Data on Combatant Activity During Afghanistan War Advance Scientific Investigation of Insurgency

Andrew C. Shaver[1,3] & Austin L. Wright[2,3]

[1] *Woodrow Wilson School of Public and International Affairs, Princeton University*

[2] *Harris School of Public Policy, The University of Chicago*

[3] *Empirical Studies of Conflict Project*

Scientific research on insurgent violence faces substantial data challenges. In this resource article, we introduce the U.S. Defense Department's accounting of combatant activity throughout the Afghanistan war (2002-2014). Conflict data of this nature—hundreds of thousands of precisely georeferenced (to the meter) and time-stamped (often to the minute) insurgent, counterinsurgent, and other wartime incidents, with corresponding weapon, tactic, and other relevant details—has not previously been available to conflict scholars and other scientists. To demonstrate how the data might be used, we analyze the spatial distribution of known insurgent and terrorist training camps throughout the Afghanistan war and describe how the observed pattern diverges from prevailing theoretical expectations.

At a global scale, powerful state governments are engaged in violent conflicts with weaker non-state political actors seeking, among other aims, to overturn existing political institutions and secure territorial independence. Despite often immense asymmetries in the military capabilities

available to both sides, these conflicts frequently persist for years. Perhaps no example better illustrates this dynamic than the current conflict in Afghanistan. There, the world's military super power, operating in concert with other sophisticated militaries, has fought a losing battle for one and a half decades against a well-established but technologically outmatched insurgent organization.

This phenomenon is not new. For as long as nation states have existed, intrepid sub-state actors have used violence—often with success—to challenge central governments. Yet, many of the conditions that make such persistent conflict possible—although widely theorized—remain uncertain due to a lack of clear empirical evidence. To address these gaps, advances in scientific research on substate conflict rely increasingly on micro conflict datasets for both causal inference and forecasting[1,2].

In this Resource Article, we introduce highly detailed data on combatant activity from the recent conflict in Afghanistan collected jointly by Afghan military and police forces and member forces of the North Atlantic Treaty Organization's (NATO) International Security Assistance Force (ISAF). At its peak level, ISAF included more than 130,000 soldiers deployed by the United States and 50 other countries. The data range from 2002 through 2014, the period formally designated by the U.S. military as Operation Enduring Freedom.

This newly declassified dataset[1] provides the precise location (accurate to several meters) and timing (often to the minute) of nearly one half million events related directly and indirectly

[1]Formally released to the authors by the U.S. Department of Defense

to violent conflict during the Afghanistan war. The dataset includes individual incidents of events ranging from insurgent violence (with corresponding details on the weaponry used, the target, and the outcome of each attack) against Afghan and ISAF forces to cases of strategic non-violent insurgent activity (including smuggling weapons, manning illicit checkpoints, training terrorists and insurgents, to name just several). The data also track individual instances of counterinsurgent success (for instance, defusing roadside bombs and discovering weapon depots).

The dataset was constructed from individual military records compiled by U.S., Afghan, and other ISAF military and police units. These reports were collected following well-established military protocol and with the use of advanced georeferencing and collation technologies, ensuring that many report details were both objectively measured and captured with a high degree of precision. This geographic feature of the dataset is particularly noteworthy. Scholars have attempted to construct numerous micro-level violence datasets reflecting battlefield activities. Yet these data sources capture only a limited subset of combatant activity. This is due, in part, to the fact that these data sources draw on media event descriptions. Specifically, a majority of existing conflict datasets (including the Uppsala Conflict Data Program Georeferenced Event Dataset (GED)[3,4], the National Consortium for the Study of Terrorism and Responses to Terrorism's Global Terrorism Database (GTD)[5], and the U.S. National Counterterrorism Center's Worldwide Incidents Tracking System (WITS) database[6] were constructed using media reports, which recent research shows can (and often do) suffer reporting biases[7]. Research on reporting biases in media-derived conflict microdata shows that (media) reporting tends to occur more regularly in high population areas. This means that acts of political violence are systematically undercounted in more remote, less

populated areas of affected countries.

In sum, the dataset we introduce in this Resource Article is unparalleled in (1) its degree of detail on conflict-related activities, (2) the breadth of activities covered, and (3) the consistent and precise nature by which it was were constructed. To our knowledge, the data are the most comprehensive and methodically collected data on insurgent activity to have ever been made publicly available to members of the academic community. These unique qualities of the dataset make it a noteworthy contribution to the broader scientific community.

To motivate potential data applications, we investigate one highly unusual feature of the data: when and where insurgents assemble training camps. Although scholarship has long debated the geography of insurgency, scarce quantitative evidence exists linking terrain ruggedness, proximity to population centers, security forces, and the strategic deployment of insurgent infrastructure. We perform an empirical investigation of these dynamics and show that the distribution of training camps follows an unexpected pattern in Afghanistan.

RESULTS

The Afghanistan War. Following the September 11, 2001 terrorist attacks on U.S. targets, the U.S. military invaded Afghanistan, where Al-Qaeda was operating under the auspices of the then ruling Taliban government. Many Al-Qaeda members who were not killed during the invasion fled the country, and the Taliban government was overthrown in approximately two weeks. For the next two years, the United States Government undertook efforts to establish a constitutional democracy

within the country. Meanwhile, exiled Taliban forces, many of whom had taken refuge behind Pakistani borders, initiated an insurgency against American troops and the nascent Afghanistan national security establishment. On August 11, 2003, NATO assumed responsibility for the country's security through the deployment of ISAF, a coalition of multinational forces (representing, at its height, 51 nations but consisting primarily of American, German, Italian, Georgian, and Romanian troops).

ISAF remained engaged in a counterinsurgency campaign against the Taliban until the end of 2014, when responsibility for Afghanistan's security affairs was formally transferred to Afghan security forces. Since then, American and other international forces have maintained a reduced troop presence in the country designed to support Afghanistan security force efforts to protect the country against Taliban attack.

Wartime Data. Throughout the ISAF campaign, ISAF and Afghan military and police forces tracked a wide variety of highly detailed data related to the conflict. Various sources reported approximately half of one million incidents of "significant activity" (SIGACT) related to the insurgency during this period. The data are typically referred to as SIGACTs. The primary reporting sources of these incidents were tactical U.S. military units associated with various combat brigades deployed across Afghanistan as well as Afghan military and police forces (hereafter collectively referred to simply as "counterinsurgents" for brevity).

