Measuring political violence in Pakistan: Insights from the BFRS Dataset

Ethan Bueno de Mesquita
Harris School of Public Policy, University of Chicago, USA

C. Christine Fair
Edmund A. Walsh School of Foreign Service, Georgetown University, USA

Jenna Jordan
Sam Nunn School of International Affairs, Georgia Institute of Technology, USA

Rasul Bakhsh Rais
Institute of Strategic Studies, Islamabad, Pakistan

Jacob N. Shapiro
Department of Politics and Woodrow Wilson School, Princeton University, USA

Abstract

This article presents the BFRS Political Violence in Pakistan dataset addressing its design, collection and utility. BFRS codes a broad range of information on 28,731 incidents of political violence from January 1988 to November 2011. For each incident we record the location, consequences, cause, type of violence and party responsible as specifically as possible. These are the first data to systematically record all different kinds of political violence in a country for such an extended period, including riots, violent political demonstrations, terrorism and state violence, as well as asymmetric and symmetric insurgent violence. Similar datasets from other countries tend to focus on one kind of violence—such as ethnic riots, terrorism or combat—and therefore do not allow scholars to study how different forms of violence interact or to account for tactical and strategic substitution between methods of contestation. To demonstrate the utility of the dataset, first we examine how patterns of tactical substitution vary over time and space in Pakistan, showing that they differ dramatically, and discuss implications for the study of political violence more broadly. Second, we show how these data can help illuminate ongoing debates in Pakistan about the causes

Corresponding author:
Jacob N. Shapiro, Department of Politics and Woodrow Wilson School, Princeton University, Princeton, NJ, USA.
Email: jns@princeton.edu

of the increase in violence in the last 10 years. Both applications demonstrate the value of disaggregating violence within countries and are illustrative of the potential uses of these data.

Introduction

Political violence in Pakistan is a central policy concern for the international community for numerous reasons. First, the country hosts numerous Islamist jihadi organizations that operate in and beyond the country. Second, as Pakistan is a nuclear-armed country with a known history of proliferation, the extensive presence of nonstate violent actors perennially stokes fears that one of these groups will acquire nuclear materials or technology. Third, many of these militant groups undertake operations in India and against Indian targets in Afghanistan. Attacks on Indian soil, and possibly against Indian assets in Afghanistan, could provoke conflict between India and Pakistan, both of which possess nuclear weapons.

Apart from these policy concerns, Pakistan's long history of violent politics offers scholars numerous opportunities to learn about the causes and consequences of intra-state violence. Political violence in Pakistan takes many forms. Competition between political parties involves frequent violent clashes between de facto party militias. Anti-state groups employ an assortment of tactics—including riots, terrorism, guerilla warfare and kidnapping—against a variety of targets that include government forces, civilian populations aligned with the state, and civilian populations more generally. Government forces also engage in a multiplicity of forms of violent repression targeted at a variety of populations, utilizing both state resources (police and military), and paramilitary proxies.

Exploring the interactions between the different kinds of violence has great potential for improving our understanding of political violence. Both the empirical and theoretical conflict literatures have a tendency to treat different forms of political violence in isolation—developing separate theoretical and empirical models for each. Rebels and governments, however, choose violent tactics strategically in response to a variety of political, economic, geographic, technological and military constraints. Thus, studying different types of political violence in isolation may lead scholars to miss important relationships across the forms of violence. This potential omission raises two important concerns. First, it might be that some existing inferences about the correlates of violence, that ignore the possibility of transitions from one form of violence to another, are not robust. Second, and perhaps more importantly, studying the interactions across various tactics might lead to a more nuanced understanding of the causes of political violence.

Unfortunately, extant datasets on political violence in Pakistan do not allow for such analyses because they focus upon certain types of political violence (e.g. terrorism) or fail to record information on militant groups' target selection and tactical choices. In an effort to address the shortcomings of existing data on Pakistan, and to provide a scholarly resource for understanding the tactical choices of rebel groups, we constructed a dataset of over 28,000 incidents of political violence in Pakistan since 1988. Unlike the Global Terrorism Database or the Worldwide Incidents Tracking System, which collect data on incidents that

meet their definitions of terrorism, we collect data on all incidents of violence that are not clearly apolitical.

For each incident we record the location, target, cause, type of violence and the responsible party as specifically as possible. This dataset is unprecedented both because of the broad scope of politically violent events it tracks in a country (e.g. riots, violent political demonstrations, terrorism, and state violence, as well as symmetric and asymmetric insurgent violence) and because of the extended time period it covers. We believe that these data will allow scholars to study a multitude of trends in political violence over space and time in Pakistan. This includes the interaction between different forms of violence and the conditions under which militants substitute one tactic or strategy of contestation for another, which is the primary focus for the analysis in this paper. Moreover, detailed geospatial data on the location of attacks provides the opportunity to investigate regional heterogeneity in violence.

The remainder of this paper proceeds as follows. The next section discusses the data and how they differ from existing resources on political violence in Pakistan. Second, we outline how the data can provide evidence for large theoretical questions about political violence. This section examines trends in national and subnational violence by focusing on theoretical insights regarding tactical substitution and the relationship between economic opportunity and mobilization. The third section uses the data to provide initial evidence on key debates within Pakistan regarding the sources of the recent increases in militant violence. The final section concludes.

Why new data on Pakistan?

There are four existing alternatives to the BFRS dataset: the Global Terrorism Database (GTD), which is maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland; the Worldwide Incidents Tracking System (WITS), which was maintained by the National Counterterrorism Center until it was discontinued in April 2012; the South Asia Terrorism Portal (SATP), which is maintained by the Institute for Conflict Management, a New Delhi-based nongovernmental organization; and the Armed Conflict Location and Event Data Project (ACLED), which is directed and operated by faculty at the University of Sussex.[1]

None of these offer the combination of definitional clarity, observational detail, spatial and temporal coverage, or comprehensive accounting of attacks available in the BFRS dataset. Both GTD[2] and WITS[3] use fairly restrictive definitions that focus upon terrorism (variously defined) that lead them to undercount attacks relative to the BFRS, which covers political violence more comprehensively. In addition, WITS and GTD rely primarily upon news aggregators such as Factiva, which exclude the major English and Urdu-language Pakistani papers (*The Dawn* and *Daily Jang*, respectively). The compilers of the SATP data on terrorism do not provide a codebook or definition of terrorism and are not transparent in their data extraction methodology. The SATP data provide incident data beginning only in 2006 and afford much less geographic information than do the BFRS data. ACLED is a cross-country dataset and only provides data on Pakistan from 2006 to 2009 with much less specificity regarding type of attack, target, location, numbers of casualties and fatalities (Raleigh et al., 2010).

Because we found extant datasets to be unsuitable for detailed subnational and across-time studies of political violence in Pakistan, we developed the BFRS dataset with the explicit goal of facilitating a more integrated approach to the study of political violence in varied

Table 1. Comparing datasets, January 2004 to December 2008

Variable	BFRS Political violence in Pakistan	BFRS Political violence excluding assassination	BFRS Political violence including political assassination	Worldwide Incident Tracking System	Global Terrorism Dataset	South Asia Terrorism Portal
Total incidents	3971[a]	2421	2825[b]	3686	1129	—
Total killed	4605[a]	2421	2987[b]	4567	3362	6991[c]
Total wounded	6523[a]	5815	6127[b]	9367	5909	—
Total nonterrorist incidents	2680	—	—	—	—	557[d]
Total killed, nonterrorist incidents	5200	—	—	—	—	1295[d]
Total wounded, other	6190	—	—	—	—	2457[d]

[a]Includes "Terrorism", "Guerilla attacks against military/paramilitary/police" and "Assassinations", where the latter category excludes selective violence attributed to the state. The BFRS definition of a terrorist attack does not match up exactly with the WITS definition, which includes some deaths that BFRS codes as "Other political violence".
[b]Includes "Terrorism", "Guerilla attacks against military/paramilitary/police" and "Assassinations", where either the event or reported cause fields are coded as "Political", and excludes selective violence attributed to the state.
[c]SATP provides breakdown by status of victim; this count includes military and civilian.
[d]SATP count appears to include only sectarian attacks involving explosives and does not provide clear coding criteria.

forms. The BFRS data are designed to allow scholars to study patterns of substitutability and complementarity across forms of political violence, as well as to help analysts account for the possibility that different forms of violence are caused by different underlying dynamics.

