

This appendix provides additional information on our coding procedures. It then presents supplementary model results and visualizations of data from the RPGMD Version 1. We display our visualizations alongside data from the PGMD of Carey et al. (2013). The two data sources use different coding methods for counting PGMs, which readers should note as they consider the comparison. Despite the differences, however, the two data sources are generally consistent in identifying which states have smaller or larger numbers of PGMs.

## Additional Detail on the Coding Procedures

The main text describes the inclusion criteria for PGMs into the RPGMD and states which variables are coded for the analysis. While the main text discusses theories behind each variable, space considerations prevent us from fully discussing their operationalization. In this section, we therefore discuss how we operationalized the variables in our analysis.

### **Variables in the dataset:**

The RPGMD variables discussed in the main text are divided into four categories: ethnic relationships, alliance relationships, violence against civilians, and context. Additional variables are included for reference and ease of data analysis.

As we worked with the ACLED events data, it became clear that the same group was not always listed in one way. For example, the Patriots Militia in Algeria is included in ACLED as both “Patriot Militia of Algerian Government” and “Patriots Militia.” This is likely a result of different news sources referring to groups in different ways, which should not be surprising. ACLED may use multiple spellings for a single group as well, but we did not find this issue with

our PGMs. Some entries, however, were entered differently. Specifically, Janjaweed has an entry with an extra space that is listed separately from the rest of the events committed by the “Janjaweed.” Therefore, we chose to include “ACLED\_multiple\_entries,” “ACLED\_number\_entries,” and “ACLED\_listings” in order to be transparent about our coding decisions. These variables provide an indicator of whether we combined multiple ACLED entries, the number of ACLED entries we combined, and the specific listings that we used from ACLED, respectively.

Next, we include indicators of whether the PGM is an umbrella group or a subgroup of an umbrella group. These indicators allowed us to aggregate or disaggregate the umbrella groups for our analysis. We chose to disaggregate the umbrella groups. Subgroups were specified when it was clear that they had their own independent agendas. This distinction is most relevant for the Mayi Mayi in the Democratic Republic of the Congo. The Mayi Mayi have a lot of subgroups, so the choice of whether to aggregate or disaggregate the umbrella group has the potential to affect results. Supporting the robustness of our results, however, model specifications with umbrella groups aggregated and disaggregated yielded similar results, as Tables A1-A4 in this appendix show.

Several indicator variables also help quickly identify PGMs that appear in the PGMD only, the RPGMD only, or both. First, “pgm\_added” is an indicator of whether we added the PGM. When this is coded as a 1, the PGM is in the RPGMD only. Second, “pre\_1997” is an indicator of whether all of the group’s activity occurred before 1997. Since 1997 is the first year included in the RPGMD, all PGMs coded as a 1 for this variable are only in the PGMD. Third,

“overlap\_period” is an indicator of whether the PGM was active during the overlap period between the PGMD and the RPGMD of 1997-2007.

To address a concern with ACLED’s standard procedure of coding multi-day events as multiple events, we also created a list of ACLED violence against civilians (VAC) that extend over multiple days. This list is available upon request. Critics would suggest that these events are duplicated in ACLED, so we used our list of multi-day VAC events to create the “Dup\_VAC” and “VAC\_noDup” variables. These variables show which groups are involved in multi-day events and provide a measure of VAC where multi-day events are treated as a single event, respectively. Results are robust to using the “VAC\_noDup” variable as a dependent variable.

*Ethnic relationships:*

We then code characteristics of PGM ethnic relationships. This begins with a dummy variable for ethnic recruiting. This variable is in line with the Carey et al. (2013) dataset. For PGMs that engage in ethnic recruiting, we code them as one of two types. One type is *Defector PGMs*. This variable is coded as 1 when the PGM comes from the same ethnic group as the rebels it is fighting and 0 otherwise.<sup>1</sup> The second type is *Rival PGMs*. This variable is coded as 1 when the PGM recruits from a different ethnic group than the one from which rebels are

---

<sup>1</sup> This coding procedure is in line with Stanton’s (2015) approach.

