Violent Non-State Actor Rivalry (ViNSAR) Dataset Codebook

Justin Conrad

University of Georgia

Kevin Greene

Princeton University

Brian J. Phillips

University of Essex

Contents

1. Introduction 1

2. Overview of Data 1

2.1. Unit of analysis 2

2.2. Dataset versions 2

2.3. Group inclusion criteria 2

2.4. Temporal inclusion criteria 3

2.5. Spatial inclusion criteria 3

2.6. Workflow and sources 4

3. Variables 5

3.1. Identifier variables 5

3.2. Field variables 5

3.3. Rivalry variables 6

3.4. Certainty variables 9

4. Acknowledgments 10

1. Introduction

This codebook describes the Violent Non-State Actor Rivalry (ViNSAR) dataset, created by Justin Conrad, Kevin Greene and Brian Phillips. When using this data, please cite the following paper:

Conrad, Justin, Kevin T. Greene and Brian J. Phillips. 2023. Introducing ViNSAR: Dyadic Data on Violent Non-State Actor Rivalry. *Journal of Conflict Resolution*.
https://doi.org/10.1177/00220027231208708

This project is supported by the National Science Foundation through award numbers 1658043 and 2001330.

2. Overview of Data

The Violent Non-State Actor Rivalry (VINSAR) dataset compiles information on intergroup rivalries and competitive behavior among non-state, militant groups. Existing quantitative studies of such competition have been limited by the availability of large-scale data. In measuring competition among militant groups, many studies have been forced to rely on simple proxy measures, such as the overall competitiveness of a country's political institutions, or simple counts of the number of organizations operating in a given theater. Some studies go further by capturing whether groups are involved in *any rivalry* in a given year, but even then, no specific information is available about the nature of those relationships.

We define rivalry as intergroup contestation over ideology, tactics and strategies, recruitment or other possible sources of disagreement. The VINSAR data offer several improvements over previous measures. First, information is coded at the *dyadic level*, indicating whether or not *specific pairs of groups* are involved in a rivalry. Second, rather than simply coding the existence of a rivalry, the VINSAR data provide detailed information on the rivalry, including the ideological relationship between each pair of organizations, as well as behavioral manifestations of the rivalry. Such manifestations can include verbal denunciations, threats of violence, or actual violence against competing groups. Third, the dataset provides this information for all pairs of groups in Africa and Asia for the time period 1990-2015. As such, the data captures changes in specific competitive relationships over time. Dyads frequently begin rivalries, intensify them, and end them within the time period covered by the dataset.

In addition to the original data collected and compiled by the project, the dataset includes a number of unique actor identifiers, which allow researchers to readily combine the data with other existing datasets on militant group characteristics and armed conflict dynamics. These

identifiers also allow researchers to subset their analysis on specific actor types, such as terrorist organizations or insurgent groups.

2.1. Unit of analysis

The unit of analysis in each version of the dataset is the *militant group dyad year* (e.g., Lord's Resistance Army (LRA)-Allied Defense Forces (ADF), 2004). The list of groups in each country was created by identifying the three types of militant groups active during the dataset's temporal period: 1) rebel/insurgent groups, 2) terrorist groups, and 3) pro-government militias.

2.2. Dataset versions

There are two versions of the dataset available in both Excel and Stata formats:

1. ViNSAR Basic Dataset – this version includes only the dyad years in which a rivalry was observed.
2. ViNSAR Full Dataset – this version includes all possible dyads in each country.

2.3. Group inclusion criteria

The first set of actors, rebel/insurgent groups, includes all non-state groups that were active in a conflict between 1990 and 2015 based on the Uppsala Conflict Data Program's (UCDP) Armed Conflict Dyadic Dataset (Harbom, Melander & Wallensteen 2008, Allansson, Melander & Themnér 2017). UCDP defines an armed conflict as a "contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths" (Gleditsch et al. 2002).

The second group of actors includes terrorist organizations identified by the Global Terrorism Database (GTD), which is compiled and managed by the National Consortium for the Study of Terrorism and Responses to Terrorism (START 2017). Groups are included in the VINSAR sample if GTD deems them responsible for ten or more terrorist attacks between 1990 and 2015. This eliminates a large number of transitory groups for which little information is available. We also exclude many GTD groups that are not "formal" groups, but represent broader groupings of individuals, such as "Students," "Hutus," and "rioters."