To illustrate the import of the varied definitional issues noted above, we illustrate the various ways in which BFRS data can be disaggregated in Table 1. The first column includes all BFRS events coded as "terrorism", "guerilla attacks against military/paramilitary/police", and "assassinations". The second column excludes assassinations, and the third includes only those cases of assassination that are political in type or reported cause. Disaggregating the data in different ways allows for more direct comparison with other datasets and shows considerable differences between the datasets, particularly in terms of overall incidents and numbers wounded. Table 1 also includes SATP data, for comparison purposes.

In Table 2, we summarize the differences between GTD and BFRS across categories of violence in more detail. In contrast to the GTD, which makes a distinction between those events that are exclusively coded as terrorism and those that could be guerilla or insurgent action, BFRS includes many more cases of guerilla or insurgent attacks. Even after we combine the two fields in GTD, it still reports far fewer attacks and lower numbers of individuals killed and arrested than does BFRS. We believe that this difference is a function of our use of a local Pakistani newspaper not included in the aggregators used by the GTD project for most of the period of our data.[4]

The BFRS data are also unique in providing detailed information on the province, district, tehsil and town/city for each incident whenever possible.[5] Using district-level data from

Table 2. Comparing GTD and BFRS, 2008–2010

|	GTD Terrorism[a]	BFRS Terrorism[a]	GTD Guerilla[c]	BFRS Guerilla[d]	GTD total	BFRS total[e]
Total incidents	1820	1632	31	537	1942	2720
Total killed	1808	1690	29	914	1926	3421
Total injured	1795	5381	27	2534	1910	8152

[a]This includes events in GTD that are unambiguously coded as terrorism.

[b]This includes only those events coded as "Terrorism" in the event field.

[c]This includes events that were coded as "Insurgency/guerilla action" in the "Doubt terrorism proper" field.
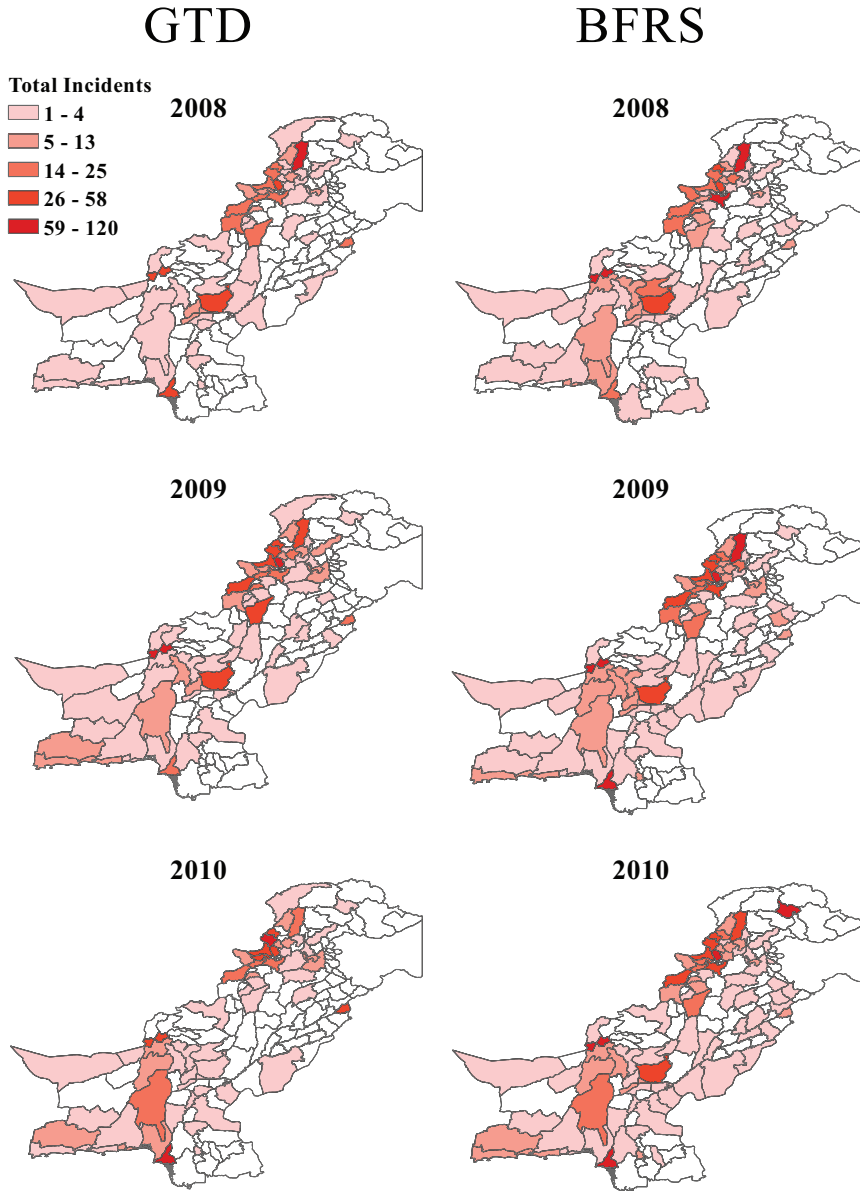
[d]This includes only those events coded as "Guerilla attacks against military/paramilitary/police" in the event field.

[e]This includes all events coded as "Terrorism", "Guerilla attacks against military/paramilitary/police" or "Assassinations" that are political in type and excludes selective violence attributed to the state.
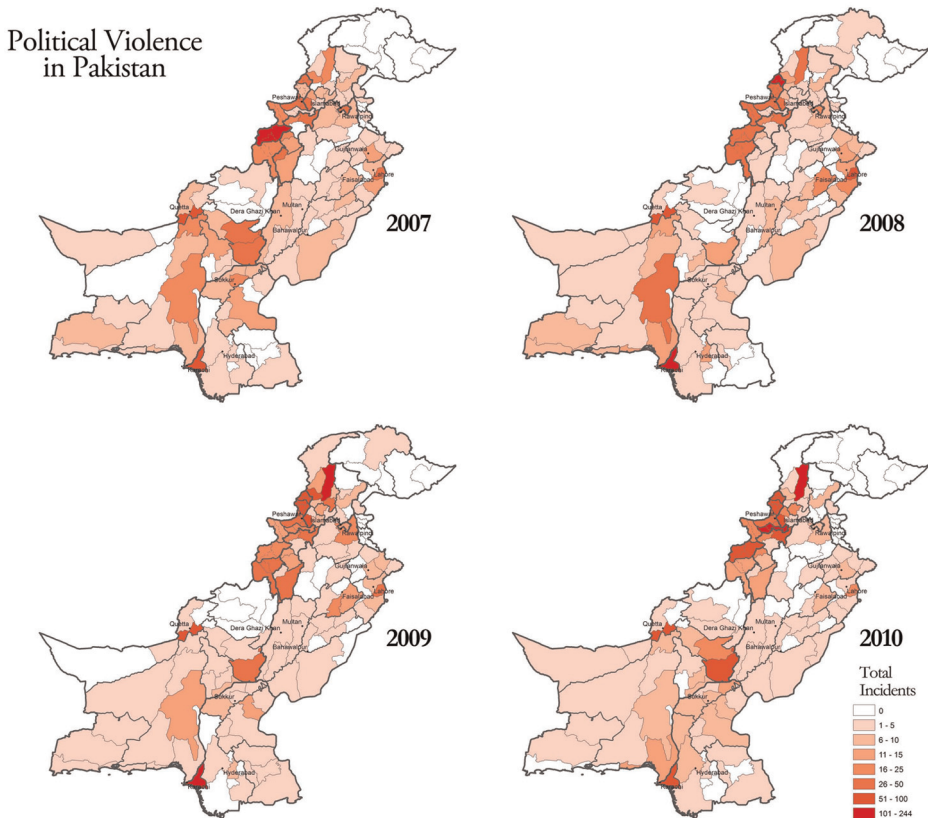
Figure 1. Yearly incidents of political violence by district (by Agency in the FATA), 2007–2010.

