Global Conflict Data: Introducing the World Handbook of Political Indicators IV

J. Craig Jenkins
Ohio State University

Charles L. Taylor
Virginia Tech University

Marianne Abbott
Ohio State University

Thomas V. Maher
Purdue University

Lindsey Peterson
Mississippi State University

ABSTRACT

The World Handbook of Political Indicators IV (WHIV) provides global cross-national daily coverage of contentious politics events – protest, violence, sanctions and relaxations – conducted by governments, insurgents, and civilians. Using auto-coding of Reuters international newswire based on KnowledgeManager,™ the dataset covers 231 countries and territories from January 1, 1990 through December 31, 2004 and includes 40 distinct event forms and a set of standard actors and targets resulting in 264,289 events. WHIV's breadth of geographic coverage, detailed atomic level event forms, and temporal specificity provides unique opportunities for studying contentious conflict globally and in specific countries. We discuss the uses of these data and the global distribution and trends in protest, political violence and governmental sanctions and relaxations.

Global Conflict Data: Introducing the World Handbook of Political Indicators IV

The use of political event data has exploded over the past two decades, as a result of new types of event data and methods for generating these data (Bernauer and Gleditsch 2012; Schrodt 2012). Recent event data projects have focused on episodic processes, such as the onset, lethality and duration of civil wars (Gleditsch et al. 2002), one-sided violence against civilians in civil wars (Eck and Hultman 2007), the behavior of states and other actors in international political crises (Wilkenfeld et al. 2010), and a combination of what we will call "atomic" level events and other discrete processes in civil war contexts (Raleigh et al. 2010). While this has generated a large literature on civil wars and related high-intensity armed conflict, it has left unaddressed other central questions, such as the driving forces behind low intensity protest and conflict, the impact of protest and political violence on democratization and political change, strategic decision-making in civil conflicts, factors influencing the selection of targets, and the dynamics of protest cycles and conflict escalation. Previous research focusing on these issues has typically focused on one country, movement or conflict at a time without considering how the processes listed above generalize or work differently across countries, regions, political regime types and the like. Addressing these questions requires a more comprehensive set of event data, including temporal data that can be integrated with non-contentious events. The *World Handbook of Political Indicators IV* (*WHIV*) dataset provides such a dataset by mapping a broad set of contentious politics events – including protest, violence, and governmental sanctions and relaxations – at the daily level with actor, target and location information in the tradition of the *World Handbook of Social and Political Indicators* (Taylor and Jodice 1983).

WHIV is a discrete daily event data set at the atomic level, by which we mean events involving actors, targets and event forms extracted from news reports on a daily 24 hour clock

basis. By *atomic*, we treat these events as the basic atoms of dyadic and multilateral political

interactions as described in these news stories. *WHIV* is derived from machine-coding of the

lead lines (first 250 words) of Reuters international newswire stories using the Virtual Research

Associates Knowledge Manager™ parser. It includes 40 event forms and covers 231 countries,

territories and world regions to identify 263,912 daily events between 1990 and 2004. Over 95

percent of these events are intra-state, i.e. the actor and target have the same national or

territorial location, but a significant share (almost 5 percent) involve transnational interactions

(i.e. different national or territorial locations for actors and targets). Other than transnational

terrorism, such transnational events are largely unstudied in existing literature. All events are

based on a *discrete* 24 hour clock day, corresponding to the normal reporting cycle of

international newswires. This means that extended campaigns and more complex *processes* (e.g.

seizures of major territory) that typically require more than a day to transpire are represented in

terms of discrete "single day" events as provided in newswire stories. The advantage of this

event data approach is that one can study the interaction among temporally proximate events in

detail using 24-hour time units. In addition to a daily event file, WHIV is also available from our

website (https://sociology.osu.edu/worldhandbook) and DataVerse https://dataverse.org/) in an

aggregated annual form and can be aggregated to fit other time units (e.g. weeks, months,

quarters), specific actors, targets and/or locations. *WHIV* events are compatible with the larger

IDEA event framework (Bond et al. 2003)[1] and thus can be integrated with IDEA measures of

non-contentious events. Less than 10% of the events in most newsfeed involve contentious

events. Reuters international newswire has the advantage during the time period in question of

being the world's largest English-language electronic newswire. It has consistent diction,

grammar and reporting structure, making it a strong basis for auto-coding.

This article first addresses the need for such a global contentious events data set. It then summarizes *WHIV*'s coding procedures and features for studying contentious politics. Finally, we discuss global trends in protest, political violence, sanctions and relaxations during the time period of January 1, 1990 to December 31, 2004. [2]

Why Contentious Event Data?

The growth of electronic news and text archives accessible through the internet coupled with the development of new machine-based search and coding methods has fueled an explosive growth of new event data projects (see: O'Brien 2010; Schrodt 2012; Hanna 2016; footnote #2 above). Recent data development has focused on particular countries or regions such as Africa (Raleigh et al. 2010; Salehyan et al. 2012), Northern Ireland (Loyle, Sullivan, and Davenport 2014), and the Middle East or "Levant" (Gerner et al. 1994; Schrodt and Gerner 1994), and critical issues such as water conflict (Bernauer et al. 2012; Gilmore et al. 2005), urban violence (Urdal and Hoelscher 2012), and anti-governmental protest and governmental instability (Banks and Wilson 2012). WHIV complements these efforts by providing global cross-national daily data on a broad range of contentious politics events. [3]

The key virtue of atomic event data is capturing the strategic interaction between groups. How else can one capture, for example, the decision by protestors to use violence to attack the government? Or the decision by authorities to use coercion or martial law to control protests? While one cannot directly tap the decision-making *per se,* one can obtain a useful set of proxy measures of strategic interaction based on news reports about contentious and other interactions. In conflict early warning research, such data are typically referred to as "dynamic" as opposed to structural measures (typically annual) that are often "too late" to provide useful early warning signal or aggregated to too coarse a time unit to be useful.

