# Quantifying the future lethality of terror organizations

Yang Yang[a,b,1], Adam R. Pah[a,b,1], and Brian Uzzi[a,b,2]

[a]Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208; and [b]Kellogg School of Management, Northwestern University, Evanston, IL 60208

Edited by Arild Underdal, University of Oslo, Oslo, Norway, and approved September 4, 2019 (received for review February 5, 2019)

**As terror groups proliferate and grow in sophistication, a major international concern is the development of scientific methods that explain and predict insurgent violence. Approaches to estimating a group's future lethality often require data on the group's capabilities and resources, but by the nature of the phenomenon, these data are intentionally concealed by the organizations themselves via encryption, the dark web, back-channel financing, and misinformation. Here, we present a statistical model for estimating a terror group's future lethality using latent-variable modeling techniques to infer a group's intrinsic capabilities and resources for inflicting harm. The analysis introduces 2 explanatory variables that are strong predictors of lethality and raise the overall explained variance when added to existing models. The explanatory variables generate a unique early-warning signal of an individual group's future lethality based on just a few of its first attacks. Relying on the first 10 to 20 attacks or the first 10 to 20% of a group's lifetime behavior, our model explains about 60% of the variance in a group's future lethality as would be explained by a group's complete lifetime data. The model's robustness is evaluated with out-of-sample testing and simulations. The findings' theoretical and pragmatic implications for the science of human conflict are discussed.**

terrorism | counter-terrorism | human conflict | organizational behavior | statistical models

The development of scientific methods for explaining and controlling violent insurgency is a major concern worldwide. The US government alone dedicates half-a-trillion dollars annually to researching, combating, and responding to terrorism (1). The cost, however, is estimated to be much larger once private institutional and consumer spending on counterterrorism is taken into account (2). A 2017 Pew survey showed that the emotional costs of violent insurgency have also burrowed into the psychology of Americans, who now rank terrorism as a top priority ahead of traditional societal priorities including education, jobs, science research, the environment, and health care (3). Similar situations exist in other nations, where increasing rates of insurgency are associated with lower levels of national investment (4).

Despite counterterrorism efforts, terrorism is increasing at a rate that outpaces security resources (5–10). From 2000 to 2015, worldwide terror attacks increased 8-fold, and 61 new groups emerged each year on average (11). At the same time, groups have become better at concealing their capabilities and resources (12–15). This dynamic has created a need to allot security resources to those terror organizations with the greatest potential for harm (12, 16, 17). In this work, we focus on terror organizations that commit acts that involve loss of life. Nonlethal groups are an important class of terror groups but differ from lethal groups in that they tend to have low activity levels—committing 2 attacks and surviving only 8 months on average (18). By contrast, lethal groups survive for up to 15 years and commit 150 attacks on average, with those lethal attacks being the most psychologically traumatic (19). This dynamic creates a need to explicitly identify and incapacitate groups that commit deadly attacks early in their lifespan. To that end, we study all groups in the Global Terrorism Database (GTD) that have committed $\geq 10$ attacks and have at least 12 mo between their first and last recorded attack in the dataset. This results in 342 terror groups that have committed 157 attacks and existed for 14.8 y on average to estimate the parameters of our model. We then use 2 additional datasets—the RAND Database on Worldwide Terrorism Incidents (RDWTI) and the 2017 GTD—for out-of-sample tests of the accuracy of our estimates (*Materials and Methods*).

Researchers have attempted to indirectly estimate a group's lethality from contextual variables such as gross domestic product (GDP) (20), income inequality, criminality, war zone activity, or alliances. For example, research generally shows that terror attacks are negatively correlated with a locale's GDP and its degree of democratic governance and positively correlated with its population size (21) and proximity to a war zone (22, 23). Related work shows that attacks at the population level of analysis are distributed as a power law (5, 8, 9) and that the number of attacks per group, ignoring the severity of attacks, is positively associated with a terror group's number of alliances, web presence, and territorial control (6, 22, 24, 25) but unrelated to group size (7). These studies have provided meaningful insights about the sociopolitical context in which terror activity is most frequent and help estimate the average attack severity within a region. However, they are less oriented toward predicting an individual terror organization's behavior. These models estimate average terror behavior within a locale rather than distinguish the potential lethality of one group from another within the region (6, 13, 26–29).

Complex system sciences show that an organization's performance broadly depends on 2 factors—capabilities and resources (30–33). In legitimate organizations, publicly reported organizational capabilities (e.g., technological skills) and resources (e.g., cash flow) are routinely used by analysts and investors to predict behavior. In covert organizations, performance should be similarly driven by the strength of capabilities and resources, but the

## Significance

**We develop and test a model that predicts the future lethality of a terror group. The model significantly increases the level of explained variance over existing models and presents early-warning signal predictions. Our early-warning model predicts the future lethality of a group using only a handful of events that occur soon after it emerges. Using the first 10 to 20 attacks or the first 10 to 20% of a group's lifetime, our early-warning model provides about 60% of the explanatory power as would having a group's complete lifetime data, which is a basis for improved counter insurgency resources.**

Author contributions: Y.Y., A.R.P., and B.U. designed research; Y.Y., A.R.P., and B.U. performed research; Y.Y. and A.R.P. analyzed data; and Y.Y., A.R.P., and B.U. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.



[1]Y.Y. and A.R.P. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: uzzi@kellogg.northwestern.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1901975116/-/DCSupplemental.

