Online Appendix: Introducing the AMAR (All Minorities at Risk) Data

July 11, 2017

1 Introduction

This data appendix provides supplementary materials for the paper introducing the AMAR Data. First we include additional detail pertaining to the simulation. Next we provide additional detail on the sampling of Selection Bias groups and weighting of the cases. We include a list of all selection bias groups and the variables coded. The following section describes the variables coded for the entire sample frame. Finally we show how the AMAR sample frame compares to and can be linked to other data for further examination of sample issues and so that a researcher can use variables coded across data.

2 Simulation

As noted in the paper, when we focus on truncation for positive values of Risk, the sampling algorithm will drop more cases from the right tail of the Risk distribution when Conflict equals zero. In that scenario we see the distribution of Risk scores for Conflict = 0 shifting slightly left, and the distribution of Risk scores for Conflict = 1 shifting slightly right. Figure A1 shows the distribution of Conflict scores (0 and 1) over values of Risk for the true relationship of approximately 0. Figure A2 shows the distribution of Conflict scores (0 and 1) over values of Risk when the data are systematically truncated more from the right tail of the Risk distribution when Conflict is 0. Among the surviving cases this type of truncation induces a spurious positive relationship between positive values of Risk and Conflict.

Figure A1: Distribution of Conflict scores over values of Risk in the full data.

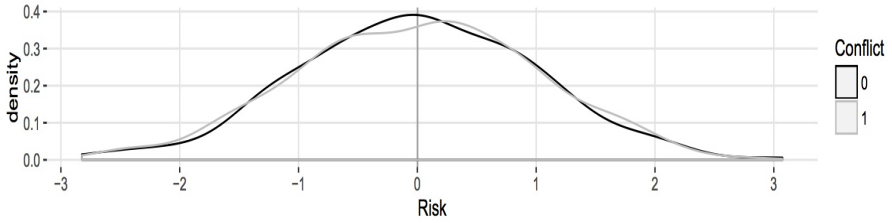
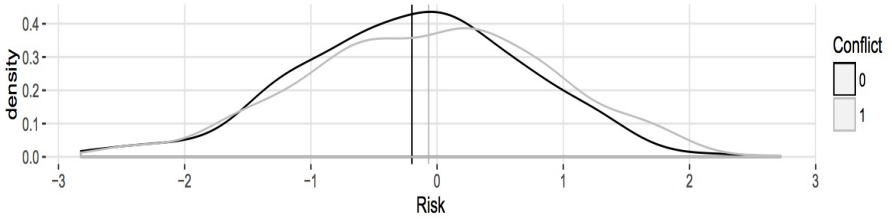


Figure A2: Distribution of Conflict scores over values of Risk in systematically truncated data.



3 Sampling Strategy

This section outlines the details of sampling strategy used to generate the 74 Selection Bias groups that make up sample segment II of the AMAR bias corrected sample data introduced in this paper (sample segment I consisting of the original MAR data).

3.1 Comparing MAR to the entire AMAR sample frame

The first task of this project was to compare MAR groups (AMAR sample segment I), to the AMAR sample frame to discover the nature of selection issues in the original MAR data. Tables A1 and A2 respectively show the number and proportions of the 1202 groups in the AMAR sample frame by region and size, and the number and proportions of 291 MAR groups as counted in the AMAR sample frame by region and size. In comparing the numbers in Table A1 and Table A2, oversampling in the MAR data of groups from Latin America and under-sampling of groups from Sub-Saharan Africa is especially notable. See

Table A1: Distribution of all AMAR sample frame groups by size and region

	Group <2%	Group 2-20%	Group >20%	Total
	137	71	28	236
Asia	11.4%	5.91%	2.33%	19.63%
	131	82	54	267
Europe	10.9%	6.82%	4.49%	22.21%
	13	3	2	18
India	1.08%	0.25%	0.17%	1.5%
Latin America and	18	32	33	83
the Caribbean	1.5%	2.66%	2.75%	6.91%
Middle East and	32	65	26	123
North Africa	2.66%	5.41%	2.16%	10.23%
	6	9	2	17
North America	0.5%	0.75%	0.17%	1.41%
Sub-Saharan	157	241	60	458
Africa	13.06%	20.05%	4.99%	38.1%
	494	503	205	1,202
Total	41.1%	41.85%	17.05%	100%

also a graphic display of this information in Figure 1 in the paper.