WHIV provides daily contentious event data at the country/territory level for 1990 through 2004 (see list below). Based on the larger IDEA event framework (Bond et al. 2003), *WHIV* uses 40 event forms tapping a variety of protests, political violence, governmental sanctions, and governmental relaxation events. Beyond more generic "protests," one can examine seven forms of protest (demonstrations, marches, altruistic protest, obstructive protest, defacements, support rallies, and strikes/boycotts), and twenty types of coercion and violence (e.g. riots, hostage seizures, political arrests, RPG attacks, aerial assaults, and so on). Following the WEIS tradition (McClelland and Hoggard 1969), event forms are formally independent of actors, although many events are in terms of natural language anchored in specific actors (e.g. "relaxations" refer to governmental actors). This broad set of event forms is of particular use when examining strategic interactions in public contentions where one is examining shifts in strategy by the involved actors. These events can also be ordered in terms of intensity and used to construct indices of conflict processes such as Goldstein (1992) weights and conflict carrying capacity (Jenkins and Bond 2001). One can also focus on particular types of actors, targets, and specific country/territorial locations. These data allow one to address such questions as:

- What types of protest and violence are used by challengers against governments and other institutional actors (Tarrow 2004; Barrie and Ketchley 2018)?

- What tactics (e.g. nonviolent protest, violent attacks) are more effective in producing social and/or political change? Does protest contribute to democratization (Kadivar and Caren 2015; Kadivar, Usmani and Bradow 2017; Kadivar and Ketchley 2018;)?

- What are the responses of authorities to various protest tactics? Do these vary by democratic or non-democratic regimes, styles of political control, temporal

5

sequence in conflict cycles, or other features of governmental legitimacy and institutions (Gilley 2006; Wada 2016; Underwood 2016; Costello 2018; Jenkins and Costello 2018)?

- Does harsh repression create a switch in protest tactics, e.g. from nonviolent to violent? Is this switch contingent on, e.g. identities of contenders, sequence in a conflict cycle, or the like (DeMerritt 2009; Rod and Weidmann 2013; Wada 2016)?

- Are human trafficking, refugee flights, other forms of crime, sexual violence, etc. linked to other types of violent civil conflict? How does political violence and other forms of contention affect economic growth and social disparities (Galster 2015)?

WHIV provides temporal flexibility. One can use daily data, or aggregate to other time units such as weeks, months, quarters and years. *WHIV* data files are provided publicly in daily and annual forms. Temporal flexibility is a major advantage when for examining the time horizons and response capacities of various actors. Rasler (1996), for example, used weeks to find that coercive state repression was initially negative on protest but, after three or more weeks, had an accelerator (i.e. positive) effect. With two or three weeks to respond, Iranian opposition groups in her study regrouped, forged new ties and mobilized new protests. For other groups and settings, the timing sequence may be different. As Shellman (2004) demonstrates, temporal flexibility is critical to testing alternative ideas and may produce different results depending on temporal aggregation. Different protest groups have different levels of tactical flexibility. Recent research has shown that the timing of repression has different effects depending on whether it occurs in a context of rising or declining dissent (Sullivan, Loyle, and Davenport

2012) and that closely timed action-reaction violence is more likely with ethnic groups that are being marginalized, in regions where poverty leaves few economic alternatives, and in urban areas where there is more robust communication networks associated with population and technology (Linke, Witmer, and O'Loughlin 2012). Using such flexibility in terms of timing allows one to make inferences about strategic calculations, such as military advantages versus retaliatory violence (Schneider, Bussmann, and Ruhe 2012).

Features of *WHIV* Data

WHIV is designed for the quantitative analysis of contentious dynamics, either by itself or in conjunction with other data. It is particularly well suited for global-comparative and time series analyses, using countries and territories as its geographic units.[4] In the following, we discuss key features of the data and how they are coded.

Event Forms. Table 1 provides the 40 *WHIV* event forms along with their frequency in the larger 15-year global dataset. *WHIV* includes only contentious events, i.e. interactions that stand outside conventional political behavior and entail an element of negative sanction and uncertainty of response (Bond et al. 2003). We also include relaxations as signals of reduced contention. These data are extracted by machine-coding, using grammatical patterns to identify actor, event form, and target along with location information and other derived features. *WHIV* events are the contentious subset of the broader IDEA event framework (Bond et al. 2003) and, as such, can be integrated with non-contentious IDEA measures and other data.[5] Events are grammatically identified by the parser, and then coded at the lowest possible level of specificity, creating a set of "cue" and "terminal level" events. For example, a protest demonstration (PDEM) is a cue category (a more general level of event form) for a reported action whose news description does not allow it to be classified at a more precise or terminal level (e.g. as protest

7

obstruction, protest march, protest defacement, or protest altruism). Despite the overlap, cue and

terminal level events are mutually exclusive in that they do not get coded in the affirmative

simultaneously (see Table 1 for all cue and terminal event forms).[6] In principle, event forms are

unlimited as to actor, although in practice some event forms (e.g. political arrests, crowd control)

are anchored in specific actors in natural language. Using the daily file allows the user to fine-

tune the selection of relevant events.

(Table 1 about here)

The most common *WHIV* events are physical assaults, small arms attacks and raids.

Support rallies are the most common protest event followed by strikes/boycotts while protest

altruism (mostly hunger strikes) is the least common. *WHIV* includes some events not covered in

other contentious events datasets, such as beatings and sexual assaults, which may not always be

political in the sense of either motivation or targeting but meet the basic criteria of

contentiousness and, in some broad sense, are political (i.e. power related). This opens up the

possibility of studying with event data methods collective rapes such as in the Balkans and

several African conflicts (Gardam and Charlesworth 2000; Gottschall 2004; Green 2004) and

street violence (e.g. brawls, murders) as a component of civil wars and other conflicts.

In addition to the meta-level categories of protest and violence, *WHIV* includes political

sanctions and relaxations. Sanctions include imposing restrictions (i.e. bans and martial law),

censorship, covert monitoring, and political arrests. Relaxations includes demining, easing military

blockades, relaxing curfews, releases and returns, relaxing censorship, releasing/returning persons

and property, and relaxing administrative sanctions. Sanctions and relaxations are less common,

with political arrests being the most common sanction, and return/release persons the most

common relaxation. Note that natural language associates these with governments and/or organizations that claim governmental authority.

