A harmonized dataset of forcibly displaced populations and their hosts

Codebook

1. HOUSEHOLD-LEVEL DATA

1.1 Sample

countrycode: Country code

countryname: Country name

hhid: Household identifier

JOR_2016_SRHCS

1) This survey has 2 duplicate individuals. In each pair, we kept the observation with the most responses. We included this note here because some HH-level variables required using individual-level data.

int_month: Interview month

BGD_2020_CBPSR1

1) We don't know specific dates for each interview. We only know the interview date range for all interviews of 21 April-20 May 2020. All respondents were assigned 5 “May”.

int_year: Interview year

BGD_2019_CBPSBL

1) Some dates in variable starttime are incorrectly listed as being from 1923, but all interview dates are from 2019.

rep_year: Reporting year where majority of survey took place

strata: Strata of the survey

survey_name: Abbreviated survey name (CCC_YYYY_NNNN)

surveyprd: Survey period (start-end years)

LBN_2015_SRHCS

1) We assumed all surveys were conducted in 2015. This has not been confirmed.

weight: Household weight

BGD_2019_CBPSBL

1) We used aa_weight which is the adult weight.

ECU_2019_EPEC

1) We used peso2 which is the individual weight.

JOR_2016_SRHCS

1) The data does not have weights as it was not designed to be representative at any level.

.

1.2 Current Assets

ALL CURRENT ASSET VARIABLES

NER_2018_EHCVM

1) We assumed that items not in the list of assets held by a household means no ownership.

ac: Ownership of air conditioner

agrilandownti: Type of ownership document for agricultural land

boiler: Ownership of boiler/water heater

cellphone: Ownership of a cellular/mobile phone

computer: Ownership of a computer

PER_2018_ENPOVE

1) The question (P113_5) asks about a computer or tablet.

cookstove: Ownership of improved cookstove

mvehicle: Ownership of a motorized vehicle

nmvehicle: Ownership of a non-motorized vehicle

oven: Ownership of oven

ownagriland: Ownership of agriculture land

ownhouse: Ownership of house

ownland: Ownership of land

UGA_2018_RHCS

1) This survey excludes agricultural land, so it is not used for this variable.

ownresland: Ownership of residential land

NER_2018_EHCVM

1) We use ownership of undeveloped land.

TCD_2018_ECOSIT4

1) We use ownership of undeveloped land.

radio: Ownership of a radio

refrigerator: Ownership of refrigerator

tv: Ownership of a television

washingm: Ownership of washing machine

1.3 Previous Assets

ALL PREVIOUS ASSET VARIABLES

BGD_2019_CBPSBL

1) The previous reference period refers to ownership in July 2017.

NER_2018_EHCVM

1) Only items currently held by a household are asked about for previous ownership.

previous_ac: Previous: Ownership of air conditioner

previous_boiler: Previous: Ownership of a boiler/water heater

previous_cellphone: Previous: Ownership of a cell phone

previous_computer: Previous: Ownership of a computer

previous_cookstove: Previous: Ownership of improved cookstove

previous_date: Date before which previous asset questions are asked about

previous_mvehicle: Previous: Ownership of a motorized vehicle

previous_nmvehicle: Previous: Ownership of a non-motorized vehicle

previous_oven: Previous: Ownership of oven

previous_ownagriland: Previous: Ownership of agriculture land

previous_ownhouse: Previous: Ownership of house

previous_ownland: Previous: Ownership of land

UGA_2018_RHCS

1) This survey excludes agricultural land, so it is not used for this variable.

previous_ownresland: Ownership of residential land

NER_2018_EHCVM

1) We use ownership of undeveloped land.

previous_radio: Previous: Ownership of a radio

previous_refrigerator: Previous: Ownership of refrigerator

previous_tv: Previous: Ownership of a television

previous_washingm: Previous: Ownership of washing machine

1.4 Demographics

hh_type: Household is from a refugee/host/IDP sample

ECU_2019_EPEC

1) We recoded estrato2 to create this variable:
estrato2 = 1 “Ecuadorian” > coded to 2 “Host”
estrato2 = 2 “Venezuelan with children” or 3 “Venezuelan without children” > coded to “5 "Venezuelan migrant".

IRQ_2015_SRHCS

1) We categorized a household as a refugee household if they were displaced from Syria (A14 = 1), or if they were born in Syria and were displaced (A24 = 1-3).
2) All those displaced from Syria and Iraq (outside KRI) (A24=3) were living in Iraq (outside Kurdistan) in 2013 (A23=1). We considered them IDP.

JOR_2016_SRHCS

1) We categorized a household as a refugee household if they were displaced from Syria (A24 = 1) and they were born in Syria (A15 = 2). We categorized a household as a host household if they were born in any country other than Syria, or if they were born in Syria but were not displaced from Syria.

LBN_2015_SRHCS

1) We categorized a household as a refugee household only if they are Syrian AND were displaced by the Syrian Crisis. Otherwise they are considered hosts. This survey did not include any IDPs.

NER_2018_EHCVM

1) We recoded variable s00q07a depending on which file it was from:
host data file Menage_s00_me_NER2018.dta
s00q07a = 1 "IDP" > coded as hh_type = 3 "IDP"
s00q07a = 2 "Refugee" > coded as hh_type = 1 "Refugee"
s00q07a = . (missing) > coded as hh_type = 4 "national sample"
refugee data file s00_me_NER2018.dta
s00q07a = 1 "IDP" > coded as hh_type = 3 "IDP"
s00q07a = 2 "Refugee" > coded as hh_type = 1 "Refugee"
s00q07a = 3 "host communities" > coded as hh_type = 2 "Host"

PER_2018_ENPOVE

1) The survey is for the Venezuelan population in Peru, classified as migrants or refugee-like by UNHCR. All respondents were classified in a new category of hh_type = 5 "Venezuelan migrant".