3.1.1 Comparing NEW and Selection Bias

So as not to replicate MAR (AMAR sample segment I) over and under-sampling of groups, in the Selection Bias random sample (AMAR sample segment II) of groups from NEW (groups not in MAR but listed in the AMAR sample frame), we drew a stratified random sample of 100 groups from NEW. The strata were defined by group size and region. The 100 initial groups for the Selection Bias sample were then randomly drawn from NEW in accordance with the proportions of groups in each cell as delineated by our strata and shown in Table A3.

Limitations in funding only allowed for the complete coding of the 40 most commonly

Table A2: Distribution of MAR groups by size and region, counted as 291 groups in the AMAR sample frame.

	Group <2%	Group 2-20%	Group >20%	Total
	16	28	7	51
Asia	5.5%	9.62%	2.41%	17.53%
	36	37	8	81
Europe	12.37%	12.71%	2.75%	27.84%
	11	2	0	13
India	3.78%	0.69%	0%	4.47%
Latin America and	9	18	7	34
the Caribbean	3.09%	6.19%	2.41%	11.68%
Middle East and	7	12	10	29
North Africa	2.41%	4.12%	3.44%	9.97%
	2	5	0	7
North America	0.69%	1.72%	0%	2.41%
Sub-Saharan	10	44	22	76
Africa	3.44%	15.12%	7.56%	26.12%
	91	146	54	291
Total	31.27%	50.17%	18.56%	100%

[a]

[a]The count of MAR groups is complicated but does not affect the sample correction because MAR groups constitute their own fully represented sample segment. In short there are 288 coded current MAR groups. Because of changes in group definition the MAR groups count as 291 in the sample frame.

Table A3: Distribution of all NEW groups by size and region

	Group <2%	Group 2-20%	Group >20%	Total
	121	43	21	185
Asia	13.28%	4.72%	2.31%	20.31%
	95	45	46	186
Europe	10.43%	4.94%	5.05%	20.42%
	2	1	2	5
India	0.22%	0.11%	0.22%	0.55%
Latin America and	9	14	26	49
the Caribbean	0.99%	1.54%	2.85%	5.38%
Middle East and	25	53	16	94
North Africa	2.74%	5.82%	1.76%	10.32%
	4	4	2	10
North America	0.44%	0.44%	0.22%	1.1%
Sub-Saharan	147	197	38	382
Africa	16.14%	21.62%	4.17%	41.93%
	403	357	151	911
Total	44.24%	39.19%	16.58%	100%

used MAR variables, annually from 1980 to 2006 for 74 of the 100 randomly drawn Selection Bias groups. Comparisons between groups in NEW in Table A3 and the 74 coded Selection Bias groups in Table A4 show that this limitation resulted in some over-representation of small groups from Asia in the Selection Bias sample segment and some underrepresentation of large groups from Europe when compared to proportions of groups in NEW. No groups from North America were coded in the Selection Bias sample segment but as Table 3 shows these constitute only around 1% of groups in NEW, and many groups from North America are already coded in MAR (AMAR sample segment I), so we don't consider this a cause for concern in the overall project.