Actors/Targets. The actor/target list ranges from generic categories (e.g. state sector) to specific groups (e.g. Christians, Muslims, Kurds). Using the *Wordnet* dictionary (Fellbaum 2010), the machine coder identifies a particular event form as well as the actor and target of the event. Multiple events may come from the sentence. In 7.88% of WHIV events, the actor is missing and, in 10.01%, the target. In news reportage, this is typical with some event forms (e.g. "A bomb exploded in a Sri Lankan market place today" provides no defined actor).[7] The daily event file provides specific actors while in the annual file events are aggregated to broad sector actors: (1) all state actors (government agents, judiciary, national executive, military, police, diplomats, etc.); and (2) all civilian actors (insurgents, political parties, religious groups, farmers, business, bankers and financiers, journalists, etc.). This means that missing actor events (which are rare) are excluded in annual files. *WHIV* also includes "unknown state" and "unknown civilian" actors where the parser cannot assign a more specific actor due to limits of the text but the sector (e.g. "civilian," "state") is evident. "Unknown actors" include missing information and other grammatical expressions, such as physical objects (e.g. buildings, bridges) that cannot be identified as to specific social sector. [8]

Time. *WHIV* currently covers the period from January 1, 1990 to December 31, 2004. Event date is based on an assumed 24 hour clock day derived from either the news story text or the date line of the news story minus 1 day. This means that some events (e.g. multi-day protests) may generate a second or third, etc. daily event based on multi-day news reportage while others get just one reported event but may have in fact lasted multiple days. We clean all multiple reports of events within 24-hour daily units based on locations. This likely introduces

9

little of an issue. As Francisco (2010: 79) reports for strikes, "the overwhelming majority last from 1 hour to 1 day." The same holds for the majority of protests, violence, political sanctions and relaxations, although it is important to recognize that some events (e.g. occupations, protest altruism) are often multi-day events (see Salehyan et al. 2012).

While most studies aggregate events to years or other larger time blocs, smaller time units are important for studying event processes, using, e.g. days (e.g. Khawaja 1993; Francisco 2009, 2010), weeks (Rasler 1996; Maher and Peterson 2007; Francisco 1996), months (Chang 2008), and quarters (Lee et al. 1996). Ultimately, the choice of appropriate time units is both theoretical (i.e. how long should a response involve?) and an empirical question (e.g. Shellman 2004; Freeman 1989).

Figures 1 and 2 provide a weekly aggregation of the global count of protest and violence (Figure 1), and a monthly aggregation of the global count of sanctions and relaxations in WHIV data (Figure 2). Violence displays greater volatility and higher counts, spiking upwards after the events of 9/11. Governmental sanctions and relaxations display greater trending. The count of relaxations is greater than sanctions, and has more spikes, particularly after 2003. While annual aggregation can be useful, Figures 1 and 2 show that aggregating also smooths what may be important temporal variation. This is an issue the analyst must make an empirically informed theoretical judgement about.

(Figure 1 and 2 about here)

Location. WHIV includes 231 countries, territories, and major world regions. Location is primarily determined by news information about the actor and target and, where unavailable as a second option, is inferred from the Reuters news report desk location. *WHIV* also includes specific territories that have been sites of contention (e.g. Corsica, Palestine, Northern Ireland)

along with major world regions (e.g. "Southeast Asia") where no further information is provided

about event location. It also contains transnational events (e.g. actor in Egypt, target in Libya).

In the daily version of the data, location is identified separately for actor and for target. In the

yearly data, location is based largely on target characteristics based on the rationale that one

knows where the, e.g. killing took place, even if you don't know the national identity of who did

it. If target location is not textually available, it is inferred from the actor location as a default. A

few events in the daily file are unlocatable and some involve transnational actor interactions with

no location in a conventional "nation" sense. While the overwhelming majority of events are

intra-state (i.e. locations of actor and target are identical), a growing share (roughly 5%) of all

events are transnational.[9] The Israel-Palestinian conflict accounts for a third of these

transnational events but, with that key conflict removed, there remains a slight uptick in

transnational contention. One should recognize that newswires do not report location as a

separate feature of events, meaning that we must rely on event and report desk reportage to infer

location. This is in fact what all analysts using news reports rely on.

(Figures 3 and 4 about here)

Figures 3 and 4 represent the average protest (Figure 3) and violence (Figure 4) per capita

for the entire period. Protest per capita (Figure 3) is greatest in western democracies. In South

America, Argentina, Suriname, French Guiana, Venezuela, and Uruguay have the highest rates

of activity. In Africa, South Africa, Zimbabwe, Cote d'Ivorie, and Morocco top the list and, in

Asia/Oceana, the Philippines, Afghanistan, Bhutan and Mongolia have the highest rates.

Political violence (Figure 4) is significantly greater and more concentrated in countries that have

experienced civil wars and international invasions (e.g. Iraq, Afghanistan, Algeria, Somalia, and

former Yugoslavia states). Overall, rates of violence are much higher than rates of protest with

peak figures between three and eighteen events per capita compared to between .25.and 1.9

protest events per capita. Most likely this follows from the "newsworthiness" of violence over

protest rather than its actual real-world incidence. Studies of news report bias have consistently

found event intensity and, in particular, violence is an important factor in news selection. The

distribution of protest and violence differ in notable ways that have the potential for fruitful

comparative analysis, and, potentially, point to interesting comparative temporal patterns as well.

Reuters News Count- A key question is the size of the "newshole" that is generating

contentious reports. Many analysts believe that this "newshole" is critical to allowing coverage

of contentious events (e.g. Oliver and Maney 2000; Earl et. al 2003). Newswires are often

thought to have a more flexible "newshole" but this should be tempered by awareness of the

reporting capacities of newswire bureaus where a daily quota of reports is often a typical reporter

assignment. To measure the "newshole," the daily *WHIV* provides a count of the total of daily

number of individual Reuters newstories by country-day which can be aggregated to the country-

year or other time unit. Thus, regardless of the number of contentious events, one can gauge the

number of underlying news stories. This is derived from the larger IDEA dataset, which is

similar to that used by King & Lowe (2003) but updated and extended with an improved parser.

