Published in ***CBRNe World*** magazine, April 2016, pp. 30 - 37,
http://www.cbrneworld.com/
This is the final pre-publication manuscript. Author affiliations are on p. 6 below.

From Factors to Actors: Enhancing the Reliability of CBRN Analysis

by Ronald Breiger, Lauren Pinson, and Gary Ackerman

Considerable debate in recent years among researchers and policymakers has focused on the growing potential of terrorist groups to acquire and use chemical, biological, radiological, and nuclear (CBRN) materials in attacks. The reliability of analyses of CBRN events is an important concern at two levels: the data on which analyses are based, and the quantitative modeling of those data. At each of these levels, recently developed procedures that we highlight in this article allow for increased nuance in threat assessment and aid in the identification of multiple threat scenarios.[1]

Data on CBRN terrorism is often misrepresented by the media and by public commentators. Limited firsthand information, conflicting reports, a tendency to count as CBRN terrorism events that upon inspection do not fit that category, and varying sympathies of event reporters all contribute to a lack of reliability in events data. We will present a significant new project under way to produce a high-quality database on CBRN events, undertaken by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) located at the University of Maryland.

However we do not believe the problem of unreliability in CBRN analysis ends with concerns about the data. Current state-of-the art research on potential adversary intent to acquire or use CBRN weapons, including but not limited to research that we and our colleagues have conducted, has been formulated as linear analysis using multiple regression models.[2] Coefficients in such models, each of which reports the effect of a predictor variable (such as the perpetrator being a lone actor) on an outcome variable (such as the possession of a CBRN weapon), are conventionally interpreted as providing a single estimate of the effect of the predictor variable on the outcome, while controlling for the other variables in the model. By way of contrast, a new stream of quantitative methods for the study of CBRN activities, developed by researchers at the University of Arizona and Ohio State University, understands each regression coefficient to be a sum across the cases (which are, in this context, the events being analyzed).[3] With respect to any linear model coefficient of substantive interest, there may be entire subspaces of cases within the sample in which the effect of the predictor variable on the outcome is null. Even worse, there may be subspaces of cases in which the effect is in the opposite direction from what was predicted. New models have thus been developed to decompose regression coefficients according to cases and clusters of cases. Thus, the novel modeling approach that we illustrate in this article might lead to the conclusion that particular organizations, or types of organizations, are largely responsible for the magnitude or significance of the regression coefficient observed, or that a predictor variable positively affects the likelihood of CBRN terrorism under some specifiable conditions, while negatively affecting it

under others. In brief, we advocate using the variables in these models to learn about the cases, a concern that is the other side of the coin to the usual intent of linear model practitioners. The new approach enhances analysts' ability to identify multiple analytical streams within a single set of data and to state the conditions under which each is likely.

The capability to move beyond single or unique interpretations, not only of CBRN data, but also of the coefficients in linear modeling of those data, is important both to government analysts and to academic researchers. Thomas Fingar, who served as the first deputy director of national intelligence for analysis and, concurrently, as chairman of the National Intelligence Council, has written that "if weighing the information differently or using different assumptions to close gaps leads logically to different conclusions, [then] that fact must be revealed to policy makers."[4] Likewise, John Padgett, a keen academic student of historical events, urges that "objectivity means the triangulated search through always biased sources for the perspective of a stable focal point."[5] We now illustrate how our concern about the reliability of CBRN data and CBRN models can improve upon some aspects of current practice.

POICN: A Database for CBRN Events involving Non-State Actors

The POICN database—Profiles of Incidents involving CBRN agents by Non-state actors—is a relational, open-source database including information on international and domestic terrorist plots, acquisitions, and attacks during the period 1990 to 2015 that involve CBRN agents. It is the most comprehensive open-source, unclassified dataset of CBRN incidents perpetrated by ideologically motivated non-state actors. The POICN database is currently comprised of 534 discrete cases coded across 162 event specific variables, 11 perpetrator variables, and 29 variables describing aspects of the agent acquisition process. The categories of events recorded in POICN include proto-plots,[6] plots, attempted acquisitions, and actual or attempted use of an agent.

In developing the POICN database, START researchers rejected the widely applied but often mistaken assumption that a case's inclusion in a dataset automatically equates to full validation of that case. A WMD data source evaluation schema (WMD-SES) was developed, consisting of a set of operationalized variables and coding instructions.[7] Multiple variables focus on capturing the intentional and accidental distortion of information regarding CBRN activities. Key metrics were developed to assess the degree of source objectivity and competence, event uncertainty and reliability, attack uncertainty, and the degree of doubt and discrepancy pertaining to many of the CBRN-specific variables in the dataset.