recruiting and 0 otherwise. All other PGMs are coded as non-ethnic. In our dataset, 83 PGMs use ethnic recruiting, where 68 PGMs are *Rival PGMs* and 15 PGMs are *Defector PGMs*.<sup>2</sup>

The determinations for these coding decisions were made based on several types of sources. First, Google searches using varying combinations of PGM names and country names yielded media sources that provided initial background information about each PGM. This information was in some cases sufficient to determine whether the PGM had engaged in ethnic recruiting and then whether it was a *Rival PGM* or *Defector PGM*. Media sources are mostly English-language sources, but when investigating events in Francophone countries we occasionally used French language sources. When these media sources were insufficient to make these determinations, we consulted NGO reports and academic publications for further in-depth information. The most valuable NGO reports tended to come from Human Rights Watch, Amnesty International, Small Arms Survey, and the Enough Project. Academic publications from journals such as *African Affairs* were used in some other cases, such as our examination of the Civil Defense Forces in Sierra Leone (Hoffman, 2007).

---

<sup>2</sup> These numbers reflect the numbers when we disaggregate PGM umbrella groups into sub-groups (e.g. Mayi Mayi). For more on this see: Stearns J. (2011) *Dancing in the Glory of Monsters*, New York: Public Affairs branch of Perseus Books Group, Stearns J. (2012) From CNDP to M23: The Evolution of an Armed Movement in Eastern Congo. London: Rift Valley Institute, Stearns J. (2013) Raia Mutomboki: The flawed peace process in the DRC and the birth of an armed franchise. London: Rift Valley Institute.

*Alliance relationships:*

Carey et al. (2013) differentiate between informal and semi-official PGMs. Informal militias are more likely to commit human rights abuses than semi-official PGMs because they allow governments to use repression for strategic benefits while evading accountability (Carey et al., 2014). We include this variable, *Government Ties*, coded 1 for semi-official PGMs and 0 for informal PGMs. While government acknowledgment of its relationship with a PGM is important, it is also necessary to assess the strength of that relationship.

Recognizing the potential for PGMs to be unreliable allies for governments, we add a dummy variable coding PGMs as a 1 if they ever fight government forces and 0 otherwise. We refer to this variable as *Ever Fight Government*. The coding for this variable is based on the ACLED coding of event descriptions and actors involved in violent events. An event must list the PGM and government forces as opposing actors or state that the PGM is fighting government forces in an event description in order to code this variable as a 1.<sup>3</sup>

*Committed PGMs* were identified from our list of PGMs who had ever fought the government. PGMs are coded as committed if: (1) Their violent events in ACLED display distinct periods of being pro-government and anti-government; and (2) Secondary sources, following the previously specified procedure of Google searches for media sources and then consulting

---

<sup>3</sup> There are certainly PGMs with the potential to fight the government who do not do so.

However, this measure is valuable precisely because armed groups may exhibit fundamentally different behavior as a result of actually being on both pro-government and anti-government sides.

NGO reports and academic articles for further information, reveal that the group had a close link to a specific leader or ethnic group and that the leader or ethnic group was not willing to sever that link.

*Opportunist PGMs* were also identified from our list of PGMs who had ever fought the government. PGMs are opportunist if: (1) Their violent events in ACLED display distinct periods of being pro-government and anti-government; and (2) Secondary sources, following the previously specified procedure of Google searches for media sources and then consulting NGO reports and academic articles for further information, reveal that the group lacked a close link to a specific leader or ethnic group or that the leader or ethnic group was willing to sever that link.

*Autonomous PGMs* were also identified from our list of PGMs who had ever fought the government. PGMs are autonomous if their violent events in ACLED do not display distinct periods of being pro-government and anti-government. Instead, they target both pro-government and anti-government groups during the same time period.

PGMs can fall into more than one of these three categories. Out of the 149 PGMs in our dataset, 59 of them have ever fought the government. There are 11 *Opportunist PGMs*, 43 *Autonomous PGMs*, and 12 *Committed PGMs*. While the Opportunist and Committed categories contain relatively few PGMs, we contend that they remain substantively relevant. Furthermore, the large and statistically significant coefficient for the Committed category when *Levels of VAC* is the dependent variable—discussed in the main text and shown in Table 4—suggests that *Committed PGMs* are a very important PGM category.