2007 to 2010, Figure 1 emphasizes the considerable variation that exists over time and space in political violence in Pakistan. Some areas in the most violent provinces are consistently peaceful. Because other databases do not include detailed subnational data on location, they

Figure 2. Terrorism in Pakistan, by district, 2008–2010: comparing GTD and BFRS.

cannot account for this regional heterogeneity. In Figures 2 and 3 we examine district level differences between the GTD and BFRS datasets for terrorist attacks between 2008 and 2010. In order to directly compare the two datasets, Figures 2 and 3 include all incidents from the GTD dataset and disaggregate the BFRS dataset to include only those events that are coded as "terrorism", "guerilla attacks against military/paramilitary/police", and political "assassinations". GTD does not provide district-level data, and to allow for geospatial
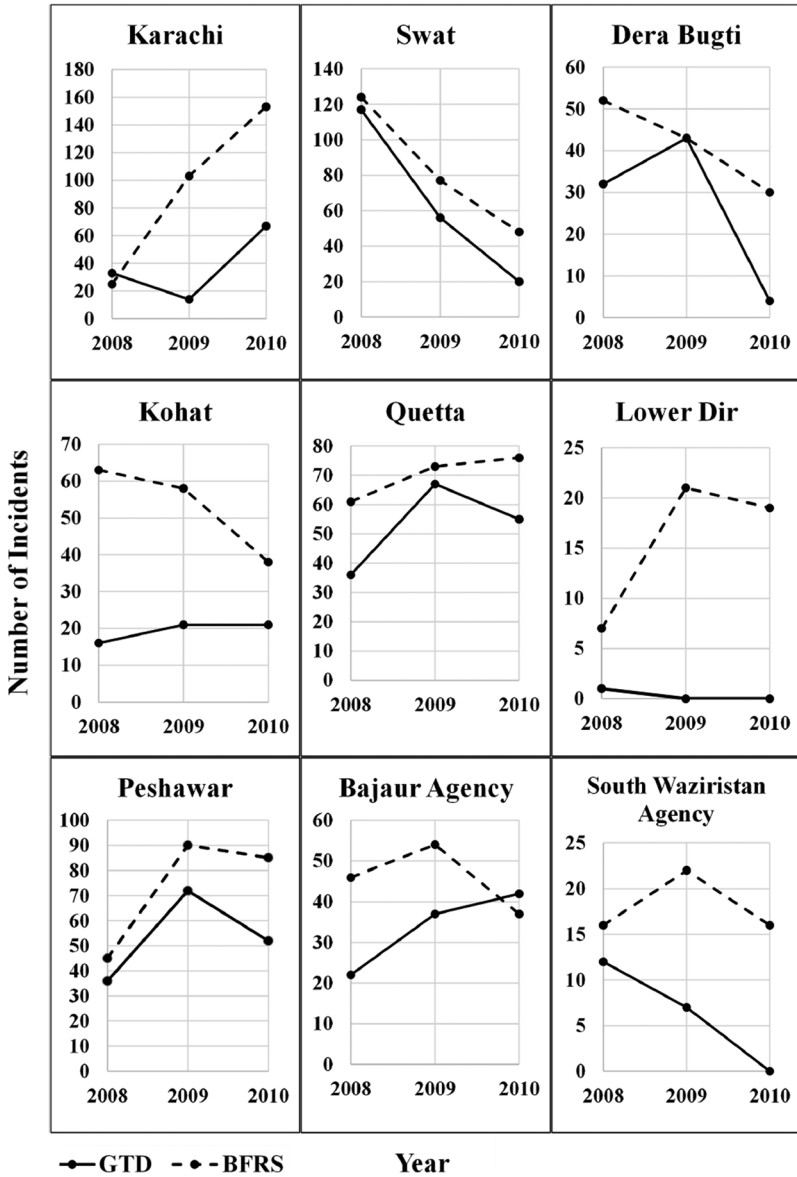
Figure 3. Comparison of the number of terrorist attacks in GTD and BFRS, by district, 2008–2010.

comparisons we geocoded the district for each attack in the GTD database. While Tables 1 and 2 exhibit numerical differences between GTD and BFRS, Figures 2 and 3 display district-level variation in the location of events. The maps indicate that GTD includes fewer attacks in some districts and excludes some entirely. This is particularly apparent for 2010, especially when looking at Figure 1, which includes different forms of political violence. Figure 3 highlights regional variation between the two dataset and compares the number of terrorist attacks by district from 2008 to 2010. Analyzing the nine districts that saw the most

variation in number of incidents between GTD and BFRS. It becomes clear that GTD significantly undercounts violence in several districts.

The BFRS Dataset: Methodology

The BFRS Dataset of Political Violence in Pakistan contains incident-level data on political violence in Pakistan, based on press reporting.[6] Our data collection model was designed to develop consistent incident-level data on the broadest possible range of violent political events over time. We define political violence as any publicly reported act that: (1) is aimed at attaining a political, economic, religious or social goal; (2) entails some violence or threat of violence—including property violence, as well as violence against people; and (3) is intentional, the result of conscious calculation on the part of the perpetrator. This may include, but is not limited to terrorist attacks, riots, assassinations, and full-scale military operations. The BFRS data capture all such events from January 1988 to November 2011 and are being continually updated.

The BFRS data are derived from press reports in *The Dawn*, the major English language newspaper in Pakistan.[7] A team operating out of the Lahore University of Management Sciences reviewed each day of *The Dawn* beginning in January 1988, recording all incidents of violence, defined as any event or incident of violence or threat of violence aimed at attaining a political, religious, economic or social goal.[8] In many cases, a single article will report multiple events, which are treated as separate observations. In order to provide a reliability check on the aggregate data, a team operating at the University of Chicago independently coded a random 10% sample, also from *The Dawn*. We discuss reliability and source bias below.

Variables

For each incident we record the date of the event, its duration, location, event type, attack type, target, the number of individuals killed and injured, whether it was successful, the cause reported in the press, and the parties involved. Below is a brief review of some of the key variables, followed by a discussion of how they were operationalized.

Event details. In the first set of variables, we record key details about each incident. First, we define six variables relating to the geographic location of the attack: (1) location (this usually refers to the smallest unit as reported in the press); (2) town or city; (3) village; (4) province; (5) district; and (6) tehsil. This detailed geographic information is unique to our dataset. Second, we record the date on which the violence began and ended in order to identify the duration of each event. Finally, we report the number of individuals killed, injured and arrested over the course of an event. When there are discrepancies within news reports on the number of people killed and injured during an incident, we report both upper and lower bounds on the number killed. We also update counts when later news reports identify a change in the consequences of an incident, for example, when people succumb to injuries after a week.

Event type and characteristics. This set of variables provides information about the attack type, the target, the party responsible, the reported motivation for the attack and whether the

police, military or paramilitary were involved. First, we define 12 broad categories for type of violence: (1) terrorism, which is defined as premeditated, politically motivated violence against noncombatant targets by subnational groups of clandestine agents; (2) riots, a violent clash between two or more nonstate groups; (3) violent political demonstration or protest, a violent mobilization of crowds in response to a political event; (4) gang-related violence; (5) attack on the state; (6) assassination, an attempt by a nonstate entity intended to kill a specific individual; (7) assassination by drone strike, an assassination carried out by an unmanned aerial vehicle; (8) conventional attacks on military, policy, paramilitary and intelligence targets, which include ambushes, direct fire, artillery, pitched battle and troop captures; (9) guerilla attacks on military, police, paramilitary and intelligence targets, which include road-side bombs, improvised explosive devices, suicide attacks and car bombs; (10) military, paramilitary or police attacks on nonsate combatants, which is violence initiated by state, federal or provincial combatants against nonstate combatants, subnational groups or clandestine agents; (11) military, paramilitary or police-selective violence, which is initiated by state, federal or provincial combatants against civilians; and (12) threat of violence, which refers to incidents in which the threat of violence is used for political purposes.

Second, we further record whether the attack was motivated by the following concerns: communal, sectarian, ethnic, tribal, Islamist, political, politico-economic, food and water, public services or fuel supply and prices. While multiple events from a single article were treated as separate observations, we created an additional category to ensure that important data were not lost. For example, if a report claimed that the total number of deaths across a location was 50, but the individual incidents only added up to 40, then we reported the difference (10 in this case) in the number killed field in a final entry coded as an "aggregated report".

Third, we identified whether the attack was successful in hitting the intended target or if it was intercepted by police or military forces. Not all recorded events are successful attacks; some are intercepted and some fail to strike their intended targets.

Fourth, we recorded the reported impetus for the event. This variable is coded according to content directly reported in the press, for example, "in response to killings, students led a protest march". Because each coder handled consecutive periods, they developed substantial subject-matter expertise and so we also include a field for the "likely cause" when our coders were able to infer it from context. A likely cause was typically included when an event was part of a long-running campaign over a particular issue in one location. For instance, press reports on inter-communal riots in Karachi in the mid-1990s often omitted the fact that ethnic conflict was driving the violence, but this was clear to our coders from the context.

