

Ebola Challenge Database

This document provides a quick introduction to the Ebola Challenge Database. The Ebola Challenge database is composed by four scenarios numbered from I to 4. Each scenario refers to a different epidemic and a different level of data quantity/quality release.

1. The first type of data file is in the format of timelines. A simple mask allows retrieving the following information:

- a) New confirmed and probable EVD cases (daily, weekly, at national and county levels)
- b) New cases among Health Care Workers (daily, weekly, at national and county levels)
- c) New deaths (daily, weekly, at national and county levels)
- d) New deaths among Health Care Workers (daily, weekly, at national and county levels)
- e) Number of new contacts traced (daily, weekly, at national and county levels)
- f) ETU occupancy – number of beds occupied, counting each bed each day (daily, weekly, at national and county levels).
- g) New suspected EVD cases (daily, weekly, at national and county levels)

Please note that only a subset of the data produced for each scenario is actually available (“Fog of war”). For most of the scenarios the above data may be not be all available in all formats. Expect to have little or no data at the daily and county level.

2. A second type of data file is in the format of a patient database. This data file contains individual patient records for all hospitalized patients with a complete medical record (patients who have been discharged, are dead or buried). The patient record database includes the following fields for each patient:

- Patient ID:** patient id number (patient A)
- Infector ID:** patient who has infected patient A
- Infection Time:** infection time
- Symptom Time:** time of symptoms onset
- Death Time:** time of death (NA if the individual survives)

Burial Time: time of burial (NA if the individual survives)
Age Group: three classes 0-14 yrs; 15-44; ≥ 45 yrs
HCW: if the patient is a Health Care Worker (1 true; 0 false)
County: County of residence of the patient
Hospital Admission Time: Time of admission to the hospital/holding center (otherwise NA if patient never goes to a hospital/holding center; N.B. an individual could have NA also because the patient goes straight to an ETU).
ETU Admission Time: Time of admission to ETU (otherwise NA if never admitted to an ETU. Note that an individual could be first in a hospital/holding center and then in an ETU)
Discharge Time: Time of discharge from the most recent health facility visited. This could be an ETU or a hospital/holding center (otherwise NA if never discharged because of death or never admitted to hospital/ETU)
Contact Tracing: 1 indicates the individual is a contact who was traced from another case; 0 if not reported from contact tracing.
Safe Burial: Safe Burial (1 if individual has received a safe burial; 0 if he has got a traditional burial; NA if the individual survived.)
New traced contacts: Number of contacts traced from the patient

Please expect incompleteness also in this dataset. Every time that a “none” is reported, the data is supposed to be lacking or not recorded.

For some scenarios in which the patient records are not available it will be possible to query some summary statistics such as the age distribution of cases etc.

3. Additional data concerning the scenarios will be available in the format of narrative and CSV files. These will be available at the web page <http://www.ebola-challenge.org/database.html>

Additional files, such as transmission trees, that simulate reporting from the field could be distributed for each scenario during the course of the challenge.

Note that the time unit is day or week as specified in each data file

In working with these datasets it is important to consider the following features that illustrate the amount of noise and uncertainties generally faced in real world situations:

- The information available in any of the data files may not be completely correct and could be modified during subsequent releases of data (cases missing, delayed reporting etc.)
- Only a fraction of the Ebola patients that are not admitted to a health facility will be identified as Ebola patients and

**EBOLA &
CHALLENGE**

reported. Also some of the cases in hospitals may not be readily reported. Data will therefore have an intrinsic level of underreporting.

- The day that it is indicated as Day=1 (Week=1) is not necessarily the first day (week) of the epidemic. On Day (Week) 1 several cases could be reported at once and in some cases from different locations. This should be considered as the first day of reporting.