

Springer Series on Evidence-Based Crime Policy

Cynthia Lum
Leslie W. Kennedy *Editors*

Evidence-Based Counterterrorism Policy

 Springer

Springer Series on Evidence-Based Crime Policy

Series Editors:

Lawrence W. Sherman, University of Pennsylvania

Heather Strang, Australian National University

Crime prevention and criminal justice policies are domains of great and growing importance around the world. Despite the rigorous research done in this field, policy decisions are often based more on ideology or speculation than on science. One reason for this may be a lack of comprehensive presentations of the key research affecting policy deliberations. While scientific studies of crime prevention and criminal policy have become more numerous in recent years, they remain widely scattered across a wide range of journals and countries. The *Springer Series on Evidence-Based Crime Policy* aims to pull this evidence together while presenting new research results. This combination in each book should provide, between two covers (or in electronic searches), the best evidence on each topic of crime policy.

The series will publish primary research on crime policies and criminal justice practices, raising critical questions or providing guidance to policy change. The series will try to make it easier for research findings to become key components in decisions about crime and justice policy. The editors welcome proposals for both monographs and edited volumes. There will be a special emphasis on studies using rigorous methods (especially field experiments) to assess crime prevention interventions in areas such as policing, corrections, juvenile justice, and crime prevention.

Published in Cooperation with the Campbell Crime and Justice Group

For further volumes:

<http://www.springer.com/series/8396>

Cynthia Lum • Leslie W. Kennedy
Editors

Evidence-Based Counterterrorism Policy

 Springer

Editors:

Cynthia Lum
Center for Evidence-Based Crime Policy
Department of Criminology, Law & Society
George Mason University
Fairfax, VA, USA
clum@gmu.edu

Leslie W. Kennedy
Rutgers School of Criminal Justice
Rutgers University
Newark, NJ, USA
kennedy@newark.rutgers.edu

ISBN 978-1-4614-0952-6 e-ISBN 978-1-4614-0953-3
DOI 10.1007/978-1-4614-0953-3
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011936394

© Springer Science+Business Media, LLC 2012

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Part I Introduction

- 1 Evidence-Based Counterterrorism Policy** 3
Cynthia Lum and Leslie W. Kennedy

Part II Data Sources for Evaluating Terrorism and Counterterrorism

- 2 Assessing and Comparing Data Sources for Terrorism Research** 13
Ivan Sascha Sheehan
- 3 Generating Terrorism Event Databases: Results from the Global Terrorism Database, 1970 to 2008** 41
Gary LaFree
- 4 Evidence-Based Intelligence Practices: Examining the Role of Fusion Centers as a Critical Source of Information** 65
Jeremy Carter and Steven Chermak

Part III Methodological Innovations for Counterterrorism Policy

- 5 Innovative Methods for Terrorism and Counterterrorism Data** 91
Michael D. Porter, Gentry White, and Lorraine Mazerolle
- 6 Introducing Group-Based Trajectory Analysis and Series Hazard Modeling: Two Innovative Methods to Systematically Examine Terrorism Over Time**..... 113
Laura Dugan and Sue-Ming Yang
- 7 A Complexity Method for Assessing Counterterrorism Policies** 151
Claudio Cioffi-Revilla

8 Analyzing Terrorism Using Spatial Analysis Techniques: A Case Study of Turkish Cities	167
Danielle M. Rusnak, Leslie W. Kennedy, Ibrahim S. Eldivan, and Joel M. Caplan	
9 The Importance of Instrument Validity in Evaluating Security Screening Programs	187
Tracy E. Costigan	
10 Translational Criminology: Using Existing Evidence for Assessing TSA’s Comprehensive Security Strategy at Airports	209
Cynthia Lum, Charlotte Gill, Breanne Cave, Julie Hibdon, and David Weisburd	
Part IV Perspectives in Evaluating Counterterrorism Policy	
11 Terrorist Finance, Informal Markets, Trade and Regulation: Challenges of Evidence Regarding International Efforts	255
Nikos Passas	
12 Evaluating the Legal Challenges and Effects of Counterterrorism Policy	281
Linda M. Merola	
13 Public Opinion Research and Evidence-Based Counterinsurgency	301
Clay Ramsay	
14 Counterinsurgency and Criminology: Applying Routine Activities Theory to Military Approaches to Counterterrorism	325
Breanne Cave	
15 Toughness vs. Fairness: Police Policies and Practices for Managing the Risk of Terrorism	353
Tom R. Tyler	
Part V Epilogue	
16 The Next Steps: A Need for a Research Infrastructure for Evaluating Counterterrorism	367
Cynthia Lum and Leslie W. Kennedy	
Biographies	377
Index	381