Finally, the list of pro-government militias (PGMs) is drawn from the Pro-Government Militias Database (PGMD) (Carey, Mitchell & Lowe 2013). Such groups are defined as organized, armed

groups that are "pro-government" or "sponsored" by the government that are also not part of the "regular" security forces (Carey, et al. 2013, p. 5).

2.4. Temporal inclusion criteria

Combining the lists of rebel/insurgent groups, terrorist organizations and PGMs provides an extensive list of possible combinations of militant groups. A potential rivalry occurs among any pair of groups that are active in the same country during the same year. For each group, the first year of activity is defined as the earliest year of activity identified in any of the three data sources. In the case of the UCDP data, this is the first year that a group meets the 25-death threshold. In the GTD, it is the first year in which a group committed a terrorist attack. And in the PGMD, it is the year listed as the date of the group's formation. For each group, the final year of activity is defined by the last year the group meets the 25-year threshold (UCDP), commits a terrorist attack (GTD) or for which there is evidence of activity (PGMD).

Important Note: Data on the activity of groups is available for the full temporal period for UCDP and GTD groups. The PGM data, however, is only available through 2007. As such, PGM groups are not included in the VINSAR data after 2007. Researchers using the VINSAR data need to consider this absence when deciding which observations/years to use in their empirical analysis.

2.5. Spatial inclusion criteria

A potential rivalry occurs among any pair of groups that are active in the same country during the same year. This means that if two groups are listed as active in the same country during the same year by *any* of the three original data sources (UCDP, GTD, PGMD), then they are a candidate for a rivalry to exist. The Full Dyadic Dataset lists all of these potential combinations.

Extraterritorial rivalries

This project is focused on identifying evidence of localized (i.e., within-country) rivalries. However, our group inclusion criteria means that some transnational or "extraterritorial" rivalries are captured. For instance, both the UCDP and GTD datasets list groups like al-Shabaab as operating in multiple countries (in the case of GTD, al-Shabaab is identified as the perpetrator of attacks in multiple countries). The VINSAR data, as a result, capture evidence of competition between al-Shabaab and other groups across multiple countries.

2.6. Workflow and sources

The data collection for this project was conducted by multiple pairs of undergraduate students, each supervised by a senior personnel member. Each team was given a unique country and tasked with coding all rivalry activities between violent groups within that state. To ensure the data collection is consistent across coders (and over time) and to make this process as clear and transparent as possible, coders were required to follow two important principles. First, coders worked independently of each other. This independence helped ensure that they collected as much unique information as possible. Each student's coding decisions were then compared at a weekly meeting, where a senior personnel member arbitrated any coding discrepancies. The second principle was that coders must maintain detailed notes as they collected and coded data. This included information about how they arrived at their coding decisions, as well as the relevant text and citations.

During their weekly meetings with senior personnel, students discussed their coding decisions. Coders were also encouraged to point out any interesting or unusual information they may have found during the data collection. After the coding for a given country was completed, a senior personnel member made the final coding decisions based on the coded data and notes from the undergraduate coders. To help ensure consistency and validity of the coding process, a random sample of the final coded data was evaluated by a member of the senior personnel and additional students who were not involved in initially coding those countries. They then compared their coding of the sample to the original coding. Any issues or inconsistencies were then resolved by the senior personnel.

3. Variables

3.1. Identifier variables

For each observation we include the year and name of the groups in each dyad. We include the GTD, UCDP or PGM identifier for a group whenever it appears in those original data sources. Please note: if the group does not appear in the given dataset in a given year, then it's corresponding identifier will be listed as "NA."

year: year of intergroup dyad

groupA: Name and/or abbreviation of Group A.

groupAGTD: Name and/or abbreviation of Group A in the GTD dataset.

groupAUCDP: Identifying group number of Group A in the UCDP dataset.

groupAPGM: Identifying group number of Group A in the PGM dataset.

groupB: Name and/or abbreviation of Group B.

groupBGTD: Name and/or abbreviation of Group B in the GTD dataset.

groupBUCDP: Identifying group number of Group B in the UCDP dataset.

groupBPGM: Identifying group number of Group B in the PGM dataset.

ccodeA: Country code for Group A location (from Correlates of War project).

ccodeB: Country code for Group B location (from Correlates of War project).