First published October 7, 2019.

concealment of capabilities and resources hampers their measurement (34–36). Our model hypothesizes that observable variables normally collected on all terror organizations can be used in original ways to estimate terror organizations' relative levels of capabilities and resources for lethality via maximum-likelihood techniques.

To address potential data issues, we conduct extensive data reliability and model robustness tests. First, we use 2 terrorism datasets—the GTD and the RDWTI database—and multiple data specifications as sensitivity analyses to ensure the robustness of the results. First, the GTD data are used to calibrate the 2 parameters of our model using cross-validation to avoid overfitting. Second, the RDWTI data are used for out-of-sample testing. Third, we test our model's sensitivity to measurement and censoring error by computing changes in the model's effects when simulated data are added to or subtracted from the reported data.

## Model Specification

We begin by separately describing the nature of capabilities and resources, testing their utility for predicting total lethality, and conducting validity tests. We then combine these 2 explanatory factors and perform multiple sensitivity analyses to demonstrate that our variables encode predictive information that is not contained in other variables or due to the specification of the data. After demonstrating the robustness of these factors, we turn to our primary contribution—using these 2 factors to predict an organization's future lethality after observing only a handful of a group's first attacks.

**Hidden Capabilities.** Capabilities are defined as organizational assets that are relatively stable over an organization's lifespan and sustain a predictable level of performance (37–39). For legitimate organizations, the link between levels of capabilities and levels of performance have been verified. For instance, fixed manufacturing and information technology capabilities can predict an organization's product reliability (40, 41). By analogy, capabilities in terror organizations include technical expertise in bomb-making, organizational know-how in armed attacks (42), or cultural-religious principles (43) that sustain their lethality. To circumvent the hidden nature of terror organizations' capabilities, our model approximates a group's capabilities by assuming that terror groups have similar types of capabilities (e.g., bomb-making capabilities) but in varying levels of strength (e.g., unsophisticated pipe bombs to sophisticated bomb know-how) (32). This assumption allows us to estimate a group's capabilities on a scale that assesses the relative strength of one terror group to all others without a need to identify specific capabilities (38, 44). In our model, the term $Q_{5,i}$ represents group $i$'s capabilities.

To operationalize $Q_{5,i}$, we use the cumulative lethality of a group's first 5 attacks. The first 5 are chosen to avoid overfitting and because it follows theory that purports that capabilities are relatively fixed over time (an assumption tested below). Using a larger number than the first 5 only increases the model's fit. Specifically, 1) we estimate the relationship between $Q_5$ and lethality for each group separately with

$$d_{i,\alpha} = Q_{5,i} p_\alpha, \quad [1]$$

where $d_{i,\alpha}$ is the number of fatalities in attack $\alpha$ by group $i$ and $p_\alpha$ represents stochastic noise in the relationship. 2) We adjust a group's $Q_{5,i}$ relative to its peer population (groups that exist before the emergence of group $i$), which becomes our final measure of group $i$'s capabilities, $Q_{5,i}$. After substitution (see *SI Appendix*, section S2.3 for derivation), this results in

$$Q_{5,i} = e^{\langle log(d_{i,\alpha})\rangle - \mu_{\hat{p}}}, \mu_{\hat{p}} = \langle log(d)\rangle, \quad [2]$$

which is estimated using maximum likelihood with the GTD dataset.

$Q_5$'s predictive utility compares well with existing models (Table 1). Based on the literature, we specified baseline regressions where future lethality (total lifetime lethality after the first 5 attacks) is regressed on sociopolitical variables (e.g., GDP, criminality) (7, 8, 26), country and decade fixed effects, as well as a direct measure of the sum of a group's first 5 kills, a quantity used to construct $Q_5$. Using the GTD data, the strictest baseline

**Table 1. Regression analysis to evaluate the relationship between our model parameters, $Q_5$ and $T_Z$, and the future lethality of a terror group, while controlling for other predictors of lethality and model specifications**

| | Baseline | | $Q_5$ | | | $T_Z$ | | $Q_5T_Z$ | | Alliances subsample | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| **$QT_Z$ model** | | | | | | | | | | | |
| $Q_5$ | | | 0.60*** | 0.50*** | 0.24* | | | 0.58*** | 0.26* | | 0.24 |
| | | | (0.10) | (0.11) | (0.11) | | | (0.093) | (0.11) | | (0.21) |
| $T_Z$ | | | | | | 0.47*** | 0.40*** | 0.46*** | 0.40*** | | 0.51* |
| | | | | | | (0.078) | (0.077) | (0.073) | (0.076) | | (0.15) |
| Control variables | | | | | | | | | | | |
| Sum of first 5 fatalities | | | | 0.0027* | 0.0034* | | 0.0039* | | 0.0028* | | -0.0017 |
| | | | | (0.0013) | (0.0014) | | (0.0012) | | (0.0013) | | (0.0045) |
| Sociopolitical variables | YES | | | | | | | | | | |
| Decade FE | YES | YES | | | YES | | YES | | YES | | YES |
| Country FE | | YES | | | YES | | YES | | YES | | YES |
| Allies number | | | | | | | | | | 0.15** | 0.19** |
| | | | | | | | | | | (0.043) | (0.042) |
| Rivals number | | | | | | | | | | 0.25 | 0.73** |
| | | | | | | | | | | (0.20) | (0.19) |
| Observations | 328 | 342 | 342 | 342 | 342 | 342 | 342 | 342 | 342 | 100 | 100 |
| $R^2$ | 0.11 | 0.37 | 0.10 | 0.11 | 0.41 | 0.10 | 0.45 | 0.19 | 0.47 | 0.12 | 0.76 |
| BIC | 1,389 | 1,639 | 1,338 | 1,340 | 1,630 | 1,339 | 1,604 | 1,307 | 1,603 | 426 | 465 |

