

# Ebola Contact Tracing Study Data – Codebook

## Document details:

Principal Investigator: Professor Helen Weiss, London School of Hygiene & Tropical Medicine

Document produced by: Lisa Danquah, Imperial College London (formerly LSHTM)

Version number: 1

Date 14.03.2019

## Persistent Identifier:

<https://doi.org/10.17037/DATA.00001069>

## Overview

This codebook describes four datasets produced as part of the Ebola Contact Tracing Study (ECTS):

1. '[summary\\_data\\_cases](#)' contains details of the 41 confirmed Ebola cases included in the study;
2. '[app\\_data\\_contacts](#)' contains details of the 646 Ebola contacts registered on the Ebola Contact Tracing App (ECT) smartphone app
3. '[paper\\_data\\_contacts](#)' describes 408 Ebola contacts who were identified from 25 Ebola cases for monitoring using the standard paper-based system for contact tracing;
4. '[main\\_analysis\\_dataset](#)' contains information on 804 Ebola contacts and their contact tracing monitoring status collected using both the ECT app and paper-based system.

Access to these datasets are restricted to protect study participants. Researchers wishing to obtain the dataset must submit an access request, outlining the research purpose in which it will be used, ethics approvals obtained, and list the specific variables requested. The request will be evaluated by a study committee and, if approved, a data transfer agreement for signing prior to being provided with the dataset. Any participant-identifiable information contained within the dataset will be removed or disguised prior to being provided.

## Dataset 1: [summary\\_data\\_cases](#)

The 'summary\_data\_cases' dataset observations on all 41 confirmed Ebola cases included in the study. Of these:

- 25 were monitored using the standard paper-based system for contact tracing that was in use throughout Sierra Leone during the epidemic.
- 16 were monitored using the Ebola Contact Tracing (ECT) Study smartphone app.

This dataset does not include information on two secret burials whose Ebola status was unconfirmed, but whose contacts were monitored using the ECT app.

Each row of the dataset refers to an Ebola case. Eight variables are described in total.

Variable Name	Variable Label	Answer Label	Answer Code	Variable Type
plk_case	Case ID number. A unique identification code to link Ebola cases in each dataset	Code is made up of the prefix PLK followed by seven numbers – the first three start with 050 e.g. PLK0501234.		String
ageyr_case	age of case (yrs)			Numeric

## Data User Guide

sex_case	gender of case			String
mortstatus_case	mortality status of case			String
n_contacts	number of contacts			Integer
onsetdate_case	onset of symptoms date			long
ct_method	contact tracing method			Numeric
		paper method	0	
		app method	1	
chiefdom_code	numbered codes for chiefdoms	Open ended	1-7	Numeric

### Dataset 2: app\_data\_contacts

The 'app\_data\_contacts' dataset contains 646 observations and 25 variables on the Ebola contacts that were registered on the Ebola Contact Tracing smartphone app. This includes contacts of 16 confirmed cases and 2 secret burials.

Please note that while this dataset contains 646 contacts, not all contacts were monitored by a contact tracer due to assignment and other technical issues, so the number of contacts assigned and monitored differs at each stage.

Each contact is featured on a single line and be aware that Ebola case information is repeated for each contact. Each contact has an individual unique PLK number starting with the prefix PLK followed by the 050 and four unique numbers and then a suffix number after a hyphen e.g. PLK0501234-01 which would indicate that this is the first contact (01) for the Ebola case of PLK0501234.

Variable Name	Variable Label	Variable Type
plk_case	Case ID number. A unique identification code to link Ebola cases in each dataset	String
plk_contact	contact id number	String
ageyr_contact	age in years of contacts	Integer
lastcont_date	date of last contact with case	String
formmetatimeEnd	form meta time end	String
formmetatimeStart	form meta time start	String
received_on	received_on	String
reg_time_start	start time registration	double
reg_time_end	end time of registration	double
receivedon_reg	registration time received on	double
rel_cont	relationship to case	String
rel_cont_other	relationship to contact – free text specified	String
contact_fever	contact had a fever	String
hcw_contact	contact was a health care worker	String
samehoh	same head of household	String
sex_contact	gender of contact	String
typeofcontact	type of contact with case	String
start_time_assign	start time assignment	double
end_time_assign	end time of assignment	double
receivedon_assign	received time of assignment	double
reg_time_start_case	start time case registered	double
reg_time_end_case	end time case registered	double
receivedon_reg_case	received time of registered case	double
chiefdom_contact	numbered codes for chiefdoms for contacts	Numeric
chiefdom_code_cases	numbered codes for chiefdoms for cases	Numeric

## Dataset 3: paper\_data\_contacts

The 'paper\_data\_contacts' dataset contains 408 observations and 22 variables contacts of Ebola cases who were monitored using the standard paper-based system. Of the 25 cases monitored using paper, only 21 cases had contacts and are included in this dataset, the remaining 4 had no contacts and are therefore not listed.