TCD_2018_ECOSIT4

1) We assumed all HHs in the host file are hosts. For refugees, in the welfare file "ehcvm_welfare_TCH2018.dta", the variable pop_group gives the details on the type of households: 1.”Refugee east” 2. “refugee south” 3. “host households <5km to camps” 4. “host households 5-15km to camps” 5.”Chandian”.

UGA_2018_RHCS

1) Used s1aq15 "15. IS HOUSEHOLD A REFUGEE HOUSEHOLD?" in HSEC1.dta to categorize as refugees or hosts (any non-refugees). Note no IDPs were included in the sample.

hhsize: Household size (# of people in household)

BGD_2019_CSPSBL

1) Instead of using the sum of HH members, we used the maximum person id in the HH (variable id).

NER_2018_EHCVM

1) For the national sample, this variable was in the welfare file “Dataout_ehcvm_welfare_NER2018.dta”. For the refugees we calculated this variable from the roster, as it was not provided in any of the refugee data directly.

live_in_camp: Live in camp or outside camp

ECU_2019_EPEC

1) There are no camps in Ecuador, so we assigned all respondents 0 “no”.

LBN_2015_SRHCS

1) We assumed this should be “yes” when
> A23 "refugee status" = 1 "yes" AND
> B2 "type of dwelling" = 3 "Informal settlement/tent" or 4 "collective center" AND
> A15 = 3 “Palestinian”.

For context: Older palestinian refugees were put in camps, but syrian refugees were not. We tried to find a nationality variable (or country of birth, when you came, displaced from) to see which type of refugee and therefore if they were in camps. However, the survey did not interview at camps so there should not be any people in camps, but it could be that some Palestinian refugees were picked up.

NER_2018_EHCVM

1) Some refugees live in camps and others live outside the camp. Data does not allow us to identify those who live in camp and those who do not.

PER_2018_ENPOVE

1) None of the respondents live in camps.

TCD_2018_ECOSIT4

1) Variable s00q06 "EA or camp" has codes, but the codes do not distinguish which HHs live in camps. However, all refugees in the survey live in camps, so all observations were assigned live_in_camp = 1 “yes”.

UGA_2018_RHCS

1) Based on the sampling frame, those in the West Nile and Southwest were in camps, those in Kampala were not.

origin_country: Household country of origin (if asked at HH-level)

BGD_2019_CBPSBL

1) We used the individual-level birth country question LS12 as a proxy. We assigned origin_country = “BGD” if all respondents were born in Bangladesh, origin_country = “MMR” if all respondents were born in Myanmar, and origin_country = “Mixed” if there was more than 1 country of origin in the HH.

urban: 1 is urban, 0 is rural

ECU_2019_EPEC

1) We made this missing because the sampling strategy was considering the prevalence of Venezuelan migrants in host communities and area was not taken into account.

ETH_2017_SPS

1) We categorized everyone as 0 “rural” because the country team said of the survey: The camps were located in very remote areas, very close to borders generally, so would be classified more as remote/rural.
We also confirmed no respondents are in Addis Ababa.

IRQ_2015_SRHCS

1) We were not able to create this variable as there were no clear urban/rural identifiers, including in the variables GOVERNORATE, B1, and camp.

PER_2018_ENPOVE

1) All respondents are urban, since the survey focused on urban areas in certain cities (where the vast majority of migrants live).

UGA_2018_RHCS

1) Used urban categorization in file “host_ea_urban.xlsx” to merge by EA for hosts outside Kampala. In-camp HHs not in this file have urban = missing. Those in Kampala were always categorized as urban.

1.5 Housing

areaspace: Area dwelling in square meters

central_acc: Access to central heating

IRQ_2015_SRHCS

1) We used asset question D1 item 14 "home heating system" even though it does not specify central heating.

JOR_2016_SRHCS

1) We used asset question D1 item 14 "home heating system" even though it does not specify central heating.

cooksource: Main cooking fuel

dweltyp: Types of Dwelling

IRQ_2015_SRHCS

1) See “IRQ_2015_SRHCS dweltyp” attached.

elec_acc: Connection to electricity in dwelling

ETH_2017_SPS

1) We used C.12 "lighting" as a proxy for electricity.

UGA_2018_RHCS

1) We used HC18 "type of lighting" as a proxy for electricity.

elechr_acc: Electricity availability (hr/day)

IRQ_2015_SRHCS

1) We used C7 even though it's only for those who answered C5 "public network", "camp network", or "solar charger".

JOR_2016_SRHCS

1) We used C7 even though it's only for those who answered C5 "public network", "camp network", or "solar charger".

LBN_2015_SRHCS

1) We used C14 even though it is only for those who answered C14 "public network".

electricity: Access to electricity in dwelling

ETH_2017_SPS

1) We used C.12 "lighting" as a proxy for electricity.

electyp: Type of lighting and/or electricity

floor: Main material used for floor

gas: Connection to gas/Usage of gas

heatsource: Main source of heating

improved_roof: Used improved material for roof

improved_stove: Household has clean cook stove

TCD_2018_ECOSIT4

1) We used s12q02 item 11 "HH ownership of gas or electric stove (hotplate)" for this variable.

kitchen: Separate kitchen in the dwelling

lightsource: Main source of lighting

rentin_agriland: Rent in agri land

NER_2018_EHCVM

1) We assumed 16A.10 “What is the tenure of this plot?” = 3 “Farming” to be rent in agriland.