Fifth, we identified the party responsible for the attacks. There are 14 categories: (1) civil/society or campaign group (these are groups that exist for a political cause, but are not a political party or organized along occupational lines); (2) foreign party (USA, India, Afghanistan or multilateral); (3) gang; (4) informal group (ethnic, Islamist/sectarian, other); (5) intelligence agency; (6) militants (ethnic, Islamist/sectarian, other); (7) military/paramilitary; (8) police; (9) political party; (10) professional union/alliance; (11) religious party; (12) student group; (13) tribal group; and (14) unaffiliated individual. Finally, we provide a more detailed description for each event, which includes a summary and any questions or uncertainties that arose during the coding process.

Intersource reliability

While the primary data were coded from the Lahore edition of *The Dawn* by a team at Lahore University of Management Sciences (LUMS), in order to ensure data quality, a team operating at the University of Chicago independently coded a random 10% sample of weeks from 1988 to 2010 from the Karachi edition of *The Dawn*.

Lahore and Karachi are the two largest cities in Pakistan, and their editions of *The Dawn* should be the most comprehensive in their coverage of events, particularly regarding important instances of political violence. The Chicago and LUMS teams coded different versions of *The Dawn* owing to availability. The University of Chicago only has the Karachi edition and we were unable to secure microfilm of the Lahore edition in Chicago or establish a research team at another institution. Because the LUMS team is based in Lahore, they only had access to the Lahore edition.

There were differences in the results between the Chicago sample and the main dataset developed at LUMS. Overall, restricting attention to the sample of days coded by the Chicago-based team, the LUMS-based team identified 2534 incidents, while the Chicago-based team identified 2314 incidents, about 8.7% fewer incidents.

There are two main reasons for this difference. The first source of the discrepancy is that there are differences in coverage between the Karachi and Lahore editions. In the average week covered by both teams, the Lahore edition reported more violence in Azad-Kashmir, Balochistan, FATA and Punjab, while the Karachi edition reported more violence in Sindh and Gilgit–Baltistan. The two editions reported similar levels of violence in KPK. We discuss the implications for analysis below, but highlight the larger point that similar biases probably exist in all press-based violence datasets that do not systematically draw on local editions.

The second source of the discrepancy is that the teams at LUMS and Chicago worked with slightly different processes. The faculty supervisor in Lahore required each coder to work through consecutive days within a one-year time period. As a result, knowledge of events over that time period would influence how the press reporting was interpreted. This is particularly salient for the reporting of small-scale events, which often consist of periodic updates rather than more detailed coverage of the events, both for reasons of political sensitivities and to conserve space in the printed edition. For example, during the mid-1990s, the state led an intense campaign against militias affiliated with the MQM party, then known as the Mohajir Qaumi Movement. An article from the December 1, 1995 issue of *The Dawn* stated, "[t]he ongoing 'terrorism' continued as armed youths that were being chased by the police took refuge in a private school". To the LUMS coder, who had been reading consecutive days of *The Dawn*, it would be clear from the context that this incident related to a clash between states forces and MQM activists. By contrast, since the University of Chicago team was coding a random sample, they worked with only a week of press reports at a time and, thus, would not have been able to infer from context that the events involved MQM activists.

To analyze the discrepancies in more detail, we use the following aggregate measures of political violence: (1) Total incidents—count of all incidents; (2) Militant attacks—all attacks by organized groups against the state, regardless of whether the target was military or nonmilitary; (3) Terrorist attacks—all incidents of premeditated, politically motivated violence perpetrated against *noncombatant* targets by subnational groups or clandestine agents; (4) Militant violence—militant attacks and terrorist attacks; (5) Assassinations—attempts (successful or failed) by nonstate entities aimed at killing a specific individual; (6) Security force actions—all attacks by state agents, including drone strikes and violence against

noncombatants; (7) Violent political demonstrations—riots and violent political demonstrations; (8) and Conventional attacks—conventional military violence, both state initiated (including violence between militaries along the Line of Control) and militant initiated, against the Pakistani military.

Using these measures, in Table 3 we summarize the total incidents and casualties for different types of attack by data source (Karachi or Lahore editions). First, we report the total number of incidents and casualties for each type of violence country-wide, and the differences between these two datasets. Second, we provide the mean weekly incidents and casualties by source and the difference in those. In the Appendix, Table A1, we illustrate the differences in means for the five main provinces (Balochistan, FATA, KPK, Punjab and Sindh) by presenting the proportional differences between sources. We find that the Lahore edition systematically under-reported violence in Sindh while the Karachi edition systematically under-reported violence in Balochistan, FATA, KPK and Sindh.

When we examine the data on a week-to-week basis, we find that the differences between sources are close to being symmetrically distributed around zero. In Figure 4, we plot the distribution of weekly differences between the Karachi and Lahore editions for the total number of incidents and total number of casualties. Both plots show roughly symmetrical distributions, with the Lahore edition reporting slightly more incidents and the Karachi edition reporting slightly more casualties in a typical week. The differences in the mean number of incidents per week or mean number of casualties per week are not statistically significant. In Figure 5, we break down these differences by provinces in which there is substantial violence. As with the country-level data, all differences cluster around zero and are roughly symmetrical.

Overall the intersource reliability checks indicate that there is most likely measurement error in the main dataset owing to reporting differences between the two versions of *The Dawn*. Aggregating incidents across sources is not a feasible solution as it is often impossible to distinguish between which short stories in one edition match longer stories in other editions and which stories represent distinct events. Given the differences that we have identified, we suggest six best practices for data users:

1. Include province fixed effects in all panel regressions (or cross-sectional regressions at the district level) to account for differences in the intensity of reporting about different regions across editions.
2. Do not rely on these data as the definitive source for the exact level of violence on any particular day. While we have uncovered no evidence of systematic differences between regions in the types of incidents reported (e.g. militant attacks in Sindh are underreported in the Lahore edition), it is also clear that some incidents that were known to reporters go unreported in each edition.
3. Consider showing robustness of results when using the Karachi-edition sample, when doing so is feasible. For example a study looking at the difference-in-differences in total violence for 1990–1995 vs 2005–2010 across some set of locations could be replicated with the Karachi sample. A study looking at monthly differences between 1998 and 2000 could not, as the number of weeks in the Karachi sample for that period is small.
4. Be wary of analysis that relies too heavily on cross-sectional differences between Punjab and Sindh. The data are better suited to looking for differential trends across regions than persistent level differences. Those level differences will reflect some combination of true differences and differences in reporting priorities and editorial decisions.

Table 3. Total incidents and casualties for different types of attack by data source

| Variable	Total incidents			Total casualties			Mean weekly incidents			Mean weekly casualties		
	Lahore	Karachi	Difference	Lahore	Karachi	Difference	Lahore	Karachi	Diffference	Lahore	Karachi	Difference
Incidents	2534	2314	220	8684	9588	−904	2.78	2.54	0.24	9.52	10.51	−0.99
Militant attacks	328	235	93	1592	1364	228	0.36	0.26	0.1	1.75	1.5	0.25
Terrorist attacks	442	199	243	1802	2006	−204	0.48	0.22	0.26	1.98	2.2	−0.22
Total militant violence	770	434	336	3394	3370	24	0.84	0.48	0.36	3.72	3.7	0.02
Assassination	854	700	154	1320	1343	−23	0.94	0.77	0.17	1.45	1.47	−0.02
Security force actions	189	278	−89	1455	1621	−166	0.21	0.3	−0.09	1.6	1.78	−0.18
Violent political demonstration	547	363	184	2165	2024	141	0.6	0.4	0.2	2.37	2.22	0.15
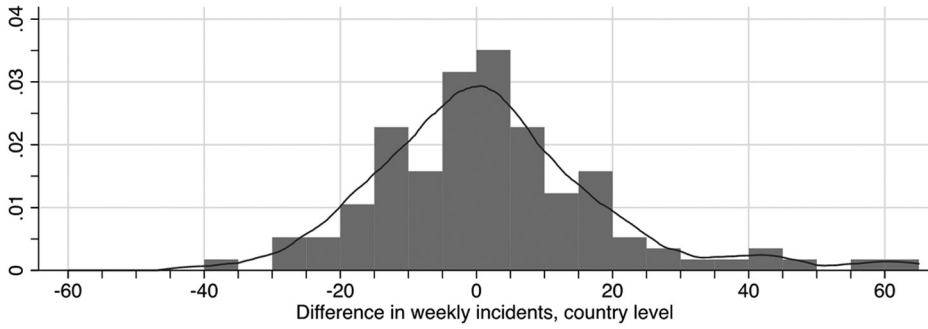
Figure 4. Distribution of weekly differences in incidents and casualties.
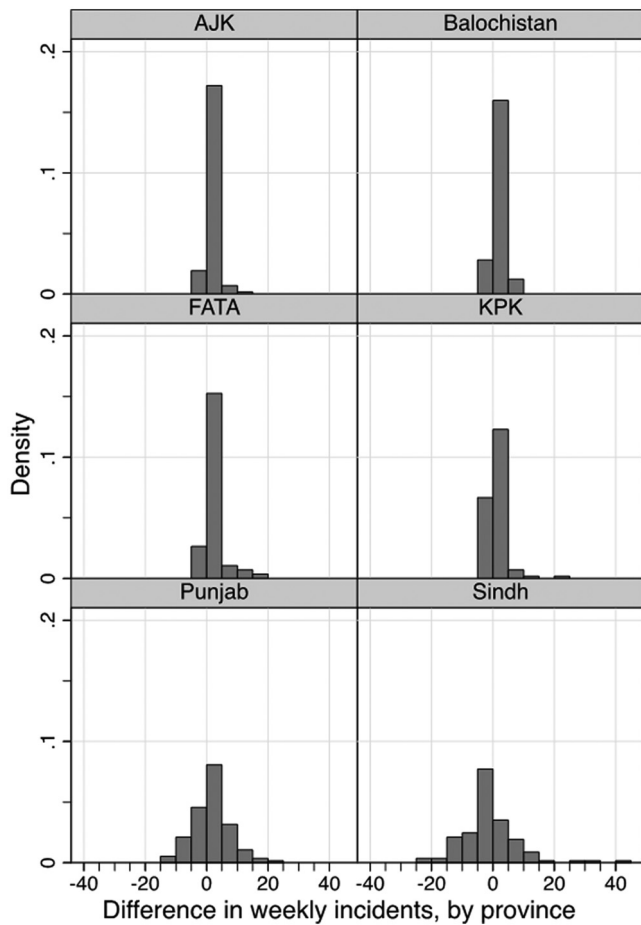