This provides some gauge of the Reuters reportage and a way of controlling for "newshole"

effects on event reportage.

News Sources and Coding. As discussed, WHIV is auto-coded by the VRA Knowledge

Manager™ parser, drawing on the first two full sentences (basically 250 words) of Reuters

newswire stories. Beginning with the larger IDEA dataset, we extracted all contentious events

and cleaned the data for false positives, duplicates, and implausible events (e.g. weather "armed

attacks," business "coups"). The first cleaning step involved removing duplicate events

generated by the parser for interstate conflict events. For instance, if the Israeli military conducted an aerial assault on a Palestinian insurgent, the parser would generate a second event that was identical in every way (event ID, etc.) with the exception that the countries were flipped (i.e. a Palestinian aerial assault on Israelis). We systematically cleaned this modest number of duplicate events by identifying cases where cross-national events were reciprocal and temporally and substantively identical.

After cleaning systematically produced duplicates, we evaluated whether actor/target combinations were plausible for each event form (i.e. is it plausible that a business takes other businesses hostage [HTAK] or engages in protest obstruction [POBS]? Or does this appear more likely to be metaphorical language used by news writers?). We identified questionable actor/target combinations, and then, with the help of a sample of event lead lines and operational information about the parser, we determined whether the actor/target combination was plausible for each event form. We removed events conducted by, targeting, or between specific groups for three broad reasons. The majority of cleaning focused on language confusion and illogical actors. These are events where the parser confuses turns of phrase for contentious events. Politicians "assaulting" opponents, businesses taking other businesses "hostage," etc, were removed. Additionally, despite the parser's attempts to omit them, a number of athletic events made it into the raw data, which we removed. In each of these cases, we reviewed several lead lines in order to ensure that we were not omitting plausible events. For instance, the above example of businesses actors engaging in protest obstruction is a genuine example where angry Hungarian investors blocked traffic to protest government policies. We are not country experts, and so we did not review actor/target/event form pairings on a country by country basis. Instead, we left all plausible pairings in place, and so we encourage users to check the cleanliness of their data,

13

particularly if they are conducting single country and small N comparison. In sum, the cleaning process removed over 14 percent of all contentious IDEA events (6% were "false duplicates"). Internal reliability checks between computer and human coders offer further support.

How reliable is the *WHIV* coding? At its most refined terminal event level, we find positive intercoder reliability across the machine, three primary investigator coders, and four student coders (Krippendorf's Alpha=.68 [Jenkins, Taylor, Abbott and Maher 2018). When coder protocol was relaxed to a three-digit level mid-level, intercoder reliability improves considerably (Krippendorf's Alpha=.88) indicating that much of the inter-coder disagreement is over coding specific types of, e.g. protests, and not about whether there was protest. We explore these points of variability in more detail in Jenkins, et al. 2018.

What Value Added?

The chief virtues of *WHIV* are its global coverage and comprehensive set of atomic contentious events. With daily data, one can aggregate to various time units and, by using actor, target and location, identify specific subsets of contentious interaction. By integrating *WHIV* events with IDEA events, one can construct general conflict indices. Table 2 compares *WHIV* against other event datasets. All have particular advantages and can be combined for analysis. *WHIV* allows one to capture the strategic interaction between rival groups. This allows one to address questions such as when groups use particular types of contention, how contenders interact, and the escalation and de-escalation of conflict. Other conflict datasets have a different event focus (e.g. the onsets, duration and lethality of armed conflicts [Gleditsch et al. 2002), a narrower event framework (e.g. Raleigh et al. 2010; Banks and Wilson 2012]) or address a specific country or world region (e.g. Salehyan et. al, 2012). The newly released GDELT dataset (Leetaru & Schrodt 2013) has the advantages of the broad CAMEO event framework, multiple

news sources (Googlenews) and subnational geo-referencing but, as of yet, does not provide a

way to separate out the simple growth of international news reportage provided in Googlenews

from actual fluctuations in contentious events. Other international event data projects, such as the

afore-mentioned ICEWS [10] and the Phoenix Data Project[11] have developed highly sophisticated

programs for integrating hundreds of diverse sources (newswires, online papers, blogs, etc.) to

construct event data and are beginning to release their data for secondary analysis. There are

complexities with addressing news selection with these multisource products (Jenkins and Maher

2016) but they indicate the direction of future international event data development.

(Table 2 about here)

Conclusion

WHIV provides a unique opportunity for the study of contentious interactions between

contending groups in a global cross-national context. It builds on a machine coding system that

has been shown to be as accurate as human coding (King & Lowe 2003; Jenkins et al. 2018) and

involves additional cleaning for duplicates, false positives and other errors that further improve

its accuracy and recall. It strengthens our ability to study low-intensity conflict, strategic

interaction between contenders, processes of escalation and de-escalation, the effect of

contention on political and social change and a range of similar issues.

We envision a number of uses for the WHIV dataset by the social science community.

The most obvious use is for global comparative analysis of political processes. To date,

comparative analyses of dissent, repression, and intra-state conflict have had to choose between

limited geographic variability (Rasler 1996; Moore 1998), limited variability in regards to event

forms, and limited temporal variability (Davenport 1995; Carey 2010; Ortiz 2007). The

availability of WHIV opens opportunities for more detailed analyses of contentious processes,