Finally putting together the AMAR sample data that consists of 288 MAR groups (sample segment I, counted as 291 groups in the AMAR sample frame due to group splits)

Table A4: Distribution of Selection Bias groups by size and region

	Group <2%	Group 2-20%	Group >20%	Total
Asia	12	5	2	19
	16.44%	6.85%	2.74%	26.03%
Europe	6	5	1	12
	8.22%	6.85%	1.37%	16.44%
India	2	0	0	2
	2.74%	0%	0%	2.74%
Latin America and the Caribbean	1	1	3	5
	1.37%	1.37%	4.11%	6.85%
Middle East and North Africa	1	3	1	5
	1.37%	4.11%	1.37%	6.85%
Sub-Saharan Africa	15	12	3	30
	20.55%	16.44%	4.11%	41.1%
Total	37	26	10	73
	50.68%	35.62%	13.7%	100%

[a]

[a]Because we do not have reliable population information on the Makonde/Yao the number of groups in this table is 73 rather than 74

and 74 Selection Bias groups (sample segment II), Table A5 shows the distribution of the 364 [1] AMAR sample data groups by size and region. In comparing Table A1 and Table A5 it is clear that overall the AMAR sample data is not fully representative of the AMAR sample frame. Most notably Sub-Saharan Africa is still underrepresented and Latin America is still over-represented. This is as expected because the stratified random sample of Selection Bias groups from NEW was not intended to ameliorate sampling issues in MAR but was intended to constitute an unbiased sample segment representative of NEW. Consequently, the AMAR sample data uncorrected still reflects some of the biases in the original MAR data though these can now be accounted for and ameliorated. Finally, Table A6 lists the 74 selection bias groups coded in the AMAR sample. Table A7 lists the the variables coded annually for the entire AMAR sample.

3.2 Weighting the AMAR sample data

A final step, therefore, is to weight the sample segments (I and II) in the AMAR sample data to bring the weight of each segment in line with the weight that segment carries in the AMAR sample frame of the population. As we note in the paper weighting is common in survey analysis, where sample segments often over- or under-represent particular population segments (men and women for example). Weights are most commonly used in descriptive analysis but recent literature notes that weights can be used in inferential analysis also (Chromy and Abeyasekera 2005).[2]

[1] The total number of groups is 364 rather than 365 because 1 group - Scheduled Tribes of East India - was coded as a MAR group but coded again in the Sample Frame because of changing group configuration.

[2] Throughout the duration of this project we have experimented with a variety of different weights including straight cell weights, cell weights, and raking weights. By and large in comparison to analysis done on uncorrected samples (such as the original MAR) our observation is that sample correction in conjunction with inclusion of weights is helpful. It is less clear how the benefits and drawbacks of each weighting scheme

Table A5: Distribution of AMAR sample groups by size and region

	Group <2%	Group 2-20%	Group >20%	Total
Asia	28	33	9	70
	7.71%	9.09%	2.48%	19.28%
Europe	42	42	9	93
	11.57%	11.57%	2.48%	25.62%
India	12	2	0	14
	3.31%	0.55%	0%	3.86%
Latin America and the Caribbean	10	19	10	39
	2.75%	5.23%	2.75%	10.74%
Middle East and North Africa	8	15	11	34
	2.2%	4.13%	3.03%	9.37%
North America	2	5	0	7
	0.55%	1.38%	0%	1.93%
Sub-Saharan Africa	25	56	25	106
	6.89%	15.43%	6.89%	29.2%
Total	127	172	64	363
	34.99%	47.38%	17.63%	100%

[a]

[a]364 is the number of 291 MAR groups counted in AMAR sample segment I and 74 Selection Bias groups counted in sample segment II. For these, population data is missing for one Selection Bias group reducing the size of the sample to 363. Fully coded the MAR sample segment contains 288 coded groups because of differences in group definition. For these we have reliable population data for 280 groups.