For each event, each source was coded for its *competence*. "Full" competence was reserved for documents with no evident internal flaws produced by authors and institutions that have proven or researched competence in the geographical and substantive domain on which they are reporting, taking into account all that is known about their history and reputation as sources. Likewise, each source was coded for its *objectivity* with respect to each event. A

newspaper that is generally measured in its approach to reporting, but, for example, is known on occasion to take a very pro-Israeli or pro-Palestinian stance on the Israeli-Palestinian conflict, is coded as evidencing only "potential" objectivity.

Each case (event) in the database was also coded for its *credibility*, providing a measure of whether the event actually took place and whether that event really constituted a CBRN event. This measure builds on those for competence and objectivity (described above) in that sources coded as not competent or not objective were excluded from the credibility measure, which evaluates the number of remaining sources for each event with reference to the extent of their mutual independence. A source is regarded as independent of another source if it does not share the same original authorship and does not rely on the same original source material. Furthermore, the lowest level of credibility is assigned to single-source events or events described by multiple, non-independent sources. The intermediate level is assigned to events reported by two independent sources not reflecting the same bias. The highest level (encompassing just over half the events in the database) is allocated to events for which there were three or more independent sources, or two independent sources with competing biases.

As illustrated above, the POICN database and the WMD source evaluation schema on which it is constructed allow the analyst to explicitly incorporate and disclose credibility and validity levels for greater flexibility in tailoring the inclusion of cases for specific analytical requirements. By facilitating more transparent analyses, the inclusion of such measures in this as well as similar datasets can result in more defensible conclusions, especially in highly-charged political and security contexts such as those surrounding terrorism and CBRN analysis.

From Factors to Actors: Modeling Heterogeneity in CBRN Events

CBRN activity is not a homogeneous phenomenon, and therefore analysts require methods to identify relatively homogeneous subsets of cases. An attempted suicide bombing of large chemical tanks at an Israeli port in 2004, the roadside burying of mustard gas containers in Grozny, Chechnya, to thwart the advance of Russian troops in 1999, and private inquiries made to a Bulgarian businessman about the possibility of acquiring spent nuclear fuel rods in 2001, are examples of reported events (taken from the POICN database) that are difficult to treat as observations sampled from a homogeneous population. Furthermore, CBRN analysts are concerned with events related by complex interdependencies—temporal, spatial, and social—that must often violate the independence of sampled observations that regression modeling assumes. For effective CBRN analysis, the analytical goal should be to let the cases be seen.

In a recently published analysis making use of the POICN database for the period 1998 – 2011, we applied the dual-to-regression modeling framework described above to shift the emphasis from variables (factors) to actors, using the variables to learn about the cases.[8] In this overview we emphasize two aspects: a profile similarity network among the cases, and the identification of multiple threat scenarios by decomposing a single regression coefficient.

