Studies in Conflict & Terrorism

ISSN: 1057-610X (Print) 1521-0731 (Online) Journal homepage: http://www.tandfonline.com/loi/uter20

Peering through the Kaleidoscope: Variation and Validity in Data Collection on Terrorist Attacks

Brandon Behlendorf, Jyoti Belur & Sumit Kumar

To cite this article: Brandon Behlendorf, Jyoti Belur & Sumit Kumar (2016) Peering through the Kaleidoscope: Variation and Validity in Data Collection on Terrorist Attacks, Studies in Conflict & Terrorism, 39:7-8, 641-667, DOI: 10.1080/1057610X.2016.1141004

To link to this article: http://dx.doi.org/10.1080/1057610X.2016.1141004

Published online: 06 Apr 2016.

STUDIES IN CONFLICT & TERRORISM
2016, VOL. 39, NOS. 7–8, 641–667
http://dx.doi.org/10.1080/1057610X.2016.1141004

Peering through the Kaleidoscope: Variation and Validity in Data Collection on Terrorist Attacks

Brandon Behlendorf[a], Jyoti Belur[b], and Sumit Kumar[c]

[a]National Consortium for the Study of Terrorism and Responses to Terrorism, Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA; [b]Department of Security and Crime Science, University College London, London, UK; [c]Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

ABSTRACT

The nature of underreporting terrorism in developing countries is often acknowledged but poorly understood. Focusing on India, we triangulate terrorist attacks captured across three media-based datasets (Global Terrorism Database, South Asia Terrorism Portal, Worldwide Incident Terrorism System) against official police records from Andhra Pradesh. Results suggest that media-based datasets capture the geographic prevalence of terrorism yet severely underestimate the frequency of violence, biasing toward lethal bombings. Considerable variation is present for attacks targeting specific classes or types of actors. Similar to other crimes, the results suggest that existing terrorism databases represent a select version of violence in these countries, discounting the prevalence and regularity of non-lethal violent activity.

Over the past decade, one of the most widely used methods of empirical analysis on insurgency, political violence, and terrorism is the collection and coding of events from open-media sources. A number of datasets on political violence have become publicly available, allowing researchers to apply statistical methodologies to identify the correlates, causal factors, and consequences of political violence.

For all of the empirical advances in the study of political violence and insurgency, little is known about the potential sources for error in existing media-based datasets. A majority of studies use a single data source of events, which contain potential biases that alter the number, type, and classification of events available for analysis.[1] Some events are viewed as more "newsworthy" and thus are more likely to be reported,[2] including large, significant events[3] with police presence[4] in proximity to a news agency[5] and located in an urban area away from most types of land-based or rural conflicts.[6] Failure to account for these biases can present political violence as primarily lethal attacks in urban areas conducted by major non-state actors.[7] In the few studies where multiple sources are used, they are often integrated

CONTACT Brandon Behlendorf bbehlend@umd.edu National Consortium for the Study of Terrorism and Responses to Terrorism, Department of Criminology and Criminal Justice, University of Maryland, 8400 Baltimore Ave., Suite 250, College Park, MD, USA.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uter.

with little consideration regarding the type and nature of data collection which produced the original set of events.[8] In these cases, events are not equivalent across samples, but are productions imbued with social and political considerations about what is and is not an "event."

To address these concerns, we examine how the data collection process can alter or shade the representation of political violence at the local level, and the implications that has for research in this area. This study compares the representative level of a specific type of political violence (terrorist attacks) captured across three media-based datasets against official police records, a source often missing from studies on terrorism and political violence. This is a good case study for how international datasets engage with domestic terrorism. The nature of Naxal insurgency is widespread and longstanding enough to be of serious national concern in India, a democratic country with free and open multilingual media. Andhra Pradesh, one of several Indian states that have been grappling with Naxal insurgency, has the most detailed records of "Naxal related cases" compiled at the state level.[9] Using the Naxalite conflict in Andhra Pradesh as our context, we find that although media-based datasets capture the geographic prevalence of political violence, they tend to severely underestimate the intensity of conflict at the local level and bias the representation by favoring lethal attacks and those involving explosives. In addition, there is considerable variation across datasets regarding specific classes or types of actors targeted by political violence, often influenced by the process in which an attack is sourced, collected, and adjudicated.

The article is structured as follows: The first section discusses the nature of data collection during conflict in both media-based and official datasets. The next two sections describe the research questions, context and background of the Naxal insurgency within Andhra Pradesh. The subsequent two sections focus on the data and methodology adopted, followed by the results and reasons for the variation between media-based datasets and official data sources. The final sections include a discussion of the implications and recommendations for future research using media-based datasets.

Nature of Data Collection on Terrorism

Media-Sourced Attack Databases

The past two decades have witnessed a sea change in the global accessibility of news on conflict and violence. Previously the purview of government efforts like the Foreign Broadcast Information Service (FBIS), the diffusion of the Internet now provides constant viewership to conflicts worldwide through news and social media. This availability has enabled the emergence of several international event-level databases that quantify terrorism. Databases such as the Global Terrorism Database (GTD), the International Terrorism: Attributes of Terrorist Events (ITERATE) dataset on transnational terrorism, the Worldwide Incident Terrorism System (WITS), the RAND Database of Worldwide Terrorism Incidents (RDWTI), and others have developed multiple perspectives on terrorism, each with their strengths and weaknesses. Across the empirical study of terrorism, the post-9/11 swell of research has relied primarily on the ITERATE or GTD datasets for their analyses, often used, "without questioning their underlying definitions, coding consistency or contrasts."[10] Key contrasts, such as the differentiation between domestic and transnational terrorism, have shaped research findings in substantive ways,[11] yet little comparative work exists to identify how these multiple views of terrorism relate to each other and to conditions "on the ground."

The challenging nature of collecting event-level data on terrorism has meant that most databases rely on print or electronic media for terrorism reports. While terrorist attacks are more likely to be reported in the media than other types of criminal activity,[12] Schmid[13] identified four problems with most terrorism databases. First, each database suffers from the inherent difficulties involved with working definitions of terrorism and terrorist acts. Wider situational, contextual, and political factors determine whether certain acts will be perceived as terrorism or not by various audiences. Where one dataset will include terrorist attacks against military targets (GTD), others will exclude those as acts of insurgency (ITERATE). Stringent inclusion criteria can also contribute to the exclusion of unclear attacks, such as those without independent confirmation or with conflicting reports, from the set of attacks.

Second, even if a working definition of terrorism is adopted, it is often inconsistently applied at different points of time for a variety of reasons. Some of these include changing political contexts surrounding specific organizations and their intersection with political violence.[14] For example, until recently, the European Union (EU) would not recognize the military wing of Hizballah as a designated terrorist organization, even though the United States designated them in 1997. It was not until a bombing in Bulgaria in 2012 was linked to the group that the EU would consider changing their designation. Other errors can arise from operationalizing difficult theoretical concepts in limited information environments (like media sources) or from the inconsistent application of definitions by the research teams themselves.[15] Definitions that distinguish terrorist acts targeting civilians from guerrilla attacks targeting military personnel[16] are often challenged in determining whether attacks against police forces constitute terrorism or insurgency, especially given the wide variance in centralization and militarization of police across countries.

Third, media reporting is not always objective; it is often politicized and biased regarding what is "newsworthy."[17] Political pressure can suppress some attacks from being widely reported, while in other cases relatively smaller attacks in distant nations may not reach a threshold for significance.[18] More importantly, the likelihood of an attack being reported could rely on whether other relevant attacks are reported nearby (which can vary from season to season), editor or reporter bias, and whether the news agency has the necessary resources to cover the attack.

Finally, only a few media-based databases collect data on domestic terrorist attacks, especially in less-developed nations. Media-based databases such as the GTD and WITS include domestic terrorist attacks in developing countries but draw mainly upon electronic and print media reports reported in widely disseminated national newspapers or international English-language sources. In many countries, including India, the likelihood of an attack being picked up by the national media would depend on the location and seriousness of the attack and the interest of the local media in reporting it.[19] Attacks that are reported in smaller, local vernacular press are unlikely to be identified and subsequently included in many international databases. Despite these problems, many have argued that newspaper sources remain a reasonably complete and continuous source for data collection, assuming that the variance provided by press bias is stable within and across sources.[20] More importantly, to draw inference on the nature of terrorism, users of these databases assume that any missing data on terrorist attacks is randomly distributed and not the product of the data collection process.

Official Data

A second potential source of data on terrorist activity, albeit underutilized, is official data collected by government counterterrorism or law enforcement agencies. Often collected by states with a long history of terrorism (Turkey, Israel, Colombia, and the United States[21]), these countries have established both systematic and/or *ad hoc* frameworks for identifying, analyzing, and reporting terrorist-related activity; frameworks with considerable within- and between-country variation. These datasets have been used to analyze the spatial distribution of terrorism,[22] build geographic profiles of offenders,[23] and assess the deterrent effect of arresting terrorists.[24] These sources could also provide rich detail on specific plots or attacks drawn from extensive investigations or court records that are never reported by media, providing an alternative perspective of terrorism as mediated through official channels.[25]

While useful, three specific limitations of official data stand out.[26] First, definitions of terrorism are often overloaded by emotional and political implications with little agreement between agencies and individuals.[27] Second, the police can often attribute nonviolent or other criminal actions arising from non-ideological motives committed by known terrorist organizations as terrorist acts. These types of non-terrorist acts, including extortion and assault, could be used for personal gain and/or to instill fear or terror in a target population, rendering it difficult to distinguish the purely acquisitive or personal crimes from terrorism in some official data collection efforts. It has also been suggested that while police archives are a good alternative to media-based datasets, they also project their own biased view of offender and offense.[28] On the other hand, previous research by Chermak and colleagues[29] comparing official data with open source data on homicides committed by right wing extremists in the United States indicated that the official data compiled by the Federal Bureau of Investigation (FBI; responsible for investigating all domestic terrorist attacks) is the smallest, as their definition of terrorism is the most restrictive as compared to datasets complied by watch-group listings, scholars, and others based on open sources. Thus, we might expect official databases to be expansive or narrow, depending on the inclusion criteria.

A third problem is that the difficult prosecutorial challenges presented by terrorism investigations could force formal criminal complaints that are for other-than-terrorism crimes. For example, in India most acts of terrorism recorded by the police are charged under relevant sections of other criminal laws, such as the Indian Penal Code of 1861, the Arms Act of 1959, and the Explosive Substances Act of 1908, and are part of the overall crime figures reported in the annual Crime in India reports published by the National Crime Records Bureau. Only a few states maintain a separate database of all Naxal-related offenses, with little coordination between the states and a national dataset on Naxal-related offenses.

