

# ClinEpiDB: Documents Transferring List

Weilu Song

April 2023

---

## Contents

<b>1</b>	<b>EDA Files</b>	<b>1</b>
1.1	Integration studies	1
1.1.1	Gates_MORDOR Study: MORDOR	1
1.1.2	Gates_AVENIR Study: AVENIR	1
1.1.3	Gates_Namibia	2
1.1.4	Gates_JILINDE	2
1.1.5	Guatemala_Cohort	4
1.1.6	ICEMR_Peru_Cross-Sectional	4
1.1.7	Gates_LLINEUP25_Cluster_RCT	4
1.1.8	ICEMR_Peru_P3	5
1.1.9	SIP- cowork	5
1.2	REDMINE Tickets	6
<b>2</b>	<b>Outreach Files</b>	<b>6</b>
2.1	UI/UX	6
2.2	Video	6
2.3	Conference and Meeting	7
<b>3</b>	<b>Others</b>	<b>7</b>

# 1 EDA Files

## 1.1 Integration studies

### 1.1.1 Gates\_MORDOR Study: [MORDOR](#)

- **Study Process Description.** MORDOR study has been released in B62, this study has been done with all reloading files. The biggest issue before official release was about the inconsistent number that was raised by study team, the outreach team fixed the number and received the release consent from study team.
- **Files Location.**
  - Study meeting notes: [Meeting notes and other docs](#)
  - Study page: [Study Page-HTML format](#)
  - Other study docs: [MORDOR-issue tracker](#)
  - [Dropbox](#): Files are backed up in dropbox, including: (a.) .rmd file-modrodr\_long\_reshape for data integration; (b.) long-format dataset; (c.) Original codebook from study team; (d.) New mordor dataset provided by study team in Aug. 2.

### 1.1.2 Gates\_AVENIR Study: [AVENIR](#)

- **Study Process Description.** The AVENIR study is currently ongoing, and it has been updated based on the latest datasets provided by the study team. All pre-loading files, including the Conversion file, valueMap, and idMap, have been updated accordingly. At present, the study is in the process of loading. The latest dataset was provided in January 2023. The most significant update for this dataset is the consolidation and simplification of all vaccination information. The study is now waiting for the reload process to be completed.

As the study team said in the email: *"...I updated the vaccination information, as well as the codebook for Round 1. The new vaccination variables are: polio\_num, dpt\_num, measles\_num, pneumo\_num, rota\_num, bcg, hepB\_num, yellow, tetanos, imovax, vitaminA, other\_vaccinenum. The variables with "num" at the end correspond to the number of vaccines received, and the variables that do not end in "num" are binary variables that correspond with whether or not the vaccine was received at all. I added these descriptions into the codebook."*

- **Files Location.**
  - Study meeting notes: [AVENIR docs - Conflunce](#)
  - Study page: [AVENIR Study page-HTML code](#)
  - Other study docs: [AVENIR Issue tracker](#)
  - [Dropbox](#): Files are backed up in dropbox, including: (a.) Rmd for last round datasets update(01242023); (b.) cd folder contains dictionary from study team; (c.) rawdataset folder and reshaped dataset folder containing the datasets accordingly.

### 1.1.3 Gates\_Namibia

- **Study Process Description.** The initial data integration and loading for the Namibia study has been completed. Currently, the study is undergoing an in-depth review to assign unique IDs to each entity. The main challenge faced by the study team is the absence of a primary key. The team has attempted to identify different combinations that could serve as a primary key, but the missing values in the primary key columns persist. This poses a significant challenge for the team since they aim to maintain consistency with their publications. To address this issue, the team is exploring alternative solutions for assigning IDs or considering separating the datasets as Cluster RCT and cross-sectional. The study's ticket is now under Danica's supervision. See the ticket in section of ticket.
- **Files Location.**
  - Study meeting notes: [Meeting Notes](#)
  - Study page: [Study page-HTML](#)
  - Other study docs: (a. )[Issue tracker on google doc](#); (b.) [Note-Question for PK](#)
  - [Dropbox](#): Files are backed up in dropbox, including:(a.) Data collection forms from study team; (b.) .rmd file for data integration; (c.) Namibia new datasets from study team (09\_2022); (d.) Namibia id and other issue checking log (. html file).