$Q_5$ and $T_Z$ together are consistently strong predictors of future lethality and add significant explanatory power ($\Delta$BIC $> 10$). Variance inflation factor statistics do not indicate ill-conditioned specifications. Sensitivity analyses in *SI Appendix* present a range of more detailed tests. Table 2 presents $Q_5$ and $T_Z$'s power as early-warning signals of a terror group's future lethality based on a fraction of a group's early behavior. The numbers in parentheses indicate the standard errors of coefficients. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

regression (Table 1, column 2) has a substantively large $R^2$ of 0.37. By comparison, regressing future lethality on just $Q_5$ shows that $Q_5$ significantly ($P < 0.001$) predicts future lethality, and $Q_5$ alone produces an $R^2$ of 10% (col. 3)—almost 30% the explained variance of the baseline model (col. 2) and comparable to the sociopolitical variables (col. 1).

To observe $Q_5$'s added explanatory value over existing models, col. 5 combines $Q_5$ and the baseline model. $Q_5$ continues to be highly significant ($P < 0.001$) when incorporated into the baseline model. Further, $Q_5$ raises the overall $R^2$ from 0.37 to 0.41, a 11% improvement in $R^2$ and has "positive support" for including $Q_5$ in the model (per Bayesian information criterion [BIC] statistic) (45).

Construct validation tests support the findings. If $Q_5$ is a proxy for relative levels of capabilities, then we should observe that high $Q_5$ groups perform better than low $Q_5$ groups. One widely recorded capability for terror groups is their attack type (46). Attack types include bombings, armed assaults, kidnappings, etc. For example, a capability in bombings requires unique know-how in detonation, explosives, clean rooms, and projectiles. Kidnapping requires capabilities in managing hideouts and telecommunications. To measure attack success in a way that does not use lethality as the measure of success, we used a separate measure of terror success: a group's attack success relative to the attack's intent. The GTD estimates a terror group's success or failure with respect to the group's intended objective. For example, if a terror group's intended objective was to destroy a power station but it fails to do so, the attack is classified as unsuccessful—even if (tragically) it was lethal because causalities resulted. Thus, the GTD "success" variable provides a separate measure of success that differs from lethality but should nonetheless correlate with $Q_5$, as estimated from our model. Consistent with validation expectations, we found that high $Q_5$ groups have significantly higher overall success rates than low $Q_5$ groups using the same capability (Kolmogorov–Smirnov [KS] test; $P < 0.001$; see *SI Appendix*, section S2.8 for details). For example, high $Q_5$ groups with capabilities in bombings more successfully reach their intended goals than low $Q_5$ groups with capabilities in bombings. We also validate that $Q_5$ is a relatively fixed organizational quantity. Consistent with the validation expectations, tests show that $Q_5$ is uncorrelated with a terror group's lifetime (*SI Appendix*, Fig. S5).

**Hidden Resources.** Together with capabilities, resources affect organizational performance (30, 39). Whereas capabilities are relatively fixed assets, resources are assets that can fluctuate unpredictably over time. Resources include such factors as a terror group's funding, leadership, or intelligence reports (32, 47). To estimate a terror organization's resources, we draw on case studies and organizational research, which shows that the timing of an organization's product releases can be a proxy for an organization's underlying resources (30, 32, 33, 48). For example, when an organization releases products more systematically, it suggests that the organization has a steady stream of resources. This pattern exists because a consistent level of resources enables an organization to plan ahead with an eye to optimizing market receptivity and product quality of a series of forthcoming products. For example, in software development, a steady flow of resources enables planning for research and development and product launches that can be brought to market in ways that mutually reinforce each product's success (49). The opposite is true for organizations with erratic product releases. In this case, organizations typically have low, unpredictable levels of resources, promoting opportunistic use of their immediate and uncertain level of resources (33).

If a terror group's product is assumed to be an attack, then the timing pattern of attacks may similarly reflect a group's resource levels. The idea of randomness in attack patterns is also consistent with game theoretic models that argue that terror organizations aim to add randomness to their attack patterns conditional on their organizational constraints (14, 15, 50, 51). For example, models find that the resource levels of local rebel units influence the timing of their attacks—"As rebels gather more resources, their attacks become temporally concentrated in a manner that is distinguishable from randomized combat" (14).

To quantify the randomness of a group's attack timing pattern, we converted group $i$'s days of attacks (e.g., 01-01-1991) to a "date" between 0.0 and 1.0, where 0.0 is the time of a group's first attack and 1.0 its last known attack to normalize groups' lifespans. This transformation preserves the interevent times and number of attacks and allows us to calculate $t$, which measures how erratic a group's interevent timing is (52). However, we are primarily concerned with how random a group's attack timing is given the number of attacks conducted. To do that, we generate $10^5$ synthetic attack patterns for each group, where each synthetic attack pattern is constructed by randomly sampling (with replacement) from the population distribution of attack interevent times, and compute a group's $T_Z$ score as

$$\mathbf{T_Z} = \left| \frac{t_{obs} - \bar{t}_{sim}}{std(t_{sim})} \right|, \quad [3]$$

where $\bar{t}_{sim}$ is the average $t_{sim}$ across all simulation runs and $std(t_{sim})$ is the SD. $T_Z$ is interpreted like a standard $Z$ score, where larger values are relatively less erratic.