Figure 5. Distribution of weekly differences in incidents and casualties by province.

5. Prioritize regression results over exact comparisons when differences across regions are small. As is well known, multivariate regression is robust to normally distributed measurement error, so long as it is uncorrelated with the treatment of interest.
6. Consider restricting attention to major events as these are more consistently reported across datasets.

Tactical choice in Pakistan

In this section, we illustrate how the BFRS data can provide evidence on larger theoretical questions about political violence. In particular, we examine patterns of tactical substitution among militant groups in Pakistan to provide evidence on the constraints or opportunities that influence whether groups choose conventional tactics, irregular tactics or withdrawal from a conflict (Bueno de Mesquita, 2013).[9] More broadly, we are interested in the links between economic opportunity, mobilization and tactical choice by rebel groups.

Understanding patterns of tactical substitution is thus critical to knowing what one should make of trends in national or subnational violence. A reduction in insurgent violence, for example, can mean a group is no longer as capable as it was in a given region, or it can mean the group has stopped contesting the region for strategic reasons, perhaps because state forces have withdrawn.

To describe patterns of tactical substitution in Pakistan we divide all militant attacks into two categories: conventional and asymmetric. Militant attacks are those attributed to organized armed groups that use violence in pursuit of pre-defined political goals in ways that: (1) are planned; and (2) use weapons and tactics attributed to sustained conventional or guerrilla warfare. Conventional attacks by militants include direct conventional attacks on military, police, paramilitary and intelligence targets such that violence has the potential to be exchanged between the attackers and their targets. Asymmetric attacks include both terrorist attacks by militants, as well as militant attacks on military, police, paramilitary and intelligence targets that employ tactics that conventional forces do not, such as improvised explosive devices.

With this distinction in mind, we constructed six variables: (1) *Militant attacks* include attacks on state targets, conventional attacks on military, paramilitary, police or intelligence targets, and guerilla attacks on military, paramilitary, police or intelligence targets; (2) *Militant asymmetric attacks* include terrorist acts carried out by militants and guerilla attacks on military, paramilitary, police or intelligence targets; (3) *Militant conventional attacks* include conventional attacks on military, paramilitary, police or intelligence targets and attacks on the state carried out by militants; (4) *Militant guerilla attacks* include guerilla attacks on military, paramilitary, police or intelligence targets; (5) *State-initiated attacks on militants* include attacks by military, paramilitary and police on nonstate combatants and assassinations carried out by unmanned aerial vehicles; and (6) *Terrorist attacks* include all events of violence coded as terrorism in the database, meaning premeditated, politically motivated violence against noncombatant targets by a nonstate group.

As a starting point, Table 4 shows that the proportion of political violence from 2000 to 2009 attributable to militant organizations varies dramatically across provinces, from a high of 67% in Balochistan, to a low of 15% in Punjab. The proportion of violence falling into the conventional or asymmetric categories is similarly varied across provinces. This suggests that patterns of tactical substitution may vary dramatically within Pakistan.

Table 4. Proportion of attacks by type across provinces

Province	Total incidents (excludes riots and assassinations)	Percentage militant	Percentage conventional	Percentage asymmetric
Balochistan (population 8.4 million)	1880	67%	10%	56%
Khyber Pakhtunkhwa (population 22.6 million)	2165	46%	12%	35%
Punjab (population 93.7 million)	1180	15%	5%	10%
Sindh (population 38.7 million)	856	28%	17%	10%
Regular provinces	6081	44%	11%	33%

Note: Provincial population estimates from 2010 by Pakistan Census Organization.
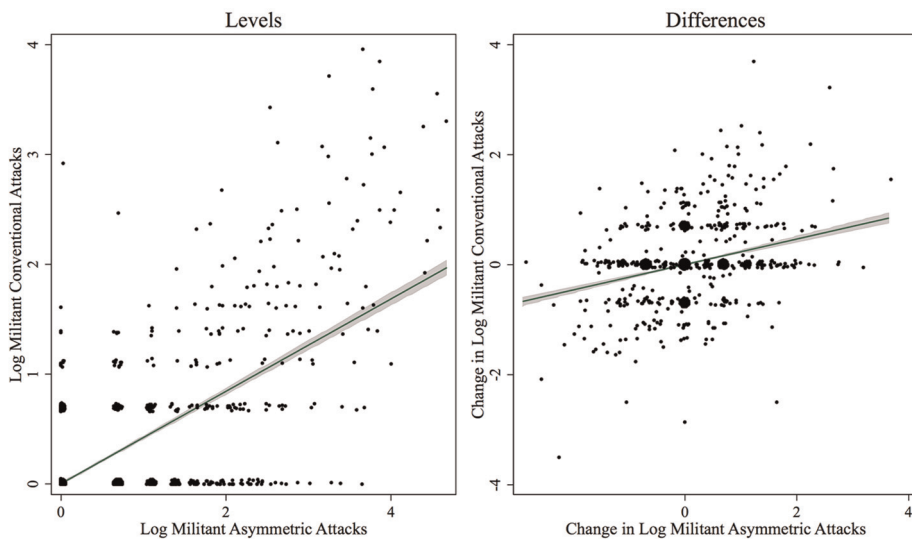
Figure 6. Conventional and asymmetric violence, 2000–2010.
Note: Data from BFRS Political Violence Dataset, 2000–2010. Unit of observation is a district year. Linear fit added.

    To investigate whether these patterns vary and to highlight the value of our subnational data, we first plot logged conventional attacks on logged asymmetric attacks, pooling data from districts across the entire country. Figure 6 shows the result, with the left panel reporting the absolute level of attacks of each kind for each district-year from 2000 to 2010 and the right panel plotting changes in conventional attacks on changes in asymmetric attacks to net out any district-specific trends.
    There is clearly a strong positive correlation at the country level, between conventional and asymmetric militant attacks in both levels and differences. This leaves open some interesting possibilities. One possible explanation is that different groups are engaged in different forms of violence, but some other variable increases all groups' capacities and, thereby, increases all forms of violence. Another possible explanation for this pattern is a technological complementarity between asymmetric and conventional violence. That is, as the capacity

of rebel groups to engage in violence increases (for as yet unknown reasons), those groups want to increase both kinds of attacks.

Of course, the dynamics of political violence are potentially quite varied across Pakistan, because different groups and different cleavages define the conflicts in each province. There is a long-running ethnic independence movement in Balochistan, while Punjab and Sindh have long suffered from significant sectarian cleavages. Hence, one might worry that there is a more nuanced picture at the local level being masked by pooling the regions.