### 1.1.4 Gates\_JILINDE

JILINDE series study contains 10 + studies. Currently, the 6 studies has been done or partially done for their pre-loading files. Including conversion file, valueMap, idMap, as well as owl files. The rest of studies have not started based on the previous study loading plan. The processed studies and their priori ordre are list below:

Table 1: JILINDE Studies List

ID	Study Name	Status	Priority
1.	gates_jilinde_demand_creation_evaluation_questionnaire	WIP	3
2.	gates_jilinde_healthcareprovider_cohort	Done	1
3.	gates_jilinde_pe_chv_cohort	WIP	2
4.	gates_jilinde_retrospective_survey	WIP	4
5.	gates_jilinde_retrospective_survey	WIP	NA
6.	gates_jilinde_AGYW awareness survey	Not start	NA
7.	gates_jilinde_AGYW PrEP and FP utilization survey	Not start	NA
8.	gates_jilinde Client Exit	Not start	NA
9.	gates_jilinde Client feedback	Not start	NA
10.	gates_jilinde Routine service delivery	Not start	NA
11.	gates_jilinde_prospective_cohort	<b>released</b>	NA
12.	gates_jilinde_costing_survey	<b>released</b>	NA

Below is the details of study process status for each JILINDE study:

#### 1. JILINED\_Prospective\_Cohort\_Study

- ***Study Process Description.*** The study has been released in B61. However, there are some tiny issues such as multivalue conflict in the RxC mosaic plot, specifically for the variable self-identified gender. This issue requires routine maintenance to be resolved. Original dataset is single, long-format table with different study timepoints for each participant.

- ***Files Location***

- Study meeting notes: [Meeting notes and questions in process](#)
- Study page: Could not find this study's HTML format study page file.
- Other study docs: [Issue tracker](#)
- [Dropbox](#): Files are backed up in dropbox, including:(a.) Raw dataset and dictionary from studyteam; (b.) rawdataset; (c.) Qualitative data files from study team; (d.) .rmd file for data integration.

## 2. *JILINED\_Cost\_Cross-Sectional*

- ***Study Process Description.*** A single format dataset, has been released in B63.

- ***Files Location in Dropbox.***

- Study meeting notes: [Meeting note](#)
- Study page: [Study Page- HTML format](#)
- Other study docs: [Question for studyteam](#)
- [Dropbox](#): Files are backed up in dropbox, including:(a.) New jilinde costing codebook; (b.) Two checking files for missing frequencies and dataset inspection; (c.) .rmd for data integration.

## 3. *JILINED\_Healthcare\_Cohort*

- ***Study Process Description.*** A study has been completed with all of its pre-loaded files (including the idMap, conversion file, and valueMap). The ontology team has generated an owl file, and all of the files are stored in the Github repository of ApiCommonData. The only issue is that the variable names are currently being referred to as X1, X2, and so on, rather than their actual names. This is because the study team labeled the variables using variable labels, rather than using R-readable variable names. To facilitate the data integration process, we temporarily changed the variable names to the X series. However, we will need to rename the original variable names in accordance with the codebook provided by the study team, as per later requirements. The code can be referred to the previous done Jilinde study: `Jilinde_prospective_cohort`.

- ***Files Location in Dropbox.***

- Study docs on Confluence: [ALL study related docs](#)
- Other study docs: [Issue tracker](#)
- Dropbox : NA, since this study will not be processed anymore.

#### 4. Other JILINDE series studys.

In addition to the studies that were addressed above, the `jilinde_awareness_survey`, `jilinde_demand_creation_evaluation`, `jilinde_demand_creation_evaluation_questionnaire`, `jilinde_pe_chv_cohort`, `jilinde_retrospective_survey` are in the same situation as `jilinde_health_cohort`. All the pre-loading can be found in the ClinEpiDB git repository.

##### 1.1.5 Guatemala\_Cohort

- **Study Process Description.** Gutatemala study has been done and pass outreach QA. Now the study has been loaded on qa.restricted website and under study team QA process. We have not heard any further release consent from them.
- **Files Location.**
  - [Docs on Conflunce](#)
  - Other study docs: [Issue tracker](#)
  - [Dropbox](#): Files are backed up in dropbox, including:(a.) workspace. Rproj; (b.) data; (c.) .rmd files from both Sheena and Weilu.

##### 1.1.6 ICEMR\_Peru\_Cross-Sectional

- **Study Process Description.** This study is currently in progress. It has been uploaded to the QA-restricted website and has passed the outreach QA. The study team has performed their own QA and identified areas for improvement in the next update. This study provided the Spanish codebook, the translation version and original version can be found here:[Codebook-Spanish and English versions](#)
- **Files Location.**
  - [Meeting note](#)
  - Study page: Study team does not provided study page information, now the study page on website is used as placeholder. The study team has drafted the study page information according to the main publication. But this draft need to be edited. [Study page draft](#)
  - Other study docs: [Issue Tracker](#)
  - [Dropbox](#). Sheena has been grouped based on the different type of files. It contains (a.) R\_output; (b.) data; (c.) R\_code.