Table A6: The 74 Selection Bias groups

Country	Group	Country	Group
Afghanistan	Aimaq	**Nepal**	Kirata/Kiranti/Kirati
Afghanistan	Brahui	**Nepal**	Sherpa
Angola	Nyaneka-Humbe	**Netherlands**	Frisians
Bangladesh	Garo	**New Zealand**	Asians
Bangladesh	Santals	**Nigeria**	Kamberi
Botswana	Kgalagadi	**Nigeria**	Plateau Chadic
Burkina Faso	Busansi	**Pakistan**	Seraiki/Saraiki
Burkina Faso	Songhay	**Papua New Guinea**	Kamano
Burma	Lahu	**Peru**	Asians
Cambodia	Chinese	**Russia**	Dargins
Chile	White/Mestizo	**Russia**	Kalmyks
China	Mongolian	**Russia**	Komi
Colombia	Mestizo/White	**Saudi Arabia**	Egyptians
Democratic Republic of Congo	Chokwe/Koko/Tshokwe	**Serbia**	Serbs
Democratic Republic of Congo	Logo/Logokuli	**Somalia**	Bantu (Non-Somali)
Democratic Republic of Congo	Lugbara	**Spain**	Valencian
Democratic Republic of Congo	Bemba/Shila	**Sri Lanka**	Sinhalese
Gabon	Kota	**Sudan**	Arab/Ja'Aliyin
Gambia	Fulani	**Swaziland**	Zulu
India	Scheduled Tribes Of East India	**Syria**	Druze
India	Syrian/Malabar Christians	**Tanzania**	Gogo
Indonesia	Pasemah	**Tanzania**	Iraqw, Mbulu
Jamaica	Mixed	**Tanzania**	Luo
Japan	Filipinos	**Thailand**	Thai
Kazakhstan	Tatar/Tartar	**Timor-Leste**	Papuan
Kazakhstan	Uighur	**Tunisia**	Berber
Kenya	Turkana	**Uganda**	Banyoro
Latvia	Lithuanian	**Uganda**	Nkole/Nkore
Latvia	Polish	**Uganda**	South Asians
Lebanon	Armenian	**United Kingdom**	Welsh
Macedonia	Turks	**Uruguay**	White/Mestizo
Malaysia	Orang Asli	**Vietnam**	Hmong
Mali	Fulani/Fulbe/Peuls	**Yemen**	Sunni Arabs
Mali	Mande	**Zambia**	Tonga-Ila-Lenje
Mali	Maures/Moors	**Zimbabwe**	Kunda/Seba
Mozambique	Makonde	**Zimbabwe**	Lozi
Namibia	Ovambo	**Zimbabwe**	Nyanja

Table A7. AMAR sample: Coded variables.

Variable Name	Variable Description	Variable Name	Variable Description
NUMCODE	Ethnic group case identifier (country code + group id)	**LEGISREP**	Group representation in legislative branch of central government
CCCODE	Country ID number: (The Correlates of War (Singer and Small) country identification number)	**EXECREP**	Group representation in executive branch of central government
COUNTRY	Country in which the group resides	**GUARREP**	Group is guaranteed representation in central government
REGION	AMAR regions	**POLGR**	Highest level of political grievance
YEAR	Year of Observation	**ECGR**	Highest level of economic grievance
GPOP	Group's population	**CULGR**	Highest level of cultural grievance
CPOP	Country's population	**KINSUP**	Any kindred group support
GPRO	Group proportion of country population	**KINMATSUP**	Kindred group material, non-military, support
LANG	Different language group	**KINPOLSUP**	Kindred group political support
CUSTOM	Different group customs (marriage, family, dress, etc.)	**KINMILSUP**	Kindred group military support
BELIEF	Different group religion	**STASUP**	Any foreign state or IGO support
RELIGS1	Specific religion: Plurality religion of group	**STAMATSUP**	Foreign state/IGO material, non-military, support
RACE	Different physical appearance	**STAPOLSUP**	Foreign state/IGO political support
GROUPCON	Group spatial distribution	**STAMILSUP**	Foreign state/IGO military support
GC119	Urban/rural distribution	**NSASUP**	Any non-state actor support
GC2	Regional base	**NSAMATSUP**	Non-state actor material, non-military, support
GC6B	Regional base--proportion of group members in regional base	**NSAPOLSUP**	Non-state actor political support
GC7	Proportion of group living outside regional base	**NSAMILSUP**	Non-state actor military support
GC10	Transnational dispersion -- kindred groups	**INTRACON**	Presence of intracommunal conflict
GC11	Transnational dispersion -- kindred groups in power	**FACTCC1**	Names of intracommunal antagonists with highest level of conflict
AUTLOST	Index of lost political autonomy, based on YEARWT, MAGN, PRSTAT	**FACTSEV1**	Severity of conflict for first pair of antagonists
YEARWT	Year of the most recent less of autonomy	**FACTCC2**	Names of intracommunal antagonists with second-highest level of conflict
MAGN	Magnitude of change	**FACTSEV2**	Severity of conflict for second pair of antagonists
PRSTAT	Group status prior to change	**FACTCC3**	Names of intracommunal antagonists with third-highest level of conflict
AUTONEND	Year/decade/century autonomy was lost	**FACTSEV3**	Severity of conflict for third pair of antagonists
TRANSYR	Year/decade/century transferred	**INTERCON**	Presence of intercommunal conflict
SEPX	Separatism index	**CCGROUP1**	Name of group with highest level of conflict
SEPKIN	Active separatism among kin groups	**CCGROUPSEV1**	Level of conflict with CCGROUP1
EMIG	Emigration for political or economic reasons	**CCGROUP2**	Name of group with second-highest level of conflict
DISPLACE	Internal displacement for political or economic reasons	**CCGROUPSEV2**	Level of conflict with CCGROUP2
POLDIS	Political discrimination index	**CCGROUP3**	Name of group with second-highest level of conflict
ECDIS	Economic discrimination index	**CCGROUPSEV3**	Level of conflict with CCGROUP3
CULPO1	Restrictions on religion	**PROT**	Protest
CULPO2	Restrictions on use of language or language instruction	**REB**	Rebellion
GOJPA	Group organization for joint political action	**REPGENCIV**	Repression of group civilian populations (those not engaging in violent or nonviolent political activities)
AUTON2	Group autonomy status	**REPNVIOL**	Repression of group members engaged in nonviolent collective action
AUTGAIN	Year group gained autonomy	**REPVIOL**	Repression of group members engaged in violent collective action
AUTPRO	Percentage of group in autonomous region		