Each incident recorded in the dataset includes a variety of details, which allow the data to be broken down in a variety of ways. The major event categories into which the data can

be classified include: **insurgent violence** – individual instances of insurgent violence including, but not limited to, direct fire (DF), indirect fire (IDF), and improvised explosive device attacks (IED); **counterinsurgent successes** – individual cases of battlefield outcomes that accrued some benefit to counterinsurgent forces including the discovery of weapons caches and the identification and diffusion (or other means of neutralizing) of IEDs and mines; **insurgent behavior** – events reflecting insurgent wartime activities including running illicit checkpoints, weapons and narcotics smuggling, weapons transport, and operating terrorist training camps; and **criminal activity** – activities with which counterinsurgents were aware but were not necessarily associated with the insurgency, including, for instance, murders and kidnappings.

Three characteristics of this incident-based dataset are particularly noteworthy. First, the dataset represents the comprehensive catalogue of combat engagements occurring during the Afghanistan war. Similar conflict datasets capture a subset of these events discussed in media reports. This dataset also represents the most complete record of non-combat events, including insurgent training camps, village elder assassinations, and kidnappings. These events are typically excluded from studies of subnational conflict. Second, by using global positioning system technologies, counterinsurgents succeeded in georeferencing all of the nearly one half million activities they reported. From the locations of roadside bomb attacks to the identified locations of terrorist training camps, precise coordinates are associated with each event, providing a high degree of spatial accuracy. Third, each event was timestamped by counterinsurgents. Events are often specific to the minute on which they occurred. When they are not, event are often rounded to the nearest fifth minute (e.g. 10:10 AM, 10:15 AM). Thus, for instance, the data can be used to identify differences

in the temporal patterns of insurgent attacks by type – whether, for instance, insurgents tended to attack with rockets and mortars with different frequencies and/or at different times than they did with small arms and rocket-propelled grenades.

In addition to details described above, for a subset of violent observations, a variety of details related to target type, target identity, outcome of attack, and actor precedence are also included. Amongst the data on insurgent attacks, incidents are associated with outcomes. Specifically, these include, but are not limited to: "Killed", "Wounded", "Destroyed", " Damaged", and "Unaffected".

Attacks are also often associated with target types and identities. Frequently, the data indicate the type of target against which attacks were carried out. This target classification includes, for instance, aircraft, bases, dismounted patrols, and military convoys. The data also indicate which group(s) came under attack. In addition to Afghan military and, separately, ISAF forces, the data distinguish between various Afghan police units including the border, uniform, local and National Civil Order police.

Another useful feature of the data involves its inclusion of a precedence variable. Specifically, the variable indicates the sequence of actors affected by given events. Thus, for instance, an insurgent attack directed against Afghan police forces that also affected Afghan civilians in the vicinity of the attack will show that both sets of actors were affected, with Afghan police forces and civilians being assigned primary and secondary precedence, respectively. The quality and completeness of this data increased during the campaign, and was systematically collected from the end of 2009 until 2014. This period includes 77.86 % of combatant activity.

Data Descriptives. The data reveal a wide variety of characteristics about the conflict. Throughout the conflict, ISAF tracked more than 200,000 incidents of violence involving insurgents, including a total of 119,908, 28,678, and 38,004 DF, IDF, and IED attacks, respectively. To put this figure in perspective, the Pentagon tracked approximately 192,340 incidents of insurgent violence throughout the Iraq War[8].

Counterinsurgents also documented a wide variety of successes. These forces discovered and neutralized 44,508 IEDs and another 1,195 mines, and they uncovered 16,500 weapons caches. They also tracked a wide variety of non-violent insurgent and criminal activities, including approximately three thousand total cases of smuggling and the operation of terrorist training camps and, separately, illegal checkpoints.

Some of the data relate to airpower. For instance, in nearly 5,500 instances ground forces received close air support – fixed and/or rotary-winged military aircraft assistance against enemy forces with whom they were proximately engaged. Separately, insurgents fired on airborne aircraft on at least 6,917 occasions.

The data also provide details related to hundreds of instances of "friendly fire." While such events are often inadvertent, insider attacks have plagued counterinsurgent efforts to pacify Afghanistan for years. This month, for instance, another seven US soldiers were shot by an Afghan soldier[9], the latest in a string of such attacks that have thus far claimed the lives of 714 counterinsurgents.[10] The data show that most friendly fire attacks were "green on green" (Afghan on other Afghan forces); yet, in nearly 100 cases, Afghan forces fired on American or other inter-

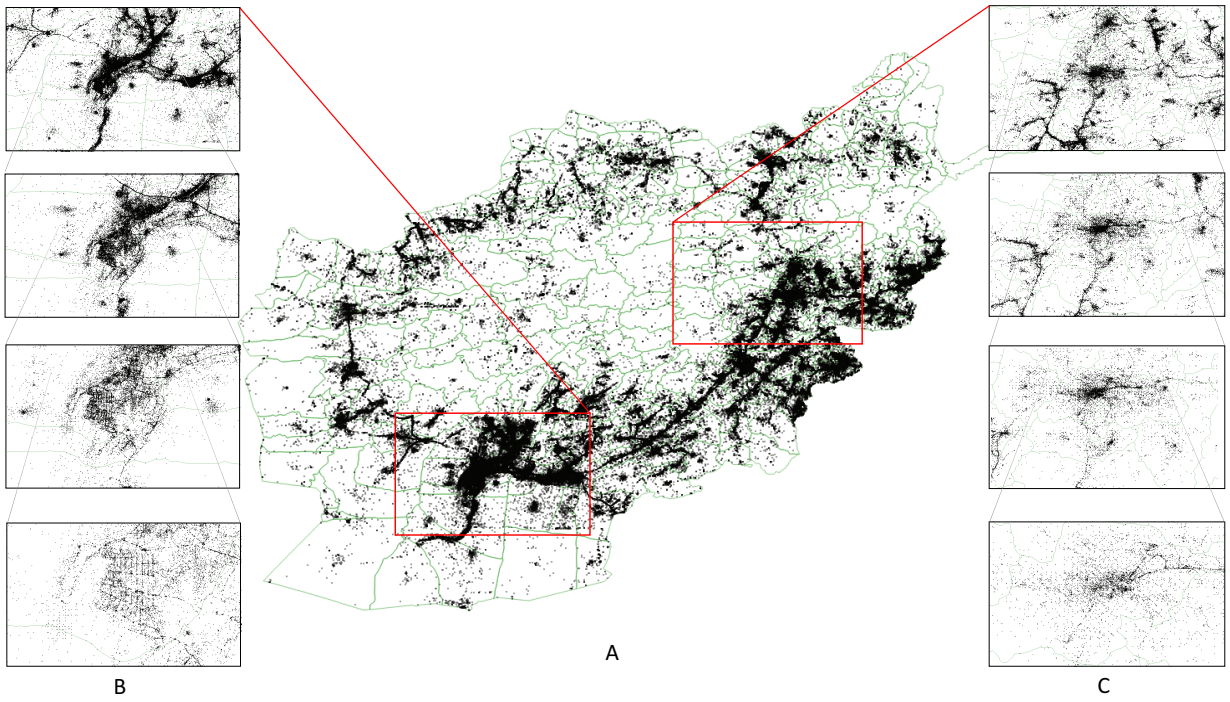
national forces in the country.