##### 1.1.7 Gates\_LLINEUP25\_Cluster\_RCT

- **Study Process Description** The aim of this study is to update the previous LLINEUP study by incorporating the datasets from the last round. The study team provided five different datasets, out of which only the Household level data 25m survey\_FINAL.dta and the Household member level data 25m survey\_FINAL.dta were utilized for the update. To maintain the integrity of the previous LLINEUP study, a new independent study called LLINUP25 was created. The folder name has been changed to `llinup_rct` in git by the ontology team. Currently, the study includes all the datasets from the previous rounds and

is undergoing in-depth QA by Danica. The whole dataset can be accessed from here: [Full datasets for all rounds](#)

- ***Files Location in Dropbox.***

- [Data processing note](#):

- Study page: This study totally copied the study page from previous LLINEUP. [LLINUP\\_25Study page](#)

- Other study docs:

- [Dropbox](#) Files are backed up in dropbox, including: (a.) derived codebook script. two versions can be found either in "llineup\_qa\_study.rmd" or the independent file "cd\_derived.R"; (b.) FromProvider folder contains the 5 datasets for the last round from the study team; (c.) Difference of variables comparison report; (d.) Danica\_merge\_conversion file template in R.

#### 1.1.8 ICEMR\_Peru\_P3

- ***Study Process Description.*** The study team uploaded all datasets to Dropbox. The initial work involved importing and diagnosing the datasets. However, no substantial work has been implemented at this stage.

- ***Files Location in Dropbox.*** There is no any useful docs to be backed up at current stage.

#### 1.1.9 SIP- cowork

- ***Study Process Description.*** Since I only assisted Sheena to convert ethiopian clander for this study. Nothing from my end should be uploaded and backup. All docs can be traced and accessed in either Dropbox and Confluence.

## 1.2 REDMINE Tickets

Table 2: ClinEpiDB EDA Tasks: Assignend to Weilu

ID	Project	Status	Subject	Last up- dated by	% Done	Assignee	Coworkers
<a href="#">49510</a>	ClinEpiDB	New	Create a tool to identify ID issues	Sheena Tomko	0	Weilu Song	Jay Humphrey
<a href="#">46484</a>	ClinEpiDB	Outreach QA	Study:AVENIR	Jay Humphrey	80	Weilu Song	Nupur Kittur
<a href="#">47469</a>	ClinEpiDB	Ontology develop- ment	Study:JILINDE retro- spective survey	Sheena Tomko	20	Weilu Song	Nupur Kittur
<a href="#">47822</a>	ClinEpiDB	Provider Contact / QA	Study:Guatemala op- portunistic infection cohort	Sheena Tomko	90	Weilu Song	Sheena Tomko
<a href="#">50039</a>	ClinEpiDB	Outreach QA	Study:Amazonia ICEMR2 Peru Cross- sectional	Jay Humphrey	80	Weilu Song	Nupur Kittur
<a href="#">47697</a>	ClinEpiDB	Ontology develop- ment	Study:JILINDE awareness survey	Sheena Tomko	20	Weilu Song	Nupur Kittur
<a href="#">47710</a>	ClinEpiDB	Outreach QA	Study:JILINDE healthcare provider cohort	Jay Humphrey	80	Weilu Song	Nupur Kittur
<a href="#">47711</a>	ClinEpiDB	Ontology develop- ment	Study:JILINDE Peer Educator and CHV cohort	Jie Zheng	20	Weilu Song	Nupur Kittur
<a href="#">48251</a>	ClinEpiDB	Ontology develop- ment	Study:JILINDE demand creation eval- uation questionnaire	Sheena Tomko	20	Weilu Song	Nupur Kittur

## 2 Outreach Files

### 2.1 UI/UX

Nothing to upload or save. The derived variable figma UI/UX design is stored in google drive.

### 2.2 Video

There are two main videos that were created based on different demands during the posion. Producing the video generated a huge backup file for Final cut pro. The latest backup file is up to 803.5 GB. After asking the group's opinion in outreach team meeting, the team decide not uploading and saving this back up file. The file will be deleted by the end of May.

→ ClinEpiDB 3 mins introduction: [Click to see video on Youtube](#)

→ ClinEpiDB 30 mins tutorial video:[Click to see video on Youtube](#)

## **2.3 Conference and Meeting**

ASTMH 2022-work

## **3 Others**

Will add as needed based on feedback from outreach team members till 5/25/2023.