Figure 7 therefore repeats the exercise from Figure 6 at the provincial level, showing that there is indeed local heterogeneity. The positive correlation between different kinds of militant violence found at the country level is evident in the three smaller areas: Balochistan, FATA, and KPK. However, this relationship is much weaker in Punjab and Sindh, the two most populous provinces.

Our conclusions in comparing across provinces are sensitive, to whether the correlation between reporting biases and economic activity is different across provinces. While there are clear and enduring level differences in reporting intensity across provinces, we have found no evidence that these differences change dramatically over time or across the covariate space. We therefore believe that these comparisons provide useful evidence on the differential patterns of violence across provinces.

Figures 6 and 7 suggest the following logic (although certainly other explanations for the observed patterns also exist). Some factor or factors shift, changing rebel groups' overall capacity or motivation to engage in violence. As a result, two things happen. First, the total level of violence—indeed the level of each type of tactic—increases. Second, the increased capacity of the rebel group leads them to increasingly direct effort toward conventional attacks—resulting in an increase in conventional attacks as a percentage of total attacks. Using the BFRS data, one can start to probe the question of what factors might underlie these trends.

One possibility commonly posited in theoretical and empirical work is that, as economic opportunity worsens, mobilization increases. This could simultaneously lead to an increase in total violence and make relatively labor-intensive, conventional attacks more attractive. Testing such a hypothesis in a rigorous way would require finding a source of exogenous variation in economic opportunity. One intriguing possibility might be to use variation in the world price of regional commodity bundles (as in Dube and Vargas's 2013 study of Colombia). However, this task is beyond the scope of this paper. Here, we make a simple first cut, studying the correlations between household income (a measure of economic opportunity), total violence, and tactical mix.

Unfortunately no reliable district-level income figures exist annually for Pakistan, but high-quality provincial-level figures are available from the annual labor force surveys. Using these we construct panel data providing the average monthly household income for each of the four main provinces from 2000 to 2010. As Figure 8 shows, the correlations suggest that income does not appear to be playing the role suggested above. At the national level, total violence is positively, not negatively, correlated with income. At the provincial level, violence is either positively correlated or uncorrelated with income. Moreover, there is no clear relationship between income and the mix of tactics. In Sindh, KPK and Balochistan, there is essentially no relationship between income and tactical mix. Only in Punjab do we see the hypothesized relationship—when income is higher, conventional tactics are a smaller percentage of total attacks.
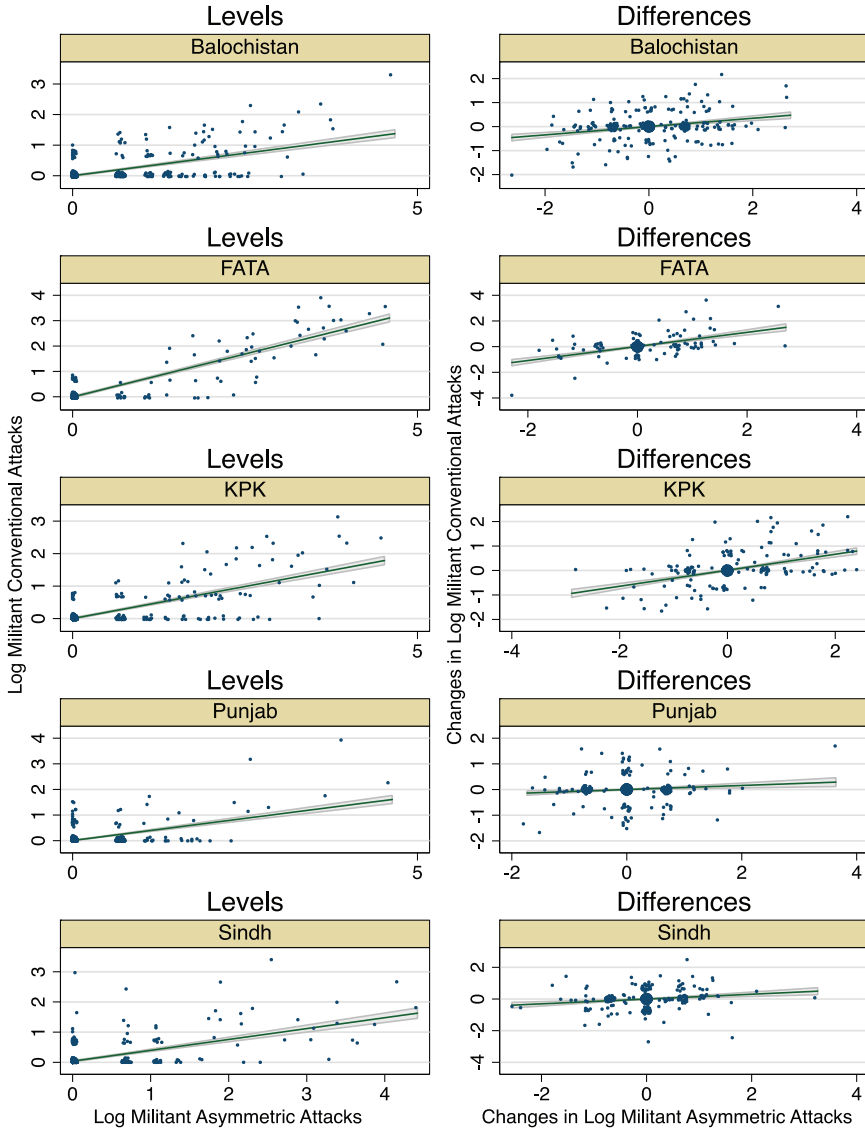
Figure 7. Tactical substitution by province, 2000–2010.
Note: Data from BFRS Political Violence Dataset, 2000–2010. Unit of observation is a district year. Linear fit added.

This analysis of the relationship between income and violence grew out of the theoretical intuition that changes in opportunity costs can lead to forms of tactical substitution. In particular, as opportunity costs go up, you may see substitution out of guerrilla warfare and into terrorist violence because insurgents cannot muster enough forces (Bueno de Mesquita, 2013). The correlations we identify are intended to highlight the need for more careful work that takes these nuances into account.
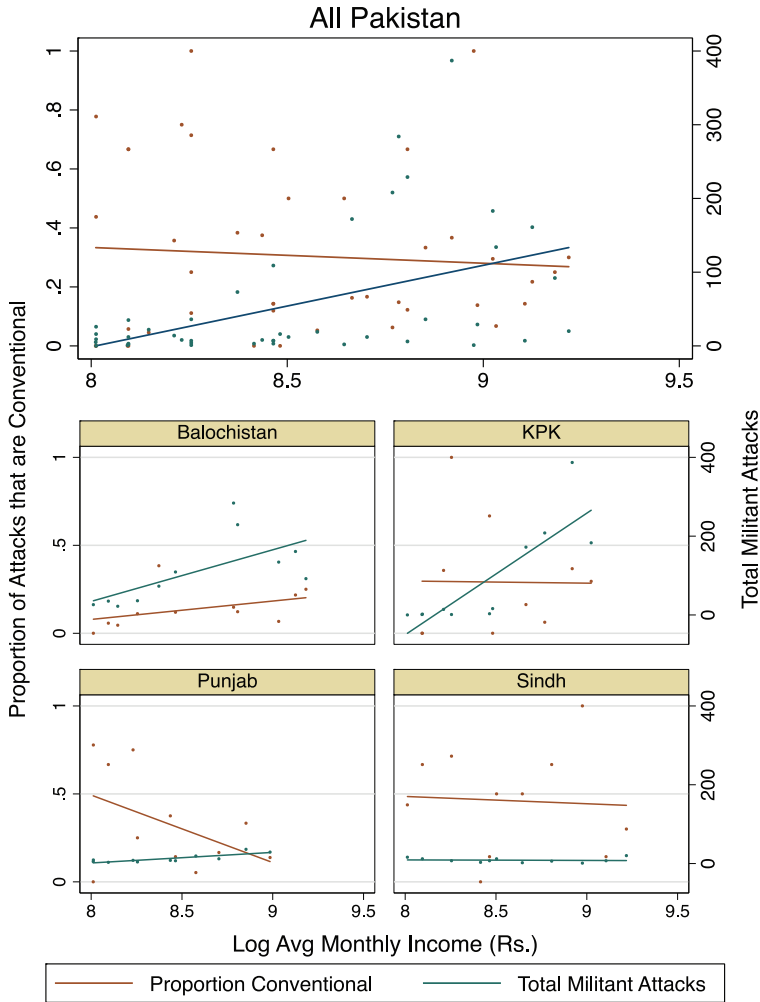
Figure 8. Income and violence across Pakistan, 2000–2010.

