****

VERSION 1: 28 January 2020

****

Information relates to data collected in 2015 in Eastern Congo in a targeted 1,120 villages in the provinces: South Kivu, Haut Katanga and Tanganyika.

Additional, updated versions of the datasets may be posted later.

The PIs can only provide limited support to users.

****

Panel information:

- Part (781 villages) of the 1,120 targeted villages in 2015 were also (targeted to be) visited in 2012.
- The 2012 data can be found here: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BSASJR
- The unique village code (IDV) in 2012 was also used in 2015, and can thus be used to merge the data from both data collection rounds.
- We also aimed to re-interview respondents from 2012.

****

Instruments:

- "**DRC2015_Instruments.pdf**" contains all instruments: surveys, protocols and the codebook related to the data.
- All variable names in the datasets have question numbers and a keyword that correspond to the instruments.
- All variables in the datasets (.dta format) are labeled, where the first word of the label indicates the survey model from where the data originated in the survey. See "**DRC2015_Overview_Label.xls**" for more info.

****

Datasets:

There are six separate datasets in Stata format: DRC2015_EC.dta, DRC2015_EM.dta, DRC2015_EE.dta, DRC2015_ES.dta, DRC2015_EX.dta, and DRC2015_Village_Info.dta.

There are three identifying variables:

- CDCCODE: Unique code at the CDC level.[1]
- IDV: Unique village ID
- IDS: Unique survey ID

The ID codes are nested: IDS = IDV*10 plus a number, IDV= CDC*10 plus a number.

Generic codes that are similar across the datasets:

- "-10": "Other"
- "-9": "Don't know"
- "-8": "Not applicable"
- "-7": "Refused to respond"

DRC2015_EC.dta

- Is the Chief survey (**E**nquete **C**hef in French)
- A total of 1,120 villages were targeted
- There are 3 types of villages (see variable "TYPE_V"):
    1. "**T2_PLUS**" are villages not visited in 2012, but were visited as a baseline for the Tuungane 2+ development program that took place in 2015/6.
    2. "**PANEL_5**" are villages that were visited/targeted to be visited in 2012. In 2012, data from 5 randomly selected households were collected in these villages. We aimed to find these same five households back again in 2015.
    3. "**PANEL_10**" are villages that were visited/targeted to be visited in 2012. In 2012, data from 10 randomly selected households were collected in these villages. We aimed to find 5 households of these 10 back again in 2015.

DRC_EM.dta

- Is the household survey (**E**nquete **M**enage in French)
- The aim was to collect five households per village, so 1,120 * 5 = 5,600
- The dataset contains 7,580 individuals: 396 PANEL_10 villages * 10 (10 households) + 385 PANEL_5 villages * 5 (5 households) + 339 T2_PLUS villages * 5 (5 households) = 7,580.

DRC2015_EE.dta

- This is the school survey (**E**nquete **E**cole in French)
- The number of observations is more than the 1,120 villages because the aim was to collect (at least) one school survey per village
- Note that one school can cover multiple villages. The dataset thus contains the following two additional variables:
    - EE_Cover: The number of villages in the dataset covered by this school
    - EE_IDV: The IDV of all the villages in the dataset covered by this school

[1] The CDC is an artificially created grouping of villages, done by the implementing partner in 2006. Part of the program operated at this level.

DRC2015_ES.dta

- This is the hospital survey (**E**nquete **S**ante in French)
- The number of observations is more than the 1,120 villages because the aim was to collect (at least) one hospital survey per village
- Note that one hospital can cover multiple villages. The dataset thus contains the following two additional variables:
    - ES_Cover: The number of villages in the dataset covered by this hospital
    - ES_IDV: The IDV of all the villages in the dataset covered by this hospital

DRC2015_Exam.dta

- This is the exam with one randomly selected child per household, among only those households that have one or more children of school-going age and gave their approval for the exam.

DRC2015_Village_Info.dta

- This file contains additional information from "the field"
- Survey teams were not always able to visit a village. For a subset of these villages, the reason for not visiting is recorded in the variable "Reason_attrition".
- In total, 781 panel villages were aimed to be visited. In addition, 339 additional villages that received the T2+ project were aimed to be visited. A few times, it occurs that these are the same villages (i.e. it was indicated both as a panel and as a T2+ village). When this occurred, the survey team only visited the village once, and the data are currently recorded in the dataset only once. These instances are indicated by Reason_attrition=="Village twice in dataset", and the variable "Villages_same" indicates which two villages (IDV) are the same.

****

Note that due to security and logistical reasons, etc. not all the targeted villages / households / schools / hospitals were visited.

****

We left out some data purposefully from these datasets:

- Anything that can identify the village, household or the individual (names, phone numbers, GPS locations, etc.). Identifiable data are available from the Chiefdom level and up.
- Variables that contain information about treatment assignment for which we hope to write future academic papers.

****

For questions or suggestions, please send an email to petervanderwindt@nyu.edu

**** END ****