Also as noted in the paper in defining the weights we follow the common strategy of accounting for the inversed sampling probability of an individual observation $W_i = \frac{1}{p_i}$. Furthermore, to account for stratification of the random sample of Selection Bias groups from NEW, the weight for each observation among the Selection Bias groups is defined as the inverse sampling probability of an individual observation $W_i = \frac{Ni}{n}$, where $n = \sum_{i=1}^{j} = n_i =$ total sample size (in this case the total sample segment II size because only one of the sample segments is stratified). N_i is the population size of stratum i, i=1,2,3...,j, and $N \sum_{i=1}^{j} = N_i =$ the total population size. Thus, groups in the Selection Bias sample segment differ according to the regional and population strata from which the observation was randomly drawn. For example, as shown in Tables A3 and A4 there are 147 groups in the population stratum of groups accounting for under 2% of a country's population in Sub-Saharan Africa. 15 Selection Bias groups were randomly drawn and coded from this particular stratum. Consequently, the inverse sampling probability weight of each Selection Bias group from this stratum is $W_{i(SSA<2\%=1)} = \frac{147}{15}$.[3]

4 AMAR sample frame coded data.

The third data contribution of this project mentioned in the paper consists of new AMAR variables that we coded for the entire AMAR sample frame of 1202 groups. These are

compare but this remains a topic for further study.

[3]This weighting procedure describes how all weights were assigned. As noted in the paper the analysis is conducted with non-politically dominant groups only as we are interested in examining the likelihood any non-politically dominant minority group engage the state in conflict and the causes of this type of conflict. The weights we use for the analysis of non-politically dominant groups are assigned the same way as weights for all groups, but the relevant numbers used for calculations account for minorities only. For example the NEW portion of non-dominant groups in the AMAR sample frame is 797 groups as 114 groups are the sole dominant groups in their country. The number of Selection Bias groups in this segment is 66.

of three kinds. The first set identifies the group and consists of variables already present in the MAR data that were expanded to account for all AMAR groups. The second set of variables is classificatory. These variables detail whether the case is an ethnic group that was originally in the MAR data or belongs to the set of NEW groups in the AMAR sample frame. Furthermore, a series of variables account for whether the original MAR cases changed in any way between the MAR data and the AMAR sample frame, either by splitting the group or merging with another group or changing the name of the group. Finally, a separate variable accounts for the 74 randomly selected groups that were fully coded for MAR variables.