*Measures of violence:*

We add a variable that captures the abusive behavior of pro-government militias: *violence against civilians*. This variable is constructed from ACLED Version 5. It only includes events that ACLED designates as violence against civilians. This variable is a count variable displaying over-dispersion (Long and Freese, 2014). While it is subject to a variety of reporting biases and urban bias (Davenport, 2010; Kalyvas, 2004), it does provide a valuable resource for comparing relative violence levels (Tarrow, 2008; Tilly, 2008). Since our dataset uses PGMs as the unit of analysis, we divide the total event count by the number of years in which the PGM was active from 1997-2014.<sup>4</sup>

To calculate the number of years in which the PGM was active, we count the number of years between the year in which the PGM's first and last activity occurred. The years in which the first and last activity occurred are included. For instance, if the PGM's first activity occurred in 1997 and its last activity occurred in 1999, then we code it as being active for three years. If the first and last activity occurred in the same year, then we code the PGM as being active for one year. If the PGM begins to operate as a rebel group, we code the last activity as the last activity while the group fought on the government's side. Thus, if a group fights on the government's side from 1997 to 1999, against the government from 2000 to 2003, and then on

---

<sup>4</sup> The cross-sectional nature of the Relational PGM dataset precludes us from analyzing temporal variation in PGM behavior, but it still allows us to analyze many aspects of the links between PGM-government and PGM-civilian relationships, context, and behavior.

the government's side again from 2004 to 2005, we code the group as being an active PGM for five years.

*Context:*

Finally, we include two contextual variables: regime type and state vulnerability. Regime type is an average of the country's polity IV score for the period under analysis. From 1997-2014, regime type exhibited minimal variation for most African states, so this measurement does not mask significant temporal variation. State vulnerability is a dichotomous measure of whether rebels have controlled territory within the country since the country obtained independence. PGM research is interested in the role of state capacity. Yet, we do not feel that traditional variables used to measure state capacity, such as GDP per capita, tax collection, infant mortality, and maternal mortality, adequately capture the components of state capacity that are most relevant for a government's decision to employ PGMs and a PGMs decision to carry out human rights abuses (Fearon and Laitin, 2003; Thies, 2004; Thies, 2005). Instead, we chose to focus on the government's perceived potential to lose control of territory. We measure this by assessing whether the government has ever lost control of part of its territory to rebels since gaining independence. Out of the 26 African states using PGMs, 18 states are coded as vulnerable.

## Supplementary Model Results

Table A1 and Table A3 show the results with PGM sub-groups aggregated into umbrella groups. Table A2 and Table A4 show the results with PGM umbrella groups disaggregated into sub-groups. Results are consistent across models using aggregated or disaggregated groups. Model 8 and Model 16 are the same models that are reported in Table 4 of the main text. Additionally, results are also robust to the exclusion of the Mayi Mayi, a particularly difficult case to code given confusion over whether several groups that identify as Mayi Mayi are rebel groups or PGMs.

These models also show that while the coefficients for the variables of whether the PGM ever fights the government are positive and significant across Tables A1-A4, disaggregating the measure into Committed, Opportunist, and Autonomous PGMs reveals that Autonomous PGMs are more likely to abuse civilians and abuse civilians at higher magnitudes. Committed PGMs are no more or less likely to carry out violence against civilians, but when they do they tend to do so at higher magnitudes. The lack of statistical significance for Opportunist PGMs with both *Presence of VAC* and *Levels of VAC* merits further study. Closer inspection may reveal that Opportunist PGMs display varying behavior from other types of PGMs on different measures, or it may be necessary to extend the geographical and temporal scope to include more cases.

