated measures**

Submitting Data Access Request

Submitting Data Access Request

ClinEpiDB
Clinical Epidemiology Resources

Please log in

Username or Email: s.shah.simpson@gmail.com

Password: ••••

[Login] [Cancel]

Forgot Password?   Register/Subscribe

Visit our partner Bioinformatics Resource Center, BV-BRC

Submitting Data Access Request

Data Access Request for GEMS1 Case Control

Data files will be available to download in a tab-delimited format with an additional data dictionary file. To process your download request, the data providers for this study require documentation of the following information.

To ensure transparency and promote collaboration within the wider scientific community, your name, organization, date of request and purpose for which the data will be used, as submitted below, will appear publicly on the corresponding study page after a request has been granted. The dataset page also contains critical methodologic information and study findings that are necessary to interpret the requested study data.

If you have any questions about a data access request please contact us at help@clinepidb.org.

Date of Request:	10/6/2022
Your Name:	Sheena Tomko
Your Email Address:	s.shah.simpson@gmail.com
Your Organization:	UPenn

Purpose for which data will be used (Note this field will be displayed publicly upon approval and should be complete but concise):

Downloading Complete Datasets

View Study Details   Browse and Subset   Visualize   **Download**   Record Notes

Data downloads for this study are protected. You have been granted access to download the data.

Full Dataset (Release 21)

Date: 2022-MAR-03
Change Log: Data were reformatted for the new version of ClinEpiDB. Some variables may have moved and/or have updated labels.

File Description	Type	Size
⬇ rls0021_GEMS_gems1_case_control_Download_Files_README	txt	0.01 MB
⬇ rls0021_GEMS_gems1_case_control_Household_repeated_measures	txt.zip	0.36 MB
⬇ rls0021_GEMS_gems1_case_control_Households	txt.zip	0.98 MB
⬇ rls0021_GEMS_gems1_case_control_OntologyMetadata	txt.zip	0.03 MB
⬇ rls0021_GEMS_gems1_case_control_Participant_repeated_measures	txt.zip	3.21 MB
⬇ rls0021_GEMS_gems1_case_control_Participants	txt.zip	2.25 MB
⬇ rls0021_GEMS_gems1_case_control_Samples	txt.zip	1.52 MB

Downloading Complete Datasets

ClinEpiDB.org

README for GEMS1 Case Control Download Files:

General notes:

• Column headers for each data file contain the variable label used on the ClinEpiDB.org website along with the Internationalized Resource Identifier (IRI) in brackets.

• Each variable on ClinEpiDB.org has been mapped to an ontology term via the IRI.
generates a unified semantic framework, where the same variables across different
to the same ontology term. IRIs can be searched on ontobee.org. Note that while mo
used in ClinEpiDB are from existing OBO Foundry ontologies, some are placeholder t
created as needed and may not be published yet.

• GEMS_gems1_case_control_ontologyMetadata.txt links variables to their original s
users can reference study data collection forms and data dictionaries to learn mor
variable.

• Delimited text files are used to store data. Each line represents a single house
participant repeated measure, etc. Each line has fields separated by a tab delimit

• If a variable has multiple values for a given participant/participant repeated m
will be in a quoted, comma-separated list.

GEMS_gems1_case_control_ontologyMetadata.txt, the ontology term association file i
columns:

GEMS_gems1_case_control_ontologyMetadata.txt, the ontology term association file includes the
columns:

• iri -- Internationalized Resource Identifier (IRI) assigned to variable (Example: "EUPATH_0010375")

• label -- Displayed as variable name on ClinEpiDB (Example: "Case or control participant")

• type -- Whether data is formatted as number, string, or date. If (Data Files:
without data

• parentLabel -- Term the variable falls directly under in the varia
"Administrative information")

• category -- Highest level parent term (Household, Participant, Pa
variable falls under (Example: "Participant")

• definition -- Study specific description for variable that is disp
label when variable is selected

• min, max, average, median, upper_quartile, lower_quartile -- Prov

• number_distinct_values -- Count of all possible values for variab

• distinct_values -- Quoted, comma-separated list of all possible va

• variable -- Column header, or variable name, from original data f
multiple variables from original data files were mapped to one onto

Data Files:

• Key identifiers:

  ○ Household_Id: unique identifier given to every household

  ○ Household_repeated_measure_Id: unique identifier given to every unique observation for every household

  ○ Participant_Id: unique identifier given to every participant

  ○ Participant_repeated_measure_Id: unique identifier given to every unique observation for every participant

  ○ Sample_Id: unique identifier given to every sample

1. GEMS_gems1_case_control_Households.txt

  • Key identifiers: Household_Id

  • 1 row for each household

Downloading Complete Datasets

ClinEpiDB.org

iri	label	type	parentlabel	category	definition	min	max	average	median	upper_quarti	lower_quarti	number_dist	distinct_valu	variable	replaces
IAO_000041	Weight	number	Anthropome	Participant r	Weight; coul	0.5	131.4	9.14999915	8.8	10.6	7.4	217		["F4B_REHY	EUPATH_0000732
OBI_000116	Age	integer	Observation	Participant r	repeated meas	0	62	18.8219339	16	26	9	63		["F6_AGE","	EUPATH_0000579
OGMS_0000	Physical examination		Participant	Participant repeated measure											
CMO_00002	Respiratory r	integer	Physical exa	Participant r	Assessed at	14	120	34.9055658	34	39	30	81		["F4B_RESP1","F5_RESP1","F7_RESP1"]	
EUPATH_00	Signs and symptoms		Participant	Participant repeated measure											
EUPATH_00	Observation details		Participant	Participant repeated measure											
EUPATH_00	Anthropometry		Participant	Participant repeated measure											
EUPATH_00	Observation	date	Observation	Participant r	Observation	28/11/2007	08/05/2011	01/08/2009	27/07/2009	18/04/2010	11/11/2008	1256		["F4B_DATE","F5_DATE","F7_DATE"]	
EUPATH_001	Height	number	Anthropome	Participant r	Height at en	17.4	117.4	77.229096	76	83.9	69.7	650		["F5_HT1","F7_HT1"]	
EUPATH_001	Dry mouth	string	Signs and sy	Participant r	Assessed at enrollment for cases							3	["Normal","S	["F4B_FIND_MOUTH","F4B_MOUTH"]	
EUPATH_001	Skin pinch	string	Signs and sy	Participant r	Assessed at enrollment for cases							3	["Normal","S	["F4B_SKIN_PINCH","F4B_SKIN"]	
EUPATH_001	Anthropometry flag		Anthropome	Participant repeated measure											
EUPATH_001	Rectal prola	string	Signs and sy	Participant r	Assessed at enrollment and 60 day follow-up							2	["No","Yes"]	["F4B_RECTAL","F5_RECTAL"]	
EUPATH_001	Bipedal eder	string	Signs and sy	Participant r	Assessed at enrollment and 60 day follow-up							2	["No","Yes"]	["F7_BIPEDAL","F4B_BIPEDAL","F5_BIPEDAL"]	
EUPATH_001	Undernutriti	string	Signs and sy	Participant r	Assessed at enrollment and 60 day follow-up							2	["No","Yes"]	["F7_UNDER_NUTR","F5_UNDER_NUTR","F4B_UND	
EUPATH_001	Skin has 'flak	string	Signs and sy	Participant r	Assessed at enrollment and 60 day follow-up							2	["No","Yes"]	["F7_SKIN_FLAKY","F5_SKIN_FLAKY","F4B_SKIN_FLA	
EUPATH_001	Weight-for-a	string	Anthropome	Participant repeated measure								2	["No","Yes"]	["F4B_WAZ_F","F7_WAZ_F","F4B_REHYD_WAZ_F",	
EUPATH_001	Height-for-a	string	Anthropome	Participant r	Calculated using mean length or height							2	["No","Yes"]	["F7_HAZ_F","F4B_HAZ_F","F5_HAZ_F"]	
EUPATH_001	Weight-for-l	string	Anthropome	Participant r	Calculated using mean length or height							2	["No","Yes"]	["F4B_FIND_WHZ_F","F4B_WHZ_F","F7_WHZ_F","	