TCD_2018_ECOSIT4

1) We assumed 16A.10 “What is the tenure of this plot?” = 3 “Farming” to be rent in agriland.

rooms: Number of habitable rooms

NER_2018_EHCVM

1) We used s11q02 "number of rooms, excluding kitchens and bathrooms" as a proxy.

TCD_2018_ECOSIT4

1) We used s11q02 "number of rooms, excluding kitchens and bathrooms" as a proxy.

UGA_2018_RHCS

1) We used HC03 "rooms household uses for sleeping".

wall: Used improved material for external walls

1.6 Non-Labor Income

cash_assistance: Household received cash assistance from government, international agencies, NGOs, religious community, etc. in the last 30 days

BGD_2019_CBPSBL

1) We used H101 "in the last year" instead of the last 30 days.

BGD_2021_CBPSR3

1) For variable ss_assist_govt_type_r3, we assumed the response options should be categorized as cash, in-kind, or both as follows:
CASH: 2. Stipend for Secondary Education Female Students 3. Old Age Allowance;
IN- KIND: 1. Vulnerable Group Feeding (VGF) 6. Open market Sales (OMS) 7. Vulnerable Group Development (VGD);
BOTH: 4. Gratuitous Relief (GR) 5. General Relief (GR)-including special support during Covid-19.

IRQ_2015_SRHCS

1) In addition to using questions F1-F4 “in-kind food assistance (PDS)”, we used questions E1-6. However, E1-6 asks about assistance from government/NGO/WFP /UNHCR, but combines cash and in-kind into a single response. We used E1-6 for both cash_assistance and inkind_assistance.

JOR_2016_SRHCS

1) In addition to using questions F1-F4 “in-kind food assistance (PDS)”, we used questions E1-6. However, E1-6 asks about assistance from government/NGO/WFP /UNHCR, but combines cash and in-kind into a single response. We used E1-6 for both cash_assistance and inkind_assistance.

LBN_2015_SRHCS

1) In addition to using questions F1-F4 “in-kind food assistance (PDS)”, we used questions E1-6. However, E1-6 asks about assistance from government/NGO/WFP /UNHCR, but combines cash and in-kind into a single response. We used E1-6 for both cash_assistance and inkind_assistance.

NER_2018_EHCVM

1) The recall period used was 12 months instead of 30 days.

TCD_2018_ECOSIT4

1) The recall period used was 12 months instead of 30 days.

UGA_2018_RHCS

1) One of the variables we used is HP10A, which mixes cash and in-kind.

inkind_assistance: Household received in kind assistance from government, international agencies, NGOs, religious community, etc. in the last 30 days

BGD_2019_CBPSBL

1) We used SS01 "in the past year" instead of the last 30 days.

BGD_2021_CBPSR3

1) For variable ss_assist_govt_type_r3, we assumed the response options should be categorized as cash, in-kind, or both as follows:
CASH: 2. Stipend for Secondary Education Female Students 3. Old Age Allowance;
IN- KIND: 1. Vulnerable Group Feeding (VGF) 6. Open market Sales (OMS) 7. Vulnerable Group Development (VGD);

BOTH: 4. Gratuitous Relief (GR) 5. General Relief (GR)-including special support during Covid-19.

ETH_2017_SPS

1) We used E.4, G.7, and D.21 “source of free food”. However, note that question D.21 is variable D_10_free_main.

IRQ_2015_SRHCS

1) In addition to using questions F1-F4 “in-kind food assistance (PDS)”, we used questions E1-6. However, E1-6 asks about assistance from government/NGO/WFP /UNHCR, but combines cash and in-kind into a single response. We used E1-6 for both cash_assistance and inkind_assistance.

JOR_2016_SRHCS

1) In addition to using questions F1-F4 “in-kind food assistance (PDS)”, we used questions E1-6. However, E1-6 asks about assistance from government/NGO/WFP /UNHCR, but combines cash and in-kind into a single response. We used E1-6 for both cash_assistance and inkind_assistance.

LBN_2015_SRHCS

1) In addition to using questions F1-F4 “in-kind food assistance (PDS)”, we used questions E1-6. However, E1-6 asks about assistance from government/NGO/WFP /UNHCR, but combines cash and in-kind into a single response. We used E1-6 for both cash_assistance and inkind_assistance.

NER_2018_EHCVM

1) The recall period used was 12 months instead of 30 days.

TCD_2018_ECOSIT4

1) The recall period used was 12 months instead of 30 days.

UGA_2018_RHCS

1) We used HP10A which mixes cash and in-kind.

pension: Household received pension benefits in the last 30 days

BGD_2019_CBPSBL

1) We used H101 "in the last year" instead of the last 30 days.

BGD_2020_CBPSR2

1) The recall period in the question is, "current income sources".

BGD_2021_CBPSR3

1) The recall period in the question is, "current income sources".

ETH_2017_SPS

1) In addition to B.63, we used C.75 = 13 "Social Security (Pension and other)", which combines social security and pension.

NER_2018_EHCVM

1) The recall period used was 12 months instead of 30 days.

TCD_2018_ECOSIT4

1) The recall period used was 12 months instead of 30 days.

UGA_2018_RHCS

1) Question CB 01 in Section 7 (HSEC7B.dta) asks about 14 different sources of HH earnings (F is pension and social security, G and H are remittances).