$T_Z$ alone significantly predicts ($P < 0.001$) a group's future lethality (Table 1, column 6) and explains comparable variance to the sociopolitical model (7, 8, 26) (Table 1, column 1). When combined with the baseline model variables, $T_Z$ remains significant ($P < 0.001$) in the presence of control variables and improves the explained variance ($R^2 = 0.45$ vs. $R^2 = 0.37$). This finding also indicates "very strong support" (45) for incorporating $T_Z$ as an independent variable in a model of lethality ($\Delta$BIC $= 35$).

The strong statistical relationship between $T_Z$ and lethality is bolstered by construct-validity tests. If $T_Z$ is a measure of latent resources, it should vary with the diversity of weaponry used by the group. The use of diverse weaponry across attacks (e.g., melee, firearms, vehicles, etc.) requires a group to have a more consistent level of resources than if only one type of weaponry is used in attacks (53). Consistent with this reasoning, the GTD data show that high $T_Z$ groups use significantly more diverse weaponry on average than low $T_Z$ groups (KS test; $P < 0.01$; *SI Appendix*, section S3.3 and Fig. S15). $T_Z$ is also not driven by the length of a group's lifespan or number of attacks (*SI Appendix*, section S3.4). These findings demonstrate that $Q_5$ and $T_Z$ measure information not captured by other variables in the model and evidentially offer some support for the hypothesis that they are proxy measure's of a terror group's relative capabilities and resources, although future research is required to decidedly confirm the connection.

**$QT_Z$ Model Robustness Checks.** Our full model uses $Q_5$ and $T_Z$ simultaneously. $Q_5$ and $T_Z$ explain about 20% of the observed variation in the data and have a positive and significant association with lethality ($P < 0.01$; Table 1, column 8). When $Q_5$ and $T_Z$ are added to the regression with control variables (Table 1, column 9) the $R^2$ rises to 47%, which is a 27% increase over the control variables on their own. $Q_5$ and $T_Z$ continue to have independent and significant effects in the combined model. The margins plot for $Q_5$ and $T_Z$ in the full regression model further demonstrates that both factors are needed to make accurate predictions (Fig. 1). Fig. 1 summarizes the composite relationship of predicted future lethality, $Q_5$, and $T_Z$ net of all control variables as specific in Table 1 (column 9). The dashed vertical and horizontal lines designate the medians of $Q_5$ and $T_Z$,

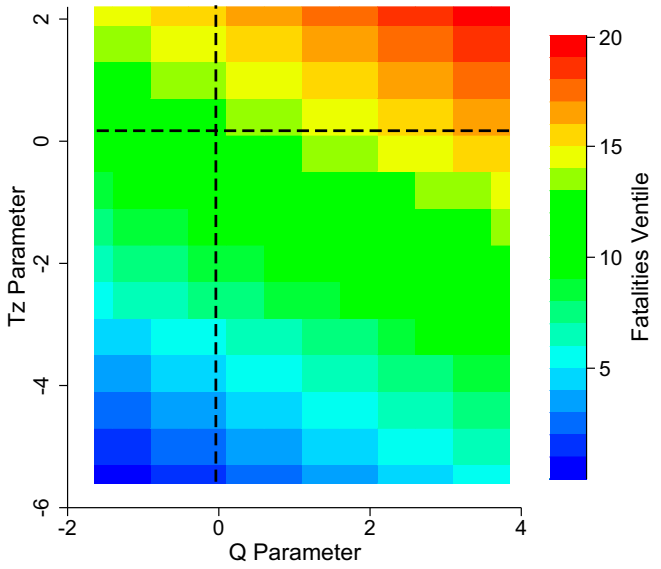


**Fig. 1.** $Q_5$ and $T_Z$ jointly predict lethality. The margins plot for $Q_5$ and $T_Z$ in the full regression model (Table 1, column 9) demonstrates that both factors are necessary to achieve the best prediction. As an example, we plot the median value (indicated with dashed lines) for both $Q_5$ and $T_Z$. Here, we find that a group's total number of fatalities could be in any one of the ventiles for the total number of casualties if $Q_5$ or $T_Z$ were used alone.

respectively, and the colors of the heat map represent ventiles of lethality. As expected, the upper right-hand quadrant has the highest predicted levels of future lethality and vice versa for the bottom left-hand quadrant. Notably, however, for most single values of $Q_5$ or $T_Z$, such as the median, a group's predicted future lethality could be in any ventiles, indicating that both $Q_5$ and $T_Z$ are needed to predict and explain a group's lethality.