The third set of AMAR variables listed in Table A8 is functional and is intended to encourage further analysis of the AMAR sample data in relation to the sample frame and to facilitate analysis across datasets. These include group proportion of national population for every group in the AMAR sample frame, and variables that account for whether the group has at any point been politically dominant. Next are variables specifically intended to facilitate analysis across data to ameliorate selection deficiencies and/or take advantage of alternate variables not coded in the AMAR project. Thus these variables account for whether the AMAR sample frame group has a match in Fearon (2003), Alesina et al. (2003), or in either of the 2 matched versions of EPR. Table A8 lists the variables coded for the entire AMAR sample frame. Detailed information on all of these variables is also included in the new AMAR Phase I codebook.

4.1 Using the AMAR sample data across datasets by way of the AMAR sample frame

One of the ideas motivating this project is to link the AMAR sample data to other data-collections on ethnic groups so that researchers can incorporate exogenous variables into

Table A8: Variables coded for the entire AMAR sample frame.

Variable Name	Variable Description
NUMCODE	Ethnic group case identifier (country code + group id)
AMAR GROUP	Full name of AMAR ethnic group
CCCODE	Country ID number (The Correlates of War (Singer and Small) country identification number
COUNTRY	Country in which the group resides
MAR PROPER	Group coded in MAR proper data (MAR Phase I-V)
SELECTION BIAS	Group coded in AMAR selection bias data
NAME CHANGED	Group's name changed from MAR to AMAR
PREVIOUS NAME	Name of group as appeared in MAR proper
SPLIT GROUP	Group split from MAR to AMAR
MERGED GROUP	Group merged from MAR to AMAR
ONE DOM GROUP	One politically dominant group in the country
ALL DOM GROUPS	Politically dominant groups
ALESINA MATCH	AMAR group matched to ethnic group in Alesina et al. 2003 data
EPR V1 MATCH	AMAR group matched to ethnic group in Ethnic Power Relations (EPR) v1 2010 data
EPR V3 MATCH	AMAR group matched to ethnic group in Ethnic Power Relations (EPR) v1 2010 data
FEARON MATCH	AMAR group matched to ethnic group in Fearon 2003 data
GPRO AMAR	Group proportion of country population for all AMAR groups

their analysis when using the AMAR sample data and use the AMAR sample frame to better understand and ameliorate selection issues across data on ethnic groups. Therefore we matched the AMAR sample frame by country and group name to groups in: Fearon (2003), Alesina et al. (2003) and EPR (2010, version 1.1) and (2012, version 3).

The results of our comparative linking to other data on ethnic groups, which is demonstrated in Table A9, suggest that there is substantial agreement about the configuration of socially relevant ethnic groups at the national level across datasets collected independently and using very different coding rules. The principal differences we found between these lists were due to differences in project definitions and/or objective, in aggregation, and in inclusion parameters. Despite these differences, we found that there was great overlap among the lists. Nearly all of the groups enumerated in these other lists were either in the main AMAR

list or listed as sub-groups of AMAR umbrella groups (for a suggestive list of subgroups see Birnir et al. 2015). Table A7 thus reveals that a loose set of commonly identified socially relevant groups is shared across datasets, even if the notion of an ethnic group is fuzzy (Birnir et al. 2015).[4]

Even though there is significant overlap between AMAR and these existing ethnic group data, due to differences in selection criteria the AMAR data contain substantially more ethnic groups. As seen in Table A9, in the case of Alesina et al., AMAR contains 477 more groups, in the case of Fearon, 410 more groups, and in the case of EPR v1.1 and v3.0, 502 and 500 more groups. We hope that researchers can use the more extensive AMAR sample frame and the bias corrected AMAR sample data, drawing on this frame to understand and ameliorate selection issues across data on ethnic groups.