Beyond these specific concerns, official data on terrorist-related activity also suffers from the traditionally accepted deficiencies of official crime data: victim reporting behavior,[30] police operational norms, legislative changes,[31] changes in formal recording rules and politicization of crime,[32] all of which affects how and what is recorded as crime.[33] These findings were replicated in New York by Green and colleagues,[34] who found that gay advocacy groups recorded much higher numbers of antigay hate crimes than the police for the same geographic area. The authors speculate that this was probably because of either different reporting standards or varying levels of commitment to recording hate crime. These factors suggest that official data also suffers from the "dark figure of unrecorded and unreported crimes,"[35] which could distort the validity of research findings supported by official data.[36]

Despite these problems, Barranco and Wisler[37] suggest that a strong candidate for improving the validity of media-based data is to compare against official data recorded by the police. Using the example of "public demonstrations," they argue that since events like public demonstrations are the objects of administrative regulation, they are more likely to be recorded. Similarly McCarthy and colleagues,[38] studying demonstrations in Washington between 1982 and 1991, found that the vast majority of demonstrations were ignored by the media, leaving only the largest ones covered. As a political conflict progresses, the use of violence may become routine, fatiguing the press into covering only the most severe events.[39] The inherently administrative character of the state when policing terrorism provides an alternative source that may capture a broader range of violence than available through open media sources, although the extent to which these two sources differ remains understudied.

Research Questions

Overall, we are interested in examining potential issues of selection bias within three publically available databases on terrorism: the GTD, the WITS, and the South Asia Terrorism Portal (SATP). Several of these sources have provided the basis for quantitative analyses previously,[40] yet to date there have been no studies comparing the quality of data collection among all three.[41] Some studies that try to improve the quality of data analysis integrate or combine multiple datasets to acquire the broadest account of data on a specific conflict, yet utilize media sources, which rely on select information migrating into the public sphere. We compare the three collections of terrorist attacks, as well as a combined version of all three, against a fourth source of event data (official police records), which the literature suggests may contain more or fewer recorded attacks than media-based data sources.

First, we focus on the nature and type of events that constitute each of the datasets, as well as the differences between them. The data from media sources draw from information filtered and processed through media exposure, with each dataset providing a different lens on the phenomena. While some collections rely on local vernacular newspapers, others use a combination of national and international sources, both of which can produce considerably different perspective on the quantity of terrorism recorded. Inclusion criteria regarding the validity of sources can also influence what attacks are captured, with some collections using more stringent requirements on the reputational quality of a source. Understanding how the number of cases reported across these databases varies can provide a baseline perspective about the nature of terrorism captured within each dataset.

Second, we examine whether this variation is spatially or temporally patterned. Each of these datasets is the result of a collection schema that is carried out over time, and each can have important variation regarding where and how often cases are collected. Third, we are interested in the typological variation among attacks reported across the four sources, and where these facets of terrorism diverge. Are media-based collections more likely to contain certain types of terrorist attacks (bombings, fatal attacks), and do differences in the environmental context of an attack (the distance from the capital, ruggedness, etc.) affect the likelihood of attacks being reported? Even though many studies of conflict employ structural factors (like poverty) as controls,[42] little research has examined whether attacks in these areas are less likely to be reported. Variation in reporting based on ecological factors surrounding an attack can have substantial impacts on theories of political violence, as it may overestimate the relationship between certain geographic or demographic factors and the prevalence of insurgency.

We address these three questions by examining the recorded prevalence and quality of terrorist attacks within a rural land-based conflict in a developing country: the Maoist insurgency in Andhra Pradesh.[43] We utilize records from 2005 to 2009, a period in which all four datasets maintain coverage of terrorist attacks within the region. Additionally, the period under review is sufficiently past the initiation of the Maoist conflict in Andhra Pradesh, and therefore should maximize the likelihood of coverage by media sources. This is especially true as the nature of the conflict in Andhra Pradesh became more central to the mission of the state government, and the publicity surrounding the counterterrorism efforts in the state drew national attention.[44]

The Maoist Conflict in Andhra Pradesh

In 2006, Prime Minister Manmohan Singh declared the Naxal movement, increasingly referred to as left-wing extremism (LWE) or Maoism, as the "single largest internal security threat" that India has faced.[45] The Maoist movement has a long history in Andhra Pradesh, fueled by a number of grievances against the state, including: lack of land reforms, labor exploitation, lack of governance, lack of development and employment opportunities, state-sponsored exploitation of natural resources at the expense of tribal populations, deepening caste and economic inequalities, and lack of access to justice.[46] Initially starting as a Communist struggle against feudal rule in the 1940s, the movement morphed into an agrarian uprising over land reforms in the 1960s. While tempered during the 1970s, the Communist movement began to regain strength in the early 1980s, spawning several splinter groups, including the People's War Group (PWG) in Andhra Pradesh, and the Maoist Communist Centre (MCC) in Bihar.[47] The PWG was committed to armed struggle against the state, including kidnapping, extortions, and the killing of civilians and political leaders.[48] The late 1980s saw a dramatic escalation of conflict after the PWG decided to target the state directly through attacks on the police, inaugurating a campaign of violence which still continues currently.[49] This cyclical upsurge of insurgent and counterinsurgent violence led to substantial casualties for security forces, who have responded with severe state repression, brutal search and cordon operations, torture, beatings and displacement of populations.[50] Eventually, the government adopted a development-focused approach to erode popular support for the Naxals as well.[51]

The 2004 merger of the PWG and the MCC to form the Communist Party of India (Maoist) ushered in a new phase in the Naxalite Movement, which assumed a pan-India formed with a centralized leadership under unified command and control. At the time, the Andhra Pradesh government called for a six-month hiatus in the conflict to conduct peace talks. However, since peace talks broke down in early 2005, the state has been uncompromising in its opposition of the Maoists. Using a number of the strategies previously mentioned, the government of Andhra Pradesh has been mostly successful in reducing the frequency of Naxalite violence within the state, evidenced by declining fatalities of security forces as well as civilians. Andhra Pradesh police have been actively involved in a rising number of arrests and surrenders of Naxals, including prominent leaders, which have weakened the movement considerably.[52]

Toward the end of 2009 the state witnessed a surge of Naxal activities in the districts of the Telangana region in a bid to regain their stronghold in these areas, but with limited success.[53] As of 2014, active Naxal presence is largely confined to two districts—a sea change from the situation in 2005 when nearly all of the 28 districts in Andhra Pradesh were considered to be highly or moderately affected.[54] While the situation continues to cause grave concern, the total number of casualties (civilians and security forces) associated with the movement has been declining nationally from 1005 in 2010 to 415 in 2012; and in Andhra Pradesh from 24 in 2010 to 13 in 2012.[55]

Data and Methodology

Given the strengths and weaknesses of terrorism data discussed above, researchers have long recognized the need for using multiple data sources to triangulate the phenomena under study.[56] To examine the salience of data collection procedures on the identification of political violence in India, we chose four event-level datasets that capture this information relevant to the Maoist conflict from 2005 to 2009. One of the sources are existing data collections is readily available for scholars (GTD), one source was previously available and used in a number of previous studies (WITS), one source requires the extraction of data from semi-unstructured text (SATP), while a fourth source comes from official records (Andhra Pradesh Police).[57] Table 1 provides a summary of these definition criteria and scope of data collection effort across the four datasets.

Global Terrorism Database (GTD)

Described elsewhere,[58] the GTD consists of data on terrorist attacks collected from open media sources, including wire services (FBIS and Reuters, among others), U.S. and foreign government reports, and U.S. and foreign newspapers (including the *New York Times*, the *British Financial Times*, the *Christian Science Monitor*, the *Washington Post*, the *Washington Times*, and the *Wall Street Journal*). It defines terrorism as "the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation."[59] The GTD has utilized a similar basic coding scheme during the entire 41 years of data collection. Compared to other media-based databases, from its inception GTD has tracked the kind of domestic terrorist attacks that have characterized localized conflicts like the Maoists in India. Yet a comparison against other sources of terrorist activity remains lacking. Relevant to this discussion, the GTD also includes attacks against the military and police, actions that are sometimes classified in other media-based databases as insurgent actions and distinct from terrorist attacks.

Worldwide Incident Tracking System (WITS)

Begun in 2004, WITS was a database of terrorist attacks collected by the National Counterterrorism Center in the United States. Initially collected for the U.S. government's yearly *Country Reports on Terrorism*, WITS contains information on a number of variables related to specific terrorist attacks, including attack types and weapons used. According to their documentation, a terrorist attack is one where "groups or individuals acting on political motivation deliberately or recklessly attack civilians/non-combatants or their property and the

Table 1. Definitional differences between datasets.

Dataset	Method of data collection	Source	Criteria for Inclusion	Events other than terrorism?	Perpetrators other than Naxal?
Global Terrorism Database 1970–current	Until 2007: retrospective Since 2008: prospective	Media articles: • national and international until 2011; • local articles added starting in 2012	For an attack to be included in the GTD all of the following (3) attributes must be present: 1. Attack must be intentional—result of conscious calculation on the part of the perpetrator 2. Attack must entail some level of violence 3. There must be sub-national perpetrators (GTD limits itself to non-state terrorism) In addition, at least two of the following criteria must be present before an attack can be included: 1. The act must be aimed at attaining a political, economic, religious or social goal. In terms of economic goals, the exclusive pursuit of profit does not satisfy this criterion 2. 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 3. The act must be outside the context of legitimate warfare activities	No	Yes
Worldwide Incident Tracking System 2004*–2011	Prospective	Media articles: national and international	Events that meet the definition criteria of 22 U.S.C.§2656f(d)(2), in which "groups or individuals acting on political motivation deliberately or recklessly attack civilians / non-combatants or their property and the attack does not fall into another special category of political violence, such as crime, rioting, or tribal violence."	No	Yes
South Asia Terrorism Portal 1992–current	Prospective	Media articles: local and regional	Reports drawn from open sources of attacks, clashes, encounters, and/or statements made by either nonstate violent organizations within South Asia or by governments targeting these organizations. Data is unstructured, so other than association with a specific group (like Naxals), there is no additional inclusion criteria.	Yes	Yes
Andhra Pradesh Police Data 2000–2010	Prospective	Citizen reports; Police reports	Reported events by citizens to local police stations, who then report to District Crime Bureaus, and then reported to State Intelligence Bureau task force for Naxal-related events. Key inclusion is whether event is Naxal-related, and the collection contains terrorist attacks, clashes, threats, extortions, kidnappings for ransom, and other events involving known or suspected Naxal-related groups or individuals.	Yes	No

*2004 data was limited in production; 2005 represented the first full year of data collection.

attack does not fall into another special category of political violence, such as crime, rioting, or tribal violence."[60] Unlike GTD, WITS does not release information on the perpetrator of the attack; rather they provide general classifications of perpetrator ideologies (Islamic Extremist, Environmental, Tribal, etc.).[61]

One limitation of both the GTD and WITS is the requirement of intentionality as a criterion for inclusion in the database. Specifically, there has to be no evidence in the attack that the attack was started by someone other than the group committing the attack. "Clashes" between security forces and terrorist organizations, common in a number of insurgent environments, do not satisfy this requirement, as it would be difficult to distinguish who initiated the attack. While some of these "clashes" may actually have been extrajudicial killings initiated by the security forces, otherwise known as "police encounters" within India,[62] the limited information on attacks contained within many media-based sources in these areas restricts the ability of the researcher to succinctly identify these cases. What is presented provides at best a conservative estimate of the actual group-initiated violence in these areas.