Contributors

Joel M. Caplan, Ph.D School of Criminal Justice, Rutgers University,
Newark, NJ, USA

Rutgers Center on Public Security, Rutgers University, Newark, NJ, USA

Jeremy Carter, Ph.D Department of Criminology and Criminal Justice,
University of North Florida, Jacksonville, FL, USA

Breanne Cave, M.A Center for Evidence-Based Crime Policy, Department of
Criminology, Law & Society, George Mason University, Fairfax, VA, USA

Steven Chermak, Ph.D School of Criminal Justice, Michigan State University,
East Lansing, MI, USA

Claudio Cioffi-Revilla, D.Sc.Pol., Ph.D Center for Social Complexity
and Department of Computational Social Science, Krasnow Institute
for Advanced Study, George Mason University, Fairfax, VA, USA

Tracy E. Costigan, Ph.D American Institutes for Research (AIR),
Washington, DC, USA

Laura Dugan, Ph.D Department of Criminology and Criminal Justice,
University of Maryland, College Park, MD, USA

Ibrahim S. Eldivan School of Criminal Justice, Rutgers University,
Newark, NJ, USA

Rutgers Center on Public Security, Rutgers University, Newark, NJ, USA

Charlotte Gill, Ph.D Center for Evidence-Based Crime Policy, Department of
Criminology, Law & Society, George Mason University, Fairfax, VA, USA

Julie Hibdon, Ph.D Center for Evidence-Based Crime Policy, Department of
Criminology, Law & Society, George Mason University, Fairfax, VA, USA

Leslie W. Kennedy, Ph.D School of Criminal Justice and the Center on Public Security, Rutgers University, Newark, NJ, USA

Gary LaFree, Ph.D National Center for the Study of Terrorism and Responses to Terrorism (START), and the Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Cynthia Lum, Ph.D Center for Evidence-Based Crime Policy, Department of Criminology, Law & Society, George Mason University, Fairfax, VA, USA

Lorraine Mazerolle, Ph.D ARC-Centre of Excellence in Policing and Security, Institute for Social Science Research, The University of Queensland, Brisbane, Queensland, Australia

Linda M. Merola, J.D., Ph.D Department of Criminology, Law & Society, George Mason University, Fairfax, VA, USA

Nikos Passas, Ph.D School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Michael D. Porter, Ph.D GeoEye Analytics, McLean, VA, USA

Clay Ramsay, Ph.D Program on International Policy Attitudes (PIPA) and the Center for International and Security Studies (CISSM), University of Maryland, College Park, MD, USA

Danielle M. Rusnak, M.A School of Criminal Justice, Rutgers University, Newark, NJ, USA

Ivan Sascha Sheehan, Ph.D Negotiation and Conflict Management, School of Public and International Affairs, University of Baltimore, Baltimore, MD, USA

Tom R. Tyler, Ph.D Department of Psychology, New York University, New York, NY, USA

David Weisburd, Ph.D Center for Evidence-Based Crime Policy, Department of Criminology, Law & Society, George Mason University, Fairfax, VA, USA

Gentry White, Ph.D ARC-Centre of Excellence in Policing and Security, Institute for Social Science Research, The University of Queensland, Brisbane, Queensland, Australia

Sue-Ming Yang, Ph.D Institute of Criminology, National Chung Cheng University, Chia-Yi, Taiwan

Part I

Introduction

Chapter 1

Evidence-Based Counterterrorism Policy

Cynthia Lum and Leslie W. Kennedy

As we approach the tenth anniversary of the terror attacks on September 11th, one fact is clear among the many unknowns about terrorism: there has been an exponential increase in spending on counterterrorism measures. For the United States, this increase in spending is not just reflected in federal homeland security measures or military efforts in Iraq, Afghanistan, and Pakistan, but also among private individuals, corporations, and public entities at the municipal, county, regional, and state levels. Most recently, the killing of Osama Bin Laden by the United States has led to a renewed increase in resource allocation at home by local law enforcement agencies (van Natta, 2011). In the scientific fields, there has also been more funding for a variety of academic and technological research and development related to terrorism and counterterrorism.