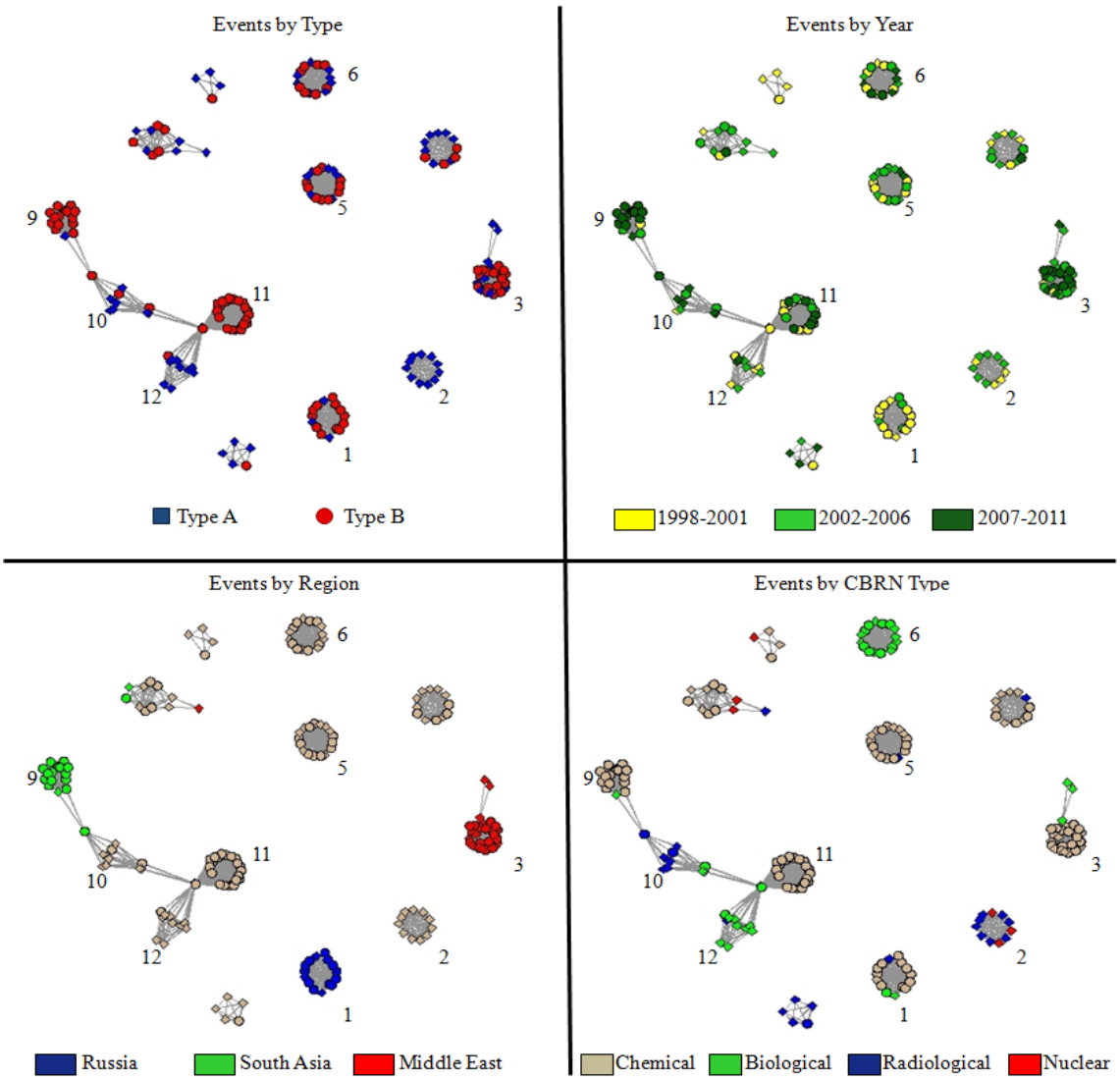

Fig. 1: A profile similarity network of CBRN events

Figure 1 depicts a network among the 175 high-credibility events from the POICN database that we analyzed. The network is reflects where there is a high degree of similarity among pairs of cases across the nine predictor variables in the conventional logistic regression model that we formulated to predict possession, rather than pursuit, of a CBRN weapon. The numerical values in the matrix of which Figure 1 is a depiction can be used to generate the predicted values of the outcome variable (i.e., what predicts possession by a non-state actor of CBRN), which is why this network of cases is "dual" to the conventional regression model which studies relations among the variables. Here we emphasize how to identify relatively homogeneous subsets of CBRN events.

The same numbering scheme is used to identify clusters in each panel of Figure 1. The upper-left panel shows that, even though only predictor variables were used in constructing this

network, the resulting clusters of cases tended to consist predominantly of cases with the same outcome: either pursuit of CBRN weapons (e.g., cluster 2, Type A, colored blue) or possession (e.g., cluster 1, Type B, colored red). Looking at cluster 1 across the four panels of the Figure, it is seen that this cluster consisted predominantly of events transpiring in Russia and the post-Soviet (NIS) countries in the early period of our data (1998 – 2001) involving possession of chemical agents. By way of contrast, cluster 2 consists predominantly of post-9/11 events (2002-2006) involving the pursuit of radiological and (to a lesser extent) nuclear materials in regions outside of Russia, the Middle East, and South Asia. (Further examination shows many of the cluster 2 events to have taken place in Europe and to have a connection to al-Qa'ida.) We argue that identification of such distinctive clusters as these two is an important move beyond conventional regression analyses that assume that each of the cases is a randomly selected representative of the same population.

We will illustrate important differences between one other pair of clusters. Both clusters 6 and 12 are composed predominantly of events involving possession of biological agents (see the upper-left and lower-right panels of Figure 1). What then distinguishes cases with respect to these clusters? Further examination indicates that all cases in cluster 6 involve lone perpetrators in the United States, and biological agents only, whereas all cases in cluster 12 involve chemical as well as biological materials and involve perpetrators with a religious extremist ideology, none of whom are lone perpetrators.[9] This contrast in clusters provides further support for our effort to move regression analysis beyond the study of variables ("factors") to allow focus on the actors and other features of CBRN events. We have illustrated how a multifaceted, compound interaction of variables can be clarified within our framework.