South Asia Terrorism Portal (SATP)

A third set of data relevant to the Maoist conflict in India comes from the South Asia Terrorism Portal. A product of the Institute for Conflict Management (ICM) in New Delhi, the SATP reviews a large number of national and local newspapers, government and nongovernment publications, and other Web-based media sources for reports on violence, insurgency, and terrorism across a number of violent conflicts in India and other South Asian countries.[63] It is a source of information for a variety of data collection efforts that use it as either a primary source of data collection[64] or as a secondary "seed" source to provide preliminary information on attacks that can be corroborated with additional primary source media.[65]

Event-level information within SATP is reported in a semi-unstructured list, which describes the action conducted, geographic information on where the event occurred, and the date of those actions. Data on these events do not reside in a structured format for analysis. Researchers interested in using this information must recode the data manually. The list of events contain a much broader array of information, including violent attacks against combatants as well as statements made by terrorist organizations and arrests of key leaders. These lists are usually grouped by an aggregate location or group conducting the action, and multiple lists can contain information on the same event. Additionally, SATP appears to be conflict-focused, containing considerable variation by date and location. For some conflicts, consistent information is available beginning in 2000 or 2001; for others (like insurgent violence in Assam, India) SATP provides information back to 1992.

Andhra Pradesh Police

To complement the three media-based data collections, and to provide a more detailed set of Naxal terrorist events in Andhra Pradesh, the fourth dataset examined was received from the Andhra Pradesh Police. This database includes every reported event of Naxal related-violence or threat recorded in individual police stations across the state. The State Intelligence Bureau has access to all recorded crimes at the state level and compiles a state-wide database of Naxal-related events as they are recorded. It does not distinguish between ideological and non-ideological events, but includes all events where the alleged perpetrators are known or suspected of being Naxalites. As mentioned above, these might include crimes committed by alleged Naxalites for personal causes, and this more expanded inclusion criterion implies that more, rather than fewer, events are captured by this data set compared to other sources. Brief facts of the case in the original data were coded to construct categories similar to the above datasets. It includes events that would not ordinarily qualify as a terrorist attack under the criteria set by GTD or WITS. However, it is extremely instructive in illustrating the types of events considered to be related to Naxal insurgency by the local law enforcement agency.

As police-recorded data, these data suffer from all the traditionally accepted deficiencies that official police data suffers from as discussed above. The original Andhra Pradesh police data demonstrates changing modes in recording practices that could skew inferences about trends and patterns.[66] Although official data provided by the Andhra Pradesh police may contain some variations in the coding strategy across time, the systematic and comprehensive efforts they undertook to collate this data at the state level provides a unique opportunity to utilize official, event-level data collected during an insurgency.

Methodology

To examine the potential sources of selection bias and case reporting variation among these four sources, we employed a four-phase process. First, we conducted a case-by-case definitional alignment between each of the data collections. To maintain uniformity of definition between datasets, and to focus on terrorist attacks by Naxalites, we applied the GTD criteria to all four datasets and removed any events which were neither terrorist in definition nor leftist in ideological motivation (see Table 1). Across all four sources, only those cases that identified Naxal actors proactively committing a violent attack to further a political, social, or economic goal were included, with differences noted in the analysis. Project coders were trained by two of the authors who have multiyear experience coding GTD data, and each attack was coded according to the established criteria and subsequently reviewed by one of the authors to ensure consistency in coding.

Second, using a combination of descriptions and case attributes, we were able to triangulate cases among the various sources. This allows us to identify those cases which overlapped across all four sources, as well as those attacks that were unique to each data collection. Each comparison was initially conducted by a project coder and then reviewed by one of the authors to ensure accuracy in the case matching process. Search windows 14 days before and after a particular attack were used, as dates for specific attacks can vary considerably when comparing media and police sources. In addition, coders erred on the side of *inclusion* when key details (e.g., such as target or number killed) were vague in one of the two matched attacks, providing a conservative estimate of the number of attacks *excluded* from media-based datasets. Contrasting details between two nearly similar attacks, such as distinctly different targets or attack types, did not meet the threshold for matching and were not linked within the collection effort. For example, considering two attacks in the same district on the same day, if both had the same target type but one reported a death and another did not, then *ceterus paribus*, they were matched. If these two attacks were reported as one targeting physical infrastructure and one targeting civilians in a marketplace, then the contrasting details would render them unmatched.

Third, we employed descriptive analysis to identify potential variations in case reporting across the sources. Both spatial and temporal aspects of variation in case reporting are considered, as well as differentiation across various attack characteristics (fatal attacks, bombings, etc.). Finally, to examine whether environmental or structural factors influence the likelihood that an attack would be found within the media-based collections, we test a number of hypothesis in predicting the likelihood using logit models to compare cases from WITS, SATP, GTD, and an integrated dataset containing all unique cases from the media-based databases against cases from the police.[67]

Results

There was considerable variation in the number of events (both terrorist attacks and other types) across the datasets. The largest collection of events within Andhra Pradesh was the reports from the Andhra Pradesh Police, which documented a total of 1,436 left-wing extremism events between 2005 and 2009. During the same period, WITS recorded 296 events (all terrorist attacks), SATP recorded 229 events (a mixture of terrorist attacks and other events), and the GTD recorded 24 (all terrorist attacks). A review of these identified a number of events, especially within the police data, which would not meet a common definition of a terrorist attack, including lootings, extortions of local civilians, and other violent acts without political motivation. Applying the definition alignment mentioned previously, Table 2 presents our final count of terrorist attacks for the four datasets, which netted a total of 806 attacks in the police data, followed by WITS with 296 attacks, SATP with 197 attacks, and GTD with 24 attacks.

Once a common definition of left-wing terrorism was applied at the event-level across all four datasets, we conducted a case-by-case comparison to triangulate overlapping attacks between the collections. Using the police data as a starting point and using descriptions and case attributes to triangulate, between 85 and 90 percent of cases across the datasets were matched to cases in the police data.[68] Across those cases in the media-based datasets that were not found in the police data, several key patterns emerge. First, the lethality of an attack had little consistent influence on its omission from the police data. In some cases (SATP), the majority of missing attacks were fatal (10 of 17 attacks), while in other cases (WITS and GTD), the proportion that were fatal was considerably lower (43 percent, or 14 of 32 attacks, within WITS; 33 percent, or 1 of 3 attacks, within GTD). Second, for both WITS and SATP, there was an overrepresentation of missing cases in Khammam district, where Naxal fighting was particularly heavy. Although attacks in Khammam constitute 16 percent of total terrorist attacks within the police data, they represent over 25 percent of non-

Table 2. Dataset descriptives.

	Police Data	WITS	SATP	GTD
# of Events (attacks and other)	1,436	296	229	24
# of Events Meeting Terrorism Criteria (attacks)	806	296	197	24
# of Terrorist Attacks Matched to Police Data	806	264	180	21
# of Matched Terrorist Attacks After Triangulation	806	256	174	19
Percent of Terrorist Attacks Reported in Police Data	100	32	22	2
# of Terrorist Attacks Unmatched	0	32	17	3
Percent of Terrorist Attacks Unmatched	0	11	9	13

matched attacks within SATP, WITS, and GTD, suggesting that police data may also suffer from some of the same limitations of reporting within heavily affected locations that face media sources. Third, while attacks in 2008 or 2009 represent 14 percent of all attacks within the police data, they constitute 38 percent of missing attacks (12 out of 32) within WITS, 64 percent of missing attacks (11 out of 17) within SATP, and 100 percent of missing attacks (3 out of 3) within GTD. Thus, there could be an exhaustion effect within official data, where event capture lessens as the duration of administrative data collection is lengthened.

In the process of triangulation, however, new information was introduced on a number of cases; information which would change whether the case met the criteria for terrorism mentioned previously. For example, an attack on 10 July 2005 in SATP mentioned that CPI (M) Janshakti cadres abducted two employees of a transport company in the district of Khammam. However when we compared that attack to police data, we found that those abductions were solely for ransom, which does not meet the criteria of terrorism applied in this study, and the attack was subsequently dropped from the comparison. To provide a conservative estimate of the number of terrorist attacks which match across the datasets, we removed those cases from the overall analysis that were no longer classified as terrorist attacks based on the new information. This removed another two attacks from GTD, six attacks from SATP, and 12 attacks from WITS.

Temporal Variation

To examine temporal variation in matching between the three media-based datasets and the police data, Table 3 presents the total number of attacks in each source for

Table 3. Variation in attack matching by year, type, and target of attack.