However, it only has 10 variables (CB01__1, CB01__2, etc.) that are not well labeled. Each HH has values 1 2 or 3 to rank the 3 most important sources, and 0 for the sources that aren't in the top 3. However, the variables' labels are cut short, so we couldn't tell to which source each variable corresponds, since there are fewer variables than sources in the questionnaire. At the time of harmonization, we had received clarification from the country team. We were only able to create this variable based on P15 and RM1, but were not able to incorporate the CB 01 variables.
2) We used variables P15 and CEE03 that are 12 month recall.

remittances: Household received remittances in the last 30 days

BGD_2019_CBPSBL

1) We used H101 "in the last year" instead of the last 30 days.

BGD_2020_CBPSR2

1) The recall period in the question is, "current income sources".

BGD_2021_CBPSR3

1) The recall period in the question is, "current income sources".

ETH_2017_SPS

1) We used C.80 "remittances in the last 12 months", which does not strictly adhere to the 30 day recall specification for this variable.

NER_2018_EHCVM

1) The recall period used was 12 months instead of 30 days.

TCD_2018_ECOSIT4

1) The recall period used was 12 months instead of 30 days.

UGA_2018_RHCS

1) Question CB 01 in Section 7 (HSEC7B.dta) asks about 14 different sources of HH earnings (F is pension and social security, G and H are remittances). However, it only has 10 variables (CB01__1, CB01__2, etc.) that are not well labeled. Each HH has values 1 2 or 3 to rank the 3 most important sources, and 0 for the sources that aren't in the top 3. However, the variables' labels are cut short, so we couldn't tell to which source each variable corresponds, since there are fewer variables than sources in the questionnaire. At the time of harmonization, we had not received clarification from the country team. We were only able to create this variable based on P15 and RM1, but were not able to incorporate the CB 01 variables.
2) We used variables P15 and RM1 that are 12 month recall.

1.7 Water and Sanitation-Hygiene (WASH)

imp_san_rec: Improved sanitation facility

imp_wat_rec: Improved water

NER_2018_EHCVM

1) See “NER_2018_EHCVM imp_wat_rec” attached for frequency of dry and rainy season water supplies.
2) Niger has two relevant questions, one about dry season water supply (s11q27a) and one about rainy season water supply (s11q27b). We made imp_wat_rec = 1 "yes" when at least one variable is:

1 tap in the housing
2 Faucet in the yard / Concession
3 Neighbor's faucet
4 Public fountain / tap
7 Covered well in the yard / Concession
8 Well covered elsewhere
9 Drilling in the concession
10 Drilling elsewhere
14 Bottled water
16 Street vendor

and imp_wat_rec = 0 "no" when neither is one of the categories above, and both are among the categories below:

5 Open well in the courtyard / Concession
6 well open elsewhere
11 Furnished source
12 Undeveloped spring
13 River / River / Lake / Dam
15 Rainwater
17 Other

open_def: Open defecation

piped: Access to piped water

NER_2018_EHCVM

1) Variable s11q22 asks about piped water directly, but since there are differing dry and wet season access variables, we must use those dry and wet season variables instead.

TCD_2018_ECOSIT4

1) Variable s11q22 asks about piped water directly, but since there are differing dry and wet season access variables, we must use those dry and wet season variables instead.

sanitation_source: Main sanitation facility

shared_san_source: Sanitation source is shared with other households

shared_water_source: Water source is shared with other households

PER_2018_ENPOVE

1) The variable P110 "water source" = "well" and "other" are excluded since their shared status is unclear.

waste: Main types of solid waste disposal

water_source: Sources of drinking water (14 categories)

ECU_2019_EPEC

1) See “ECU_2019_EPEC water_source” attached.

NER_2018_EHCVM

1) There are dry and wet season variables. We categorized based on the lowest quality for each HH.

TCD_2018_ECOSIT4

1) There are dry and wet season variables. We categorized based on the lowest quality for each HH.

watertype_quest: Type of water questions used in the survey

2. INDIVIDUAL-LEVEL DATA

2.1 Sample

countrycode: Country code

countryname: Country name

hhid: Household identifier

int_month: Interview month

BGD_2020_CBPSR1

1) We don't know specific dates for each interview. We only know the interview date range for all interviews of 21 April-20 May 2020. All respondents were assigned 5 “May”.

TCD_2018_ECOSIT4

1) This survey had interview date for the national sample, but not for refugees or hosts.

int_year: Interview year

BGD_2019_CBPSBL

1) Some dates in variable starttime are incorrectly listed as being from 1923, but all interview dates are from 2019.

TCD_2018_ECOSIT4

1) This survey had interview date for the national sample, but not for refugees or hosts.

pid: Personal identifier

JOR_2016_SRHCS

1) This survey has 2 duplicate individuals. In each pair, we kept the observation with the most responses.

rep_year: Reporting year where majority of survey took place

strata: Strata of the survey

IRQ_2015_SRHCS

1) Since the sampling unit was HH, the strata must be defined at the HH level.

survey_name: Abbreviated survey name (CCC_YYYY_NNNN)

surveyprd: Survey period (start-end years)

weight: Household weight

BGD_2019_CBPSBL

1) We used aa_weight which is the adult weight.

ECU_2019_EPEC

1) We used peso2 which is the individual weight.

JOR_2016_SRHCS

1) The data does not have weights as it was not designed to be representative at any level.