**Table A1.** Marginal Effects from Logit Models of Presence of Violence Against Civilians -- Aggregated Groups

	Model 1		Model 2		Model 3		Model 4	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Vulnerable State	-0.28*	0.14	-0.32*	0.14	-0.26*	0.13	-0.30*	0.13
Ever Fight the Government	0.28**	0.09	0.29**	0.08				
Opportunist					0.05	0.20	0.01	0.19
Committed					0.10	0.15	0.07	0.15
Autonomous					0.35**	0.11	0.39**	0.11
Government Ties	-0.02	0.10	-0.02	0.10	-0.02	0.11	-0.04	0.10
Polity	-0.03**	0.01	-0.03**	0.01	-0.03**	0.01	-0.03**	0.01
Ethnic Recruiting	-0.05	0.09			-0.02	0.09		
Defector PGM			-0.16	0.13			-0.20	0.13
Rival PGM			-0.01	0.10			0.05	0.10
N	108		108		108		108	

+p<0.10, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001

**Table A2.** Marginal Effects from Logit Models of Presence of Violence Against Civilians-- Disaggregated Groups

	Model 5		Model 6		Model 7		Model 8	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Vulnerable State	-0.31*	0.14	-0.36*	0.15	-0.29*	0.13	-0.34*	0.14
Ever Fight the Government	0.28***	0.07	0.28***	0.07				
Opportunist					-0.03	0.15	-0.05	0.14
Committed					0.11	0.15	0.09	0.14
Autonomous					0.37***	0.08	0.38***	0.08
Government Ties	-0.06	0.10	-0.06	0.10	-0.01	0.10	-0.02	0.10
Polity	-0.03**	0.01	-0.04***	0.01	-0.03**	0.01	-0.03**	0.01
Ethnic Recruiting	-0.01	0.08			0.02	0.08		
Defector PGM			-0.17	0.13			-0.18	0.13
Rival PGM			0.02	0.08			0.07	0.08
N	149		149		149		149	

+p<0.10, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001

**Table A3.** Marginal Effects from Hurdle Models of Levels of Violence Against Civilians -- Aggregated Groups

	Model 9		Model 10		Model 11		Model 12	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Vulnerable State	-2.33	2.07	-2.86	2.06	-0.65	2.10	-1.45	2.03
Ever Fight the Government	6.67*	2.68	6.67*	2.62				
Opportunist					4.41	4.06	3.40	3.76
Committed					10.27*	5.23	9.41*	4.76
Autonomous					4.69*	2.27	5.51*	2.34
Government Ties	-5.17+	2.75	-5.14+	2.65	-4.85+	2.75	-4.86+	2.62
Polity	-0.12	0.19	-0.12	0.18	-0.13	0.19	-0.14	0.18
Ethnic Recruiting	-3.49+	2.06			-3.76+	2.21		
Defector PGM			-7.02*	3.33			-7.55*	3.47
Rival PGM			-1.88	1.86			-1.89	1.96
N	108		108		108		108	

+p<0.10, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001

**Table A4.** Marginal Effects from Hurdle Models of Levels of Violence Against Civilians-- Disaggregated Groups

	Model 13		Model 14		Model 15		Model 16	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Vulnerable State	-2.13	1.63	-2.74	1.63	-0.67	1.65	-1.44	1.59
Ever Fight the Government	5.15**	1.73	5.03**	1.65				
Opportunist					1.66	2.41	1.15	2.25
Committed					8.00*	3.65	7.63*	3.37
Autonomous					4.42**	1.59	4.57**	1.52
Government Ties	-4.58*	2.06	-4.42*	1.96	-4.16*	2.07	-3.93*	1.92
Polity	-0.14	0.14	-0.17	0.13	-0.11	0.14	-0.15	0.13
Ethnic Recruiting	-2.46+	1.43			-2.39	1.46		
Defector PGM			-5.77*	2.45			-6.23*	2.54
Rival PGM			-1.53	1.31			-1.29	1.30
N	149		149		149		149	

+p<0.10, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001

## Visualizations of RPGMD Version 1 (1997-2014) vs. PGMD (1997-2007)

Since there are many different PGMs included in the PGMD and RPGMD, we also include choropleth maps here to display the comparison between the two data sources for each year 1997-2007, 2008-2014 for the RPGMD, and country totals for the 1997-2007 overlap time period.

Broadly, the choropleth maps show that the data sources are pretty consistent in identifying which states have the most PGMs and the states have the fewest PGMs. The most notable differences are the PGMD's inclusion of more PGMs from Algeria, Libya, and Ghana and the RPGMD's inclusion of more PGMs from Somalia and South Africa. Other differences between the two data sources are minor.





