An addendum to this fact: This massive increase in capacity building has not been matched by evaluation and assessment regarding the cost-effectiveness of those expenditures. The many efforts to detect, prevent, deter, and reduce the risk of terrorism-related violence at home and aboard have rarely been scientifically evaluated using rigorous methods. Lum, Kennedy, and Sherley (2006) discovered this state of the research when conducting a Campbell Collaboration systematic review of counterterrorism research. The Campbell review sought to determine “what we know” about the types of counterterrorism strategies and tactics that can reduce the likelihood of, or damage from, terrorism events. However, after reviewing thousands of research articles on terrorism and counterterrorism, the comprehensive search only produced *seven* evaluation studies of at least moderate scientific rigor, primarily advanced by a few scholars (see, e.g., Enders & Sandler, 1993, 2000,

C. Lum (✉)

Center for Evidence-Based Crime Policy, Department of Criminology,
Law & Society, George Mason University, Fairfax, VA, USA
e-mail: clum@gmu.edu

L.W. Kennedy

School of Criminal Justice, Rutgers University, Newark, NJ, USA

2006; Enders, Sandler, & Cauley, 1990). Most of the analysis focused on the use of metal detectors at airports to prevent hijacking. Some research also suggested that interventions (e.g., military ones) could increase terrorism in the short run, while in the long run have little significant effect on levels of terroristic violence.

Lum and colleagues also emphasized that the only way to determine the cost-effectiveness of counterterrorism interventions, as with crime prevention more generally, was through scientifically rigorous evaluations. Yet evaluation research on counterterrorism interventions in the 10 years since September 11th (and 5 years after the Lum et al. review) remains dismal compared to crime prevention research or even other areas of terrorism studies (i.e., etiological or case studies on terrorism or groups). Lum recently reported to the National Academies that evaluations of police interventions outnumber those on security and counterterrorism more than tenfold (National Research Council, 2010). Only recently has the Transportation Security Agency (TSA) requested more scientific and comprehensive assessments of interventions to ensure airport security (see Lum et al., this volume). Setting aside evaluations of outcome effectiveness, there are also very few studies of the implementation of security measures to understand how they work and how their functioning could be enhanced.

Modern democracies are supposed to be accountable for their use of public resources, and the lack of a balance between implementation and assessment reflects negatively on them. How nations choose to prevent terroristic violence both at home and abroad, and how they evaluate those interventions, matters not only to outcomes but also to how citizens view the legitimacy of government actions. Thus, the decision to commit large amounts of funds to antiterrorism strategies must be coupled with the responsibility of assessing whether programs are effective, ineffective, or harmful – that is, interventions should be *evidence-based*. The term “evidence-based” – as a descriptor for policy and decision-making – means that choices to implement interventions, like those that attempt to counter terrorism, are based on scientific and analytic knowledge that rigorously examines their impact on outcomes (Cullen & Gendreau, 2000; Davies, Nutley, & Smith, 2000; MacKenzie, 2000; Nutley & Davies, 1999; Sherman, 1998; Sherman, Farrington, Welsh, & MacKenzie, 2002; Weisburd, Lum, & Petrosino, 2001). This approach to decision-making not only provides both scientific and fiscal justification for programs but also appeals ethically (see Chalmers, 2003), and can temper rash responses to crises and moral panics (Lum, 2009).

In theory, the principles underpinning evidence-based policy can also support counterterrorism (Lum & Koper, 2011). These principles include pursuing methodologically rigorous evaluations of interventions; systematically reviewing research; disseminating, translating, and using research to inform practice; engaging in partnerships that foster evaluation (i.e., between practitioners and researchers); and expanding the collection of high-quality data. In practice, especially with concerns such as terrorism and random violence, there are many challenges to fostering such an approach (see Petrosino, Boruch, Soydan, Duggan, & Sanchez-Meca, 2001; Van Brunschot & Kennedy, 2008). While this collection is governed by the presumption that such assessments are not only scientifically valid, useful, and

necessary, but also possible, the reality of such an assumption is questionable. Even the most liberal governments are resistant to self-evaluation, especially regarding subject matters (terrorism) and actions (counterterrorism), that bureaucrats may believe are beyond the reach of assessment because of national security concerns (whether real or imagined). This tug-of-war between high-quality governance and national security lies at the heart of the current lack of evaluations of counterterrorism policy, and it is likely one major reason why research in etiology of terrorism has thrived, while evaluation research in counterterrorism has not.

The chapters within this volume, individually and collectively, underscore the principles of an evidence-based approach while exploring these challenges to scientific evaluation of counterterrorism interventions. Specifically, we assembled a set of papers that together would move toward two goals of evidence-based crime policy. First, the papers would speak to the basic issues of creating an evaluation and scientific base for counterterrorism interventions. Second, each chapter would contribute to research innovation in this area by expanding and creating depth in our thinking about evaluating counterterrorism measures.

Innovation in Research

Three strategies can create innovation in research for subjects (like counterterrorism) that need a stronger evidence base. First, the range, depth, and quality of available data to measure outcomes and to evaluate interventions can be scrutinized and improved. Observations and data, whether qualitative or quantitative, are the foundations of analysis and provide important access into both theoretical understanding and assessment. Second, innovation in research also is provoked by challenging and developing the methods used in evaluation. The use of the myriad of available rigorous scientific methods, from modeling to experimentation, helps us see complex relationships, including causal ones that are not immediately obvious. And finally, research innovation occurs when new perspectives are offered. Idea creation can help inspire others, including new generations of researchers, to consider alternative ways of looking at problems. Such innovation can expand the evidence base of topics that seem difficult for science to touch.