and more rigorous hypothesis testing. Additionally, the detailed nature allows for improved modeling and testing of conflict escalation processes. Evaluating how changes in dissent and repression tactics (Lichbach 1987) happens at the daily, weekly, or monthly level, as well as how these processes vary across countries and regions offers a more fine-grained lens for evaluating how both war and peace breakout. *WHIV* also includes several unique event forms (e.g. sexual assaults, street assaults) which have been neglected in event analysis. Finally, we envision that this dataset will be useful for scholars interested in using contentiousness as an independent variable for analyses of a variety of processes. Political contentiousness has an impact on economic development, fiscal crises, democratization, and corruption as well as other important outcomes. Scholars may also be interested in the relationship between norm diffusion and state behavior. There are a number of additional potential uses of the WHIV data, and we hope that scholars find it a useful tool for pushing the field of international conflict forward.
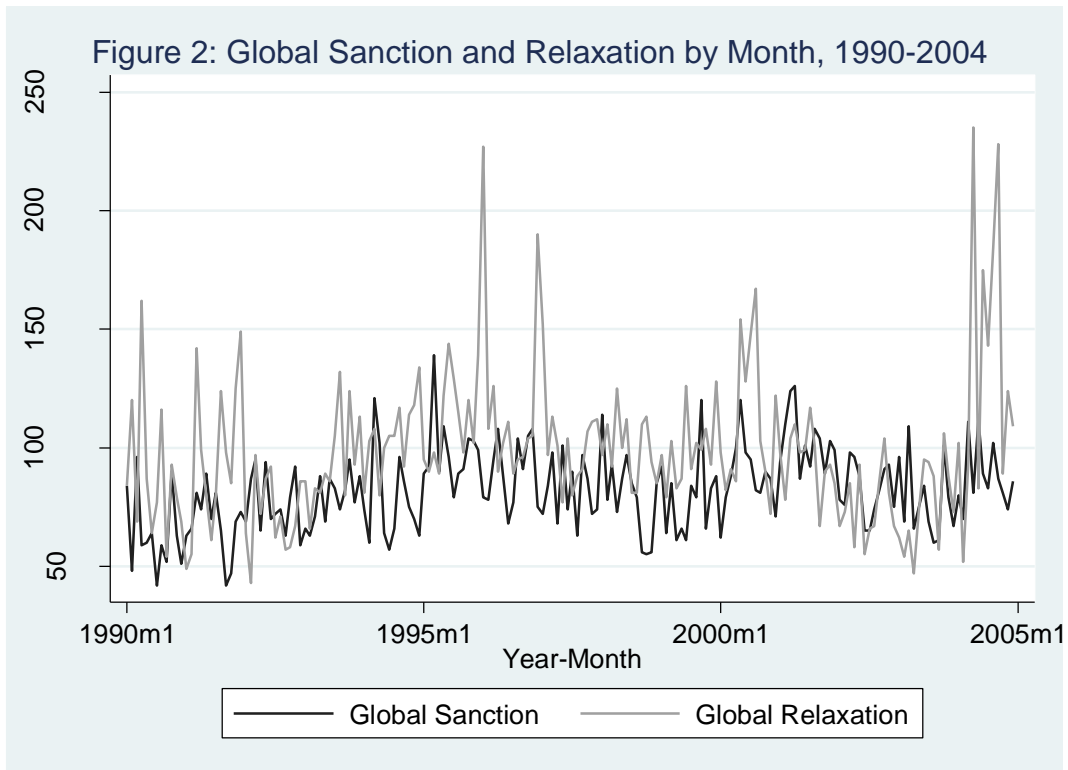
References

Banks, Arthur S. and Kenneth A. Wilson. 2012. "Cross-National Time-Series Data Archive."

Retrieved (http://www.databanksinternational.com).

Barrie, Christopher and Neil Ketchley. 2018. "Opportunity without Organization: Labour

Mobilization in Egypt after the 25th January Revolution." *Mobilization*

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2997517 (accessed 3.29.18)

Bernauer, Thomas et al. 2012. "Water-Related Intrastate Conflict and Cooperation (WARICC):

A New Event Dataset." *International Interactions* 38(4):529–45.

Bernauer, Thomas and Nils Petter Gleditsch. 2012. *New Event Data in Conflict Research*. Taylor

& Francis.

Bond, Doug, Joe Bond, Churl Oh, J. Craig Jenkins, and Charles Lewis Taylor. 2003. "Integrated

Data for Events Analysis (IDEA): An Event Typology for Automated Events Data

Development." *Journal of Peace Research* 40(6):733–745.

Costello, Matthew. 2018. "Oil and Gas Rents and Civilian Violence in the Middle East and

North Africa, 1990-2004: Resource Curse or Rentier Peace?" *Social Sciences* 7(3);

doi:10.3390/socsci7030039.

DeMerritt, Jacqueline Hope Rubin. 2009. Delegating Death: A Strategic Logic of Governmental

Killing. PhD Dissertation, Florida State University.

Eck, Kristine and Lisa Hultman. 2007. "One-Sided Violence against Civilians in War: Insights

from New Fatality Data." *Journal of Peace Research* 44(2):233–246.

Galster, Kirk. 2015. "Poverty and Conflict: Can Economic Development Prevent Conflict?"

Journal of Interdisciplinary Conflict Science I:1-35.

https://nsuworks.nova.edu/cgi/viewcontent.cgi?referer=https://scholar.google.com/&https
redir=1&article=1000&context=jics (accessed 3.29.18)

Gardam, Judith Gail and Hilary Charlesworth. 2000. "Protection of Women in Armed Conflict."
Human Rights Quarterly 22(1):148–166.

Gerner, Deborah J., Philip A. Schrodt, Ronald A. Francisco, and Judith L. Weddle. 1994.
"Machine Coding of Event Data Using Regional and International Sources."
International Studies Quarterly 91–119.

Gilley, Bruce. 2006. "The Meaning and Measure of State Legitimacy." *European Journal of
Political Research* 45:499-525.

Gilmore, Elisabeth, Nils Petter Gleditsch, Päivi Lujala, and Jan Ketil Rød. 2005. "Conflict
Diamonds: A New Dataset." *Conflict Management and Peace Science* 22(3):257–272.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and H\a avard
Strand. 2002. "Armed Conflict 1946-2001: A New Dataset." *Journal of Peace Research*
39(5):615–637.

Goldstein, Joshua S. 1992. "A Conflict-Cooperation Scale for WEIS Events Data." *Journal of
Conflict Resolution* 36(2):369–385.

Gottschall, Jonathan. 2004. "Explaining Wartime Rape." *Journal of Sex Research* 41(2):129–
136.

Green, Jennifer L. 2004. "Uncovering Collective Rape: A Comparative Study of Political Sexual
Violence." *International Journal of Sociology* 34(1):97–116.

Hanna, Alex. 2016. Automated Coding of Protest Event Data: Development and Applications.
PhD Dissertation. University of Wisconsin-Madison.

