

# Codebook

## Ethnicity of Refugees (ER) Dataset

Seraina Rügger

Center for Comparative and International Studies, ETH Zürich, Switzerland

E-mail: [ruegger@icr.gess.ethz.ch](mailto:ruegger@icr.gess.ethz.ch).

October 10, 2018

*Version 2018.1*

## **Overview**

The Ethnicity of Refugees (ER) dataset encompasses the ethnic composition of refugee stocks between neighboring countries worldwide for the years 1975 to 2017. The dataset covers all refugee groups in neighboring countries and countries in proximity to each other (maximal distance between country borders  $\leq 950$  km) that consist of at least 2'000 refugees. Up to the three largest ethnic groups of each country-dyadic refugee population are identified and their share of the total refugee stock indicated. Information on countries of asylum and countries of origin as well as refugee stocks is based on data from the United Nations High Commissioner for Refugees (UNHCR, 2014) and the United Nations Relief and Works Agency for Palestine Refugees (UNRWA, 2010).

The definition of “ethnicity” is based on the Ethnic Power Relations (EPR-ETH) dataset (Cederman, Wimmer and Min, 2010; Vogt et al., 2015) that identifies all politically relevant ethnic groups in a country and records the level of access to state power by their representatives. The definition of ethnicity includes ethno-linguistic, racial and ethno-religious groups. Coders used the EPR-ETH dataset as a source to identify the ethnic groups living in a refugee sending country. In some cases, we identified refugees who belong to an ethnic group that is considered politically irrelevant and is, thus, not recorded in the EPR dataset, nevertheless this information was coded in the refugee dataset.

The Ethnicity of Refugees (ER) data is available on the GROWup data portal: <http://www.icr.ethz.ch/data/er>. When using this dataset in your research, please cite the following reference:

- Rügger, Seraina and Heidrun Bohnet. 2018. “The Ethnicity of Refugees (ER): A new dataset for understanding flight patterns.” *Conflict Management and Peace Science* 35(1): 65-88.

## **Description of variables**

- **CoA**

Description: Country of asylum

Type: String

Source: UNHCR, UNRWA

- **ccode\_coa**

Description: GWcode of the country of asylum

Type: Integer

Source: GW

- **CoO**

Description: Country of origin

Type: String

Source: UNHCR, UNRWA

- **ccode\_coo**

Description: GWcode of the country of origin

Type: Integer

Source: GW

- **year**

Description: Year

Type: Integer

- **totalrefugees**

Description: Total number of refugees and people in refugee-like situations

Type: Integer

Value range:  $\mathbb{N}$

Source: UNHCR, UNRWA

- **minimal\_distance**

Description: Minimal distance between country of origin and country of asylum in kilometers

Type: Integer

Value range:  $\mathbb{N}$

Source: Cshapes (Weidmann, Kuse and Gleditsch, 2010)

- **groupname1**

Description: Name of the first ethnic refugee group

If applicable groupname according to EPR-ETH list of ethnic groups  
99 "Missing"

Type: String

Source: Several, see variable sources

- **GWgroupid1**

Description: GWgroupid of the first ethnic refugee group

Type: Integer

Source: EPR-ETH

- **groupname2**

Description: Name of the second ethnic refugee group

If applicable group name according to EPR-ETH list of ethnic groups  
99 "Missing"

Type: String

Source: Several, see variable sources

- **GWgroupid2**

Description: GWgroupid of the second ethnic refugee group

Type: Integer

Source: EPR-ETH

- **groupname3**

Description: Name of the third ethnic refugee group

If applicable group name according to EPR-ETH list of ethnic groups

99 "Missing"

Type: String

Source: Several, see variable sources

- **GWgroupid3**

Description: GWgroupid of the third ethnic refugee group

Type: Integer

Source: EPR-ETH

- **groupshare1\_num**

Description: Categorical size of the first ethnic refugee group

1 "Dominant"

2 "Majority"

3 "Minority"

Type: Integer

Value range: [1,3]

- **group1share\_multiplier**

Description: Estimated share of the first ethnic refugee group

Type: Integer

Value range: [0.05,0.95], see page 6.

- **group1\_size**

Description: Estimated number of refugees belonging to the first ethnic refugee group

$\text{totalrefugees} * \text{group1share\_multiplier}$

Type: Integer

Value range:  $\mathbb{N}$

- **groupshare2\_num**

Description: Categorical size of the second ethnic refugee group

1 “Dominant”

2 “Majority”

3 “Minority”

Type: Integer

Value range: [1,3]

- **group2share\_multiplier**

Description: Estimated share of the second ethnic refugee group

Type: Integer

Value range: [0.05,0.95], see page 6.

- **group2\_size**

Description: Estimated number of refugees belonging to the second ethnic refugee group

$\text{totalrefugees} * \text{group2share\_multiplier}$

Type: Integer

Value range:  $\mathbb{N}$

- **groupshare3\_num**

Description: Categorical size of the third ethnic refugee group

1 “Dominant”

2 “Majority”

3 “Minority”

Type: Integer

Value range: [1,3]

- **group3share\_multiplier**

Description: Estimated share of the third ethnic refugee group

Type: Integer

Value range: [0.05,0.95], see page 6.