For the conventional linear modeling of these 175 events, we found a negative regression coefficient for the effect of religious extremist ideology on CBRN possession. The conventional interpretation would be that religious extremist ideology is negatively associated with possession (as distinct from pursuit) of CBRN agents. However, our modeling framework allows us to decompose this coefficient as a sum across clusters. Within cluster 9, the contribution to the overall coefficient is positive. Why is this? From Figure 1 we see that cluster 9 events take place in South Asia, predominantly in the most recent period (2007 – 2011), and feature possession of chemical agents. Additional examination indicates that a typical cluster 9 case is this one: "In November 2008, two men on motorcycles sprayed acid from water pistols on a group of female students on their way to a high school for young women on the outskirts of Kandahar, Afghanistan; four of the young women received serious burns, and two of them were blinded. Afghan government accounts state that the perpetrators confessed that a senior Taliban official was offering a sum of money for every injured girl." We hypothesize that monitoring of religious extremism by the authorities in that time and place was relatively weak, and therefore religious extremists were more free to use acid in the commission of heinous crimes such as this one. We speculate that this could be a reason why the effect of a predictor variable (religious extremism) on the outcome (possession of a CBRN agent, as distinct from pursuit of one) can differ within a

sample in ways that analysts should be able to uncover. Our modeling framework provides a counterpart to the conventional linear models within which cases are usually analyzed.

Conclusion

The various uncertainties and complexities inherent in understanding CBRN terrorism call for a fresh approach both to the production of databases and the formulation of quantitative methods that allow for more nuanced description and analysis than those that have often served as the basis for research and policy. We have highlighted our recent efforts aimed at furthering a more nuanced, and hence a more accurate and productive, understanding of the CBRN activities of non-state actors.

Ronald Breiger is Professor of Sociology at the University of Arizona and Editor for Social and Political Science of the journal *Network Science*.

Lauren Pinson is a National Science Foundation Graduate Research Fellow and PhD student in political science at Yale University.

Gary Ackerman is Director of the Unconventional Weapons and Technology Division at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), a Department of Homeland Security Center of Excellence led by the University of Maryland.

Notes

[1] This research was supported by grant HDTA1-10-1-0017 from the Defense Threat Reduction Agency. The views expressed in this article are our own, and do not represent those of DTRA or the U.S. Government. We thank our colleagues Victor Asal, H. Brinton Milward, R. Karl Rethemeyer, David Melamed, and Eric Schoon.

[2] Victor Asal and R. Karl Rethemeyer, "Islamist Use and Pursuit of CBRN Terrorism" in Gary Ackerman and Jeremy Tamsett (eds.), *Jihadists and Weapons of Mass Destruction* (Boca Raton, FL: CRC Press, 2009); Victor Asal et al., "CBRN Activity and Attacks by Insurgent Organizations," *START Research Brief*, College Park, MD (October 2015), https://www.start.umd.edu/pubs/START_BAAD_CBRN_ResearchBrief_Oct2015.pdf

[3] Ronald Breiger et al., "Application of a Profile Similarity Methodology for Identifying Terrorist Groups that use Or Pursue CBRN Weapons" in John Salerno et al. (eds.), *Social Computing, Behavioral-Cultural Modeling and Prediction* (Berlin: Springer, 2011), doi: 10.1007/978-3-642-19656-0_5; David Melamed et al., ""Using Organizational Similarity to Identify Statistical Interactions for Improving Situational Awareness of CBRN Activities" in Shanchieh Jay Yang et al. (eds.), *Social Computing, Behavioral-Cultural Modeling and Prediction* (Berlin: Springer, 2012), doi: 10.1007/978-3-642-29047-3_8; Ronald Breiger and David Melamed, "The Duality of Organizations and their Attributes: Turning Regression Modeling 'Inside Out'." *Research in the Sociology of Organizations* 40 (2014), doi: 10.1108/S0733-558X(2014)0000040013.

[4] Thomas Fingar, *Reducing Uncertainty: Intelligence Analysis and National Security* (Palo Alto, CA: Stanford University Press, 2011), p. 110.

[5] John F. Padgett, Triangulating on Causal Process," *American Politics Research* 39 (2011), p. 466, doi: 10.1177/1532673X10396310

[6] An event is labeled a proto-plot when the available sources do not present any evidence of an actual plot but rather mention events that may lay the groundwork of an actual plot (such as the discovery of a chemical weapons manual or knowledge of a terrorist group hiring a scientist with a CBRN relevant specialty).

[7] Gary Ackerman and Lauren Pinson, "Speaking Truth to Sources: Introducing a Method for the Quantitative Evaluation of Open Sources in Event Data," *Studies in Conflict & Terrorism*, awaiting publication.

[8] Ronald Breiger and Lauren Pinson, "A New Approach for the Identification of Multiple Threat Scenarios to Counter CBRN Networks," in Luke M. Gerdes (ed.), *Illuminating Dark Networks: The Study of Clandestine Groups and Organizations* (New York: Cambridge University Press, 2015).

[9] Breiger and Pinson (2015) coded the event as biological if it was also chemical; as radiological if it was also biological or chemical, and as nuclear no matter what other materials might be involved.