Because the data include the outcome of violent engagements between insurgents and counterinsurgent forces, we can calculate the number of both fatal and injurious events (instances in which one or more individuals was killed and, separately, wounded) suffered by all sides throughout the conflict. While data on the deaths of American soldiers has been released by the Pentagon, until now, there has been no official release of these numbers by relevant governments with which we are aware. We offer these details here for the first time.

Throughout the ISAF campaign, the Taliban suffered the greatest losses of any combatant group, experiencing 12,967 fatal events. Similarly, Afghan military and police forces suffered a total of 9,902 fatal incidents, respectively. Civilians were caught in the crossfire or directly targeted by insurgents on 13,731 cases. As a result, civilians suffered 6,184 fatal incidents.

Although the data indicates only deadly incidents, because we know from the data the number of fatal events that ISAF suffered (1,451) and, from outside sources, the actual number of ISAF soldiers killed by the end of 2014 (3,405)[11], we know that an average of approximately 2.35 soldiers were killed per fatal incident. Using this same rate, we roughly estimate total deaths suffered by other parties to the conflict. Under this approach, we calculate that approximately 30,500 Taliban fighters were killed between 2001 and the end of 2014. This number, however, is likely conservative because only cases of battlefield deaths are likely to have been captured in the Pentagon data. Wounded Taliban fighters who succumbed to their injuries in the hours, days, or weeks following engagements with ISAF or Afghan forces are unlikely to have been tracked (whereas the

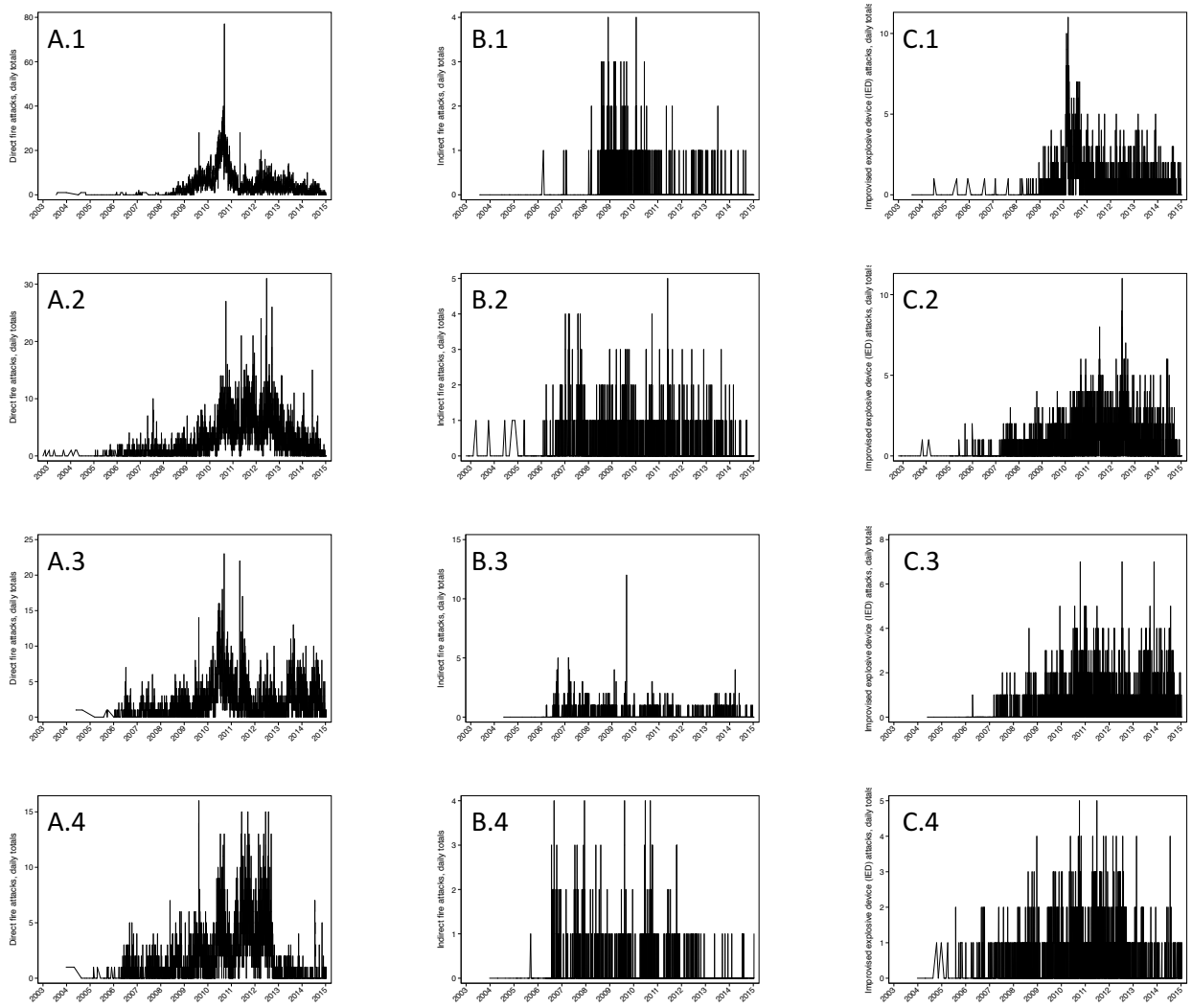
data would have been updated to reflect any such subsequent deaths among ISAF forces). Similarly, we calculate that Afghan security forces have suffered some 23,237 deaths during this same period.