	Police Data	WITS		SATP		GTD	
	#	#	%	#	%	#	%
Total Matched Attacks	806	256	32	174	22	19	2
2005	451	168	37	69	15	4	1
2006	144	18	13	31	22	2	1
2007	98	36	37	37	38	1	1
2008	65	21	32	21	32	5	8
2009	48	13	27	16	33	7	15
Fatal Attacks	309	160	52	123	40	12	4
1 killed	284	141	50	104	37	6	2
2 killed	14	8	57	8	57	1	7
3 killed	7	7	100	7	100	3	43
4 killed	1	1	100	1	100	1	100
5+ killed	3	3	100	3	100	1	33
Non-Fatal Attacks	497	96	19	51	10	7	1
Bombings	133	52	39	31	23	5	4
Non-Fatal	124	45	36	23	19	1	1
Fatal	9	7	78	8	89	4	44
Non-Bombings	670	204	30	143	21	14	2
Type of Target							
State Leaders / Officials	18	10	56	5	28	4	22
Village Leaders / Officials	42	20	48	14	33	2	5
Political Party Activists	77	32	42	26	34	2	3
Police / Security Forces	136	52	38	40	29	8	6
Other Targets	533	142	27	89	17	3	1

years between 2005 and 2009. The number of terrorist attacks in 2005 was substantially higher in all four datasets, reflecting the increased counterterrorism efforts of the Andhra Pradesh government in subsequent years. Factoring in the declining rate of terrorism across the years, there was considerable variation in the number of terrorist attacks which could be matched to attacks in the police data. From 2005 to 2009, as terrorism decreased in the state, both SATP and GTD increased their coverage of attacks. For GTD, there was a considerable increase in 2008 and 2009 which coincided with a change in data collection strategy.[69] In contrast, there was a considerable decline from 2005 to 2006 in both matched attacks (from 162 attacks to 18 attacks) as well as total attacks (from 193 attacks to 21 attacks) within the WITS data. Interestingly enough, starting in 2007 their count of attacks in Andhra Pradesh closely resembles the count of attacks within SATP data as well, which corresponds to a possible shift in data collection strategy which included SATP as a source for their data. Overall, while the three media-based datasets represented only a share of the total attacks reported within the police data, the temporal trends of the three datasets were roughly similar across the periods, suggesting that differences may be one of magnitude rather than type of attack (see Figure 1).

Variation by Type of Attack

When case comparisons are conducted by type of attack, two patterns clearly emerge. First, fatal attacks are overrepresented in media-based data collection efforts. Nearly half of all fatal

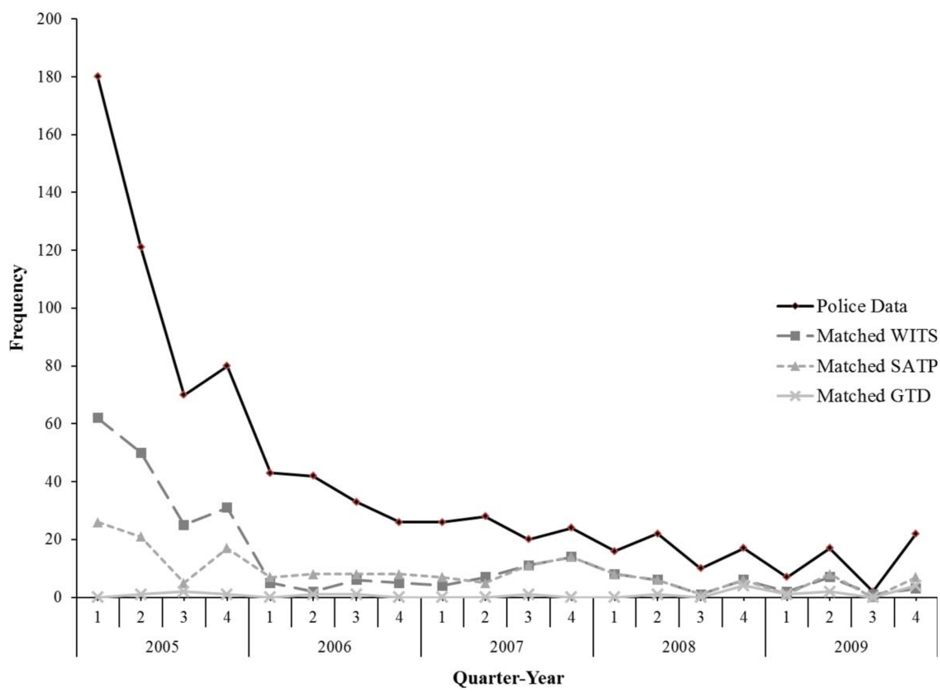

Figure 1. Quarterly trends in data reporting and matching, by dataset, 2005–2009.

terrorist attacks reported by the police were also reported by WITS, while 40 percent of the fatal attacks were reported by SATP. Disaggregated by number of individuals killed in the attack, the pattern is more apparent. Most of the terrorist attacks with at least two individuals killed are found within the media-based datasets, which suggests that the most lethal attacks are consistently filtered into the media. Among attacks with only one fatality reported by the police, each of the datasets exhibited considerable variation, ranging from two percent of attacks in the GTD to 48 percent covered in WITS. Regardless of the variation between datasets in reporting fatal attacks, the percent of non-fatal attacks reported is substantially low, ranging from 1 to 19 percent of the attacks reported by the police, suggesting that the landscape of terrorism reported in media-based data collections may over-represent the average lethality of an attack, especially in situations where the overall frequency of terrorist attacks (lethal and non-lethal) is much higher.

Second, this overrepresentation may be limited only to specific types of attacks that could attract substantial media attention. For example, bombings by left-wing terrorists in Andhra Pradesh are not sizably more represented within each dataset than other types of attacks. When the bombing results in a fatality, however, the likelihood that it will be recorded in media-based datasets increases dramatically. Of the nine fatal terrorist bombings in Andhra Pradesh from 2005–2009 reported in the police data, WITS recorded seven, SATP recorded eight, and GTD recorded four attacks, respectively. Furthermore, when considering various types of targets of attacks, we see that roughly 20 to 40 percent of police-reported attacks were recorded in WITS or SATP. The GTD was distinct in their lower reported rate of terrorist attacks targeting local political leaders, security forces,[70] or political party activists.

Spatial Variation

Finally, we consider whether variation exists in the geographic distribution of reported attacks. Comparing total, fatal, and bombing attacks across administrative districts[71] in Figure 2, we see that terrorist attacks reported by the police are distributed throughout the majority of Andhra Pradesh, representing the broad nature of terrorist activity among the Maoists during this period. Fatal attacks were concentrated in a smaller number of districts, while bombings occurred most often in Vishakapatnum, Khamam, and Warangal districts in the center and east of the state. Comparing across datasets, we find that WITS, SATP, and the police data all report a number of the same districts as top location for terrorist activity. For example, two of the top three districts in each dataset were Khammam and Vishakapatnum, consistent with the conflict over land rights in forested areas central to the ideology of the Maoists in these two districts.[72] In contrast, both WITS and SATP underrepresent terrorist attacks in several other districts, including Karimnagar (3 percent of WITS and 4 percent of SATP cases, respectively) compared to attacks reported by the police (9 percent). This district in particular served as the site of an important strategic development in the counterterrorism efforts against the Maoists in Andhra Pradesh,[73] and underrepresenting the level of terrorist violence in the district would hinder efforts to identify successful outcomes of experimental policies like the one implemented. Nevertheless, while there were geographic variations in the proportion of terrorist attacks in one district compared to another, with noted exceptions the differences across datasets geographically were one of magnitude within-district rather than variation between districts.[74]

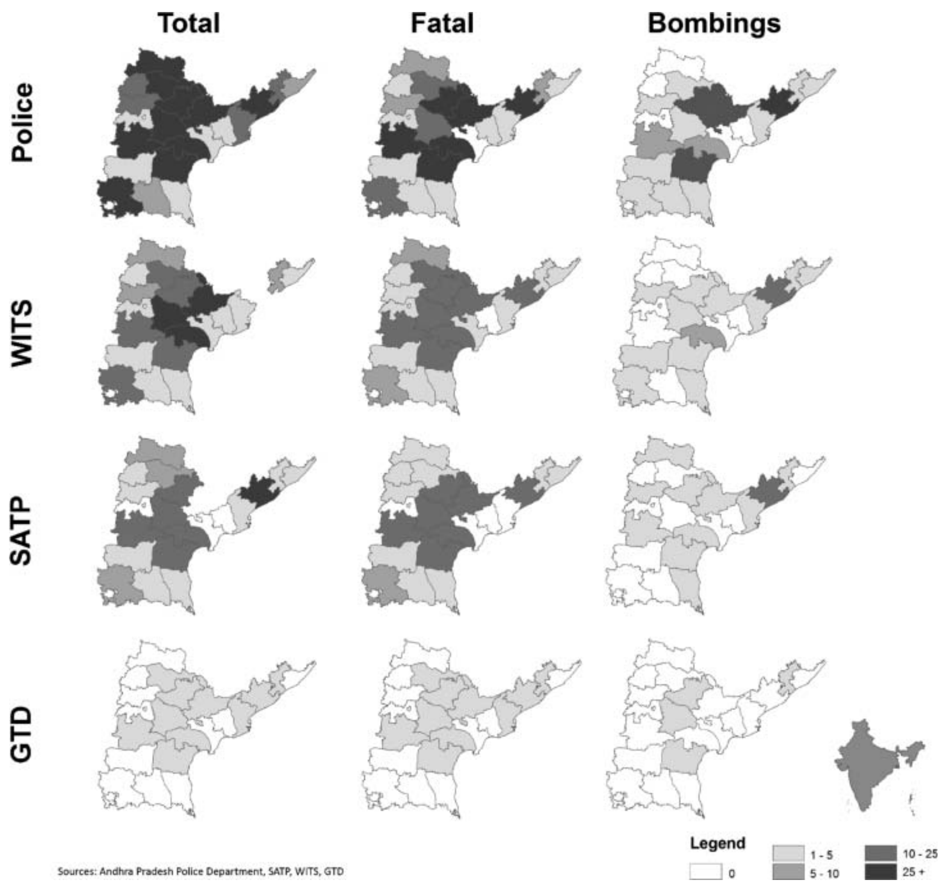

Figure 2. Spatial variation in reported terrorist attacks, by dataset and type of attack.

Predicting the Likelihood of Reporting

We find a moderate yet substantive level of variation in reporting terrorist attacks across the media-based datasets when compared to official records maintained by the State Police. To parse the specific leverage that these variations simultaneously exert, we employ logistic regression models to examine the odds that different types of variation would have on the likelihood of matched attacks being reported in each of the datasets. Beyond the types of variation listed above, we included a number of predictors which may influence the nature of newspaper coverage or readership in each district where an attack occurred. These factors include the district-level percent of population (1) in *scheduled castes or tribes*; (2) living below the *poverty* line; and (3) with at least a middle school *education*. We also include a number of environmental factors at the district-level traditionally associated with the occurrence of insurgency or conflict, including (4) the percent of the land area with greater than 6 percent slope (*ruggedness*); (5) the percent of land area covered with forests (*forest area*); and (6) the logged *distance* between the center of the district and the state capital of Hyderabad. Finally, we examine whether previous attacks in a district draw media attention to that location, which may increase the likelihood of a subsequent attack being captured by a media-based dataset, by including a measure of (7) whether there was an attack in the same district within 30 days prior to the attack under study (*previous attack*).