The Data, Methods, and Perspectives sections of this volume reflect this research innovation and provide reasonable, scientific, and neutral suggestions in improving the evidence base of counterterrorism. When discussing terrorism data, LaFree and Sheehan detail and compare the utility of major terrorism data sources, including the Global Terrorism Database (GTD). While the types of data available are interesting, there are two important takeaway points from both chapters regarding how we interpret terrorism trends across time: How the data is collected and also the motivations behind that data collection matter when measuring the effects of a counterterrorism intervention. Further, unlike in crime prevention studies, where data are relatively plentiful and provided with fewer obstacles, terrorism research relies on a more tenuous supply of data, whose sources are often shrouded in secrecy,

sometimes unnecessarily. As Sheehan notes, validation between datasets can also be difficult, as the definitions, political motivations, and quality of efforts behind different data collection efforts can vary significantly. While the National Consortium for the Study of Terrorism and Responses to Terrorism (START Center)¹ at the University of Maryland has made major strides in the collection and open publication of terrorism incident data, as LaFree notes, other types of information remain more mysterious. Sheehan provides a comparison across terrorism data sources in his chapter.

Related to the discussion of developing data sources is data sharing for prevention purposes, which has become a major concern post September 11th. Carter and Chermak discuss both the advantages and challenges in intelligence flow, sharing and use of preventative data, and analysis with fusion centers. Their survey indicates that while there is much interest in cooperation among law enforcement entities, the quality, sharing, and use of fusion center data remain less well established, and the effectiveness of fusion centers still remains unevaluated. All three chapters emphasize one major point: that the scrutiny of terrorism-related data and their use is a central exercise in increasing our knowledge of the impact of counterterrorism efforts.

Even with better availability and quality of data, determining how to analyze rare events and evaluate interventions that attempt to prevent rare events are major challenges to evidence-based counterterrorism, as the section on methodology indicates. Terrorism event trend analysis can be used in evaluating counterterrorism (see Lum et al., 2006), and four articles in this section focus specifically on exploring new methods to evaluate terrorism trends over time. Porter's change-point analysis shows how combining both frequency and impact of attacks presents important nuances that can result in different conclusions about the nature of terrorism and counterterrorism over time than previously thought. Dugan and Yang's group-based trajectory approach also provides an alternative to previously used interrupted time series and vector autoregressive models. They argue that trajectory modeling can reveal important variations in group activity that may not be seen using other approaches. Cioffi-Revilla, using a social complexity framework, suggests that the nonnormal distribution of terrorism over time requires rethinking models of terrorist attacks and how to counter them. Finally, Rusnak and her colleagues shift the focus to spatial trends, looking for the factors in the environment (in this case major features in Turkey) that would, when combined, increase the risk of terrorism attack. This approach provides a basis of examining the environmental backdrop of terrorism, making forecasts of events less dependent on previous incidents and more informed by the important correlates of the event.

The fact that terrorism is considered a rare event often leads to a focus on statistical modeling of the effects of counterterrorism interventions as opposed to other evaluation methods, such as, experimental designs. However, as Lum et al. and Costigan discuss in their chapters, experimental evaluation and assessment, as well as the concerns about testing and assessing the instrument validity of specific security measures and interventions, are also important for those building the evidence-base

¹ See <http://www.start.umd.edu/start/>

of counterterrorism interventions. Costigan's detailed discussion of instrument validity and reliability for screening systems to prevent air travel violence is an important step in evaluating the effectiveness of screening technologies, especially since these are widely used. Lum and colleagues, also focusing on air transportation security, suggest a "translational criminological" approach to preliminary evaluation. That is, if evaluations of existing practices are not readily available to judge programs, or if evaluations are not immediately possible, perhaps a way to preliminarily assess, and then to build an agenda for evaluation, is to apply existing criminological knowledge to counterterrorism measures in a systematic way. All of the discussions in the methods section of this volume present interesting and unique approaches and considerations if more opportunities for evaluation arise for counterterrorism interventions.

The contributions in the data and methods section of this volume all point to the need for more creative thinking to encourage the building of an evidence base for counterterrorism evaluation. However, new and creative ideas still must fall within the confines of good science and structured thinking about developing facts and knowledge. To bring this point home, we offer a series of articles that discuss fresh, alternative, and varied perspectives on scientific approaches to knowledge about counterterrorism interventions. These chapters provide a unique addition to the counterterrorism literature and are intended to broaden thinking about types of counterterrorism interventions that can be evaluated, as well as from what perspective they might be assessed.