The data reveal how concentrated the violence was. Although violence affected a large number of the country's districts (397 of 398) (Figure 1A), most violent attacks perpetrated throughout the conflict occurred in two general areas of the country: the Hilmand-Kandahar region and the area surrounding Kabul. We plot these regions in Figures 1B and 1C respectively.

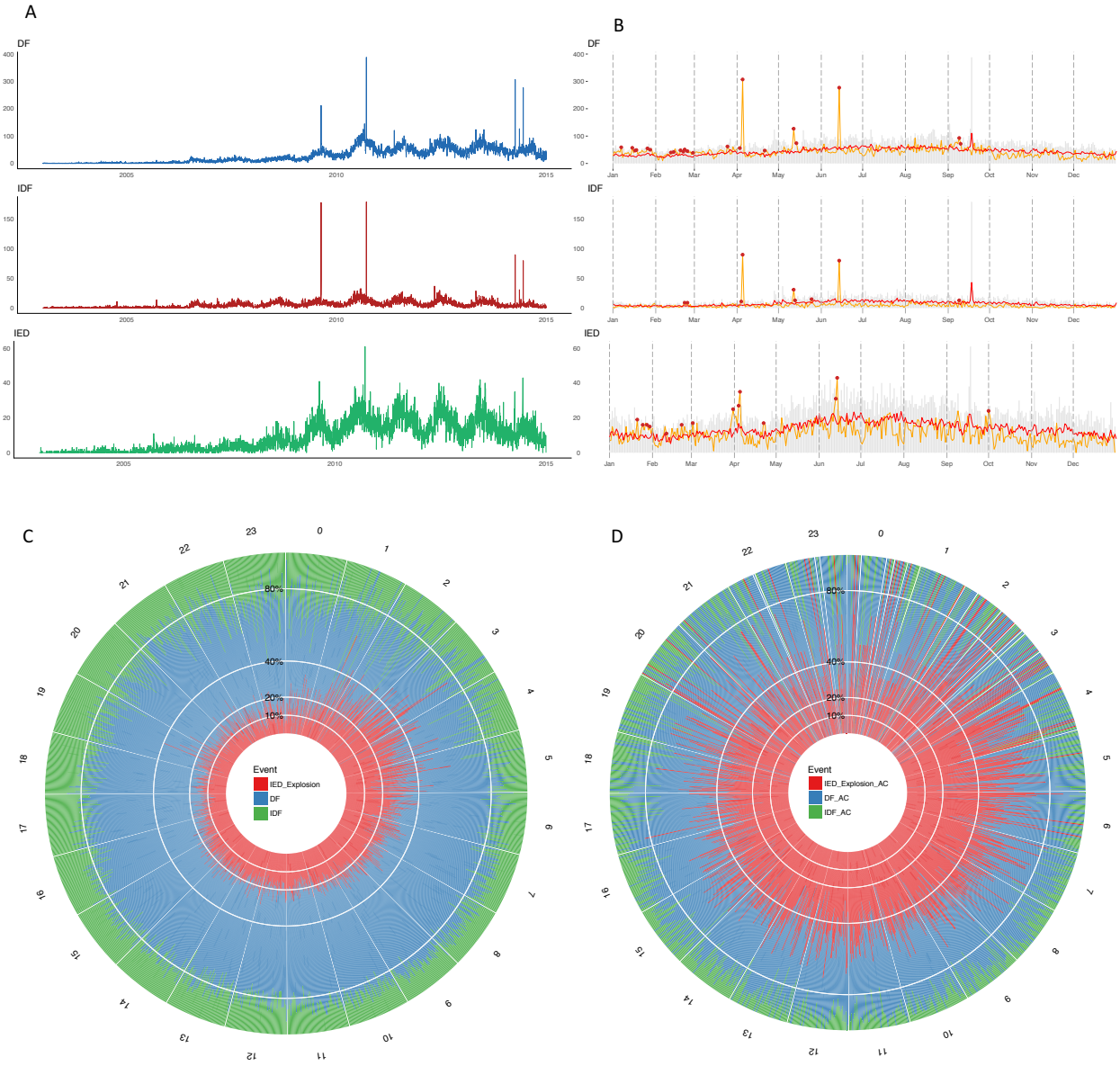
Interesting patterns in the data also emerge when they are viewed from various spatial and temporal perspectives. For instance, although violence at the national level in Afghanistan fol-

lowed a particular trend (Figure 3A), violence followed different patterns across various areas of the country. Although the progression of war is often considered by analysts at the level of the country, the data reveal that within a given conflict, insurgent activity develops along different timelines and at substantially different levels of intensity across different areas of such country. Such heterogeneity is depicted in the individual violence plots of four different Afghan districts (Figures 2A.1-2A.4). This Figure also reveals that rebels in the same district adopt different technologies of combat over time (Figures 2A.1-2C.1), at certain points in the campaign carrying out a large number of direct fire attacks while switching to IED attacks at other times.