Additionally, there have been several recent efforts to merge data from SATP and GTD or WITS together into an integrated dataset for hypothesis testing.[75] The hope has been to create a comprehensive dataset which captures more of the variation in violence at the attack level, with multiple independent sources contributing cases that are not captured in the other databases. Merging data drawn from different sources and through different collection protocols could also amplify potential sources of bias rather than address them. To evaluate the veracity of this claim, we combined unique attacks drawn from SATP, GTD, or WITS into a single dataset, and examined whether the factors above influenced variations in this dataset compared to police records.

The results confirm that compared to police data, several of the descriptive patterns of variation among the datasets also increase the odds of reporting in a media-based dataset considerably. Table 4 includes odds ratios from logistic regression models[76] with year dummies at the dataset level in which a binary variable for reporting in a dataset (0 = non-reported, 1 = reported) is regressed against specific types of attacks, as well as environmental and geographic factors. Compared to other types of attacks in the police data, fatal attacks were more than six times as likely to be reported in a media-based dataset as non-fatal attacks, strongly suggesting their over-representation in these collections. Additionally, bombings were nearly three times as likely to be reported in a media-based dataset as other types of attacks for all but the GTD.

Considering environmental variation in attack reporting, we find that as the percentage of ruggedness in a district increases in Andhra Pradesh, there is a greater likelihood that the attack will end up in WITS or SATP. Contrasted to the level of ruggedness within a district, we find that attacks in heavily forested areas have a lower likelihood of being captured in at least one of the databases (WITS). Particularly for the WITS dataset, as attacks occur in a specific district, they also increase the odds that subsequent attacks within 30 days will be recorded within the data.

Table 4. Logistic regression models predicting reporting in datasets.

	WITS		SATP		GTD^{+}		Integrated dataset	
Predictor	OR	S.E.	OR	S.E.	OR	S.E.	OR	S.E.
Fatal Attack	6.53***	(1.33)	9.36***	(−2.32)	5.20**	(−3.11)	6.87***	(−1.33)
Bombing	2.81***	(0.77)	2.73***	(−0.80)	3.29	(−2.37)	3.31***	(−0.81)
Targeted Police and / or Security Forces	1.30	(0.34)	2.15**	(−0.56)	4.98**	(−3.00)	1.53	(−0.36)
Targeted State Leaders or Officials	3.79*	(2.22)	0.93	(−0.51)	9.67**	(−8.39)	2.05	(−1.15)
Targeted Village Leaders or Officials	1.72	(0.70)	2.38*	(−1.01)	3.55	(−3.13)	1.89	(−0.71)
Targeted Political Party Activists	1.48	(0.47)	2.46**	(−0.70)	2.26	(−1.78)	1.36	(−0.39)
% of Pop in Scheduled Caste/Tribe	1.09**	(0.03)	1.03	(−0.03)	1.07	(−0.09)	1.07*	(−0.03)
% of Pop with Below Poverty Line Card	0.98	(0.02)	1.02	(−0.02)	0.97	(−0.05)	0.98	(−0.02)
% of Pop with at least Middle School Education	0.99	(0.04)	0.96	(−0.05)	1.02	(−0.13)	0.98	(−0.04)
% of District with greater than 6% Slope	1.05***	(0.01)	1.03*	(−0.02)	1.03	(−0.04)	1.05***	(−0.01)
% of District covered with Forest	0.92***	(0.02)	0.97	(−0.02)	0.92	(−0.05)	0.94***	(−0.02)
Distance to Capital (logged)	1.34	(1.01)	1.19	(−0.99)	1.31	(−3.16)	0.91	(−0.67)
Previous Attack w/in 30 days in Same District	1.66*	(0.42)	1.64	(−0.45)	0.83	(−0.48)	1.61*	(−0.39)
Number of observations	806		806		806		806	
Adjusted R2	0.202		0.210		+		0.182	

***$p < .001$, l**$p < .01$, *$p < .05$ + GTD model run with Penalized Maximum Likelihood Estimations; no Adjusted R^2 provided; OR = Odds Ratio .

In addition, several predictors influenced the likelihood of reporting for one dataset but not the other, highlighting the role that production strategies within each of the datasets can influence inclusion. Attacks which targeted local officials (including village leaders), attacks against political party activists, and attacks which targeted police or security forces were more likely to be reported in SATP than other types of attacks. The primary governmental representatives targeted by the Naxalites in Andhra Pradesh during the study period were the police, with over 130 attacks across the five years. Given the use of local media outlets and newsletters by staff at SATP, it is reasonable to assume that these types of sources would favor reporting on attacks against local officials or party members since these types of attacks will be of more interest to local readers who favor local newspapers and newsletters. In contrast, attacks against state officials and leaders were more likely to be recorded within both WITS and GTD, as these types of attacks tend to draw regional and even national coverage in sources used by these data collection efforts.

Finally, testing a strategy of data integration of all the unique cases within the three datasets we find that bombings, fatal attacks, those where the proportion of scheduled populations is high, and those which occur in rugged areas would be more likely to be reported in an integrated media-based dataset than other types of attacks. By complimenting the data collection strengths of each of the underlying datasets, the results suggest that integrated approaches obscure the over-representation of several specific types of attacks relevant to local conflict dynamics, including attacks against local officials, police/security forces, and political party activists. Overall, results from the integrated dataset were not qualitatively different than results obtained from either the WITS or the SATP data, except for magnitude. While WITS reported 256 matched attacks, SATP reported 174 attacks, and GTD reported 19 attacks, the integrated dataset contained 293 unique terrorist attacks across the three datasets. This only accounted for roughly 36 percent of the total reported attacks contained in the police records, and only a 14 percent increase in the number of attacks as compared to WITS.

Discussion

What Remains Missing from Media-Based Datasets on Terrorism

The results above suggest that within the local dynamics of the Naxal insurgency of Andhra Pradesh, media-based datasets of terrorist attacks fail to capture a substantial amount of those attacks. At best, almost two-thirds of terrorist attacks recorded by the police (513 attacks) were missing from any of the three datasets evaluated here. While many of these attacks were not lethal, they included strategies and tactics similar to terrorist campaigns in Ireland or Spain, including the killing of suspected police informants and the destruction of infrastructure and government property. For example, Naxals in nine separate occasions used explosives or arson to destroy telephone infrastructure within a two-year period, none of which were captured by the media-based datasets. In addition, over 100 attacks involved the violent targeting of business owners selling liquor, election workers manning polls, contractors working construction sites, and civilians violating other Naxal proscriptions, only 18 of which were captured by media-based datasets. Such attacks resonate with the local populace and law enforcement, and their exclusion means media-based datasets may underestimate the frequency at which violent organizations use terrorism at the local level to coerce a population and government. Often these attacks are not reported in national or international English media, the primary source for most media-based datasets.

The Broader Context of Terrorism within an Insurgency

Beyond missing terrorist attacks, the results from our comparison highlight additional types of events often excluded from media-based datasets on terrorism. Official data collection efforts regularly contain a broader array of actions than datasets focused on terrorism. Nearly 40 percent of the original Naxal events categorized as aggressive acts by the police in Andhra Pradesh (631 events) fell outside the common definitions of terrorism used in this study, highlighting the wider ecology of aggression in which terrorist attacks can be found. Of these, almost half (297 events) refer to cases of extortion or looting committed by Naxal actors using threat or force to raise funds for their political arm. While these events were excluded from our analysis due to their primary economic motive (akin to common robbery), police interviewed by the second author in 2011 in Andhra Pradesh reported little operational differentiation between these acts and other forms of violent aggression by Naxals. According to them, these crimes do provide a window into how insurgency affects local residents, suggesting that the requirements for political motive embedded within datasets on terrorism may be difficult to determine in the context of violent domestic conflicts.

Alongside purely economic crimes committed by Naxal actors, another fifty percent (321 events) of the excluded events recorded in the police data refer to "exchanges of fire" between Naxalites and security forces. Difficulties in determining whether the non-state actor intentionally committed and/or planned the attack led to their exclusion from this comparison against media-based terrorist attacks, although the fog of war can often cloud the documentation of "who shot first" within internal conflicts. Naxalite documents clearly indicate paramilitary and security forces to be legitimate targets in the revolutionary war and specifically mention "wiping out the enemy in vulnerable areas."[77] On the other hand, the police have been known to target Naxals under the pseudonym of these "exchange of fire" events.[78] In these circumstances, the exclusion of "exchange of fire" events to fit within conventional definitions of terrorism may mask the broader interchange of violence between the state and violent organizations that are prevalent within insurgencies.

Implications for Future Research on Terrorism

Our results suggest that media-based datasets on terrorism exhibit considerable variation in their coverage of domestic terrorist attacks in developing countries. When compared to official data sources, terrorist attacks within media-based datasets are disproportionately lethal and explosive, underrepresenting a considerable range of violent, politically motivated activity which meet their criteria for inclusion. These biases in sample selection can have substantive consequences for the internal and external validity of research on terrorism,[79] requiring scholars to rigorously consider these variations in their research design.[80]

Our findings indicate that although previous studies have cautioned against the use of official police data in terrorism research because it might be restrictive,[81] using official police data can be rewarding, albeit when used with caution. All data are biased along some dimension, and the presence of attacks within the media-based datasets that were not found within the official data suggests that these biases must also be addressed when using official data. Police data suffer from some of the same reporting limitations in heavily contested areas that face media sources, and the increasing likelihood of missing cases in later years of data collection suggest that researchers using official data need to consider the role that administrative exhaustion can exert with official data collection efforts. Moreover, the broader array of events captured within official data requires additional scrutiny, depending on the question being answered. One recommendation is that researchers in areas of domestic terrorism could foster mutually beneficial relationships with concerned law enforcement agencies to access official data sources which might not be open access. Further, they can use this data, along with qualitative studies on terrorism, to consider whether and how the larger open source terrorism databases capture or reflect the nature, shape and manifestation of a broader insurgency and its impact on the local populace. These efforts should not proceed without skepticism of inclusion criteria and reporting practices for these official data, as the misuse of official data can bias inference as much as the under-reported open source data.

Our research also indicates that the more rugged the terrain, the more likely it is that the attack will be included in the media-based database.[82] One explanation is that in rugged areas, given their difficult terrain and lower population density, terrorist attacks are more infrequent and therefore more "newsworthy." This has important implications, since many studies of insurgency find increased levels of conflict in more rugged areas.[83] If these studies use data collected from open media sources, the initial findings from a single conflict here suggest that they are more likely to capture attacks in rugged regions. Thus, the relationship between terrain and conflict intensity may be a product of the data collection strategy rather than an underlying causal factor.