For example, Passas' chapter shows how a traditional emphasis on examining financial institutions has led to a general neglect of scientific evaluation of other measures, such as counterterrorism finance. Examining informal markets, as well as pursuing trade transparency and regulation, may be just as, if not more, important as a counterterrorism measure than law enforcement activities. Ramsay's and Cave's articles both focus on what might be considered a very specific type of counterterrorism intervention – counterinsurgency and military interventions – to quell political violence and stabilize states. Ramsay and Cave apply common social science methods (surveys) and theories (opportunity, routine activities) respectively, and by doing so, they emphasize the contradictions to theory building, evaluation, and research given the unique subject matters at hand (insurgency and counterinsurgency). Indeed, terrorism and political violence often contradict traditional moral perspectives of the etiology and (lack of) support for violent crimes, which suggests that countering these types of social violence requires rethinking the scope of traditional social science theories. Whether survey or polling research can assist with such evaluations is, as Ramsay discusses, itself an open question.

The legitimacy of the state and the law is a tenuous assumption in times of terrorism and insurgency. Merola advises the legal and law and society communities that much can be gained from adding social science and evaluation research approaches to understanding the impact of legal changes brought on by counterterrorism. As Tyler suggests, being more sensitive to the complexities of social relationships and also society's relations with the state and legal processes when engaging in prevention approaches can make the difference between evaluating

success or failure, especially in future support for terroristic violence as a means to an end.

These ideas regarding data, methods, and perspectives are not specific to terrorism, but are generally important in criminal justice, policing, and security policies. And, similar to policing and security, moving forward in counterterrorism policy requires a stronger focus on the measurement and evaluation of the effectiveness of interventions; the rationale behind evaluation; lessons learned from other areas of research; the obstacles to good evaluation; and the resources needed to affect these changes. Given the range of issues covered in this volume, where does this lead us in developing a future agenda around evaluation? In our epilogue, we outline ideas for developing a research infrastructure that might support more generation of research evidence in this area. Research infrastructures are necessary across many research agendas and provide support for conducting research and promoting the use of science in practice. This is especially the case with counterterrorism policy, as we still have many opportunities to test and assess interventions as they are in their early stages of development.

References

- Chalmers, I. (2003). Trying to do more good than harm in policy and practice: The role of rigorous, transparent, up-to-date evaluations. *Annals of the American Academy of Political and Social Sciences*, 589, 22–40.
- Cullen, F. T., & Gendreau, P. (2000). Assessing correctional rehabilitation: Policy, practice, and prospects. In J. Horney (Ed.), *Policies, processes, and decisions of the criminal justice system: Criminal justice 3* (pp. 109–175). Washington: National Institute of Justice.
- Davies, H. O., Nutley, S., & Smith, P. C. (2000). *What works: Evidence-based policy and practice in public services*. London: Policy Press.
- Enders, W., & Sandler, T. (1993). The effectiveness of antiterrorism policies: A vector-autoregression-intervention analysis. *The American Political Science Review*, 87(4), 829–844.
- Enders, W., & Sandler, T. (2000). Is transnational terrorism becoming more threatening? *Journal of Conflict Resolution*, 44, 307–332.
- Enders, W., & Sandler, T. (2006). *The political economy of terrorism*. New York: Cambridge University Press.
- Enders, W., Sandler, T., & Cauley, J. (1990). UN conventions, terrorism, and retaliation in the fight against terrorism: An econometric evaluation. *Terrorism and Political Violence*, 2(1), 83–105.
- Lum, C. (2009). *Translating police research into practice. Ideas in American Policing*. Washington: Police Foundation.
- Lum, C., Kennedy, L., & Sherley, A. (2006). Are counter-terrorism strategies effective? The results of the Campbell Systematic Review on counter-terrorism evaluation research. *Journal of Experimental Criminology*, 2(4), 489–516.
- Lum, C., & Koper, C. (2011). Is crime prevention relevant to counter-terrorism? In B. Forst, J. Greene, & J. Lynch (Eds.), *Criminologists on terrorism and homeland security*. New York: Cambridge University Press.
- MacKenzie, D. L. (2000). Evidence-based corrections: Identifying what works. *Crime and Delinquency*, 48(4), 457–471.