When plotted across time, the data reveal that insurgents carried out the greatest number of attacks on election days (Figure 3A). Figure 3B also shows that conflict in Afghanistan is highly seasonal, with most of a given year's conflict taking place during the spring and summer months. In Figure 3C, we plot the distribution of attacks across minutes of the day throughout the conflict. While indirect combat encounters were common across the day, IEDs are most frequently used

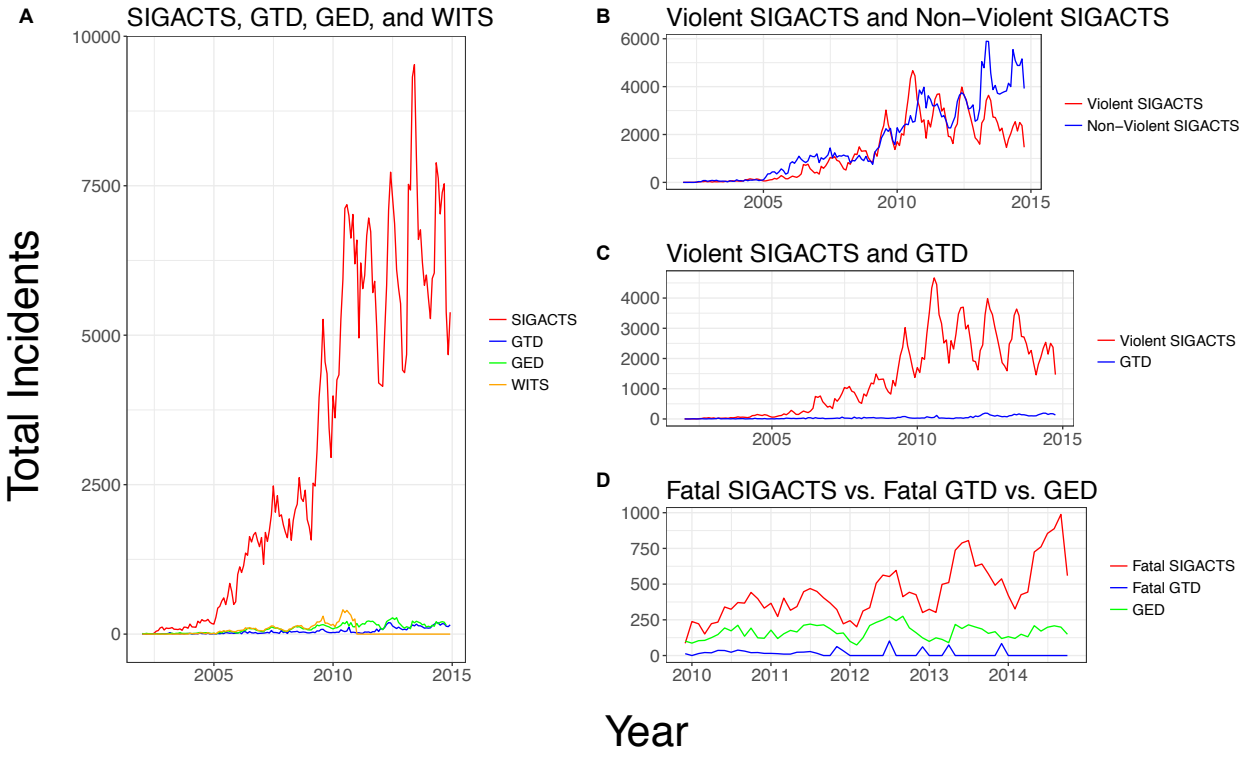
between 1:00 AM and 7:00 AM local time. Figure 3D shows the minute-by-minute patterns of combat casualties. Notice that IEDs are the most deadly weapon insurgents used, and remain a consistent threat through the day.



Comparisons with Existing Datasets. A variety of subnational conflict datasets have been developed by scholars to facilitate micro-empirical studies of conflict. Notable existing subnational datasets that cover the Afghanistan war include, as previously referenced, the GTD, GED, and WITS datasets, which are based on media reports. For events covering the Afghanistan War, SIGACTs is far larger than the other datasets, containing a total of 487,313 observations relative to 7,612, 9,383, and 15,092 observations for GTD, WITS, and GED, respectively (see Figure 4A). However, these numbers are not directly comparable. Although all three datasets track individual incidents of political violence, they differ from one another and the SIGACTs dataset in their focus and coverage.

GTD and WITS both focus primarily on terrorist violence but also track instances of political violence that can also be classified as insurgent or guerrilla activity. WITS, however, restricts coverage of insurgent violence to only particular types thereof. In particular, WITS restricts such attacks to insurgent attacks on civilians and police forces. Attacks on military forces are included only if one or more civilian or civilian actor (e.g. civilian government officials) is affected as well. GED focuses more generally on incidents of political violence; however, unlike the first two datasets, GED covers only events (estimated) to have resulted in one or more fatalities. It is important to note that these other datasets have a wider scope, and track violence across multiple conflict theaters.[2] Our objective is to highlight the important details captured with an in-depth data collection effort focused on a single, long-run conflict.

[2]These existing data sources represent a critical resource to the scientific community and have facilitated the study of various topics requiring cross-country comparisons.



Although GTD is the most comprehensive in the types of events it covers, it nonetheless covers a far smaller number of events than SIGACTs. When SIGACTs is subset to include only incidents of insurgent violence, it contains 208,217 observations. When GTD is limited to events listed as insurgent events, it drops to 1123 observations. Figure 4C compares violent SIGACTs with all GTD insurgent activities. For WITS, when the data only attacks on police forces and, separately, attacks on military forces also affected civilians are included, the number of remaining observations is 2,688. Finally, over the course of the period during which SIGACTs consistently

tracks fatal events, SIGACTs includes a total of 26,482 fatal events compared with GED's 10,252 and 4,126 fatal events tracked by GTD (see Figure 4D). A second major advantage of the SIGACTs dataset is its inclusion of non-violent conflict related events (Figure 4B). These new data allow us, for example, to study the location of illegal checkpoints used by insurgents to tax traffic along major roadways and the routes used for smuggling weapons and narcotics from Afghanistan to neighboring countries. Data of this type, while relevant to the study of subnational conflict, are largely missing from the scientific study of conflict.