The dataset contains demographic information on the case. Please note that that as cases had multiple contacts, demographic and other case level information is repeated for each contact of the same case. Any case level information is denoted at the end with `_case`. The same is true for contact information that is denoted with `_contact`. Contact level demographic information, geographic information, relationship to contact, date of last contact and contact monitoring information e.g. whether the contact tracing monitoring form was returned at the end of the 21 day monitoring period are also featured. Please note that not all contacts (n = 408) had daily contact tracing monitoring forms returned at the end of the 21-day monitoring period. The variable `drfmiss_contact` refers to whether the daily contact tracing monitoring form was returned at the end of the study period (157 contacts had forms returned and 251 did not)

Each contact is featured on a single line and be aware that Ebola case information is repeated for each contact. Each contact has an individual unique PLK number starting with the prefix PLK followed by the 050 and four unique numbers and then a suffix number after a hyphen e.g. PLK0501234-01 which would indicate that the this is the first contact (01) for the Ebola case of PLK0501234.

Variable Name	Variable Label	Answer Label	Answer Code	Variable Type
<code>plk_case</code>	case id number			String
<code>ageyr_case</code>	age of case (yrs)			Numeric
<code>sex_case</code>	gender of case			String
<code>mortstatus_case</code>	mortality status of case			String
<code>dateofdeath_case</code>	date of death of case			long
<code>onsetdate_case</code>	onset of symptoms date			long
<code>occup_case</code>	occupation of case			String
<code>plk_contact</code>	contact id number			String
<code>sex_contact</code>	gender of contact			String
<code>ageyr_contact</code>	age in years of contact			Numeric
<code>rel_cont</code>	relationship to case			String
<code>lastcont_date</code>	date of last contact with case			long
<code>date_firstvst</code>	date of first contact tracing visit			long
<code>timefirstvst_contact</code>	time of first contact tracer visit			String
<code>drfmiss_contact</code>	daily reporting form missing			Integer
		returned	0	
		not returned	1	
<code>datecompleted_21days</code>	date completed 21 day monitoring			String
<code>nodays_monitored</code>	number of days monitored by contact tracer			Numeric
<code>typeofcontact</code>	type of contact with case			String
<code>hcw_contact</code>	contact was a health care worker			String
<code>no_contacts</code>	number of contacts for each case			Numeric
<code>chiefdom_code_cases</code>	numbered codes for chiefdoms for cases			Numeric
<code>chiefdom_code_cont</code>	numbered codes for chiefdoms for contacts			Numeric

## Dataset 4: mainanalysis\_data

The 'main\_analysis\_dataset' dataset contains 804 observations and 26 variables. This is the dataset that has been used for the main analysis of the Ebola Contact Tracing Study. This dataset combines information including the laboratory confirmation times for Ebola cases from Public Health England. The data focuses on contacts including the 157 contacts who were monitored via the paper-based systems whose contact tracing daily reporting forms were returned at the end of the study period and 647 contacts monitored with the ECT app. The reason for 647 contacts instead of 646 as indicated above for the app\_based contacts is that there was one Ebola case (PLK0508936) that had no contacts but is featured as a line in the dataset. This dataset also contains the app times including registration, assignment and visitation times (including time of first visit by the contact tracer) and time of first visit information for the paper-based contacts. Other app metadata is featured in this dataset. This dataset does not feature information on the 250 paper-based contacts whose contact tracing daily monitoring forms were not returned at the end of the study period.

Each contact is featured on a single line and be aware that Ebola case information is repeated for each contact. Each contact has an individual unique PLK number starting with the prefix PLK followed by the 050 and four unique numbers and then a suffix number after a hyphen e.g. PLK0501234-01 which would indicate that this is the first contact (01) for the Ebola case of PLK0501234.

Variable Name	Variable Label	Answer Label	Answer Code	Variable Type
plk_case	case id number			String
batchresultdatetime	PHE batch result date time			double
batchreportdatetime	PHE batch report date time			double
plk_contact	contact id number			String
chiefdom_code_cont	numbered codes for chiefdoms for contacts	Numbered from 1-7		Numeric
lastcont_date	date of last contact with case			double
total_daysmont	number of days contact traced (monitored)			Numeric
first_visit_seen	date and time of first seen visit			double
first_visit	date and time of first visit			double
health_statuspaper	health status of contact at the end of contact tracing			Numeric
		well	1	
		sick	2	
		not_seen	3	
		no_information	4	
total_seen_visits	total number of seen visits by the contact tracer			Numeric
health_status1	1 health_status1_			Numeric
		well	1	
		sick	2	
		not_seen	3	
		no_information	4	
health_status2	1 health_status2_			Numeric
		well	1	
		sick	2	
		not_seen	3	
		no_information	4	
reg_time_start	start time of registration			double
reg_time_end	end time of registration			double
reg_time_start_case	start time case registered			double
reg_time_end_case	end time case registered			double

Data User Guide

start_time_assign	start time assignment			double
end_time_assign	end time of assignment			double
t_first_contact_result	time to first use of method for contact visit (hours)			Numeric
t_conf_reg	time from confirmation to registration completion (hours)			Numeric
t_reg_assign	time from registration completion to assignment (hours)			Numeric
t_assign_visit	time from assignment to first visit of contact using method (hours)			Numeric
t_conf_assign	time from confirmation to assignment (hours)			Numeric
ct_method1	contact tracing method1			Numeric
		paper method	0	
		app method	1	
montfull_21	monitored for the full 21 days			Numeric
		no	0	
		yes	1	