- National Research Council, The National Academies. (2010). *Field evaluation in the intelligence and counterintelligence context: Workshop summary*. Washington: The National Academies Press.
- Nutley, S., & Davies, H. T. O. (1999). The Fall and Rise of Evidence in Criminal Justice. *Public Money and Management*, 19, 47–54.
- Petrosino, A., Boruch, R., Soydan, H., Duggan, L., & Sanchez-Meca, J. (2001). Meeting the challenges of evidence-based crime policy: The Campbell Collaboration. *Annals of the American Academy of Political and Social Sciences*, 578, 14–34.
- Sherman, L. W. (1998). *Evidence-based policing. Second Invitational Lecture on Ideas in Policing*. Washington: Police Foundation.
- Sherman, L. W., Farrington, D. P., Welsh, B. C., & MacKenzie, D. L. (Eds.). (2002). *Evidence based crime prevention*. London: Routledge.
- van Brunschot, E., & Kennedy, L. W. (2008). *Risk balance and security*. Thousand Oaks: Sage.
- van Natta, D. (2011). Cities nationwide heighten vigilance on terror. *The New York Times*. Retrieved May 11, 2011, from http://www.nytimes.com/2011/05/14/us/14threat.html?_r=2&hpw.
- Weisburd, D., Lum, C., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *Annals of the American Academy of Political and Social Sciences*, 578, 50–70.

Part II
Data Sources for Evaluating Terrorism
and Counterterrorism

Chapter 2

Assessing and Comparing Data Sources for Terrorism Research

Ivan Sascha Sheehan

Introduction

A journey of a thousand miles begins with a single step. Lao Tzu

Much of the early work on terrorism was based on “small-*n*” qualitative case studies. Little by little narrative chronologies were added. Today, large-*n* quantitative databases of terrorist events containing thousands and even tens of thousands of events and a wide range of variables are only a click away on the Internet. These databases have provided enormous opportunities for terrorism researchers to identify cases and test hypotheses that are relevant to the field. But how good is the quality of the data? And how should terrorism researchers go about choosing between competing datasets?

One of the assumptions behind this chapter is that insights gained from the study of small-*n* data may be relevant to the development of standards to assess and compare large-*n* terrorism datasets. Ever since Geddes’ paper on “how the cases you choose affect the answers you get” (Geddes, 1990), small-*n* qualitative researchers have quite self-consciously tried to improve the quality of small-*n* data. One result has been the growth of a large body of scholarship around the concept of “best practices” norms and standards to maximize transparency, reliability, and validity in this kind of data (Brady & Collier, 2004).

At the same time recognition has been growing that large-*n* datasets, including data on political events and processes, are often riddled with the same problems that plague small-*n* data. Collier, Brady, and Seawright (2004), for example, have drawn attention to the problems large and small-*n* researchers both face in making contextually sensitive judgments in terms of coding. Others have gone a step further suggesting that large-*n* datasets should be able to convey the kind of “detailed knowledge

I.S. Sheehan (✉)
School of Public and International Affairs,
University of Baltimore, Baltimore, MD, USA
e-mail: isheehan@ubalt.edu

and sensitivity to context” that are the hallmarks and strength of case-oriented studies (Munck & Verkuilen, 2002). Still others have argued that improving data quality in databases such as *Polity IV* requires heightened critical attention to questions normally raised by qualitative researchers on subjects such as data construction and what Herrera and Devesh call “data supply chains”:

Who produced the data? Why? What were the producer’s incentives and capabilities? Did they work as an independent agency or were they influenced by external actors? (2007, p.366)

These observations have led to increased calls for *shared* standards to evaluate small-*n* and large-*n* research. In this context, John Gerring (2001, 2010) introduced the concept of a “criterial framework” that bridges small-*n*/large-*n*, qualitative/quantitative chasms. Drawing on this concept, Evan Lieberman (2010) has shown that normative criteria commonly associated with the evaluation of small-*n* studies (e.g., citation transparency and handling of issues of uncertainty) can potentially improve political datasets such as *Polity IV*, the *Annual Survey of Freedom* (Freedom House index), the *Minorities at Risk* dataset (MAR), and the *Uppsala Conflict Data Program* dataset (UCDP).

In this chapter I will build on the concept of a “criterial framework” by extending it to large-*n* terrorism datasets and by proposing a best practices framework to help users evaluate the validity and reliability of a range of terrorism datasets. I will begin by discussing why we need norms or best practice standards to compare and evaluate quantitative terrorism data sources. I will then make a case for extending the concept of a “criterial framework,” such as the one described by Gerring (2001, 2010), to terrorism data. I will highlight challenges and problems that occur in applying such a framework to five of the most well-known and respected terrorism events datasets in the field. All five of the databases selected are translations of narrative records (usually news reports) into numerical data in the form of counts, indexes, or dummy variables indicating the presence or absence of a phenomenon related to a coded terrorism event. In selecting the databases I have chosen ones that are publically available on the Internet and could be considered elite or the best in the field. I will conclude with proposals for implementing best practices in terrorism databases.