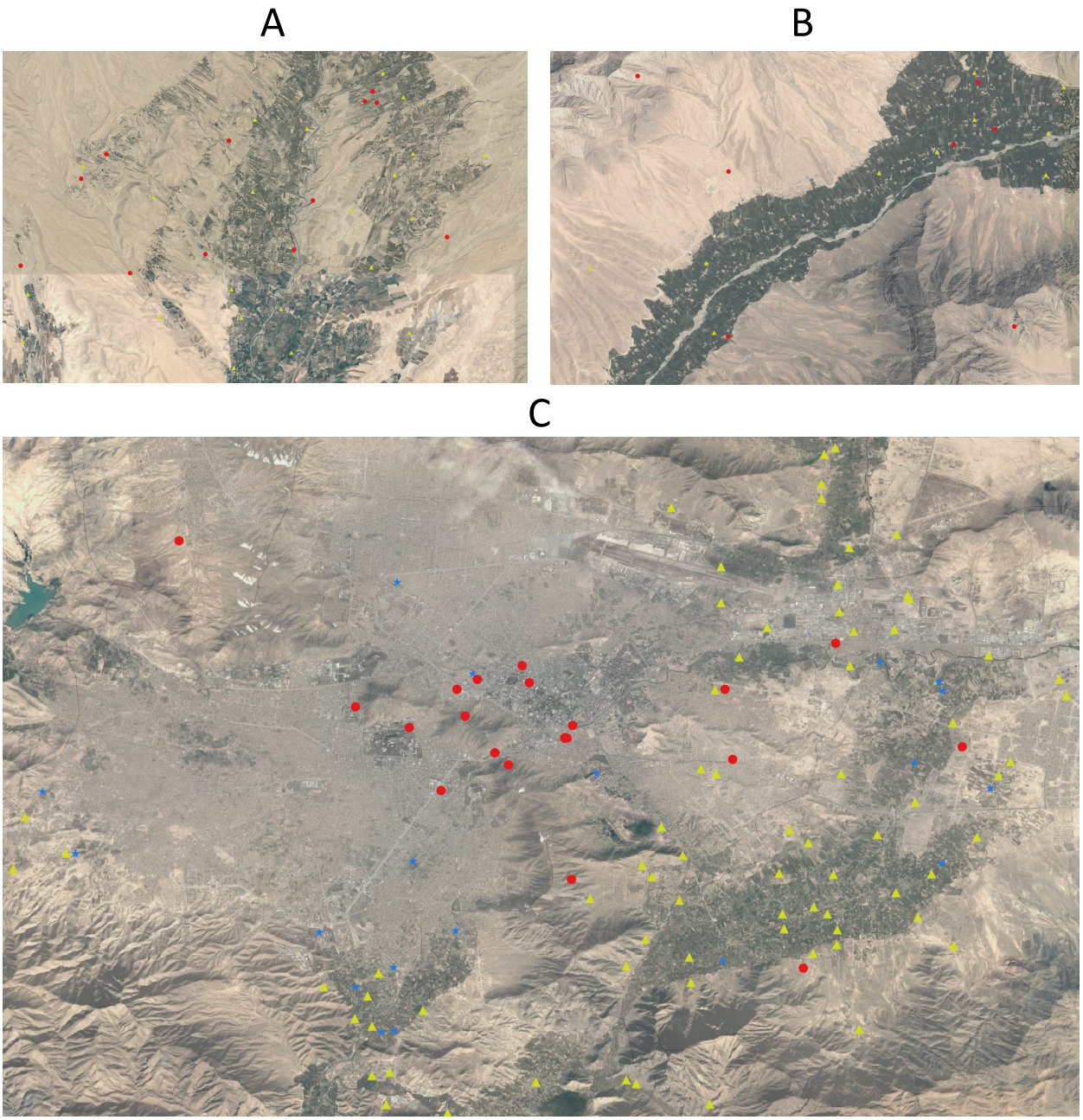
A final notable feature of the SIGACTs data is that, unlike these other datasets, it is not based on media reports. [7] finds that micro-conflict datasets are biased because of the differential rates of violence reporting inherent in media coverage. Specifically, [7] theorizes and finds evidence that areas with higher population density and cell phone access are more likely to generate media reports of political violence. While the WITS, GED, and GTD datasets rely on media coverage for the identification of violent events, the data generating process by which the SIGACTs dataset was produced ensures no difference in reporting likelihood as a function of demography or geography.

Data Application: Training Rebels in Plain Sight. To demonstrate how the data might be used, we analyze the spatial distribution of identified terrorist training camps operated by the Taliban and other violent political organizations throughout the Afghanistan war.

Scholarship in the social sciences has long asserted a direct link between rugged terrain and the likelihood of subnational conflict.[12,13] Rough terrain, it has been argued, enables otherwise weak rebel forces to train their fighters, consolidate control, and launch sophisticated attacks

against state militaries. Afghanistan is home to some of the most rugged terrain in the world. The general consensus among academics and commentators has been that the Taliban succeeded in withstanding the U.S.-led invasion of Afghanistan by securing refuge and training fighters along the rugged, snowcapped border with Pakistan.

Our data provides a unique opportunity to investigate where the Taliban planned their attacks and trained their forces. Did rebels, for instance, plan and train under the cover of rugged terrain, deep in the mountainous regions of Afghanistan? Or, on the other hand, did they accumulate their fighting capacity near urban areas, where they could seek logistical support and hide in plain sight? To investigate these questions, we combine our conflict microdata with administrative statistics on the locations of villages and government security outposts, and satellite-derived information on terrain ruggedness.



We begin by visualizing the terrain and population features proximate to rebel encampments. In Figure 5A, we plot the location of camps near Dehjawze Barekzi, in the northern region of Kandahar province. Notice that the camps (red dots) are located both in the mountain surrounding the

urban valley, and also along the outskirts of several villages. In Figure 5B, we study a similar map in Kala Kala, located within the Hilmand province. Here, a number of facilities are located within the densely populated valley region. In Figure 5C, we map training locations within the administrative boundaries of the national capital, Kabul. Throughout the war, more than a dozen logistics sites were established by the Taliban in the capital city. Several of these sites were established within a kilometer of security force outposts.

We next perform several quantitative assessments of the characteristics of training camps. Our data reveals that, on average, insurgent facilities were roughly 2 kilometers away from the nearest population settlement. In Figure 6A, we demonstrate that the land surrounding rebel training camps was no more (but probably less) rugged than villages. This finding indicates that rebel facilities were no better protected by rough terrain than the average village.