2.2 Demographics

age: Age of individual (continuous)

BGD_20YY_CBPSR? (YY = year, ? = round: BGD_2019_CSPSR1, BGD_2020_CBPSR2, BGD_2021_CBPSR3)

1) We don't have the age of respondents in R?, but we know their age at BL. We added the time between BL and R? to their age (rounded to the nearest whole year). We used the following formula for age in R?:
age in R? = round((median date of the R? interviews - date of baseline interview)/365,1) + age at baseline.
Note: For the R1 interview dates we only have the range for all interviews (21 April - 20 May 2020), so we used May 5, 2020 as the date for all R1 interviews.
Note: Median date of baseline interviews: May 26, 2019

NER_2018_EHCVM

1) We used s01q04a “age” for those who reported their age. If that was missing, we used the formula age = interview year - birth year (s01q03c).

agecat: Age of individual (categorical)

birth_country: Country of birth

NER_2018_EHCVM

1) We used nation "What is your nationality?", since people refer to their birth country as their nationality.

PER_2018_ENPOVE

1) All the respondents should have been born in Venezuela, since the survey is focused on the Venezuelan population in Peru. Variable P302 (country of birth) uses some sort of numeric country codes, but we assumed the survey should only contain people born in Venezuela and assigned all respondents to birth_country = “Venezuela”.

displaced_forcibly: Person is forcibly displaced

BGD_2019_CBPSBL

1) We only used the most recent migration reason.

ECU_2019_EPEC

1) The responses to p02_21 were not all clearly “yes” or “no”. However, we decided to consider the following categories as forcibly displaced: 5 “food scarcity” ,6 “imminent health risks”. 8 “direct threat to leave home”, 9 “extortion”, 10 “presence of armed groups”, 11 “widespread violence/insecurity”, 12 “fear of being attacked or persecurted”, 13 “fear or forced recruitment”, 15 “they took me”.

ETH_2017_SPS

1) See “ETH_2017_SPS displaced_forcibly” attached. In addition, we also included the following responses from I.2: 1 “Armed conflict in my village” 2 “Armed conflict in nearby villages” 3 “Increased crime, violence and insecurity but not armed conflict” 4 “Ethnic/political/religious discrimination and persecutions”.

NER_2018_EHCVM

1) See “NER_2018_EHCVM displaced_forcibly” attached.

PER_2018_ENPOVE

1) The survey is for the Venezuelan population in Peru, classified as migrants or refugee-like by UNHCR. All respondents were forcibly displaced.

TCD_2018_ECOSIT4

1) See “NER_2018_EHCVM displaced_forcibly” attached, since TCD_2018_ECOSIT4 uses the same codes.

displaced_month: Displacement - month

BGD_2019_CBPSBL

1) We only used the most recent migration reason.

NER_2018_EHCVM

1) We used arrival date.

TCD_2018_ECOSIT4

1) We did not create this because the survey only has s01q20 "year of arrival", which does not specify month.

UGA_2018_RHCS

1) We used arrival date.

displaced_origin: Displacement - origin

BGD_2019_CBPSBL

1) We only used the most recent migration reason.

IRQ_2015_SRHCS

1) When a respondent was displaced from Iraq and another country, we assigned them 2 “from another country".

displaced_reason: Reason for displacement

BGD_2019_CBPSBL

1) We only used the most recent migration reason.

IRQ_2015_SRHCS

1) For respondents who were living in Syria in 2010 (A22 = 1) or Iraq (outside of Kurdistan) in 2013 (A23 = 1), we cateogorized their reason for displacement as “voilence”.

PER_2018_ENPOVE

1) The survey is for the Venezuelan population in Peru, classified as migrants or refugee-like by UNHCR. All respondents were assigned displaced_reason = "Venezuelan crisis".

displaced_year: Displacement - year

BGD_2019_CBPSBL

1) We only used the most recent migration reason.

NER_2018_EHCVM

1) We used arrival date.

UGA_2018_RHCS

1) We used arrival date.

female: 1 is female, 0 is male

hasID: Person has some form of ID

IRQ_2015_SRHCS

1) We used A17 "which ID does this person have?", which is only asked for displaced people.

JOR_2016_SRHCS

1) We used A27 "Does individual have any of these IDs?", which was only asked to displaced people.

PER_2018_ENPOVE

1) We did not create this because the only relevant question, 306 “ID to enter Peru”, is close but not sufficient.

TCD_2018_ECOSIT4

1) We created this even though the only relevant question, s01q06a "respondent has any form of ID", is only asked to the refugee and hosts sample. It is not asked to the general and nationally-representative sample.

idp: Person is displaced within country from locality of origin

BGD_2019_CBPSBL

1) All respondents are migrants, and are not IDPs. People born in Bangladesh (HM05 = 1) are still refugees because they were in Myanmar when they had to flee.

ECU_2019_EPEC

1) This survey did not include any IDPs, so all respondents were assigned 0 “no”.

ETH_2017_SPS

1) This survey did not include any IDPs, so all respondents were assigned 0 “no”.

JOR_2016_SRHCS

1) This survey did not include any IDPs, so all respondents were assigned 0 “no”.

LBN_2015_SRHCS

1) This survey did not include any IDPs, so all respondents were assigned 0 “no”.

PER_2018_ENPOVE

1) The survey is for the Venezuelan population in Peru, classified as migrants or refugee-like by UNHCR. It did not include IDPs, so all respondents were assigned idp = 0 “no”.

TCD_2018_ECOSIT4

1) In the welfare file "ehcvm_welfare_TCH2018.dta", if pop_group = 3-5 (host or Chadian) and forcibly displaced (harmonized variable displaced_forcibly = 1) then IDP = 1 “yes”. Otherwise IDP = 0 “no”.

