

The MMAD Repressive Actors Dataset Codebook*

Espen Geelmuyden Rød, Jan Rustemeyer, and Sabine Otto

Uppsala University

Department of Peace and Conflict Research

Introduction

This document outlines the coding procedure and variables in the MMAD Repressive Actors Dataset (MMAD-RA). The MMAD-RA includes information on the actors of repression for every protest in the Mass Mobilization in Autocracies Database Version 1.0 (Rød and Weidmann, 2019)

The downloadable datasets are available in two formats: event report level and event level. The MMAD-RA only contains information on event reports in the MMAD v1.0 where a repressive actor is reported as present at a protest event.

Definition of protest in the MMAD

The MMAD defines a protest as ‘a public gathering of at least 25 people with an expressed political motivation either opposing or supporting a) central, regional, or local government, or b) other non-governmental institution’ (Weidmann and Rød, 2019, Chapter 4).

Universe of cases and definition of repressive actor in the MMAD-RA

We define repressive actors as organizations or groups with the capacity for violent repression of protest (e.g. police, military, militia). A repressive actor must be present at a protest event to be included in the data.

The universe of cases for MMAD-RA is protest events in the MMAD where repressive actors are coded as present at the event. We use the ordinal variable “level of official security forces engagement” from MMAD to delimit the universe of cases. Event reports where the variable takes the values 0 = explicit report of no presence, 1 = reports of presence, 2 = reports of physical intervention. Includes crowd dispersal, arrests, and beatings but excludes lethal intervention,

*Research funded by the Swedish Research Council grant nr. 2018-01222.

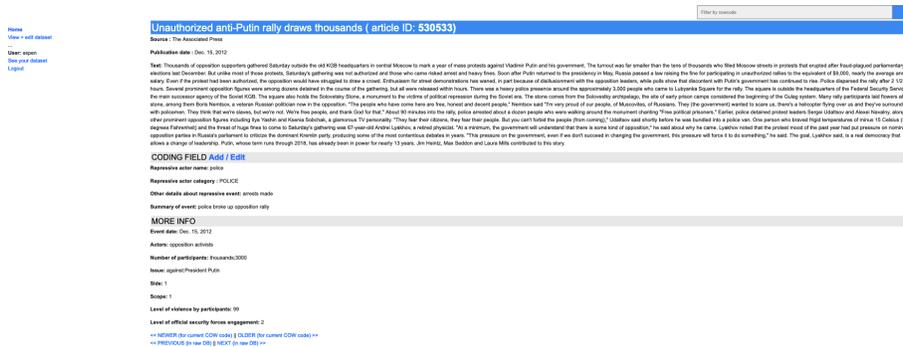


Figure 1: MMAD-RA online coding platform

and 3 = reports of lethal intervention are included in the MMAD-RA. Event reports are excluded if the variable takes the value NA = no report of official security forces engagement.

Sources and coding procedure

The MMAD-RA obtains information from previously coded event reports in the MMAD. An event report is an observation of a protest event reported by a media source. From the MMAD, event reports with official security forces engagement coded as present (1), as physical intervention (2), or as lethal intervention (3), are included in the MMAD Repressive Actors. The reports stem from three sources: Associated Press (AP), Agence France Presse (AFP) and BBC Monitoring (BBCM). Further details on the extraction of information for the MMAD are available in Weidmann and Rød (2019, Chapter 4) and the codebooks (<https://mmadatabase.org/>).

Online coding platform

All coding takes place on an online coding platform that displays the relevant MMAD event reports.¹ Event reports can be sorted by country. Coders are presented each article sequentially, ordered by date. Each article is presented along with the option of adding information on repressive actors based on it. Coders decide to add information on repressive actors or not (if the article does not report relevant information). Figure 1 shows the coding view for a single event report where information on repressive actors has been added.

¹The coding platform was developed by Andra Negus

Variables

Variables included from the MMAD The following variables from the MMAD are included in the data (Rød and Weidmann, 2019):

- Event record id (id)
- Event date (event_date)
- Location (location)
- COW code (cowcode)
- State name (state_name)
- Actors (actors_MMAD)
- Issue (issue_MMAD)
- Side (side_MMAD)
- Scope (scope_MMAD)
- Participant violence (part_violence_MMAD)
- Repressive actor engagement (sec_engagement_MMAD)
- Number of participants (num_participants_MMAD)

Repressive actor name (repressive_actor_name) (string) The name of the repressive actor as reported in the source. When the news report involves multiple repressive actors, we note down the different actors, using a semicolon to separate them. We do not include descriptive labels irrelevant to repression. For example, in report on a protest in Cameroon, **Cameroonian police** should be **police**, as the property **Cameroonian** is not relevant for the repression mentioned in the news report. Meanwhile, **traffic police** should be coded as **traffic police**, given that **traffic** (police) refers to a specific type of police unit, which might behave differently than regular **police** during repression.