- **group3\_size**

Description: Estimated number of refugees belonging to the third ethnic refugee group

`totalrefugees*group3share_multiplier`

Type: Integer

Value range:  $\mathbb{N}$

- **sources**

Description: Sources

Reference to reports from the UNHCR, USCRI, several NGOs,  
conflict narratives, news articles and others

Type: String

- **quote**

Description: Quote or reference phrase for the coding decision

Type: String

## Estimation of ethnic refugee group sizes and examples

Since precise numbers of refugees from each ethnic group often are unavailable, we indicate whether a certain ethnic group within a refugee flow was dominant, a majority or a minority. Reports on refugees frequently give approximate evidence such as “more than 31’000 [people] from Afghanistan, mostly Hindus, fled to India during the rise of the Taliban in the 1990s” (U.S. Committee for Refugees and Immigrants, 2009). We applied the rule that if the refugee population consisted of one *dominant* ethnic group, then we multiplied the size of the refugee stock, i.e. the number obtained from the UNHCR, by the factor 0.95, since there is confidence that at least 95 percent of the refugees belong to the concerned group. If there was one *majority* ethnic group within the refugee population, we multiplied it by 0.65. If several ethnic groups were identified within a refugee movement, whereas the coding rules allow for maximally three ethnic groups, the multiplying factors were readjusted according to the rules displayed in Table 1. The total share is mostly below 1 in order to account for uncertainty.

**Table 1:** Share of ethnic refugee group within refugee stock

Largest group	Share in %	Second group	Share in %	Third group	Share in %
Dominant	0.95				
Dominant	0.95	Minority	0.05		
Dominant	0.9	Minority	0.05	Minority	0.05
Majority	0.65				
Majority	0.6	Minority	0.3		
Majority	0.6	Minority	0.3	Minority	0.05
Minority	0.3				
Minority	0.3	Minority	0.1		
Minority	0.3	Minority	0.1	Minority	0.05

The shares were defined according to the approximately 50 cases with precise information on the ethnic refugee group sizes: Dominant ethnic groups comprise on average 92 percent, majority groups 59 percent and minorities 19 percent of the refugee population. The case examples are reasonably random because they are from different years and different world regions, including Albania, Armenia, Bangladesh, Djibouti, Egypt, Ethiopia, Greece, Hungary, Iran, Jordan, Lebanon, Libya, Pakistan, Syria, Thailand and Uganda. For instance, the UNHCR (2003) mentions in a report that “the Iraqi refugees in the Syrian Arab Republic are mostly of Arab ethnicity, 70% are Shias originating from the southern part of Iraq, around 15% are Sunnis and the remainder are Kurds [and others]”.



## **UNHCR data - GW country-code inconsistencies**

In some cases, the UNCHR used different definitions of country borders or dates of independence than Gleditsch & Ward. These cases and how we treated them are listed below.

- French and British oversea territories coded as French or British.
- Macao: until 1999 Portugal, then China.
- Hongkong: until 1997 Great Britain, then China.
- Occupied Palestinian Territory has the Israeli GWcode.
- Tibetan refugees have the GWcode from China.
- Western Sahara has the Moroccan GWcode.
- Stateless refugees have no GWcode, they are excluded from the dataset.
- Refugees with country of origin “various” are excluded from the dataset.

## **References**

Cederman, Lars-Erik, Andreas Wimmer and Brian Min. 2010. “Why Do Ethnic Groups Rebel? New Data and Analysis.” *World Politics* 62(1):87–119.

UNHCR. 2003. “Preliminary Repatriation and Reintegration Plan for Iraq.” Online: <http://reliefweb.int/report/iran-islamic-republic/unhcr-preliminary-repatriation-and-reintegration-plan-iraq>. Accessed: 18.5.2015.

UNHCR. 2014. “UNHCR Statistical Online Population Database.” Online: [www.unhcr.org/statistics/populationdatabase](http://www.unhcr.org/statistics/populationdatabase). Accessed: 22.10.2014.

UNRWA. 2010. “Number of registered refugees.” Online: <http://www.unrwa.org/etemplate.php?id=253>. Accessed: 1.10.2010.

U.S. Committee for Refugees and Immigrants. 2009. “World Refugee Survey 2008.” Online: <http://www.refugees.org/article.aspx?id=2114&subm=19&ssm=29&area>About%20Refugees>. Accessed: 1.10.2010.

Vogt, Manuel, Nils-Christian Bormann, Seraina Rügger, Lars-Erik Cederman, Philipp Hunziker and Luc Girardin. 2015. "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Dataset Family." *Journal of Conflict Resolution* 59(7):1327–1342.

Weidmann, Nils B., Doreen Kuse and Kristian Skrede Gleditsch. 2010. "The Geography of the International System: The CShapes Dataset." *International Interactions* 36(1):86–106.