[4] The total number of groups in the Fearon 2003 paper is 822. We received an updated version of the 2003 data from James Fearon, which consists of 858 groups; this version of the data was used for the AMAR match.

The total number of groups in AMAR is 1202. Of those 288 are current MAR groups that account for 291 AMAR groups. The AMAR lists groups in current nation states only.

Per EPR, version 1.1 "includes annual dataon over 733 groups" (at www.icr.ethz.ch/data/other/epr_old). However, the MASTER_EPR_v1.1 version of the data (at thedata.harvard.edu/dvn/dv/epr/faces/study/StudyPage.xhtml?globalId=hdl:1902.1/11796&tab=files&studyListingIndex=0_a777931694382ee99a6e0f5576cb) includes 731 unique groups after duplicates are removed. We also downloaded and removed duplicate entries for the EPR_groupyea˚_v1.1 version of the data, and found 728 unique groups (this version of the data is missing the Kpelle and Kru in Liberia and the Northern Hill Tribes in Thailand, which appear in the MASTER version of the data). Therefore, the total number of groups evaluated for this match is 731, and not 733.

If Alesina et al. (2003) or EPR subsumed groups in a category called other or omitted the group name we could not match those groups with named AMAR groups.

Table A9: comparing ethnic groups across Alesina, EPR, Fearon and AMAR.

		Alesina	EPR v1.1 2010	EPR v3.0 2012	Fearon
TOTAL IN LISTS	Total number of groups (AMAR 1202)	1054	731	758	858
TOTAL MATCHED	Total number matched with AMAR	693 66%	656 90%	677 89%	778 91%
Thereof	Full match (full congruence)	505 73%	513 78%	524 77%	637 82%
	Match, but at different level of aggregation (group is match to AMAR sub-group of aggregate group)	98 14%	98 15%	104 15%	65 8%
	Match, but group combined with another group into one aggregate in AMAR	45 7%	6 1%	6 1%	40 5%
	Match, but group is listed as two or more groups in AMAR	45 6%	39 6%	43 6%	35 4%
TOTAL NOT MATCHED	Total number not matched	361 34%	75 10%	81 11%	80 9%
Thereof	Not matched because don't meet AMAR population threshold criteria	80 24%	66 88%	67 83%	22 28%
	Not matched because don't meet AMAR ethnic criteria	6 2%	2 3%	3 4%	3 4%
	Not matched because lack of available data	4 1%	0 0%	0 0%	0 0%
	Not matched because countries don't meet AMAR population threshold, or former communist states not coded in AMAR	133 37%	6 8%	11 14%	55 69%
	Not matched because group names were not provided or group was in "other category"	129 36%	1 1%	0 0%	0 0%
	Not matched because coded in another country or for other similar reasons	9 2%	0 0%	0 0%	0 0%
TOTAL MATCHED + NOT MATCHED	Total AMAR matched + AMAR not matched	1054 100%	731 100%	758 100%	858 100%
TOTAL IN AMAR ONLY	Total groups in AMAR only	477	502	500	410

References

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, Romain Wacziarg. 2003. " Fractionalization," *Journal of Economic Growth*, Springer, vol. 8(2): 155-94.

Birnir, Jóhanna. K., Jonathan Wilkenfeld, James D. Fearon, David D Laitin, Ted R. Gurr, Dawn Brancati, Stephen M. Saideman, Amy Pate, Agatha S. Hultquist. 2015. Socially relevant ethnic groups, ethnic structure, and AMAR. *Journal of Peace Research.* 52 (1):110-115.

Chromy James R. and Savitri Abeyasekera 2005. "Chapter XIX: Statistical analysis of survey data." In *Studies in Methods Household Sample Surveys in Developing and Transition Countries Section E: Analysis of Survey Data. Department of Economic and Social Affairs Statistics Division.* New York: United Nations. At unstats.un.org/unsd/hhsurveys/ sectione_new.htm

EPR v1 and v3 at thedata.harvard.edu

Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth.* 8(2):195-222.