Robustness tests for a group's alliances, definition of attacks, fraction of nonlethal attacks, cross-validation/overfitting, out-of-sample predictions, and error of measurement uniformly reinforce the strength of the above conclusions. 1) After adding hand-curated, network-based information about a group's alliances for 100 of the 342 GTD organizations (22), the $T_Z$ variables remain statistically significant ($P < 0.01$; Table 1, column 11) and improve the explained variance. 2) When we define attacks made by the same group on the same day as one attack instead of separate attacks, $Q_5$ and $T_Z$ remain significant ($P <$ 0.01 for both; *SI Appendix*, section S4.1). 3) When we control for the fraction of nonlethal attacks that a group commits, $Q_5$ and $T_Z$ remain significant and their addition improves model fit ($\Delta$BIC $> 10$; *SI Appendix*, section S4.2). 4) Three-fold cross-validation produces $R^2 = 0.18$, which compares favorably to the full model result of $R^2 = 0.19$ (*SI Appendix*, section S4.3). 5) We conducted an out-of-sample test with groups that appear in both the GTD and RDWTI data. We trained the model ($Q_5\,T_Z$ model) on GTD data for the groups and predicted the future lethality of the group as recorded in the RDWTI dataset. Our out-of-sample $R^2$ was 0.14, which compares favorably to the in-sample $R^2$ of 0.13 (in-sample is prediction of future lethality as recorded in the GTD for the subset of groups; *SI Appendix*, section S4.4). 6) A second out-of-sample testing was conducted on the 2017 GTD dataset to observe whether our model's predictions agree with future behavior. We found high agreement with an $R^2$ of 0.49 (*SI Appendix*, section S4.5). Further, training the $Q_5\,T_Z$ model on data up to 2010, up to 2005, and up to 2000 and predicting 2017 all work as well ($R$ = 0.46, 0,48, and 0.49), with $Q_5$ and $T_Z$ both significant ($P < 0.05$) in each model. 7) We evaluated the possible impact of missing or misattributed attacks in the GTD using synthetic data in an effort to address errors or inaccuracies in our data. To do this, we devised 3 different conditions, one where we randomly added attacks to a group's history, the second where we randomly deleted attacks, and the third when we swapped attacks between groups in the GTD. For each one of these conditions, we created 100 synthetic datasets and recalculated the $Q_5$ and $T_Z$ parameters. We then predicted the actual total number of kills for each group using the $Q_5$ and $T_Z$ parameters that were fit on the synthetic noisy data and found that the model performance was still robust, even when 3 out of 10 events are due to noise ($R^2 = 0.21$, 0.23, and 0.18 at 30% noise for add, remove, and swap, respectively; *SI Appendix*, section 4.7). The same synthetic data methodology can act as a check on right-censoring (54) by adding up to 30% more data to the attack measurements of right-censored organizations under the assumption that the right-censored organizations exist for hypothetical periods into the future. These tests suggest that measurement and censoring inaccuracies on the scale of more than half of the events decrease model performance.

**Early-Warning Predictions.** Accurately predicting a group's future lethality is most important soon after they emerge if security professionals are to engage in proactive targeting of these groups. A good early-warning model should have predictive utility that is similar to when the full data are available but with using only the handful of events that occur right after a group emerges.

To conduct our analysis, we calibrated $T_Z(x)$ using 2 different specifications, each of which restricts the data in an unique way, while keeping the construction of $Q_5$ unchanged. In the first specification, we considered a fixed number of first attacks for each group, beginning with its first 10 attacks and then its first 20 and 30 attacks. In the second specification, we used training data based on a percentage of a group's lifespan, which leads to a specification that includes all attacks that occurred within either the first 10, 20, or 30% of a group's lifespan. Table 2 demonstrates that the $Q_5\,T_Z$ model provides early-warning signals of a terror group's future lethality based on a few first attacks. Our analysis indicates that 1) $Q_5\,T_Z$ improves on the predictive utility of the control variables alone at all levels and 2) the $Q_5\,T_Z$ model's predictions are consistent at all early-warning sampling frames ($P < 0.01$; $t$ test; SEs calculated with bootstrap). Importantly, we achieve improvements in $R^2$ that are 20% more than the baseline model. For example, using the first 20 attacks or the first one-fifth of a group's lifetime provides over 60% of the explanatory power on average as having the complete lifetime data of a group ($R^2$ of 0.29 for first 20 attacks divided by

**Table 2. Table regresses a terror group's future lethality on values of $Q_5$ and $T_Z$ at different stages in a group's infancy**

| Model | Lifespan-based | | | Number of attacks-based | | |
|---|---|---|---|---|---|---|
| Variables | $x$ = 10% | $x$ = 20% | $x$ = 30% | $x$ = 10 | $x$ = 20 | $x$ = 30 |
| $Q_5$ | 0.62** | 0.58*** | 0.63*** | 0.48** | 0.45*** | 0.48*** |
| | (0.15) | (0.13) | (0.13) | (0.15) | (0.17) | (0.16) |
| $T_Z(x)$ | 0.20 | 0.35* | 0.28* | -0.09 | 0.21 | 0.58* |
| | (0.13) | (0.12) | (0.11) | (0.12) | (0.12) | (0.15) |
| FE | YES | YES | YES | YES | YES | YES |
| $R^2$ | 0.44 | 0.46 | 0.40 | 0.34 | 0.29 | 0.34 |
| $\Delta R^2$ | 16.70% | 17.90% | 20.20% | 9.60% | 7.4% | 17.2% |
| $\Delta$ BIC | 17 | 23 | 24 | 3 | 1 | 14 |

$Q_5$ is always based on the first 5 attacks and $T_Z$ is based on a group's first 10, 20, or 30% of lifespan data or a group's first 10, 20, or 30 attacks as designated in the column headings. FE refers to country- and decade-fixed effects being included in the regression. Improvements in $R^2$ are up to 30% more than the baseline model. Using the first 20 attacks or the first one-fifth of a group's lifetime provides over 60% of the explanatory power on average as having a group's complete lifetime data. The numbers in parentheses indicate the standard errors of coefficients. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

$R^2$ of 0.47 for all attacks $\approx 61\%$). These results suggest that the $Q_5\,T_Z$ model makes meaningful theoretical and pragmatic contributions to terrorism studies.