Relevant to the Naxal conflict, our research also revealed that attacks in forested areas are less likely to be included in the media-based databases as compared to less-forested areas. One explanation for this may be "media exhaustion" resulting from the continual and routine occurrence of Naxal conflict in these areas. Grievances resulting from land reform disputes between the Naxals and the Indian government provide the central narrative for which Naxal violence has been committed, and districts which are heavily forested also hold the primary relevance for the Naxal political strategy. We suggest that the "mundane" nature of terrorism in these regions (threats, beatings, or destruction of government property) leaves fewer and fewer unique cases of terrorism reducing the likelihood of reporting. Other explanations for low reporting include lack of media access to these more remote locations, or the general instability within the district, which may prevent reporters from filing on these attacks or coerce their silence out of fear for reprisals.

Another reason for low reporting is the fact that these forest areas are Maoist strongholds and so it is possible that they might have greater control on media coverage—especially reporting attacks that might reduce their popularity and eventually lead their core supporters to move away. Some accounts suggest that by highlighting the targeting of local infrastructure and individuals, popular support for Naxals in Andhra Pradesh was substantially reduced.[84] Even though the purpose of terrorism is to draw attention to their cause through violence,[85] we suggest that Naxals might prefer to censor media reporting of their activities as part of a broader media management campaign, especially if they involved "mundane" attacks against civilians. The very fact that the terrorist attacks are plentiful, "mundane," and not reported in the media means that the actual extent of "terror" on the ground is not properly reflected where analysis rely on media-based datasets alone.

A final implication is that the substantial exclusion of terrorist attacks (and even those that fall on the definitional fringe) limits the ability to examine the true impact of counterterrorism or counterinsurgency efforts at the local level.[86] What appears as a successful counterterrorism effort (by reducing the number of fatal attacks in a given year) may actually hide the subsequent tactical transference by violent organizations towards softer targets which do not attract as much extra-local attention. Failure to capture these less-lethal attacks can provide a false basis for selecting between counterterrorism strategies, ultimately influencing the ability to build an evidence-based framework for counterterrorism.

Conclusion

Our research suggests that studies on domestic terrorism which use data drawn from media sources may be biased due to an inconsistency between the occurrence of terrorism and its reporting within media sources. Even when the definition of terrorism is standardized across different datasets which draw from different media sources, the results may represent only a part of the actual universe of terrorist attacks. However, we are not suggesting that government data should be the gold standard by which scholars should evaluate their datasets. Not only do few countries maintain such datasets, but even when they exist, scholars are rarely given access to such information. Moreover in countries like India, the security forces remain under substantial political influences,[87] leaving it entirely possible that terrorism may be over- or undercounted in government records. This might be especially true of the period for which the data were analyzed, as it was a very important phase in the State's effort to curb insurgency in the region and it is possible that this affected how the data were recorded. Where possible, to overcome this limitation multiple sources available to researchers should be used to triangulate results across the various data collections. And sometimes (like in this case) the integrated dataset may still not be fully representative to the kinds of terrorist-related activity on the ground. In those cases, it is advisable for researchers to proceed with caution, especially while making predictions and policy recommendations based on available media-based data.

Future studies should also consider the mechanism by which events are covered, reported, editorialized and subsequently picked up by media-based datasets. This may be especially true for places where media reporting is either difficult due to contextual factors (like restricted access to conflict areas) or where electronic media presence remains limited. Creating an "error profile," which Chermak and colleagues find "identifies possible sources within the data collection methodology that may bias the results through non-sampling errors"[88] for each of the media-based datasets would provide a surer footing for future efforts to understand the causes and correlates of terrorism, especially if the collection managers themselves developed these profiles drawn from a nuanced understanding of their specific dataset.

Acknowledgments

Authors are listed in alphabetical order to represent equal authorship attribution. The authors thank Anupam Dalvi, Deepak Kukade and Amanda King for their research assistance in coding the terrorism data, and Shane Johnson, Gary LaFree, Emmanuel Teitelbaum, three anonymous reviewers, and the participants of the Workshop on the Maoist Insurgency for their comments and review.

Funding

Support for this research was provided by the Science and Technology Directorate of the U.S. Department of Homeland Security through the National Consortium for the Study of Terrorism and Responses to Terrorism (Award #N00140510629) and from the National Science Foundation (Award #ACI-1255793 and SES-1343123). The second author also thanks the Andhra Pradesh Police for their help and the Leverhulme Trust for Early Career Fellowship (Award #ECF/2009/0086) for funding this research. No funding from the Andhra Pradesh Police was received to support this project.

Notes

1. Sven Chojnacki, Christian Ickler, Michael Spies, and John Wiesel, "Event Data on Armed Conflict and Security: New Perspectives, Old Challenges, and Some Solutions," *International Interactions* 38 (2012), pp. 382–401.
2. M. Herbert Danzger, "Validating Conflict Data," *American Sociological Review* 40 (October 1975), pp. 570–584; Peter Hocke, "Determining the Selection Bias in Local and National Newspaper Reports on Protest Events," in Dieter Rucht, Ruud Koopmans, Friedhelm Niedhardt, Mark R. Beissinger, Louis J. Crishock, Grzegorz Ekiert, Olivier Fillieule, et al., eds., *Acts of Dissent: New Developments in the Study of Protest* (Lanham, MD: Rowman & Littlefield Publishers, 1999), pp. 131–163; John D. McCarthy, Clark McPhail, and Jackie Smith, "Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991," *American Sociological Review* 61 (June 1996), pp. 478–499.
3. Jose Barranco and Dominique Wisler, "Validity and Systematicity of Newspaper Data in Event Analysis," *European Sociological Review* 15 (September 1999), pp. 301–322; Carol Mueller, "International Press Coverage of East German Protest Events, 1989," *American Sociological Review* 62 (1997), pp. 820–832; Pamela E. Oliver and Gregory M. Maney, "Political Processes and Local Newspaper Coverage of Protest Events: From Selection Bias to Triadic Interactions," *American Journal of Sociology* 106 (September 2000), pp. 463–505.
4. Oliver and Maney, "Political Processes and Local Newspaper Coverage of Protest Events."
5. McCarthy, McPhail, and Smith, "Images of Protest."
6. Stathis. N. Kalyvas, "Civil Wars," in Carles Boix and Susan Stokes, eds., *The Oxford Handbook of Comparative Politics* (Oxford, UK: Oxford University Press, 2009), pp. 416–434.
7. Johan Galtung and Marie A. Ruge, "The Structure of Foreign News: The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers," *Journal of Peace Research* 2 (March 1965), pp. 64–90; Christian Davenport, "State Repression and Political Order," *Annual Review of Political Science* 10 (June 2007), pp. 1–23; Clionadh Raleigh, Andrew Linke, Håvard Hegre, and Joakim Karlsen, "Introducing ACLED: An Armed Conflict Location and Event Dataset Special Data Feature," *Journal of Peace Research* 47 (September 2010), pp. 651–660.
8. For example, see Bridget Rose Nolan, "The Effects of Cleric Statements on Suicide Bombings in Pakistan, 2000–2010," *Studies in Conflict & Terrorism* 36 (March 2013), pp. 219–234.
9. A state is a self-contained unit for law and order purposes in India.
10. Walter Enders, Todd Sandler, and Khusrav Gaibulloev, "Domestic Versus Transnational Terrorism: Data, Decomposition and Dynamics," *Journal of Conflict Resolution* 48 (May 2011), p. 320.
11. Bruce Hoffman, "The Confluence of International and Domestic Trends in Terrorism," *Terrorism and Political Violence* 9 (Summer 1997), pp. 1—15; Albert J. Jongman, "Trends in International and Domestic Terrorism in Western Europe, 1968–1988," *Terrorism and Political Violence* 4 (1992), pp. 26–76; Luis De La Calle and Ignacio Sánchez-Cuenca, "What We Talk About When We Talk About Terrorism," *Politics and Society* 39 (September 2011), pp. 451–472; Matthew C. Wilson and James A. Piazza, "Autocracies and Terrorism: Conditioning Effects of Authoritarian

Regime Type on Terrorist Attacks," *American Journal of Political Science* 57 (October 2013), pp. 941–955.
12. Gary LaFree, "The Global Terrorism Database: Accomplishments and Challenges," *Perspectives on Terrorism* 4 (March 2010), pp. 24–46.
13. Alex Schmid, "Statistics on Terrorism: The Challenge of Measuring Trends in Global Terrorism," *Forum on Crime and Society* 4 (December 2004), pp. 49–70.
14. Thomas J. Badey, "Defining International Terrorism: A Pragmatic Approach," *Terrorism and Political Violence* 10 (Spring 1998), pp. 90–107.
15. Stephen M. Chermak, Joshua D. Freilich, William S. Parkin, and James P. Lynch, "American Terrorism and Extremist Crime Data Sources and Selectivity Bias: An Investigation Focusing on Homicide Events Committed by Far-Right Extremists," *Journal of Quantitative Criminology* 28 (March 2012), pp. 191–218.
16. Boaz Ganor, "Defining Terrorism: Is One Man's Terrorist Another Man's Freedom Fighter?" *Police Practice and Research* 3 (December 2002), pp. 287–304.
17. Danzger, "Validating Conflict Data"; Galtung and Ruge, "The Structure of Foreign News"; H. Tony Harcup and Dierdre O'Neill, "What is News? Galtung and Ruge Revisited," *Journalism Studies* 2 (2001), pp. 261–280.
18. Galtung and Ruge, "The Structure of Foreign News"; McCarthy, McPhail, and Smith, "Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991"; Nils B. Weidmann, "The Higher the Better? The Limits of Analytical Resolution in Conflict Event Datasets," *Cooperation and Conflict* 48 (December 2013), pp. 567–576.
19. J. Craig Jenkins and Charles Perrow, "Insurgency of the Powerless: Farm Worker Movements (1946–1972)," *American Sociological Review* 42 (April 1977), pp. 812–829.
20. Ruud Koopmans, "The Use of Protest Event Data in Comparative Research: Cross-National Comparability, Sampling Methods and Robustness," in Dieter Rucht, Ruud Koopmans, Friedhelm Niedhardt, Mark R. Beissinger, Louis J. Crishock, Grzegorz Ekiert, Olivier Fillieule, et al., eds., *Acts of Dissent: New Developments in the Study of Protest* (Lanham, MD: Rowman & Littlefield Publishers, 1999), pp. 90–110; Barranco and Wisler, "Validity and Systematicity of Newspaper Data in Event Analysis."
21. Brent L. Smith and Kelly R. Damphousse, "Terrorism, Politics, and Punishment: A Test of Structural Contextual Theory and the Liberation Hypothesis," *Criminology* 36 (February 1998), pp. 67–92; Brent L. Smith and Kelly R. Damphousse, "Patterns of Precursor Behaviors in the Lifespan of a U.S. Environmental Terrorist Group," *Criminology and Public Policy* 8 (August 2009), pp. 475–496.
22. Danielle M. Rusnak, Leslie W. Kennedy, Ibrahim S. Eldivan, and Joel M. Caplan. "Analyzing Terrorism Using Spatial Analysis Techniques: A Case Study of Turkish Cities," in Cynthia Lum and Leslie W. Kennedy, eds., *Evidence-Based Counterterrorism Policy* (New York: Springer, 2006), pp. 167–185; Andrew Zammit-Mangion, Michael Dewar, Visakan Kadirkamanathan, and Guido Sanguinetti, "Point Process Modelling of the Afghan War Diary," *Proceedings of the National Academy of Sciences* 109 (July 2012), pp. 12414–12419; Michael Townsley, Shane D. Johnson, and Jerry H. Ratcliffe, "Space-Time Dynamics of Insurgent Activity in Iraq," *Security Journal* 21 (July 2008), pp. 139–146; Alex Braithwaite and Shane D. Johnson, "Space–Time Modeling of Insurgency and Counterinsurgency in Iraq," *Journal of Quantitative Criminology* 28 (March 2012), pp. 31–48.
23. D. Kim Rossmo and Keith Harries, "The Geospatial Structure of Terrorist Cells," *Justice Quarterly* 28 (April 2011), pp. 221–248.
24. Syed Ejaz Hussain, *Terrorism in Pakistan: Incident Patterns, Terrorists' Characteristics, and the Impact of Terrorist Arrests on Terrorism*, PhD diss., University of Pennsylvania, 2010. Publicly Accessible Penn Dissertations, Paper 136.
25. Brent L. Smith, *Terrorism in America: Pipe Bombs and Pipe Dreams* (Albany: State University of New York Press, 1994); Brent L. Smith and Kelly R. Damphousse, "Punishing Political Offenders: The Effect of Political Motive on Federal Sentencing Decisions," *Criminology* 34 (August 1996), pp. 289–321.