Jenkins, J. Craig and Doug Bond. 2001. "Conflict-Carrying Capacity, Political Crisis, and

    Reconstruction A Framework for the Early Warning of Political System Vulnerability."

    Journal of Conflict Resolution 45(1):3–31.

Jenkins, J. Craig and Matthew Costello. 2018. "Oil States, State Capacity and Rebellion: Does

    Rentierism Breed Rebellion or Peace?" Dept of Sociology, Ohio State University.

Jenkins, J. Craig, Charles L. Taylor, Marianne Abbott, and Thomas Maher. 2018. "Coding

    Reliability in a Large Autocoded Event Dataset: The World Handbook of Political

    Indicators IV Dataset." Dept. of Sociology, Ohio State University.

Kadivar, Mohammad Ali and Neal Caren. 2015. "Disruptive Democratization: Contentious

    Events and Liberalizing Outcomes Globally, 1990–2004." *Social Forces* 94(3):975–996.

Kadivar, Mohammad Ali, Adaner Usmani, Benjamin Bradlow. 2017. "The Long March:

    Contentious Mobilization and Deep Democracy." Research Paper #2017-38. Watson

    Institute for International and Public Affairs, Brown University.

Kadivar, Mohammad Ali and Neil Ketchley. 2018. "Sticks, Stones and Molotov Cocktails:

    Unarmed Collective Violence and Democratization." Watson Institute for International

    and Public Affairs, Brown University.

Linke, Andrew M., Frank D. W. Witmer, and John O'Loughlin. 2012. "Space-Time Granger

    Analysis of the War in Iraq: A Study of Coalition and Insurgent Action-Reaction."

    International Interactions 38(4):402–25.

Loyle, Cyanne E., Christopher Sullivan, and Christian Davenport. 2014. "The Northern Ireland

    Research Initiative: Data on the Troubles from 1968 to 1998." *Conflict Management and
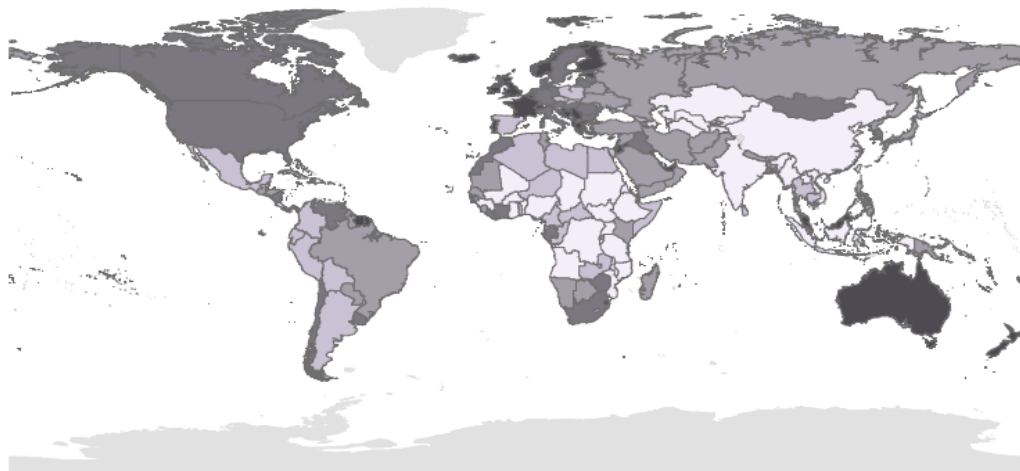
    Peace Science* 31(1):94–106.

McClelland, Charles A. and Gary D. Hoggard. 1969. "Conflict Patterns in the Interactions Among Nations." Pp. 711–24 in *International Politics and Foreign Policy: A Reader in Research and Theory*, edited by J. N. Rosenau. New York: Free Press.

O'Brien, Sean P. 2010. "Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research." *International Studies Review* 12(1):87–104.

Raleigh, Clionadh, Andrew Linke, Haavard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset." *Journal of Peace Research* 47(5):651–660.

Rasler, Karen. 1996. "Concessions, Repression, and Political Protest in the Iranian Revolution." *American Sociological Review* 132–152.

Rod, Espen Geelmuyden and Nils G. Weidmann. 2013. "Protesting Dictatorship: The Mass Mobilization in Autocracies Database."
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.716.2625&rep=rep1&type=pdf

Salehyan, Idean et al. 2012. "Social Conflict in Africa: A New Database." *International Interactions* 38(4):503–511.

Schneider, Gerald, Margit Bussmann, and Constantin Ruhe. 2012. "The Dynamics of Mass Killings: Testing Time-Series Models of One-Sided Violence in the Bosnian Civil War." *International Interactions* 38(4):443–461.

Schrodt, Phil A. 2012. "Precedents, Progress, and Prospects in Political Event Data." *International Interactions* 38(4):546–569.

Schrodt, Philip A. and Deborah J. Gerner. 1994. "Validity Assessment of a Machine-Coded Event Data Set for the Middle East, 1982-92." *American Journal of Political Science* 825–854.

Sirkeci, Ifran and Jerry Cohen. 2016. "Cultures of Migration and Conflict in Contemporary Human Mobility in Turkey." *European Review* 24(3):381-396.

Shellman, Stephen M. 2004. "Time Series Intervals and Statistical Inference: The Effects of Temporal Aggregation on Event Data Analysis." *Political Analysis* 12(1):97–104.

Sullivan, Christopher M., Cyanne E. Loyle, and Christian Davenport. 2012. "The Coercive Weight of the Past: Temporal Dependence and the Conflict-Repression Nexus in the Northern Ireland 'Troubles.'" *International Interactions* 38(4):426–442.

Tarrow, Sidney. 2004. *Power and Movement*. New York: Cambridge University Press.

Taylor, Charles L. and David Jodice. 1983. "World Handbook of Social and Political Indicators." *New Haven*.

Underwood, Patrick. 2016. Conflict and Stability in the Neoliberal Era: Explaining Urban Unrest in Latin America. PhD Dissertation. Seattle: University of Washington.