## Discussion

The $Q_5\,T_Z$ model estimates latent properties of terror groups from attack-timing data that predict the total future lethality of a group. Further, the $Q_5$ and $T_Z$ parameters are additive and complementary to other approaches in the field that leverage sociopolitical attributes at the country level or ones that incorporate relationship data between terror groups. Importantly, we find that our estimation of these parameters and their predictive ability is robust—with promising predictive performance even when we test out-of-sample or on noisy synthetic data—and that it is capable of providing a meaningful signal shortly after a group becomes active.

Despite the fact that terror groups do their best to obscure their operational activities and organizational strength (5, 7, 34, 35), the $Q_5\,T_Z$ model is uniquely able to predict their future lethality given their observable activities. Advancing work focused on conflict zones (5), robustness tests demonstrate that our predictions are generalizable across time and sociopolitical contexts.

The $Q_5\,T_Z$'s model features position it between theoretical literature and data sources. On a theoretical basis, we demonstrated how group-level variation in the size and timing of attacks could be a proxy for the group's hidden capabilities and resources for destruction. A particularly remarkable finding is that when a group's attacks are timed in a less random manner, their lethality is significantly higher than their random counterparts, and this lethality grows as their execution becomes more uniform. The model's group-level estimates from attack data situate it between theoretical work on the sociopolitical determinants of terror and case-centered work on alliances and ideologies. The relative ranking of group "strength" that $Q_5$ and $T_Z$ provide is also a complement to the Uppsala work that focuses on ranking groups' strengths relative to their opposing state actor (55, 56). Nevertheless, before strong conclusions are drawn, future research should study the link between $Q$, $T_Z$, and hidden capabilities and resources in greater detail. In particular, more investigation is needed on how organizational and contextual factors, such as the strength of opposing security forces, or locale-based factors, such as government oppression or the media, motivate terrorist activity and sympathies in way that systematically correlate with $Q$ and $T_Z$.

From a security and policy perspective, we see 3 major approaches to preventing and minimizing the harms of terrorism: hardening targets, emergency preparedness, and proactive targeting. The first approach focuses on improving physical or virtual barriers that deter would-be attacks through, for example, border patrol, surveillance, or dark web analytics (57). The second approach attempts to contain the carnage and havoc that arises in an attack's aftermath by better coordinating emergency response personnel and institutions (58). Our method contributes to the third approach. Whereas the first 2 approaches are broadly defensive, the third approach is proactive. Our model improves the ability of analysts and counterterrorism efforts to anticipate and incapacitate those terror groups by guiding scarce resources to those groups most likely to be most destructive. Moreover, because our model has distinctive early-warning benefits, which allow highly lethal groups to be proactively disarmed early in their lifespan, it minimizes the unintended consequences of a proactive policy. For example, a downside of a proactive policy is that preemptive counterterrorism against mature and visible terror groups can unintentionally increase grievances and recruitment (59).

A next step in research is to uncover specific types of capabilities and resources, their links to lethality, and how capabilities can be destabilized and resource transfers disrupted (60). Similarly, our focus on predicting the total future lethality of a group highlights potential next steps in research. How lethal the next attack will be is an important theoretical and practical question. Similarly, the needs to accurately predict the future lethality of a group on a short time scale or on the actions of lone wolves remain problems on the frontier (6). Nevertheless, due to the general scientific methodology used to derive it, the $Q_5\,T_Z$ model appears to have the potential to address generic organizational behavior questions not just of illicit organizations but also legal organizations, such as start-ups or privately held firms where public data are lacking. For example, generalizing the model to types of illegitimate activity, such as organized crime or gangs, can aid efforts to improve societal health and safety and lower security costs that drain resources from other productive activities. Finally, we take note that one might argue that these results are self-defeating because they provide terror groups with new information deception. However, as shown elsewhere (6), insurgents that attempt to alter the appearances of their capabilities and resources are likely to unwittingly make organizational changes that undermine other aspects of their behavior (33, 61, 62).

## Materials and Methods

**Data.** We use data from 3 sources: 1) the GTD of 2014, 2) the updated GTD of 2017, and 3) the RDWTI database. We use the 2014 GTD primarily for model building and the 2017 GTD and RDWTI data for model verification. The GTD is a database on terrorist events around the world from 1970 through 2014, which includes more than 140,000 incidents. A detailed summary of the data and the source data can be found in ref. 11. The RDWTI database covers the time period from 1968 through 2009 (63). It records data similar to the GTD with around 40,000 incidents.

**ACKNOWLEDGMENTS.** Funding for this research has been generously given by the Northwestern University Data Science Initiative and Northwestern Institute for Complex Systems, US Army Research Laboratory and US Army Research Office Grant W911NF-15-1-0577, and Army Research Laboratory Network Science CTA under Cooperative Agreement W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the social policies, either expressed or implied, of the US government.