Several points are worth noting here. First, as discussed above, these correlations should not be over-interpreted. We have done nothing to address the problems associated with interpreting the obviously endogenous relationship between income and violence as a causal one. Second, we have looked at only one possible factor that might explain the relationship between total violence and the share of violence that is conventional. The BFRS data create the possibility of repeating this exercise with a variety of economic, political, social or other factors that might account for the trends in violence. Third, all of our analyses highlight the importance of taking regional heterogeneity seriously, a possibility opened up by sub-national data of the sort we provide.

Application to Pakistan-specific issues

Debates within Pakistan over the nature and the causes of the recent increase in political violence have generally proceeded without any systematic data. This problem is not unique to Pakistan; efforts to understand recent events in Iraq have suffered similar problems.[10] This section highlights a number of ways in which our data can contribute to these contemporary debates.

There is a general perception in academic and policy-making circles that Pakistan has become increasingly violent over the past decade. First, a variety of new groups have emerged and are targeting state security forces, ordinary citizens and political rivals. Second, the political will and capacities of the Pakistani state to defeat these groups and end political violence appear to be on the decline. Worse, on several occasions the Pakistani state made tacit agreements to "cede" various kinds of control to Islamist militants operating in FATA (2004, 2005) and in Swat (2008), among several other informal deals (Khattak, 2012). While, none of these deals brought peace, they did expand the political space for Islamists and Islamist militants by "effectively providing them a sphere of influence in the tribal areas and some settled districts of [KPK], including Swat" (International Crisis Group, 2009). Worse, the deals strengthened the links between Pakistan's militant groups and international organizations such as al-Qaeda and reinforced the efficacy of Islamist violence as a means of coercing the state. At the time of writing, the Pakistan's leadership is considering yet another round of negotiations with militants who demand implementation of Sharia across the country—even though militants were the primary beneficiaries of past deals rather than the state.

These arguments about Pakistan's descent into a quagmire of violence lack nuance. In Figure 9, we plot the annual per capita casualties from four kinds of political violence—riots and violent political demonstrations, terrorist attacks, militant attacks and assassinations—from 1988 to 2010 for each of Pakistan's four major provinces. Three facts stand out. First, political violence has not increased since 2005 in Punjab or Sindh, the two provinces that housed 79% of Pakistan's population in 2010. Second, the rate of terrorist attacks and militant attacks began increasing in Balochistan between 2002 and 2005, several years before the increase in KPK. Third, the nature of political violence in Balochistan has shifted substantially from the early 1990s, with terrorist attacks taking on a new prominence.

Conclusion

In this article, we introduce the BFRS dataset, which provides incident-level data on over 28,000 violent political events in Pakistan from January 1988 to May 2011. These data are intended to facilitate better research on patterns of violence in Pakistan and should be useful for testing theories about political violence, particularly those that take into account anti-government forces' abilities to make strategic choices regarding which tactics to use at different times.

Our initial analysis provides evidence that, as groups' overall engagement in violence increases, they tend to allocate a larger share of their efforts to conventional attacks. This pattern is true across much of Pakistan. A common argument is that such increases in militant capacity occur when the economy worsens because groups are better able to recruit fighters. In line with previous work on Afghanistan, Iraq and the Philippines (Berman et al., 2010), we find preliminary evidence for the opposite. At the national level, greater household income is
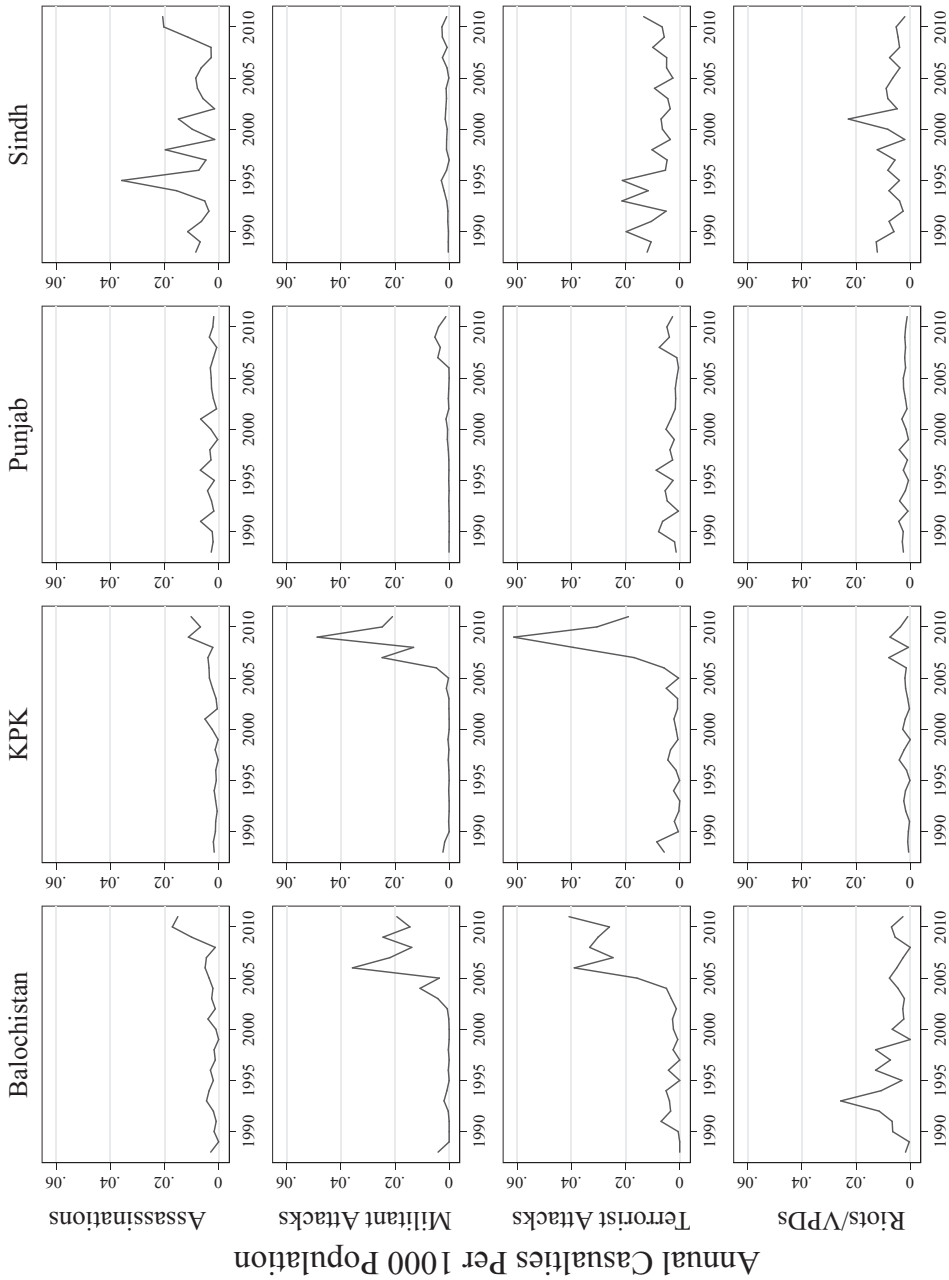
Figure 9. Consequences of nonstate violence in Pakistan, 1988–2010.

associated with more attacks (not fewer) and the proportion of attacks that are conventional in nature appears to be unrelated to income in three provinces and decreasing in one.

We also demonstrate that disaggregated subnational data are useful for providing insight into broad arguments being made in current policy debates. The BFRS data allow analysts to identify how trends in different kinds of political violence vary across regions, offering the potential for more informed discussions about why Pakistan continues to suffer such high levels of politically motivated unrest.