26. Gary LaFree, Nancy A. Morris, and Laura Dugan, "Cross-National Patterns of Terrorism: Comparing Trajectories for Total, Attributed and Fatal Attacks, 1970–2006," *British Journal of Criminology* 50 (July 2010), pp. 622–649.
27. In the context of this study, the second author conducted interviews with police officers in various states that found considerable variation in the threat perception of Naxals. While some officers conceded that Naxal acts were terrorist-like but Naxalites themselves were not terrorists, others were convinced that there was no difference between Naxalism and terrorism. This might affect how events are perceived and whether and how they are recorded and reported if attempts to aggregate official data occur at the national level.
28. Koopmans, "The Use of Protest Event Data in Comparative Research"; Clayton J. Mosher, Terance D. Miethe, and Dretha M. Phillips, *The Mismeasure of Crime* (Thousand Oaks, CA: Sage Publications, 2002).
29. Chermak, Freilich, Parkin, and Lynch, "American Terrorism and Extremist Crime Data Sources and Selectivity Bias."
30. The victim's propensity to report crimes is influential in shaping the police recording of crimes. While analyzing crime it is important to understand changing social processes underlying the reporting and non-reporting of certain crimes, which can impact how crime trends and patterns are analyzed. There is also some evidence to indicate the public's dislike of the police has a negative impact on reporting rates. It might be that fear of the police and/or the Naxals could have prevented people from reporting events for official data collection. See Ziggy MacDonald, "Revisiting the Dark Figure: A Microeconometric Analysis of the Under-Reporting of Property Crime and Its Implications," *British Journal of Criminology* 41 (January 2001), pp. 127–149; Roger Tarling and Katie Morris, "Reporting Crime to the Police," *British Journal of Criminology* 50 (May 2010), pp. 474–490.
31. Since Naxal insurgency is treated as a law and order problem in India, local state governments make all policing-related decisions. According to the Indian Constitution, "Law and Order" and "Policing" are state subjects (i.e., under the purview of state governments). Political interests are therefore at the forefront of anti-Naxal policies and politicians have a vested interest in the way Naxal-related events are recorded. Over the years, regional political parties have had a cyclical relationship with the Naxals—wooing them around election time and then forcefully repressing them when levels of violence become unacceptable. Thus, the possibility that official data might be affected by state government dictate cannot be ignored.
32. The police can be subject to political pressure to present data in particular ways depending on the inclination of the political party in power in each state and its attitude toward the Naxals. Just as crime data are the barometer of "the moral health of the nation," terrorism statistics indicate the security status of the state. The data are "indices of organizational processes" and depending on police practices, events will be recorded (or not) in particular ways. There have been instances where the police have deliberately manufactured data or not recorded events (a phenomenon known as "burking" in India) either due to police bias that deems some events as not worth recording as opposed to others; or because they have a low probability of detection; or as a result of corruption or due to political interferences. This reporting bias could be a reflection of the regime type and how it views the problem. See Michael Maguire, "Crime Data and Statistics," in Robert Reiner, Rod Morgan, and Mike Maguire, eds., *The Oxford Handbook of Criminology* (Oxford, UK: Oxford University Press, 2007), pp. 241–301; Konstantinos Drakos and Andreas Gofas, "The Devil You Know But Are Afraid to Face: Underreporting Bias and its Distorting Effects on the Study of Terrorism," *Journal of Conflict Resolution* 50 (October 2006), pp. 714–735.
33. Maguire, "Crime Data and Statistics."
34. Donald P. Green, Dara Z. Strolovitch, Janelle S. Wong, and Robert W. Bailey, "Measuring Gay Populations and Antigay Hate Crime," *Social Science Quarterly* 82 (June 2001), pp. 281–296.
35. Clive Coleman, *Understanding Crime Data: Haunted by the Dark Figure* (Philadelphia, PA: Open University Press, 1996).
36. Drakos and Gofas, "The Devil You Know But Are Afraid to Face."
37. Barranco and Wisler, "Validity and Systematicity of Newspaper Data in Event Analysis."

38. McCarthy, McPhail, and Smith, “Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991.”
39. Jennifer Earl, Andrew Martin, John D. McCarthy, and Sarah A. Soule, “The Use of Newspaper Data in the Study of Collective Action,” *Annual Review of Sociology* 30 (August 2004), pp. 65–80.
40. Gary LaFree, Laura Dugan, and Raven Korte, “The Impact of British Counterterrorist Strategies on Political Violence in Northern Ireland: Comparing Deterrence and Backlash Models,” *Criminology* 47 (February 2009), pp. 17–45; Will Bullock, Kosuke Imai, and Jacob N. Shapiro, “Statistical Analysis of Endorsement Experiments: Measuring Support for Militant Groups in Pakistan,” *Political Analysis* 19 (Autumn 2011), pp. 363–384; Prakarsh Singh, “Impact of Terrorism on Investment Decisions of Farmers Evidence from the Punjab Insurgency,” *Journal of Conflict Resolution* 57 (January 2013), pp. 143–168.
41. Others have provided a summative overview of the methodological differences between GTD, WITS, RAND, and ITERATE databases. See Ivan Sascha Sheehan, “Assessing and Comparing Data Sources for Terrorism Research,” in Cynthia Lum and Leslie W. Kennedy, eds., *Evidence-Based Counterterrorism Policy* (New York: Springer, 2006), pp. 13–40.
42. Alberto Abadie, “Poverty, Political Freedom, and the Roots of Terrorism,” *American Economic Review* 96 (May 2006), pp. 50–56; James A. Piazza, “Poverty, Minority Economic Discrimination, and Domestic Terrorism,” *Journal of Peace Research* 48 (May 2011), pp. 339–353.
43. While we acknowledge there are considerable differences in the definitions provided for terrorism and insurgency, event-level databases like the GTD and SATP focus on the act of violence and its immediate political goal, rather than the long-term strategy of the actor perpetrating the violence. While some see terrorism as a strategy within insurgency, others see them as qualitatively different. For the purposes of this study, we restrict the event comparisons to intentional and proactive actions committed by violent nonstate actors to promote a political or economic goal through the use or threat of force or violence. These could include one-off bombings typical of terrorism writ large, as well as the use of terrorism as a tactic within an insurgency. Other types of events commonly found in insurgencies where it is unclear that the violent nonstate actor was the initiator of the event, like exchanges of fire and encounters, are not considered in this study. See Ariel Merari, “Terrorism as a Strategy of Insurgency,” *Terrorism and Political Violence* 5 (Winter 1993), pp. 213–251; De La Calle and Sanchez-Cuenca, “What We Talk About When We Talk About Terrorism.”
44. J. K Achuthan, “Tackling Maoists: The Andhra Paradigm,” *India Defence Review* 25 (April–June 2010), p. 2; D. M. Mitra, *Genesis and Spread of Maoist Violence and Appropriate State Strategy to Handle It* (New Delhi: Ministry of Home Affairs, Bureau of Police Research and Development, 2011).
45. Manmohan Singh, “Prime Minister’s Concluding Remarks,” Speech at the 2nd Meeting of the Standing Committee of Chief Ministers on Naxalism, New Delhi, India, 13 April 2006.
46. Pratul Ahuja and Rajat Ganguly, “The Fire Within: Naxalite Insurgency Violence in India,” *Small Wars and Insurgencies* 18 (June 2007), pp. 249–274.
47. Rajat Kujur, “The Naxal Movement in India: A Profile,” *Institute of Peace and Conflict Studies Research Paper* 15 (September 2008).
48. Mitra, *Genesis and Spread of Maoist Violence and Appropriate State Strategy to Handle It.*
49. Jairus Banaji, “The Ironies of Indian Maoism,” *International Socialism* 128 (October 2010), pp. 129–148.
50. K. Balagopal, “Maoist Movement in Andhra Pradesh,” *Economic and Political Weekly* 41 (July 2006), pp. 3183–3187.
51. M. Shashidhar Reddy, “A Political Approach to the Naxalite Problem,” in P. V. Raman, ed., *The Naxal Challenge: Causes, Linkages and Policy Options* (New Delhi, India: Dorling Kindersley, 2008), pp. 39–61; P. V. Ramana, “Measures to Deal with Left-Wing Extremism/Naxalism,” *Institute for Defence Studies and Analysis Occasional Paper* No. 20 (2011).
52. Ajai Sahni, “Anti-Maoist Strategy: Utter Disarray,” *South Asia Intelligence Review* 9 (July 2010).
53. K. Srinivas Reddy, “Maoist Make Their Presence Felt in Telangana Again,” *The Hindu*, 13 November 2011. Available at http://www.thehindu.com/news/cities/Hyderabad/maoists-make-their-presence-felt-in-telangana-again/article2624001.ece