1. A. Belasco, *Cost of Iraq, Afghanistan, and Other Global War on Terror Operations Since 9/11* (CRS Report for Congress, Congressional Research Service, Washington, DC, 2009). pp. 1–100.
2. M. R. Powers, Z. Shen, Colonel blotto in the war on terror: Implications for event frequency. *J. Homel. Secur. Emerg. Manag.* **6**, 1435 (2009).
3. Pew Research Center, After seismic political shift, modest changes in public's policy agenda. http://www.people-press.org/2017/01/24/after-seismic-political-shift-modest-changes-in-publics-policy-agenda/ (2017). Accessed 12 January 2018.
4. A. Abadie, J. Gardeazabal, Terrorism and the world economy. *Eur. Econ. Rev.* **52**, 1–27 (2008).
5. N. Johnson *et al.*, Pattern in escalations in insurgent and terrorist activity. *Science* **333**, 81–84 (2011).
6. N. F. Johnson *et al.*, New online ecology of adversarial aggregates: Isis and beyond. *Science* **352**, 1459–1463 (2016).
7. A. Clauset, K. S. Gleditsch, The developmental dynamics of terrorist organizations. *PLoS One* **7**, e48633 (2012).
8. A. Clauset, R. Woodard, Estimating the historical and future probabilities of large terrorist events. *Ann. Appl. Stat.* **7**, 1838–1865 (2013).
9. A. Clauset, M. Young, K. S. Gleditsch, On the frequency of severe terrorist events. *J. Confl. Resolut.* **51**, 58–87 (2007).
10. D. Helbing *et al.*, Saving human lives: What complexity science and information systems can contribute. *J. Stat. Phys.* **158**, 735–781 (2015).
11. NationalConsortium for the Study of Terrorism and Responses to Terrorism (START), Global Terrorism Database [Data file]. https://www.start.umd.edu/gtd (2017). Accessed 12 January 2018.
12. R. Powell, Defending against terrorist attacks with limited resources. *Am. Pol. Sci. Rev.* **101**, 527–541 (2007).
13. J. C. Bohorquez, S. Gourley, A. R. Dixon, M. Spagat, N. F. Johnson, Common ecology quantifies human insurgency. *Nature* **462**, 911–914 (2009).
14. K. Sonin, J. Wilson, A. L. Wright, Rebel capacity, intelligence gathering, and the timing of combat operations (Centre for Economic Policy Research Discussion Paper DP13155, Centre for Economic Policy Research, London, UK, 2018).