3.2. Field variables

primary_field: This variable contains the name of the primary "field" of the groups, regardless of whether there is a rivalry. A group's field captures the broader social movement to which it belongs, and therefore indicates a group's central motivation or ideology. Groups can be coded as one of the following eight categories:

- Pro-government
- Anti-government
- Ethnic
- Religious
- Islamist
- Left-wing

- Right-wing
- Other

If the group is not a part of a clearly identifiable field in the country or region, no field is listed for that group.

primary_kw: Provides keywords with additional information about the group's Primary Field. For instance, a group coded as "Islamist" may include the keyword "Sunni" to further distinguish it from other groups.

secondary_field: Many groups belong to multiple sets (representing a particular ethnic group but also leftist), so in these cases, a secondary field may also be listed. Refer to "Primary Field" for the possible categories.

secondary_kw: See Primary Keyword.

samefield: This dichotomous variable is coded 1 if the two groups in a dyad are from the same Primary Field, regardless of whether or not a rivalry occurs. The variable is coded 0 otherwise.

3.3. Rivalry variables

rivalry: This is the main variable, coded 0 if no rivalry is observed for the dyad-year (i.e., coders were unable to find evidence of a rivalry), and 1 if a rivalry is observed. In the Basic Dataset, all observations are coded as 1. A rivalry is present if there are verbal denouncements or threats between the groups in the dyad, or if there is violence between them. A dyad is *not* coded for a rivalry if the groups simply co-exist in similar geographic or ideological space, theoretically competing for resources. Rivalry is *observed* contestation, verbal or physical.

In some cases, direct evidence may exist that a rivalry occurred in years *y1* and *y3,* but not in *y2.* In these cases, we impute the existence of a rivalry in a "gap" year if sources reasonably imply that the rivalry continued. We also impute the level of rivalry as the highest level observed in the other years. These imputed observations are coded as a '5' on the certainty classifications (see below), so that they can be easily identified.

denounce: This binary variable is coded as 1 if either group or its representatives (e.g, leader or spokesperson) publicly criticized, condemned or denounced the other. Any message credibly attributed to the group or its members is included. This includes criticism of actions, policies, leaders, members and/or supporters linked to the other group.

denouncesym: This variable captures to what extent the rivalrous behavior is symmetric or asymmetric. Are both groups engaged in the behavior, or is it one-sided? Dyads not in a rivalry

are coded 0; if only group A engages in the behavior, the variable is coded 1; if only group B engages in the behavior, the variable is coded 2; if both groups are engaged in some degree of the behavior (i.e., symmetric), it is coded 3.

threat: This binary variable is coded as 1 if either group or their representatives (e.g, leader or spokesperson) publicly threatened violence against the other group. Non-violent threats, such as a threat to walk away from negotiations, are not included, but they may be codeable as denouncements. Threats can be directed towards leaders, members and/or supporters linked to a group. We also code a group as engaging in threats whenever violence is *implied* by their statement. For instance, if military commanders of a group publicly pledge to protect a village from another group, we consider this an implicit threat of violence.

threatsym: This variable captures to what extent the rivalrous behavior is symmetric or asymmetric. Are both groups engaged in the behavior, or is it one-sided? Dyads not in a rivalry are coded 0; if only group A engages in the behavior, the variable is coded 1; if only group B engages in the behavior, the variable is coded 2; if both groups are engaged in some degree of the behavior (i.e., symmetric), it is coded 3.

violence: This binary variable is coded as 1 if there is evidence of either group using violence against the other, or joint violence between the two groups. Violence can be committed against leaders, members and or/supporters of a group. In the latter case, evidence must exist that the supporters were targeted specifically because of their association with the group.