Developing Standards for Terrorism Data

Why Do We Need Them?

There are practical as well as methodological reasons for developing standards to assess and compare large-*n* terrorism datasets.

Practical Reasons

Terrorism datasets are unique in that most of them were first developed, maintained, and used *outside* universities in the intelligence and defense communities. First

published in 1975, the *RAND Terrorism Chronologies*, which later became the basis for the well-known *RAND-MIPT* terrorism database, now subsumed into the *Rand Database of Worldwide Terrorism Incidents* (RDTWI), were originally developed by policy analysts such as Brian Michael Jenkins under a defense department grant for intelligence purposes (Fowler, 1981). The RAND chronologies were subsequently used by CIA analyst, Edward Mickolus, to produce the *International Terrorism: Attributes of Terrorist Events* dataset (ITERATE), a dataset that later formed the basis for the CIA's annual terrorism report, *Patterns of Global Terrorism* (Schmid, 1983, p. 257). Much of the *Global Terrorism Database* (GTD) originated with reports made by Pinkerton Global Intelligence Service (PGIS), a private global security firm (La Free, 2010), and the *World Incident Tracking System* (WITS) is a replacement for the previous annual *Patterns of Terrorism* report put out by the U.S. National Terrorism and Counter Terrorism Center (NCTC) (Wigle, 2010).

Because of these origins, *outside* an academic environment, terrorism data were not always subjected to the kinds of rigorous norms in terms of collection or coding that are usually expected in academia (Schmid, 2004). This situation led to considerable embarrassment when two Princeton scholars reviewed the data tables at the end of the State Department's annual *Patterns of Global Terrorism* report for 2003 and found that the numbers in the tables did not add up and that the conclusion of the report, namely that global terrorism had decreased that year, was in error and that terrorism had actually increased. When they subsequently published that information in an op-ed piece in the *Washington Post* (Krueger & Laitin, 2004a) and in an article in *Foreign Affairs* (Krueger & Laitin, 2004b), the State Department admitted that the report was wrong and retracted it.

Today large numbers of the users of terrorism data, however, are academic scholars. Many of these academics were nurtured in programs that emphasized the importance of best practices in collecting data and while they may agree with Gerring that the objects of social science often “refuse to lie still in the manner of rocks, animals, cells and atoms” (Gerring, 1999, p. 393) there is a much greater expectation that terrorism data should be based on solid norms and that it should be subjected to evaluation and questioning.

Terrorism datasets differ from other political and social science data in another important way. Since much of the data is derived from media sources in real time, and since its developers have frequently used different definitions and coding rules, no one dataset is completely comprehensive or exhaustive and there is a great deal of variability across datasets. For example, in a previous comparison of transnational terrorism events data from two terrorism databases, ITERATE and RAND-MIPT, this author found several large discrepancies in quarterly events counts for the time period 1993–2004. As shown in Fig. 2.1, there were distinct differences in counts at the outset of the series in 1993 and again between 2001 and 2002 and in 2004. A likely explanation for these differences was that the databases operationalized what constituted a “transnational” event very differently. In general, transnational terrorist events are viewed as ones that involve perpetrators and victims from different countries.¹ However, what constitutes a country may be differently defined

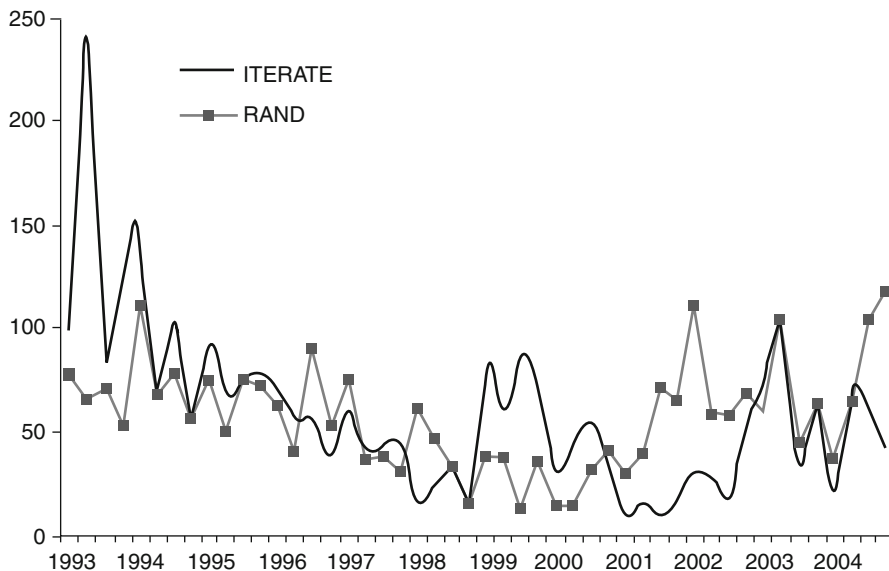


Fig. 2.1 ITERATE vs. RAND-MIPT: quarterly number of transnational terrorist incidents (1993–2004). *ITERATE* International Terrorism: Attributes of Terrorist Events; *RAND-MIPT* RAND-Memorial Institute for Prevention of Terrorism Database