Repressive actor dichotomous variables

- **Police and militarized police** (police and milpolice): We select police in case the actor is identified as the regular police force. In the cases where special police units such as riot/militarized police or gendarmerie are deployed, we code these as militarized police.
- **Military** (military): We code as military if the name of the repressive actor is part of the military. This includes names such as army, soldiers or military.

- **Militia** (militia): In reports where non-state actors are the repressive actor, we code the militia actor type. Repressive actor names such as militants, loyalists or (pro-...) fighters are included in this actor type. In instances where the report references a specific group (e.g., the Basij in Iran), we employ the Pro-Government Militias Database (Carey and Mitchell, 2017) to confirm if the actor is to be coded as the militia actor type.
- **Ambiguous security force** (secforc_amb): We code the actor type as ambiguous security force if the report mentions actors with names such as security forces or security personnel without providing more information on the repressive actor name.
- **Other** (other): We code as other when the report mentions actors that do not fit into the other actor type, such as border guards or mercenaries.
- **Missing** (actor_missing): In reports where the actor type cannot be determined we code as missing. For instance, these include cases where the report mentions how 'authorities' were present at a protest event.

Other details on repressive actor (other_details_on_repressive_actor) (string) Any other comments related to the repressive actor mentioned in the news report. If applicable, these include comments on the unit or organization the repressive actor belongs to and the actor's equipment and vehicles.

Number of people injured (number_of_people_injured) (string) Estimate of the number of people injured. Coded as an *integer number*, a *phrase* or left blank. When a number is provided, we code it in digits. If phrases such as "a couple", "dozens" or "hundreds" describe the number of people injured, we code the exact phrase. When news reports employ phrases such as **more than/ around / ca** (...a specific number), we code the specific number. In cases where no information is available, it is left blank. A semicolon (;) separates different estimates on the number of people injured when there is more than one number in the report. As such, we code the number of people injured in the following example as **30;100**:

• "Police charged batons at a rally of Jamaat-e-Islami at the north gate of Baitul Mukarram Mosque in the capital, leaving at least 30 Jamaat workers injured. ...However, Jamaat publicity secretary Tasnim Alam claimed that around 100 activists were injured in police beating at the rally that was organised by Jamaat Dhaka city unit" (BBC Monitoring South Asia, November 12, 2008).

Number of people killed (number_of_people_killed) (string) Estimate of the number of people killed. Coded in a similar way as the 'number of people injured' variable. Can be an *integer number*, a *phrase* or left blank. When a number is provided, we code it in digits. If phrases such as "a couple", "dozens" or "hundreds" describe the number of people killed, we code the exact

phrase. When news reports employ phrases such as **more than/ around / ca** (...a specific number), we code the specific number. A semicolon (;) separates different estimates on the number of people killed when there is more than one number in the report.

Other details about repressive event (other_details_about_repressive_event)

(string) If the news report mentions details on the repressive event, we add these under this variable. Among others, these details include the use of weapons, property destroyed and arrests made. When the news report states that people are injured/killed who represent a actor different to both the repressive actor and the protest actor coded in the MMAD, we add this here. The following example is an excerpt from a report describing an anti-regime protest in Myanmar: "Two people were arrested trying to flee after a 20-minute standoff with armed police and soldiers who reportedly used tear gas at the Strand Road in central Yangon, but another 5,000 monks and 15,000 citizens managed to march along Merchant Street in the downtown core, past the US Embassy and towards the Sule Pagoda before beginning to disperse" (BBC Monitoring Asia Pacific, September 26, 2007). Under the 'Other details about repressive event' variable for this event-report we add that arrests were made and tear gas was used.

Summary of event (summary_of_event) (string) A summary of the event, outlining the mobilization issue and the repressive actor(s) involved.

From words and phrases to numeric values

The two variables which measure the number of people injured and the number of people killed in an event-report often contain words or phrases describing the numeric value. In order to account for these instances in our data, we transform the word/phrase into a numeric value if a numeric estimate exists, following Merriam-Webster Dictionary. See Table 1 for an overview. When the word/phrase can take on multiple numeric values, we select the lowest number. For instance, the word 'several' translates to "an indefinite number more than 2 and less than many" (Merriam-Webster). Hence, we code 'several' as indicating 3.

Data citation

Espen Geelmuyden Rød, Jan Rustemeyer, and Sabine Otto. 2023. "Introducing the MMAD Repressive Actors Dataset". *Research and Politics*. In Press.

References

Carey, Sabine C and Neil J Mitchell. 2017. "Progovernment militias." *Annual Review of Political Science* 20:127–147.

String indicator	Numeric value
a couple	2
several	3
(a) dozen	12
two dozen	24
dozens	24
scores	40
hundreds	300
several hundred	300
a few	2
a number (of)	5
some	5
many, many more	25
numerous	25

Table 1: Overview of translation from words and phrases to numeric values.

Rød, Espen Geelmuyden and Nils B Weidmann. 2019. “Coding Instructions for the Mass Mobilization in Autocracies Database, version 1.0.”. Codebook.

Weidmann, Nils B and Espen Geelmuyden Rød. 2019. *The Internet and political protest in autocracies*. Oxford University Press.