violencesym: This variable captures to what extent the rivalrous behavior is symmetric or asymmetric. Are both groups engaged in the behavior, or is it one-sided? Dyads not in a rivalry are coded 0; if only group A engages in the behavior, the variable is coded 1; if only group B engages in the behavior, the variable is coded 2; if both groups are engaged in some degree of the behavior (i.e., symmetric), it is coded 3.

deaths: Where available, this variable is the number of deaths reportedly resulting from rivalry-based violence in that year.

deaths_ordinal: This is an ordinal variable which is coded 0 if there is no evidence of deaths directly occurring as a result of intergroup competition. The variable is coded as 1 if evidence

suggests that between 1 and 49 people were killed. This includes any deaths attributable to intergroup conflict (as opposed to militant group-government conflict or other kinds of violence). This may include deaths of leaders, members and/or supporters of a group that allegedly resulted from intergroup competition. The highest value, 2, is reserved for cases where the evidence suggests that 50 or more people were killed as a result of intergroup competition. If there were conflicting estimates of deaths, the lower (minimum) count was used. Researchers should note that the "0" code indicating "no evidence of deaths" does not mean that deaths did not occur. Rather, it only means that the VINSAR coders were unable to find direct evidence of deaths. The *Deaths* variable does not identify the perpetrator or target of the fatalities - it is a dyadic-level count.

violencelocation: This is a qualitative measure indicating the location(s) where violence occurred, if the *Violence* variable is coded 1. Location can be listed as a particular city, cities, a state/province, or region ("the north of the country"). In some cases, the violence might be widespread throughout a country, but a more specific location is used when possible.

rivalrymax: This ordinal variable codes the maximum 'level' that the rivalry reached in the given year. It is coded 0 if there is no evidence of a rivalry present, 1 if either group denounced the other, 2 if either threatened the other, or 3 if there was intergroup violence. The group-year is coded for the highest level the rivalry achieved. For example, if there is evidence of denouncement and violence, the ordinal variable is coded 3 for violence. While this is an ordinal variable in the sense of increasing seriousness going from denouncement to threat to violence, note that the difference between categories are not even. In other words, a change from 0 to 1 is not the same as a change from 2 to 3.

extraterritorial: This dichotomous variable is coded as 1 if either group in the dyad was operating outside of its acknowledged home country. A home country is typically where the group was founded, regardless of the extent of its activities in other countries. Despite heavy activities in Mali, for instance, Algeria is considered the home country for Al Qaeda in the Islamic Maghreb. Nevertheless, some groups may be considered as having more than one home country. This is the case for groups operating in the Kashmir region, for instance.

3.4. Certainty variables

This is a set of variables measuring how certain the project team was about (1) the existence of the rivalry (**rivalrycertainty**) and (2) the rivalry level (**rivalrymaxcertainty**). If there are multiple reasons for low certainty, the variable is coded with multiple codes, separated by a comma. For example, if there is only one source and that source seems biased, the certainty variable is coded as "2,3". The certainty categories are nominal, and are as follows:

1: *Higher certainty*. The coding is based on two or more credible and independent sources such as academic articles, NGO reports, or media sources.

2: *Lower certainty (one source)*. The coding is based on only one source, but the source is credible.

3: *Lower certainty (bias)*. The coding is only based on a source or sources that are considered biased or non-independent. This category also captures codings that are based on information that is explicitly unverified.

4: *Lower certainty (identity)*. The coding is based on at least one credible source, but the source uses generic language so it is not clear if the text is referring to a specific group or groups.

5. *Lower certainty (timing)*. The start or end year of the rivalry is based on an imprecise source or when it must be inferred, for example one that indicates a rivalry began "in the 1990s" or "for several years."

4. Acknowledgments

We thank **Victor Asal, Karl Rethemeyer**, and **Suzanne Weedon Levy** for helpful discussions on the project and support with data.

The following individuals made significant contributions to the collection of this data. We are grateful for their support.

Melanie Bannout

Cheich Biteye

Stephanie Cannon

Samantha Daly

Qutaiba Enaya

Maria Garcia

Ahmad Hill

Nahim Kashani

Diana Maruri

Alexandria Norman

Austin Parker

Yesika Sorto Andino

Morgan Ward