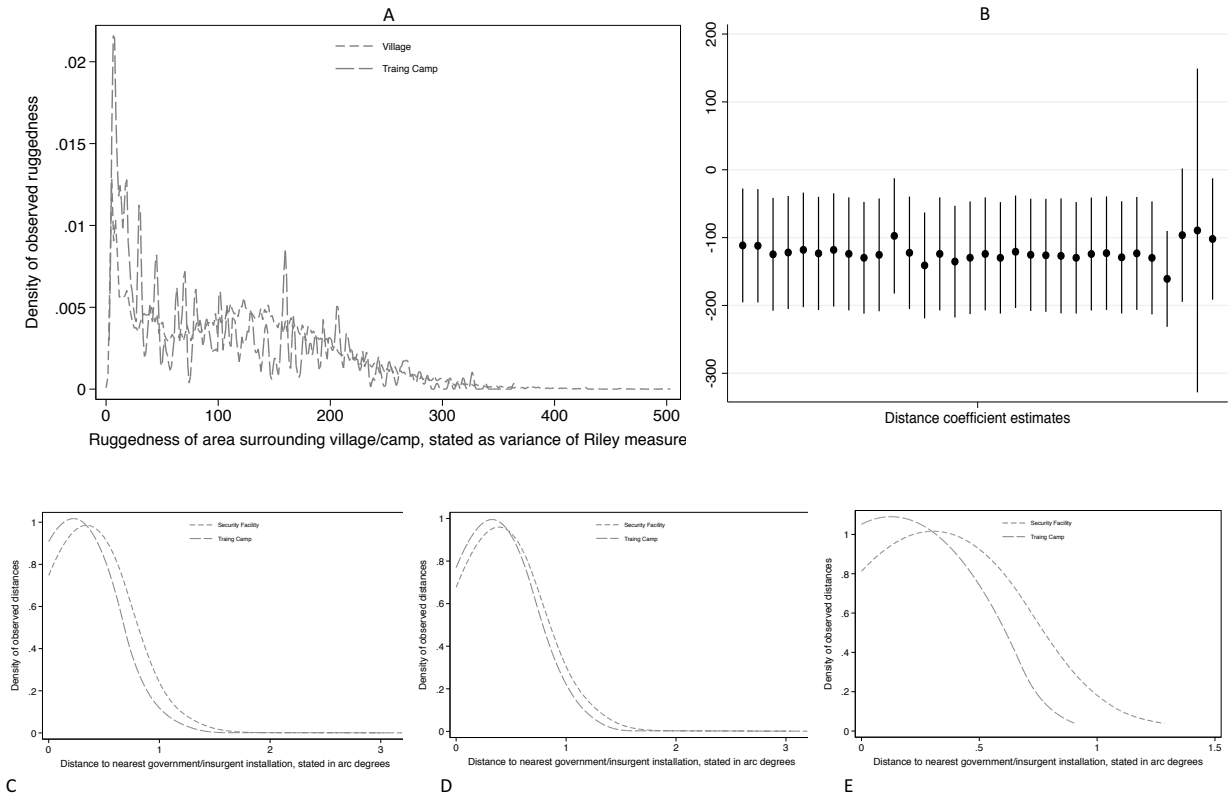
We also consider whether there is a tradeoff between terrain ruggedness, which could hinder rebel mobility, and proximity to the nearest village. These population settlements could provide essential logistical, financial, and recruitment support to insurgents. To examine this tradeoff, we utilize an ordinary least squares (OLS) regression. We regress the ruggedness of the area surrounding the training camp on the distance between each training camp and the nearest village:

$$Ruggedness_{tc} = \alpha + \beta_1 VillageProximity_{tc} + \epsilon_{tc} \quad (1)$$

If insurgents are willing to limit their force mobility in order to gain access to support from civilians, we expect to find a negative coefficient. The results of these regressions are plotted in Figure 6B, where we sequentially exclude one province from the sample. Notice that the results are

robust and substantially large with one exception. This exception, the Kandahar province, makes up roughly 20% of the sample of training camps, and excluding this province reduces the analytical power of our test significantly. Importantly, however, the coefficient remains consistent when Kandahar is removed from the sample. These results convey the importance of village proximity in shaping where training camps are established and highlight the tradeoff between terrain ruggedness and access to potential safe havens.

We conclude by addressing an important, unanswered question about the distribution of combatants in the Afghanistan war: are villages closer, on average, to rebel or government outposts? In an asymmetric conflict, it is expected that government forces outmatch rebels in their proximity to civilians. Our data indicate otherwise. In Figure 6C, we study the distribution of distances between all villages and the closest rebel and government security force encampments. Notice that the density of the training camp curve is further to the left than the security facility curve, indicating that villages are closer to rebel camps than government outposts. In Figure 6D we recalculate these distributions and confirm that the result holds even in districts that experienced below median levels of violence during the war. Figure 6E studies villages in districts with above median levels of violence. Naturally, the core result—that villages are closer to rebels than government units—is even stronger in these intensely contested districts.



Selection biases could influence the correlation between training camp locations, terrain ruggedness, and village proximity. It could be the case, for example, that counterinsurgents are more likely to identify camps within or proximate to urban centers than those situated in desolate, rugged areas of the country. Given the advanced drone and satellite monitoring technologies available to counterinsurgents (U.S. forces, in particular), we consider this unlikely. Rebel structures and suspicious activity are more easily detected through remote sensing techniques when situated within otherwise barren areas. It could also be the case that counterinsurgents are more likely to become aware of insurgent operations in urban areas if civilians provide information about their

whereabouts. Yet training camps are likely to be situated when and where insurgents are confident that civilians are unwilling to provide intelligence to counterinsurgents. Whatever the case may be, this cursory exploration of the data provides motivation for future work on Taliban tactics.

Data Limitations. Although the data offer scholars the opportunity to study empirically a variety of wartime dynamics that they previously could not, we wish to note several data limitations. First, and most obviously, because the data were generated during the Afghanistan war, inferences drawn from the data may be limited in their external validity. The conflict is unique in various respects, but shares characteristics with a number of recent insurgencies and is likely to yield insights on future counterinsurgent engagements.