Acknowledgement

Funding

Notes

1. There is another dataset on Pakistan, which is maintained by the Pakistan Institute for Peace Studies (PIPS), an Islamabad nongovernmental organization. PIPS began collecting terrorism incident data in 2006. Previously, PIPS published summary statistics on these data and made them publicly available. Now researchers must purchase a subscription to the dataset. Unfortunately, PIPS is not transparent about coding methodology or the definitions it uses for data extraction.
2. To be included as a terrorist attack in the GTD, an event must involve the threat or actual use of "illegal force and violence by a nonstate actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation". Inclusion in the GTD also requires that all three of the following attributes be present: (1) the event must be the intentional result of a perpetrator acting consciously; (2) it must involve some degree of violence or threat of violence against people or property; and (3) the perpetrators must be subnational actors. Unlike the GTD, our data include cases in which the state commits the act of terrorism. Additionally, at least two of three other criteria must be met for inclusion in the GTD. The first criterion is that "The act must be aimed at attaining a political, economic, religious or social goal". The second of three criteria is that "there must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims". Third and finally, the act "must be outside the context of legitimate warfare activities". GTD further uses a filtering mechanism in order to distinguish between attacks that are clearly acts of terrorism and those can be interpreted as insurgent or guerilla violence, internecine conflict action or mass murder; Global Terrorism Data Base (2012).
3. WITS, which was maintained by the National Counterterrorism Center before it was discontinued in April 2012, employed a definition of terrorism that was prescribed in its congressional reporting statute, 22 U.S.C. § 2656f (d)(2). According to this statute, "the term 'terrorism' means premeditated, politically motivated violence perpetrated against noncombatant targets by subnational groups or clandestine agents". While definitions used by WITS and GTD seem somewhat similar at first blush, there are important differences. WITS, for example, does not require that the

perpetrators seek to engage an audience apart from the victims, nor does it require that the act be outside of legitimate war-time activities; US Department of State, Office of the Coordinator for Counterterrorism (2012).

4. The major English-language daily in Pakistan, *The Dawn*, was only added to Factiva in November 2012.

5. In Pakistan, the province is the first subnational level of administration. In turn, each province comprises numerous districts (*zillah*). Within each district are numerous tehsils. Below the tehsil is the "union council", which we do not include because it is rarely included in press reports.

6. The database and related codebook can be found online through the Empirical Studies of Conflict Project (ESOC). See https://esoc.princeton.edu/files/bfrs-political-violence-pakistan-dataset

7. The primary data were derived from the Lahore edition of *The Dawn*.

8. The terms "event" and "incident" are used interchangeably.

9. For evidence of rebels choosing strategically under various organizational and tactical constraints see Shapiro (2013) and Staniland (2014).

10. For examples of efforts to remedy this state of affairs see Biddle et al. (2012) and Weidmann and Salehyan (2012).

References

Berman E, Callen M, Felter JH and Shapiro JN (2010) Do working men rebel? Insurgency and unemployment in Afghanistan, Iraq, and the Philippines. *Journal of Conflict Resolution* 55(4): 496–528.

Biddle S, Friedman JA and Shapiro JN (2012) Testing the surge: Why did violence decline in Iraq in 2007? *International Security* 37(1): 7–40.

Bueno de Mesquita E (2013) Rebel tactics. *Journal of Political Economy* 121(2): 323–357.

Dube O and Vargas JF (2013) Commodity price shocks and civil conflict: Evidence from Colombia. *Review of Economic Studies* 80: 1384–1421.

Global Terrorism Data Base (2012) Codebook: Inclusion Criteria and Variables, p. 6. Available at: http://www.start-dev.umd.edu/gtd/downloads/Codebook.pdf

International Crisis Group (2009) Pakistan: The militant Jihadi challenge. *Asia Report* 164: 5.

Khattak D (2012) Reviewing Pakistan's peace deals with the Taliban. *CTC Sentinel*. Available at: https://www.ctc.usma.edu/posts/reviewing-pakistans-peace-deals-with-the-taliban.

Raleigh C, Linke A, Hegre H and Karlsen J (2010) Introducing ACLED—Armed Conflict Location and Event Data. *Journal of Peace Research* 47(5): 1–10.

Shapiro JN (2013) *The Terrorist's Dilemma: Managing Violent Covert Organizations*. Princeton, NJ: Princeton University Press.

Staniland P (2014) *Networks of Rebellion: Explaining Insurgent Cohesion and Collapse*. Ithaca, NY: Cornell University Press.

US Department of State, Office of the Coordinator for Counterterrorism (2012) National Counterterrorism Center: Annex of Statistical Information, Country Reports on Terrorism 2011 Report. Available at: http://www.state.gov/j/ct/rls/crt/2011/195555.htm

Weidmann NB and Salehyan I (2012) Violence and ethnic segregation: A computational model applied to Baghdad. *International Studies Quarterly* 57(1): 52–64.

Appendix

Table A1

Incidents / Assassinations / Violent political demonstrations

Statistic		Count			Casualties		
Outcome	Geographical unit	Lahore	Karachi	Difference/Lahore	Lahore	Karachi	Difference/Lahore
Incidents							
Mean per week	All Pakistan	2.78	2.54	0.086	9.52	10.51	-0.104
	Balochistan	1.53	0.8	0.477	4.04	2.68	0.337
	FATA	2.91	1.82	0.375	18.6	16.92	0.09
	KPK	2.41	2.04	0.154	9.84	11.32	-0.15
	Punjab	6.39	5.24	0.18	23.38	23.77	-0.017
	Sindh	8.04	9.81	-0.22	18.01	27.18	-0.509
Assassinations							
Mean per week	All Pakistan	0.94	0.77	0.181	1.45	1.47	-0.014
	Balochistan	0.19	0.12	0.368	0.33	0.23	0.303
	FATA	0.32	0.24	0.25	0.52	0.51	0.019
	KPK	0.48	0.53	-0.104	0.89	0.97	-0.09
	Punjab	2.78	1.75	0.371	4.26	4.16	0.023
	Sindh	3.68	3.41	0.073	5.52	5.75	-0.042
Violent political demonstrations							
Mean per week	All Pakistan	0.6	0.4	0.333	2.37	2.22	0.063
	Balochistan	0.21	0.04	0.81	0.76	0.25	0.671
	FATA	0.24	0.18	0.25	1.1	1.74	-0.582
	KPK	0.45	0.31	0.311	1.54	1.1	0.286
	Punjab	1.74	0.81	0.534	8.27	6.23	0.247
	Sindh	2.1	1.77	0.157	7.04	8.21	-0.166

Militant attacks / Security force actions / Conventional military violence

Statistic		Count			Casualties		
Outcome	Geographical unit	Lahore	Karachi	Difference/Lahore	Lahore	Karachi	Difference/Lahore
Militant attacks							
Mean per week	All Pakistan	0.36	0.26	0.278	1.75	1.5	0.143
	Balochistan	0.28	0.26	0.071	0.46	0.46	0
	FATA	0.79	0.46	0.418	5.2	4.16	0.2
	KPK	0.44	0.32	0.273	3.26	2.3	0.294
	Punjab	0.43	0.38	0.116	2.56	1.96	0.234
	Sindh	0.24	0.35	-0.458	0.87	1.66	-0.908
Security force actions							
Mean per week	All Pakistan	0.21	0.3	-0.429	1.6	1.78	-0.112
	Balochistan	0.04	0.06	-0.5	0.3	0.09	0.7
	FATA	0.83	0.53	0.361	8.64	8.22	0.049
	KPK	0.2	0.2	0	1.32	1.89	-0.432
	Punjab	0.42	0.64	-0.524	2.04	2.21	-0.083
	Sindh	0.17	0.98	-4.765	0.46	1.75	-2.804
Conventional military violence							
Mean per week	All Pakistan	0.11	0.11	0	0.66	0.64	0.03
	Balochistan	0.13	0.14	-0.077	0.29	0.14	0.517
	FATA	0.32	0.24	0.25	3.4	2.47	0.274
	KPK	0.16	0.15	0.063	0.66	0.76	-0.152
	Punjab	0.11	0.13	-0.182	0.45	0.32	0.289
	Sindh	0.12	0.22	-0.833	0.39	1.39	-2.564

Terrorist attacks / Total militant violence

Statistic		Count			Casualties		
Outcome	Geographical unit	Lahore	Karachi	Difference/Lahore	Lahore	Karachi	Difference/Lahore
Terrorist attacks							
Mean per week	All Pakistan	0.48	0.22	0.542	1.98	2.2	-0.111
	Balochistan	0.75	0.12	0.84	2.11	1.19	0.436
	FATA	0.5	0.22	0.56	2.29	1.49	0.349
	KPK	0.61	0.33	0.459	2.59	4.33	-0.672
	Punjab	0.66	0.28	0.576	5.4	6.2	-0.148
	Sindh	1.23	0.75	0.39	3.07	4.18	-0.362
Total militant violence							
Mean per week	All Pakistan	0.84	0.48	0.429	3.72	3.7	0.005
	Balochistan	1.03	0.39	0.621	2.57	1.66	0.354
	FATA	1.29	0.68	0.473	7.49	5.65	0.246
	KPK	1.05	0.65	0.381	5.85	6.63	-0.133
	Punjab	1.09	0.66	0.394	7.96	8.16	-0.025
	Sindh	1.46	1.11	0.24	3.94	5.84	-0.482