EUPATH_001	Height-for-a	string	Anthropome	Participant r	Calculated using median length or height							2	["No","Yes"]	["F4B_MED_HAZ_F","F5_MED_HAZ_F","F7_MED_HA	
OBI_000150	Study timepo	string	Observation	Participant repeated measure								6	["Enrollment","Enrollmen	BFO_0000015	
CMO_00001	BMI	number	Anthropome	Participant r	BMI, using m	1.222	501.3583	15.6017479	15.4015	16.6231	14.3153	22618		["F4B_OUT_	EUPATH_0000661
EUPATH_00	BMI-for-age	number	Anthropome	Participant r	BMI-for-age	-15.3	273.91	-0.6067295	-0.61	0.23	-1.46	1259		["F4B_Last_BMIZ","F5_BMIZ","F4B_OUT_BMIZ","F4	
EUPATH_00	MUAC-for-ag	number	Anthropome	Participant r	MUAC-for-ag	-7.49	15.47	-0.4986604	-0.47	0.22	-1.16	883		["F5_MUAC_AGEZ","F7_MUAC_AGEZ","F4B_MUAC_	
EUPATH_00	Weight-for-a	number	Anthropome	Participant r	Weight-for-a	-10.81	14.95	-1.2771825	-1.25	-0.43	-2.08	1087		["F4B_OUT_WAZ","F7_WAZ","F5_WAZ","F4B_REHY	
EUPATH_00	Weight-for-l	number	Anthropome	Participant r	Weight-for-l	-14.4	19.71	-0.7168334	-0.72	0.13	-1.56	1240		["F4B_REHYD_WHZ","F5_WHZ","F7_WHZ","F4B_WH	
EUPATH_001	Individuals related to part		Participant	Participant repeated measure											
EUPATH_001	Caregiver plu	number	Anthropome	Participant r	Caregiver plu	0.4	145.1	61.7592587	60.6	68.1	53.5	794		["F4B_REHYD_WT_CHILD","F4B_FIND_WT_CHILD","F	
EUPATH_001	Caregiver we	number	Anthropome	Participant r	Caregiver we	18.6	135.9	53.4441958	52.2	59.4	45.4	768		["F4B_OUT_WT_CARE","F7_WT_CARE","F5_WT_CA	
EUPATH_001	Height, 2nd	number	Anthropome	Participant r	Height at en	17.5	117.3	77.2453081	76	83.9	69.8	646		["F7_HT2","F5_HT2"]	
EUPATH_001	Height, 3rd	number	Anthropome	Participant r	Height at en	17.6	117.4	77.2555671	76	83.9	69.8	644		["F7_HT3","F5_HT3"]	
EUPATH_001	Mean height	number	Anthropome	Participant r	Mean height	12.07	120.07	76.7270231	75.4	83.37	69.23	1803		["F5_HEIGHT","F7_HEIGHT","F4B_HEIGHT"]	
EUPATH_001	Median heig	number	Anthropome	Participant r	Median heig	12.1	120	76.7264813	75.4	83.3	69.2	667		["F5_HEIGHT_MEDIAN","F7_HEIGHT_MEDIAN","F4B	
EUPATH_001	Length- or h	number	Anthropome	Participant r	Length- or h	-26.37	12.57	-1.3989768	-1.37	-0.57	-2.19	1026		["F7_MED_HAZ","F5_MED_HAZ","F4B_MED_HAZ"]	

Generating Subset to Download

Generating Subset to Download

Generating Subsets to Download

View Study Details   Browse and Subset   Visualize   **Download**   Record Notes

Data downloads for this study are protected. You have been granted access to download the data.

My Subset
Configure and download one or more tabular files of the filtered dataset

5,835 of 22,567 Households

11,421 of 43,573 Household repeated measures

5,835 of 22,567 Participants

15,623 of 60,958 Participant repeated measures

5,835 of 22,567 Samples

Generating Subsets to Download

ClinEpiDB.org

WASH Benefits Kenya Cluster Randomized Trial
Unnamed Analysis

Participant repeated measures
3,248 of 40,722 records selected

(1) Choose variables

select all | clear all | expand all | collapse all

Find a variable

(2) View table and download

▼ ☐ **Participant repeated measure**
  ▼ ☐ Observation details
    🔒 ☰ Study timepoint ☆
    ☐ ☰ Included in diarrhea analysis ☆
    ☐ ☰ Observation status ☆
  ▼ ☐ Anthropometry
    ☐ 📊 Age at anthropometry ☆
    ☐ 📊 Anthropometry year ☆
    ☐ 📊 Anthropometry month ☆