54. "26 Districts Highly Naxal-Hit in Country: Govt," *Hindustan Times*, 12 April 2013. Available at http://www.hindustantimes.com/delhi/26-districts-highly-naxal-hit-in-country-govt/story-qMPbnPE2IjlmkkV1AEtZ6H.html; Sudeep Chakravarti, *Red Sun: Travels in Naxalite Country* (New Delhi, India: Penguin Global, 2009).
55. Government of India, *Annual Report 2012–13* (New Delhi: Ministry of Home Affairs, Departments of Internal Security, State, Home, Jammu & Kashmir Affairs, and Border Management, 2013).
56. For example, Davenport and Ball compared Guatemala conflict information taken from newspapers, human rights reports, and interviews and found considerable difference depending on source, suggesting that researchers should use multiple sources of information when available. See Christian Davenport and Patrick Ball, "Views to a Kill Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977–1995," *Journal of Conflict Resolution* 46 (June 2002), pp. 427–450; Jenkins and Perrow, "Insurgency of the Powerless: Farm Worker Movements (1946–1972)"; Marwan Khawaja, "Resource Mobilization, Hardship, and Popular Collective Action in the West Bank," *Social Forces* 73 (September 1994), pp. 191–220; Mark R. Bessinger, "National Violence and the State: Political Authority and Contentious Repertoires in the Former USSR," *Comparative Politics* 30 (July 1998), pp. 401–422.
57. Although a considerable portion of the quantitative terrorism literature is published using the ITERATE data, the domestic nature of the Maoist Insurgency in Andhra Pradesh would render most of its activity outside the definitional scope of the ITERATE data, which focuses on transnational terrorism. Accordingly, the study does not compare these data sources against the ITERATE data.
58. Gary LaFree and Laura Dugan, "Introducing the Global Terrorism Database," *Terrorism and Political Violence* 19 (2007), pp. 181–204; Erin Miller, "Patterns of Onset and Decline among Terrorist Organizations," *Journal of Quantitative Criminology* 28 (March 2012), pp. 77–101; Gary LaFree, Laura Dugan, and Erin Miller, *Putting Terrorism in Context: Lessons from the Global Terrorism Database* (New York: Routledge, 2015).
59. National Consortium for the Study of Terrorism and Responses to Terrorism (START), *Global Terrorism Database [data file]*, 2012. Available at http://www.start.umd.edu/gtd
60. U.S. National Counterterrorism Center, *Worldwide Incidents Tracking System (WITS)*, 2011. Available at http://wits.nctc.gov
61. Information on the perpetrator is included in the summary of the attack, but not in an easily identifiable variable within the dataset that was available to the public. Coders for this project reviewed the perpetrator within the summary of an attack and coded for that perpetrator.
62. Jyoti Belur, "Why Do The Police Use Deadly Force? Examining Police Encounters in Mumbai," *British Journal of Criminology* 50 (March 2010), pp. 320–341.
63. South Asia Terrorism Portal (SATP), 2011. Available at http://www.satp.org/
64. William G. Axinn, Dirgha Ghimire, and Natalie e. Williams, "Collecting Survey Data During Armed Conflict," *Journal of Official Statistics* 28 (June 2012), pp. 153–171; Oliver Vanden Eynde, "Targets of Violence: Evidence from India's Naxalite Conflict," *Job Market Paper* (July 2011).
65. The Uppsala Conflict Data Project out of the Peace Research Institute of Oslo, as well as WITS (during some portion of their data collection effort) and the GTD (between 2008 and 2011), all utilize SATP in this manner. See Kristine Höglund and Magnus Öberg, *Understanding Peace Research: Methods and Challenges* (New York: Routledge, 2011).
66. A striking example being the inclusion of 23 events of "burning of national flag" in the original data as a criminal offense under the Prevention of Insults to the National Honor Act (1971) in August 2004 and August 2005 in only two districts of the state. Flag-burning events do not appear again in the dataset either before or after this period. It cannot be the case that the Maoists suddenly began burning the National flag as a form of protest and intimidation, and just as suddenly stopped doing so. A more likely explanation is that particular police officers decided to record them as additional offenses against the Naxals, but subsequent changes in recording practices deemed these events no longer worthy of inclusion.
67. Regarding GTD, given the few observations that were matched to the police data, the estimates within conventional logit models would suffer from small-sample bias. Although others have

addressed this problem using rare event logit models originally proposed by King and Zeng in their 2001 paper, recent Monte Carlo simulations conducted by Leitgöb suggest that these methods overcorrect the bias as the number of observations gets smaller. Therefore, we use Penalized Maximum Likelihood Estimation, originally proposed by Firth, to address issues of small-sample bias within the GTD models alone. Additionally, these logit models do not include any cases that were in WITS, SATP, or GTD which were not found in the police data used. See Gary King and Langche Zeng, "Logistic Regression in Rare Events Data," *Political Analysis* 9 (February 2001), pp. 137–163; David Firth, "Bias Reduction of Maximum Likelihood Estimates," *Biometrika* 80 (March 1993), pp. 27–38; Christopher Zorn, "A Solution to Separation in Binary Response Models," *Political Analysis* 13 (Spring 2005), pp. 157–170; Heinz Leitgöb, "The Problem of Modeling Rare Events in ML-Based Logistic Regression: Assessing Potential Remedies via Monte Carlo Simulations," Paper presented at the 2013 European Survey Research Association, Reykjavik, Iceland, July 2013.

68. Between 9 and 13 percent of left-wing terrorist attacks in WITS, SATP, and GTD were not found within the police data. Overall, there appeared to be no definitive pattern to the missing attacks across any of the datasets. Of the 17 SATP attacks could not be located, ten attacks each recorded a single fatality. Over a quarter of the missing SATP attacks were from the Khammam district, where Naxal fighting was particularly heavy. Similarly, of the 32 missing WITS attacks, 14 had at least one fatality and a number were missing from Khammam district, including 14 missing from 2005 and 11 from 2008. A total of three GTD attacks could not be located in the police data, all within the last 14 months of the time period, with one of the missing GTD attacks starting in Andhra Pradesh yet ending in Chhattisgarh. One potential reason for missing data between the datasets were differences in reported location and timing of attacks between sources. Attacks that occur in remote locations of the State may be reported to the police more rapidly than to the media. Although the authors provided a two-week window before and after the attack to suggest equivalency between datasets, there may yet remain differences outside this window in the recorded date of the attack which could translate into unmatched cases.
69. Beginning in April 2008, START utilized a different data collection partner (Institute for the Study of Violent Groups), which used a prospective approach for capturing sources on terrorist attacks, which contrasted against archival methods used for data between 1998 and 2007. See http://www.start.umd.edu/gtd/about/History.aspx for more information.
70. Security forces include civil police, armed police, and paramilitary forces.
71. There are 23 administrative districts in Andhra Pradesh that differ in size, but individually cover a geographical area of approximately 10,000 square kilometers, with an average population density of around 250 persons per square kilometer.
72. K. Balagopal, "Land Unrest in Andhra Pradesh-III: Illegal Acquisition in Tribal Areas," *Economic and Political Weekly* 42 (October 2007), pp. 4029–4034.
73. The district was site of the Karimnagar Experiment, which contained several strategies to gain public support against the Maoist while providing a number of methods for reintegration of former Maoist rebels into society, including the public burning of all paper records on Maoists in the district by the local police. See K. S. Reddy, "Revolutionary and Counter-Revolutionary Strategies of the Naxalites and the State," in P. V. Raman, ed., *The Naxal Challenge: Causes, Linkages and Policy Options* (New Delhi, India: Dorling Kindersley, 2008), pp. 90–104.
74. This is also confirmed through comparative Spearman rank order correlations between the police data, WITS, and SATP; all of which were above 0.9. Correlations were also run between the datasets and the GTD, and correlations of 0.7 and higher were found for all pairs, suggesting a high level of correlation for between-district variation across all of the datasets.
75. Joseph F. Gomes, "The Political Economy of the Maoist Conflict in India: An Empirical Analysis," *World Development* 68 (April 2015), pp. 96–123.
76. Robust standard errors were used, clustered at the district, for all but the GTD model. Penalized Maximum Likelihood Estimation models currently do not support clustered standard errors.
77. Communist Party of India (Maoist) Central Committee, *Strategy and Tactics of the Indian Revolution*, September 21, 2004. Available at http://www.bannedthought.net/India/CPI-Maoist-Docs/Founding/StrategyTactics-pamphlet.pdf

78. N. Venugopal, “Fake Encounters: Story from Andhra Pradesh,” *Economic and Political Weekly* 42 (October 2007), pp. 4106–4111.
79. Richard A. Berk, “An Introduction to Sample Selection Bias in Sociological Data,” *American Sociological Review* 48 (June 1983), pp. 386–398.
80. Chermak, Freilich, Parkin, and Lynch, “American Terrorism and Extremist Crime Data Sources and Selectivity Bias.”
81. Ibid.
82. Additional sets of models substituting population density for ruggedness as well as including it alongside ruggedness were also specified for all four datasets. For both sets of models, population density had no significant effect on any outcome, and the effect of ruggedness remained positive and significant.
83. Cristiana C. Brafman Kittner, “The Role of Safe Havens in Islamist Terrorism,” *Terrorism and Political Violence* 19 (2007), pp. 307–329; Lars-Erik Cederman, “Articulating the Geo-Cultural Logic of Nationalist Insurgency,” in Stathis Kalyvas, Ian Shapiro, and Tarek Masoud, eds., *Order, Conflict, and Violence* (Cambridge, MA: Cambridge University Press, 2008), pp. 242–270.
84. Balagopal, “Maoist Movement in Andhra Pradesh.”
85. Drakos and Gofas, “The Devil You Know But Are Afraid to Face.”
86. Braithwaite and Johnson, “Space–Time Modeling of Insurgency and Counterinsurgency in Iraq.”
87. Arvind Verma, “Cultural Roots of Police Corruption in India,” *Policing: An International Journal of Police Strategies & Management* 22 (1999), pp. 264–279.
88. Chermak, Freilich, Parkin, and Lynch, “American Terrorism and Extremist Crime Data Sources and Selectivity Bias,” p. 215.