15. K. Hausken *et al.*, "Defending against terrorism, natural disaster, and all hazards" in *Game Theoretic Risk Analysis of Security Threats*, V. M. Bier, M. N. Azaiez, Eds. (Springer Science & Business Media, 2008), vol. 128, pp. 65–97.
16. M. N. Azaiez, V. M. Bier, Optimal resource allocation for security in reliability systems. *Eur. J. Oper. Res.* **181**, 773–786 (2007).
17. V. Bier, S. Oliveros, L. Samuelson, Choosing what to protect: Strategic defensive allocation against an unknown attacker. *J. Public Econ. Theory* **9**, 563–587 (2007).
18. G. LaFree, L. Dugan, E. Miller, *Putting Terrorism in Context: Lessons from the Global Terrorism Database* (Routledge, New York, NY, 2014).
19. J. A. Updegraff, R. C. Silver, E. A. Holman, Searching for and finding meaning in collective trauma: Results from a national longitudinal study of the 9/11 terrorist attacks. *J. Personal. Soc. Psychol.* **95**, 709–722 (2008).
20. W. Enders, G. A. Hoover, T. Sandler, The changing nonlinear relationship between income and terrorism. *J. Confl. Resolut.* **60**, 195–225 (2016).
21. R. C. Oka *et al.*, Population is the main driver of war group size and conflict casualties. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E11101–E11110 (2017).
22. V. Asal, R. K. Rethemeyer, The nature of the beast: Organizational structures and the lethality of terrorist attacks. *J. Politics* **70**, 437–449 (2008).
23. J. A. Piazza, Rooted in poverty?: Terrorism, poor economic development, and social cleavages. *Terror. Political Violence* **18**, 159–177 (2006).
24. V. Asal, R. K. Rethemeyer, Researching terrorist networks. *J. Secur. Educ.* **1**, 65–74 (2006).
25. V. Asal, P. Harwood, Search engines: terrorism's killer app. *Stud. Confl. Terror.* **31**, 641–654 (2008).
26. Q. Li, Does democracy promote or reduce transnational terrorist incidents?*J. Confl. Resolut.* **49**, 278–297 (2005).
27. M. H.-R. Hicks *et al.*, The weapons that kill civilians—deaths of children and noncombatants in Iraq, 2003–2008. *N. Engl. J. Med.* **360**, 1585–1588 (2009).
28. V. Asal, P. Gill, R. K. Rethemeyer, J. Horgan, Killing range: Explaining lethality variance within a terrorist organization. *J. Confl. Resolut.* **59**, 401–427 (2015).
29. W. H. Press, Strong profiling is not mathematically optimal for discovering rare malfeasors. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1716–1719 (2009).
30. M. A. Peteraf, The cornerstones of competitive advantage: A resource-based view. *Strateg. Manag. J.* **14**, 179–191 (1993).
31. S. L. Newbert, Empirical research on the resource-based view of the firm: An assessment and suggestions for future research. *Strateg. Manag. J.* **28**, 121–146 (2007).
32. P. B. Overgaard. The scale of terrorist attacks as a signal of resources. *J. Confl. Resolut.* **38**, 452–478 (1994).
33. B. Wernerfelt, A resource-based view of the firm. *Strateg. Manag. J.* **5**, 171–180 (1984).
34. D. E. Long, "Understanding terrorist behavior" in *The Anatomy of Terrorism* (Free Press, New York, NY, 1990), pp. 30–56.
35. J. N. Shapiro, "The terrorist's dilemma" in *Managing Violent Covert Organizations* (Princeton University Press, Princeton, NJ, 2013), pp. 26–61.
36. S. Krishnamurthy, *Grant Wardlaw Political Terrorism: Theory, Tactics and Countermeasures* (Cambridge University Press, 1982), pp. xii, 218.
37. L. A. N. Amaral *et al.*, Econophysics: Can statistical physics contribute to the science of economics?*Comput. Phys. Commun.* **121**, 145–152 (1999).
38. R. Sinatra, D. Wang, P. Deville, C. Song, A.-L. Barabási, Quantifying the evolution of individual scientific impact. *Science* **354**, aaf5239 (2016).
39. S. Aral, P. Weill, It assets, organizational capabilities, and firm performance: How resource allocations and organizational differences explain performance variation. *Organ. Sci.* **18**, 763–780 (2007).
40. R. G. Schroeder, K. A. Bates, M. A. Junttila, A resource-based view of manufacturing strategy and the relationship to manufacturing performance. *Strateg. Manag. J.* **23**, 105–117 (2002).
41. R. D. Banker, I. R. Bardhan, H. Chang, S. Lin, Plant information systems, manufacturing capabilities, and plant performance. *Manage. Inf. Syst. Q.* **30**, 315–337 (2006).
42. J. Qin, Y. Zhou, E. Reid, G. Lai, H. Chen, Analyzing terror campaigns on the internet: Technical sophistication, content richness, and web interactivity. *Int. J. Hum. Comput. Stud.* **65**, 71–84 (2007).
43. S. Baez *et al.*, Outcome-oriented moral evaluation in terrorists. *Nat. Hum. Behav.* **1**, 118 (2017).
44. D. Wang, C. Song, A.-L. Barabási, Quantifying long-term scientific impact. *Science* **342**, 127–132 (2013).
45. A. E. Raftery, Bayesian model selection in social research. *Sociol. Methodol.* **25**, 111–163 (1995).
46. R. Liscouski, W. McGann, The evolving challenges for explosive detection in the aviation sector and beyond. *CTC Sentinel.* **9**(5), 1–6 (2016).
47. K. R. Conner, C. K. Prahalad, A resource-based theory of the firm: Knowledge versus opportunism. *Organ. Sci.* **7**, 477–501 (1996).
48. M. Spence, Signaling in retrospect and the informational structure of markets. *Am. Econ. Rev.* **92**, 434–459 (2002).
49. R. G. Cooper, "Profitable product innovation: The critical success factors" in *The International Handbook on Innovation*, L.V. Shavinina, Ed. (Elsevier, Amsterdam, The Netherlands, 2003), pp. 139–157.
50. S. Krishnan, A. Pedahzur, B. Jenkins, "Suicide terrorism-opportunistic tactic or strategic campaign?" (Western Political Science Association 2011 Annual Meeting Paper, Western Political Science Association, Sacramento, CA, 2011; https://ssrn.com/abstract=1766792). Accessed 5 January 2019.
51. J. Wolfendale, Terrorism, security, and the threat of counterterrorism. *Stud. Confl. Terror.* **30**, 75–92 (2007).
52. K.-I. Goh, A.-L. Barabási, Burstiness and memory in complex systems. *Europhys. Lett.* **81**, 48002 (2008).
53. FinancialAction Task Force, Terrorist financing (Financial Action Task Force, Paris, France, 2008). http://www.fatf-gafi.org/publications/methodsandtrends/documents/fatfterroristfinancingtypologiesreport.html. Accessed 5 January 2019.
54. J. Orbe, E. Ferreira, V. Núñez-Antón, Censored partial regression. *Biostatistics* **4**, 109–121 (2003).
55. N. P. Gleditsch, P. Wallensteen, M. Eriksson, M. Sollenberg, H. Strand, Armed conflict 1946-2001: A new dataset. *J. Peace Res.* **39**, 615–637 (2002).
56. L. Themnér, P. Wallensteen, Armed conflicts, 1946–2011. *J. Peace Res.* **49**, 565–575 (2012).
57. T. Sandler, K. Siqueira, Global terrorism: Deterrence versus pre-emption. *Can. J. Econ./Revue canadienne d'économique* **39**, 1370–1387 (2006).
58. A. S. Khan, S. Morse, S. Lillibridge, Public-health preparedness for biological terrorism in the USA. *Lancet* **356**, 1179–1182 (2000).
59. B. P. Rosendorff, T. Sandler, Too much of a good thing? The proactive response dilemma. *J. Confl. Resolut.* **48**, 657–671 (2004).
60. L. Glowacki *et al.*, Formation of raiding parties for intergroup violence is mediated by social network structure. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12114–12119 (2016).
61. L. A. N. Amaral, B. Uzzi, Complex systems—A new paradigm for the integrative study of management, physical, and technological systems. *Manage Sci* **53**, 1033–1035 (2007).
62. B. Uzzi, R. Lancaster, The role of relationships in interfirm knowledge transfer and learning. *Manage Sci* **49**, 383–399 (2003).
63. RAND Corporation, RAND Database of Worldwide Terrorism Incidents. http://www.rand.org/nsrd/projects/terrorism-incidents.html (2009). Accessed 19 December 2017.