Urdal, Henrik and Kristian Hoelscher. 2012. "Explaining Urban Social Disorder and Violence: An Empirical Study of Event Data from Asian and Sub-Saharan African Cities." *International Interactions* 38(4):512–528.

Wada, Takeshi. 2016. "Rigidity and Flexibility in the Repertoires of Contention." Mobilization 21(4):449-468.

Wilkenfeld, Jonathan, Michael Brecher, Joseph Hewitt, Kyle Beardsley, and Pelin Eralp. 2010. "International Crisis Behavior Dataset: Version 10 (1918-2007)." *Center for International Development and Conflict Management*. Retrieved (http://www.cidcm.umd.edu/icb/data/).

Table 1: Descriptive Statistics for WHIV Contentious Event Forms		Daily File	Annual File		
Event form & category name		**Frequency**	**Mean**	**Min**	**Max**
Protest					
PDEM:	Protest demonstration	8,884	0.79	0	76
PALT	Protest altruism	1,122	0.10	0	18
PMAR	Protest procession	4,325	0.384	0	47
POBS	Protest obstruction	2,980	0.265	0	21
PPRO	Protest defacement	1,745	0.155	0	19
SRAL	Rally support	20,439	1.817	0	348
STRI	Strikes & boycotts	15,547	1.382	0	114
Violence					
ABDU:	Abduction	222	0.02	0	8
HTAK	Hostage taking	5,825	0.517	0	61
JACK	Hijacking	917	0.082	0	20
CBRU:	WMD attack	16	0.001	0	3
CBIO	Chem/bio attack	121	0.011	0	6
PASS:	Physical assault	52,015	4.62	0	597
BEAT	Beating	1,485	0.132	0	42
CORP	Corporal punishment	2,846	0.253	0	57
SEXA	Sexual assault	316	0.028	0	10
MAIM	Torture	1,374	0.122	0	19
RAID:	Armed action	35,873	3.182	0	403
AERI	Missile attack	7,799	0.691	0	223
ASSA	Assassination	1,925	0.171	0	31
CONC	Crowd control	4,775	0.424	0	50
COUP	Coups & mutinies	718	0.064	0	18
GRPG	Artillery attack	16,724	1.486	0	672
MINE	Mine explosion	1,179	0.105	0	70
PEXE	Small arms attack	36,613	3.251	0	401
SBOM	Suicide bombing	1,331	0.118	0	191
VBOM	Vehicle bombing	1,739	0.155	0	273
RIOT	Riot	2,856	0.254	0	45
Sanction					
BANA	Impose restriction	3,843	0.342	0	39
CENS	Censorship	1,292	0.115	0	25
MONI	Covert monitoring	1,547	0.138	0	32
POAR	Political arrest	8,108	0.721	0	72
Relaxation					
DMIN	Demining	489	0.043	0	10
EMSA	Ease military blockade	146	0.013	0	8
RCUR	Relax curfew	278	0.025	0	8
RELE:	Release/return	2,313	0.206	0	42
RRPE	Release/return persons	12,976	1.153	0	114
RRPR	Release/return property	1,175	0.104	0	11
RPOL	Relax censorship	104	0.009	0	4
RSAN	Relax admin. sanction	307	0.027	0	9

Total	264,289	100

Figure 1: Global Protest and Violence by Week, 1990-2004

Figure 1: Global Protest and Violence by Week, 1990-2004

4

Figure 2: Global Sanction and Relaxation by Month, 1990-2004

Figure 2: Global Sanction and Relaxation by Month, 1990-2004
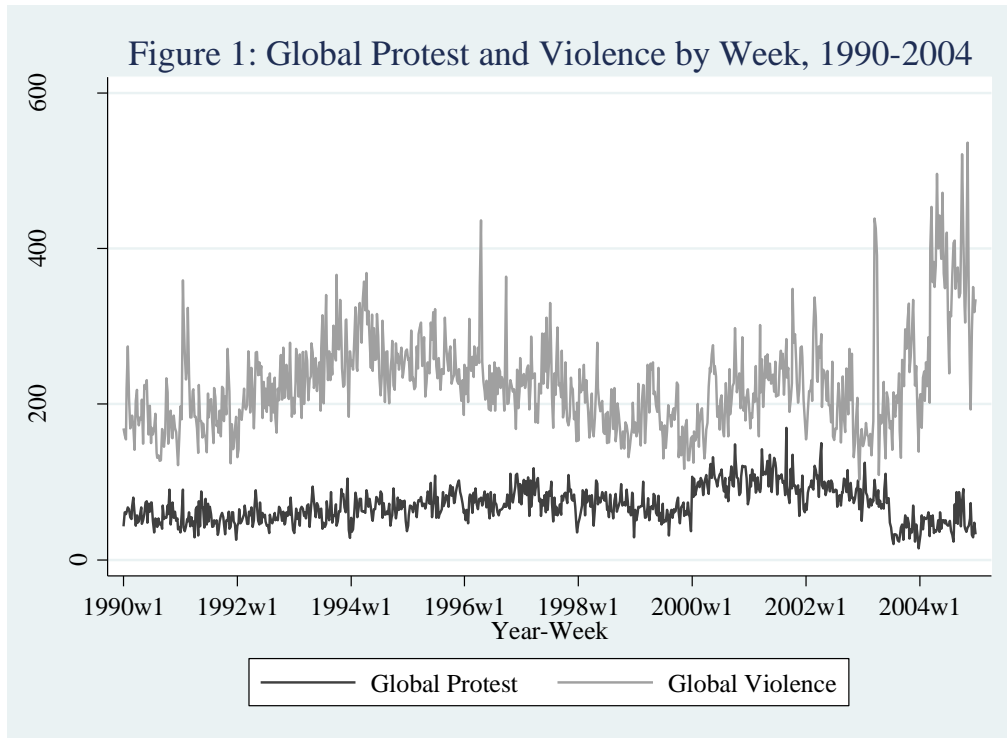
## WHIV Political Protest Events per Capita, 1990-2004

Legend

Political Protest per Capita

- missing or no pop. data
- .00-.03 events per capita
- .03-.06 events per capita
- .06-.10 events per capita
- .10-.25 events per capita
- .25-1.9 events per capita