Second, for some dataset variables, the nature of their collection ensures that we are unlikely to know the universe of all incidents. Core event types like direct, indirect, and explosive device attacks are unaffected. However, other insurgent activities included in the data, such as rebel recruitment by the Taliban, are limited to instances where such activity was reported to or discovered by security forces. As such, we do not observe all attempts to recruit fighters, and the data should interpreted and investigated with this caveat in mind.

Third, as discussed earlier, some of the more granular details captured by our data were most consistently included in military records as the conflict progressed. In some cases, the casualties associated with each event and who was targeted were left blank. In the future, we aim to explore these limitations further.

DISCUSSION

A great deal remains unknown about the dynamics of subnational conflict. In Afghanistan, for instance, where does the Taliban train its militants and otherwise stage attacks? Through what routes did the Taliban smuggle weapons, narcotics, and other contraband? More generally, how do development aid and urban investment shape the expansion or contraction of insurgent operations?

The Afghanistan SIGACTs data promises to provide researchers with the ability to rigorously study a wide variety of pressing questions to which statistical analyses could not previously be applied. As such, this dataset represents a tremendous opportunity for members of the social and physical science communities to work together to explore the nature of terrorism, insurgency, civil war and other forms of internal instability. We offer these data to both communities for the first time through this Resource Article as a means of facilitating such collaboration.

The data will be shared through several platforms. The data will be available directly through *Nature Human Behaviour*. The data will also be posted in various formats to facilitate various research uses (for instance, at different levels of spatial and temporal aggregation) on the authors' academic websites. There are no limitations imposed on the sharing or use of this data.

1. Mikulaschek, C. & Shapiro, J. N. Lessons on Political Violence from America's Post–9/11 Wars. *Journal of Conflict Resolution* (2016).

2. Berman, E. & Matanock, A. M. The empiricists' insurgency. *Annual Review of Political Science* **18**, 443–464 (2015).

3. Sundberg, R. & Melander, E. Introducing the UCDP georeferenced event dataset. *Journal of Peace Research* **50**, 523–532 (2013).

4. Croicu, M. & Sundberg, R. UCDP georeferenced event dataset codebook version 4.0. *Journal of Peace Research* **50**, 523–532 (2015).

5. LaFree, G. & Dugan, L. Introducing the Global Terrorism Database. *Terrorism and Political Violence* **19**, 181–204 (2007).

6. Wigle, J. Introducing the Worldwide Incidents Tracking System (WITS). *Perspectives on Terrorism* **4** (2010).

7. Weidmann, N. B. A closer look at reporting bias in conflict event data. *American Journal of Political Science* **60**, 206–218 (2016).

8. Shaver, A. & Bollfrass, A. Affective motivations for violence in insurgent warfare. Tech. Rep., Working Paper (2016).

9. Nordland, R. 7 U.S. Soldiers Wounded in Insider Attack in Afghanistan. *The New York Times* (2017).

10. Ahmad, J. Analysis of 'Insider Attacks' in Afghanistan. *Real Clear Defense* (2017).

11. icasualties. Iraq Coalition Casualties Count (2017). URL `http://icasualties.org`.

12. Galula, D. *Counterinsurgency warfare: theory and practice* (Greenwood Publishing Group, 2006).

13. Shaver, A., Carter, D. B. & Shawa, T. W. Terrain ruggedness and land cover: Improved data for most research designs. *Conflict Management and Peace Science* (2016).

Acknowledgements We are grateful to Luke Condra, Kyle Pizzey, and Jacob Shapiro for comments on this project.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to A.S. and A.W. at (email: ashaver@princeton.edu) and (email: austinlw@uchicago.edu).

Figure 1 Spatial distribution of violence during Afghanistan war. (A) Country-level distribution. Black dots represent individual events. (B) Regional-level distribution, Hilmand and Kandahar provinces. (C) Regional-level distribution, Kabul area.

Figure 2 Temporal distribution of violence, by type and district. (A) Direct fire attacks, daily totals. (B) Indirect fire attacks, daily totals. (C) Improvised Explosive Device (IED) attacks, daily totals. (1, 2, 3, 4) indicate Nad Ali, Nahri Sarraj, Sangin, Zhari districts, respectively. In order, these were the four most violent districts during the Afghanistan war.

Figure 3 Temporal trends in data. (A) Time series plot of direct fire (blue), indirect fire (red), and improvised explosive device (green) events throughout the campaign. The four most violent days during the conflict were the four competitive national election days (in 2009, 2010, and two rounds in 2014). (B) Time series plot of direct fire, indirect fire, and IED events by week throughout the campaign. Red line indicates the average intensity by week; yellow line tracks the 2014 violence levels by week. (C) Plot of minute-by-minute violence as a ratio of total violence. (D) Plot of minute-by-minute casualties as a ratio of total casualties.

Figure 4 Comparing conflict microdata. (A) Time series of total incidents included in SIGACTs (red), GTD (blue), GED (green), WITS (yellow). (B) Time series of violent (red) and non-violent (blue) events reported in SIGACTs. (C) Time series of violent SIGACTs

(red) and all GTD events (blue). (D) Time series of fatal events in SIGACTs (red), GTD (blue), and GED (green).

Figure 5 Rebel training camp locations. (A) Camps (red dots) and villages (yellow triangles) in Dehjawze Barekzi. (B) Camps (red dots) and villages (yellow triangles) in Kala Kala. (D) Camps (red dots), villages (yellow triangles), and security outposts (blue stars) in Kabul.

Figure 6 Terrain, population centers, and training camp locations. (A) Smoothed kernel density plot of distribution of villages (short dash) and training camps (long dash) at varying levels of terrain ruggedness. (B) Regression coefficient estimates of terrain ruggedness on distance to nearest village. Coefficients from sequential samples excluding one province at a time. Widest confidence interval is sample without Kandahar. (C) Smoothed kernel density plot of distribution of village distance to security facilities (short dash) and training camps (long dash). (D) Smoothed kernel density plot of distribution of village distance to security facilities (short dash) and training camps (long dash) in districts with below median aggregate violence during conflict. (E) Smoothed kernel density plot of distribution of village distance to security facilities (short dash) and training camps (long dash) in districts with above median aggregate violence during conflict.