UGA_2018_RHCS

1) This survey did not include any IDPs, so all respondents were assigned 0 “no”.

marstat: Marital status

migrant: Person is a migrant - lived in another locality

BGD_2019_CBPSBL

1) All respondents are migrants, and are not IDPs. Even if they were born in Bangladesh, we were instructed to code them as migrants since the survey only interviewed refugees.

PER_2018_ENPOVE

1) The survey is for the Venezuelan population in Peru, classified as migrants or refugee-like by UNHCR. All respondents were assigned migrant = 1 “yes”.

nationality: Citizenship

BGD_2019_CBPSBL

1) We used ethnicity (ls13_ethnicity: Bengali, or Rohingya) as a proxy for citizenship. For ethnicity = Rohingya, we created nationality = “Rohingya (stateless)” since they are not recognized as Myanmar citizens.

JOR_2016_SRHCS

1) Variable A16 is nationality and has labels 1-12 in the questionnaire, but has unlabeled values 1-300 in the data. We were not able to obtain the value labels for these 1-300 values in the data, so we just assumed values 1-12 match the questionnaire, and made nationality missing for the unlabeled responses 13-300.

PER_2018_ENPOVE

1) We did not create this because the only relevant question, 302 “country of birth”, is close but not sufficient.

reltohead: Relationship to the head of household

2.3 Education

ALL EDUCATION VARIABLES

IRQ_2015_SRHCS

1) The education questions were only answered by one "MAIN RESPONDENT ON BEHALF OF RANDOMLY SELECTED CHILD (BETWEEN THE AGES OF 4-18, LIVES WITH THE HOUSEHOLD)".

JOR_2016_SRHCS

1) The education questions were only answered by one "MAIN RESPONDENT ON BEHALF OF RANDOMLY SELECTED CHILD (BETWEEN THE AGES OF 4-18, LIVES WITH THE HOUSEHOLD)".

LBN_2015_SRHCS

1) The education questions were only answered by one "MAIN RESPONDENT ON BEHALF OF RANDOMLY SELECTED CHILD (BETWEEN THE AGES OF 3-18)".

dropout_age: Age person dropped out of school

BGD_2020_CBPSR2

1) Dropout age is not asked directly. We only know the dropout age for children whose parents answered e03a "When did {child_name} stop studying?" with resposnes 2 "6 months ago/during the lockdowns", 3 "1 year ago", 4 "2 years ago", 5 "3 years ago".

BGD_2021_CBPSR3

1) Dropout age is not asked directly. We only know the dropout age for children whose parents answered e03a "When did {child_name} stop studying?" with resposnes 2 "6 months ago/during the lockdowns", 3 "1 year ago", 4 "2 years ago", 5 "3 years ago".

dropout_primary: Person dropped out of primary school

BGD_2019_CBPSBL

1) It was impossible to identify currently enrolled students in their first year of primary school, so they were assigned a missing value for dropout_primary. There was a question about the highest class completed, but not a question about which class the student was currently attending.

dropout_reason: Reason for dropping out of school

BGD_2019_CSPSBL

1) We used ls19/ls53 = 2 "Age (too old/ too young)" as a proxy for dropout_reason = 2 "Considered too young".

ECU_2019_EPEC

1) We used p03_16 = 1 "Age" as a proxy for dropout_reason = 2 "Considered too young".

2) The data did not allow us to identify individuals who dropped out of primary school, so they were excluded from this variable.

NER_2018_EHCVM

1) The data did not allow us to identify individuals who dropped out of primary school, so they were excluded from this variable.

TCD_2018_ECOSIT4

1) The data did not allow us to identify individuals who dropped out of primary school, so they were excluded from this variable.

dropout_secondary: Person dropped out of secondary school

ECU_2019_EPEC

1) For those who are not currently studying and whose highest level completed is primary, we don't know if they started and dropped out of secondary school, so they are excluded from this variable.

ETH_2017_SPS

1) For people with B38 = "finished primary school" and who didn't finish any secondary school, we can't know for sure whether or not they started secondary school.

NER_2018_EHCVM

1) For those who are not currently studying and whose highest level completed is Primary, we don't know if they started and dropped out of secondary, so they are excluded from this variable.

TCD_2018_ECOSIT4

1) For those who are not currently studying and whose highest level completed is Primary, we don't know if they started and dropped out of secondary, so they are excluded from this variable.

edlev: Highest level of education attained

BGD_2019_CBPSBL

1) See “BGD_2019_CBPSBL education” attached.

ETH_2017_SPS

1) See “ETH_2017_SPS edlev” attached.

UGA_2018_RHCS

1) See “UGA_2018_RHCS edlev” attached.

educ_religious: Has some religious education

ECU_2019_EPEC

1) We made this missing for those attending a non-religious school at the time of the survey, as we did not know their previous education. This creates a biased sample, as we are including those currently in religious school but excluding those who are not.

2) In addition to individuals who reported currently attending a religious school, we included those who were asked about the main reason for choosing a school and responded with, “it’s a religious school.”

ETH_2017_SPS

1) We made this missing for those attending a non-religious school at the time of the survey, as we did not know their previous education. This creates a biased sample, as we are including those currently in religious school but excluding those who are not.

IRQ_2015_SRHCS

1) We made this missing for those attending a non-religious school at the time of the survey (K14!=4), as we did not know their previous education. This creates a biased sample, as we are including those currently in religious school but excluding those who are not.

JOR_2016_SRHCS

1) We made this missing for those attending a non-religious school at the time of the survey (K14!=4), as we did not know their previous education. This creates a biased sample, as we are including those currently in religious school but excluding those who are not.