Figure 3: WHIV Protest Events per Capita, 1990-2004

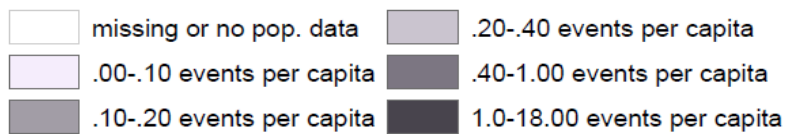WHIV Political Violence Events per Capita, 1990-2004

Legend

missing or no pop. data		.20-.40 events per capita
.00-.10 events per capita		.40-1.00 events per capita
.10-.20 events per capita		1.0-18.00 events per capita
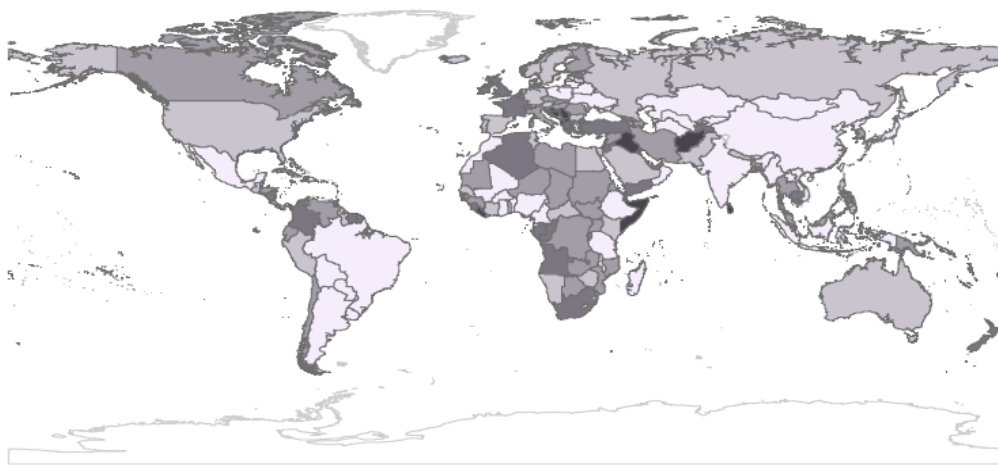
Figure 4: WHIV Violence Events per Capita, 1990-2004

Table 2: Selected Publicly Available Conflict Data Sets.

Dataset	Year Coverage	Geographic Coverage	Event Data Type	Major Event Types	# of Event Forms	Information Source	Geo-Referenced?	Citation
World Handbook of Political Indicators IV	1990-2004	Global	Discrete	Protest, political violence, sanctions, and relaxations	40	Reuters Newswire	No	Jenkins, Taylor, Abbott, Maher, and Peterson 2012
UCDP/PRIO-Uppsala Armed Conflict v.4	1946-2011	Global	Episode	Intra and inter state conflict	4	Multi-Source (Factiva Online)	No	Gleditsch, Wallensteen, Eriksson, Sollenberg, and Strand 2002
ACLED	1997-2011	Africa	Process	Battles, base establishments, rioting/protesting, and violence	6	Multi-Source (IRIN, Factiva, etc.)	Yes	Raleigh, Linke, Hegre, and Karlsen, 2010
SCAD	1990-2011	Africa	Process	Demonstrations, riots, strikes, and violence	5	Associated Press and Agence France Press	Yes	Salehyan, Hendrix, Case, Linebarger, Stull, and Williams. 2012
KEDS	1979-2011	Regional	Discrete	International political event data	235	Reuters/AFP Newswire	No	Schrodt and Gerner, 2010
GDELT	1979-2013	Global	Discrete	Intl. political event data	235	Multi-source (AFP, Google News)	Yes	Leetaru and Schrodt, 2013
Cross National Time Series (Banks)	1919-2011	Global	Annual Count	Domestic Conflict	8	Unknown	No	Banks and Wilson, 2012
PRODAT	1950-2002	Germany	Discrete	Protest	22	Frankfurter Rundschau and Süddeutsche Zeitung	No	Rucht and Teune, 2002
YANKDAT	1960-1995	United States	Discrete	Protest	21	New York Times	No	McAdam, McCarthy, Olzak, and Soule, 2010

Endnotes

[1] IDEA overlaps with the CAMEO event framework with a small number of event form and actor/target exceptions.

[2] For other comprehensive global and regional events data sets that address similar event forms, see the Integrated Crisis Early Warning System (ICEWS) (https://dataverse.harvard.edu/dataverse/icews) , the Social Conflict Analysis Database (SCAD) (https://www.strausscenter.org/scad.html), and the Armed Conflict Location and Event Data Project (ACLED) (https://www.acleddata.com/).   Each project has its own unique advantages in terms of event coverage, time and country/regional coverage along with distinct methodology of construction.

[3] In this sense, it is most comparable to ICEWS, which has the advantage of multi-source coding and coverage of both contentious and non-contentious event forms.

[4] Other event data projects such as ACLED and ICEWS have more fine-grained geocoding.

[5] The original IDEA data, also known as King and Lowe data, is available through their Dataverse site (http://gking.harvard.edu/data?dvn_subpage=/faces/study/StudyPage.xhtml?globalId=hdl:1902.1/FYXLAWZRIA). *WHIV* is a subset of this data that has been cleaned for non-events, false positives, and other errors. For further questions about the comparison between King/Lowe data and *WHIV*, please see the FAQ page on the *WHIV* website.

[6] Definitions for each of the event forms can be found in the codebook (available at https://sociology.osu.edu/worldhandbook).

[7] An expert can of course make reliable attributions as to the identity of such actors, which can be done in customized uses of *WHIV* data.

[8] Due to copywrite limitations, we cannot provide the literal texts of these actors and targets, making it impossible to make educated attributions as to the identity of these actors and targets.

[9] All state to state war events conducted by government actors and targeting government actors are excluded from the *WHIV* dataset.

[10] ICEWS; http://www.lockheedmartin.com/us/products/W-ICEWS/W-ICEWS_overview.html

[11] http://phoenixdata.org