NER_2018_EHCVM

1) We made this missing for those attending a non-religious school at the time of the survey, as we did not know their previous education. This creates a biased sample, as we are including those currently in religious school but excluding those who are not.

TCD_2018_ECOSIT4

1) We made this missing for those attending a non-religious school at the time of the survey, as we did not know their previous education. This creates a biased sample, as we are including those currently in religious school but excluding those who are not.

educat4: Highest level of education completed (4 categories)

BGD_2019_CBPSBL

1) See “BGD_2019_CBPSBL education” attached.

ETH_2017_SPS

1) See “ETH_2017_SPS educat7” attached.

educat5: Highest level of education completed (5 categories)

BGD_2019_CBPSBL

1) See “BGD_2019_CBPSBL education” attached.

ETH_2017_SPS

1) See “ETH_2017_SPS educat7” attached.

educat7: Highest level of education completed (7 categories)

BGD_2019_CBPSBL

1) See “BGD_2019_CBPSBL education” attached.

ETH_2017_SPS

1) See “ETH_2017_SPS educat7” attached.

educy: Years of completed education

noschool_reason: Reason for never attending school

BGD_2019_CSPSBL

1) We used ls19/ls53 = 2 "Age (too old/ too young)" as a proxy for noschool_reason = 2 "Considered too young".

read: Individual can read

school: Currently enrolled in or attending school

write: Individual can write

2.4 Labor

ALL LABOR VARIABLES:

IRQ_2015_SRHCS

1) The labor questions were only answered by one "RANDOMLY SELECTED INDIVIDUAL (BETWEEN THE AGES OF 15 AND 64, LIVES WITH THE HOUSEHOLD)". The questionnaire also says a second "RANDOMLY SELECTED INDIVIDUAL (BETWEEN THE AGES OF 15 AND 64) WHO ANSWERS SECTIONS MM THROUGH TT", but that data doesn't appear to be incorporated into the source data we used.

JOR_2016_SRHCS

1) The labor questions were only answered by one "RANDOMLY SELECTED INDIVIDUAL (BETWEEN THE AGES OF 15 AND 64, LIVES WITH THE HOUSEHOLD)".

LBN_2015_SRHCS

1) The labor questions were only answered by one "RANDOMLY SELECTED INDIVIDUAL (BETWEEN THE AGES OF 15 AND 64, LIVES WITH THE HOUSEHOLD)". The questionnaire also says a second "RANDOMLY

SELECTED INDIVIDUAL (BETWEEN THE AGES OF 15 AND 64) WHO ANSWERS SECTIONS MM THROUGH TT", but that data doesn't appear to be incorporated into the source data we used.

empstat: Employment status, primary job (7-day ref period)

healthins: Health insurance (7-day ref period)

LBN_2015_SRHCS

1) We used O10 "Does your work provide social insurance or health insurance?".

industrycat10: 1 digit industry classification, primary job (7-day ref period)

ALL SURVEYS

1) See “industrycat10 ISIC_4rev to GMD” attached.

industrycat4: 4-category industry classification, primary job (7-day ref period)

job_2: Has second job

ETH_2017_SPS

1) We did not create this because although B.74 is the relevant variable, it is not clearly a second job, as they could have been at different times.

NER_2018_EHCVM

1) We did not create this because although s04q28a asks about a secondary job, it is only about secondary jobs in the last 12 months, and only if the person is not currently employed (see filter before 4.27).

TCD_2018_ECOSIT4

1) We did not create these because although s4q28b asks about secondary job, it is only about secondary jobs in the last 12 months, and only if the person is not currently employed (see filter before 4.27).

UGA_2018_RHCS

1) We did not create this variable because although N08 "HH members involved in HH enterprise" could identify second jobs for some people, it is only asked to certain people: those whose main job was something else (identified in Section 8) AND also worked in the HH enterprise.

laborincome: Total annual individual labor income in all jobs, incl. bonuses, etc.

BGD_2020_CBPSR1

1) We could not create this because, although we do have main job income from L07, L08, and L10, we do not have income from bonuses, etc.

BGD_2020_CBPSR2

1) We could not create this because, although we do have main job income from L12 and L14, we do not have income from bonuses, etc.

LBN_2015_SRHCS

1) We used O14 "primary job income in past month" as there were no questions about any other income.

UGA_2018_RHCS

1) We do not have income from all jobs, we only have and only used LB5 "income from primary job for paid employees".

lfstat: Labor force status

linc_nc: Total annual wage income in all jobs, excl. bonuses, etc.

BGD_2020_CBPSR1

1) Income was only reported for the main job of wage workers. It was not reported for any additional jobs or those who are self-employed. We just used income for the main job of wage workers.

BGD_2020_CBPSR2

1) Income was only reported for the main job of wage workers. It was not reported for any additional jobs or those who are self-employed. We just used income for the main job of wage workers.

BGD_2021_CBPSR3

1) Income was only reported for the main job of wage workers. It was not reported for any additional jobs or those who are self-employed. We just used income for the main job of wage workers.

JOR_2016_SRHCS

1) The source income variables (O15 O30 P21 P30) all include bonuses, so we did not create this variable because we could not exclude bonuses.

LBN_2015_SRHCS

1) We used O14 "primary job income in past month" as there were no questions about any other income.

lstatus: Labor status (7-day ref period)

minlaborage: Labor module application age (7-day ref period)

BGD_20YY_CBPSR? (where YY = year, ? = round: BGD_2019_CSPSR1, BGD_2020_CBPSR2, BGD_2021_CBPSR3)

1) Minimum age should be baseline minimum age + time since interview for R1, R2, etc. We rounded the following to the nearest year, as minlaborage should only be whole numbers (years):
R? minlaborage = round((median date of the R? interviews - the median date of the BL interviews (26 May 2019))/365,1) + 16 (minaborage at BL)
For the R1 interview dates we only have the range for all interviews (21 April - 20 May 2020), so we used May 5, 2020 as the date for all R1 interviews.

ngo: Employed at an NGO

BGD_2020_CBPSR2

1) This question was only asked for wage workers, not self-employed or own-account workers.

njobs: Total number of jobs

BGD_2021_CBPSR3

1) This question was only asked for wage workers, not self-employed or own-account workers.

ECU_2019_EPEC

1) The relevant variables (4.06) only tell you if the respondent has 1, or 2+ jobs. But if the respondent has more than 1 job, we can't tell exactly how many they have. We assumed 2+ to be 2 jobs.

ETH_2017_SPS

1) We did not create this because although B.74 is the relevant variable, it is not clearly a second job, as they could have been at different times.

IRQ_2015_SRHCS

1) The relevant variables (O25, O32, P25, P32) only tell you if the respondent has 1, 2, or 3+ wage or HH enterprise jobs. But if the respondent has more than 2 jobs, we can't tell exactly how many they have. We assumed 3+ to be 3 jobs.

JOR_2016_SRHCS

1) The relevant variables (O25, O32, P25, P32) only tell you if the respondent has 1, 2, or 3+ wage or HH enterprise jobs. But if the respondent has more than 2 jobs, we can't tell exactly how many they have. We assumed 3+ to be 3 jobs.

LBN_2015_SRHCS

1) The relevant variable M4 only tells us if the respondent has 1 or 2+ jobs, but if the respondent has more than 1 job, we can't tell exactly how many they have. We assumed 2+ to be 2 jobs.

UGA_2018_RHCS

1) We did not create this variable because although N08 "HH members involved in HH enterprise" could identify second jobs for some people, it is only asked to certain people: those whose main job was something else (identified in Section 8) AND also worked in the HH enterprise.

nlfreason: Reason not in the labor force (7-day ref period)

occup: 1 digit occupational classification, primary job (7-day ref period)

BGD_2019_CBPSBL

1) See “BGD_2019_CBPSBL occup” attached.

PER_2018_ENPOVE

1) The variable P605C4 for question 605 has 4-digit codes; the four digits seem to match the ISCO-88 occupation codes (see https://warwick.ac.uk/fac/soc/ier/research/classification/isco88/). Based on that, we will map them to occup using the first digit.

occup_2: 1 digit occupational classification, secondary job (7-day ref period)

ocusec: Sector of activity, primary job (7-day ref period)

ocusec_2: Sector of activity, secondary job (7-day ref period)

t_wage_total: Annualized total wage for all jobs (7-day ref period)

BGD_2020_CBPSR2

1) We could not create this because, although we do have main job income from L12 and L14, we do not have income from bonuses, etc.

LBN_2015_SRHCS

1) We used O14 "primary job income in past month" as there were no questions about any other income.

UGA_2018_RHCS

1) We do not have income from all jobs, we only have and only used LB5 "income from primary job for paid employees".

unempl_reason: Reason for challenge in finding work

unempldur_l: Unemployment duration (months) lower bracket (7-day ref period)

BGD_2020_CBPSR2

1) This question was only asked for respondents who were not currently working but had worked since January 2020.

BGD_2021_CBPSR3

1) This question was only asked for respondents who were not currently working but had worked since January 2020.

ECU_2019_EPEC

1) We did not create this variable because the only relevant variable, 4.35 "time looking for a job", is not sufficient to determine unemployment.

ETH_2017_SPS

1) We assumed >1 year = 13 months.

unempldur_u: Unemployment duration (months) upper bracket (7-day ref period)

BGD_2020_CBPSR2

1) This question was only asked for respondents who were not currently working but had worked since January 2020.

BGD_2021_CBPSR3

1) This question was only asked for respondents who were not currently working but had worked since January 2020.

ECU_2019_EPEC

1) We did not create this variable because the only relevant variable, 4.35 "time looking for a job", is not sufficient to determine unemployment.

ETH_2017_SPS

1) We assumed >1 year = 13 months.

want_to_work: Individual is interested in working or setting up a business

BGD_2019_CBPSBL

1) We used LS27 “Did (...) look for work in the past 7 days?” = 1 “Yes” for want_to_work = “yes”, and LS 28 “Why was (...) not available/did not look for work?” = 10 “Not interested in work” for want_to_work = “no”.

whours: Hours of work in last week, primary job only (all ref periods)

JOR_2016_SRHCS

1) It is unclear if response M3 = "98" is actual hours worked or "refused to answer". We assumed it was actual hours worked.

LBN_2015_SRHCS

1) It is unclear if response M3 = "98" and "99" are actual hours worked or "refused to answer"/"missing". We assumed they were actual hours worked.

whours_all: Hours of work in last week, all jobs (all ref periods)

BGD_2019_CSPSBL

1) We used hours worked in 1st and 2nd jobs only, as no other hours were reported.

JOR_2016_SRHCS

1) It is unclear if response M3 = "98" is actual hours worked or "refused to answer". We assumed it was actual hours worked.

LBN_2015_SRHCS

1) It is unclear if response M3 = "98" and "99" are actual hours worked or "refused to answer"/"missing". We assumed they were actual hours worked.
2) We used M3 "hours worked in main job in last 7 days" as there were no questions about any other hours worked.