when a territory is still disputed. A likely explanation for the discrepancy at the outset of the series in 1993 is that *ITERATE* counted more events that started in one of the newly formed Soviet states and ended in another or targeted people as transnational whereas *RAND* excluded those events under the assumption that they were not yet transnational. By contrast, closer inspection of the data for the period 2001–2002 and for 2004 suggests that *RAND* counted more incidents associated with

¹When a terrorist incident in one country involves victims, targets, institutions, or citizens of another country, it is considered transnational or international and is included. The 9/11 hijackings, for example, are included as transnational terrorist incidents for at least three reasons. First, the perpetrators came from outside of the United States. Second, the victims were from over 80 countries. And third, the incidents had worldwide economic and security ramifications. The bombings of the US embassies in Kenya and Tanzania on August 7, 1998, as well as the suicide car bombings aimed at British and Jewish targets in Istanbul, Turkey on November 20, 2003, are similarly included as transnational terrorist incidents since they involve perpetrators and victims from different countries. On the other hand, the bombing of the Murrah Federal Building in Oklahoma City by Timothy McVeigh is not included since it is considered to be a purely domestic event. Similarly, bombings by the IRA in Northern Ireland are not included as transnational terrorist acts. However, IRA attacks in England would be included.

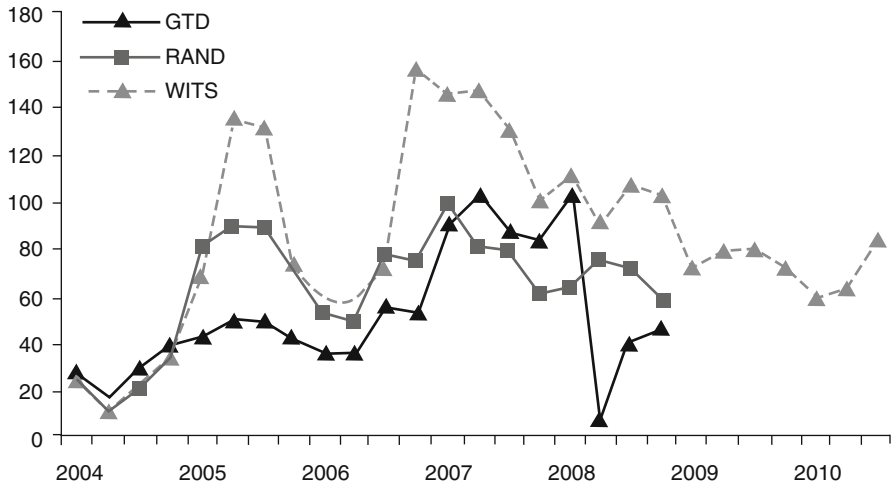


Fig. 2.2 GTD, RAND, and WITS (2004–2008): quarterly number of suicide attacks. *GTD* Global Terrorism Database; *RAND* Rand Database of Worldwide Terrorism Incidents (RDTWI); *WITS* Worldwide Incidents Tracking System

Palestinian uprisings against Israel (Intifada) as transnational whereas ITERATE, making the assumption that such events were domestic, did not include them (Sheehan, 2007, 2009).

Similar disparities are evident in more recent data. Figure 2.2, for example, shows the quarterly number of suicide attacks for more recent data on international and domestic terrorism events from three current terrorism datasets (the GTD, RAND, and WITS) for the time period 2004–2008. In this case the RAND and WITS datasets both show much higher frequencies of suicide attacks at almost every time point compared to the GTD.

Disparities like these can be a source of consternation for terrorism researchers. Just looking at the plots, we do not know if the disparities are a function of differences between the two databases in the ways they define terrorism, in the sources they use, in coding rules, or something else. And without more information, it is not possible to tell if one or the other dataset capturing more “true” events or including more “false” ones.

Part of the problem is that terrorism databases often tap into different data sources and information. The plot below (Fig. 2.3) of international terrorist incidents from ITERATE and RAND-MIPT for the years 1993–2004 illustrates this point.

The plot above shows that *unique* incidents, ones covered in only one of the two databases, outnumbered *overlapping* ones at almost every quarterly period shown. Overall for the time period, overlapping incidents constituted only about one third

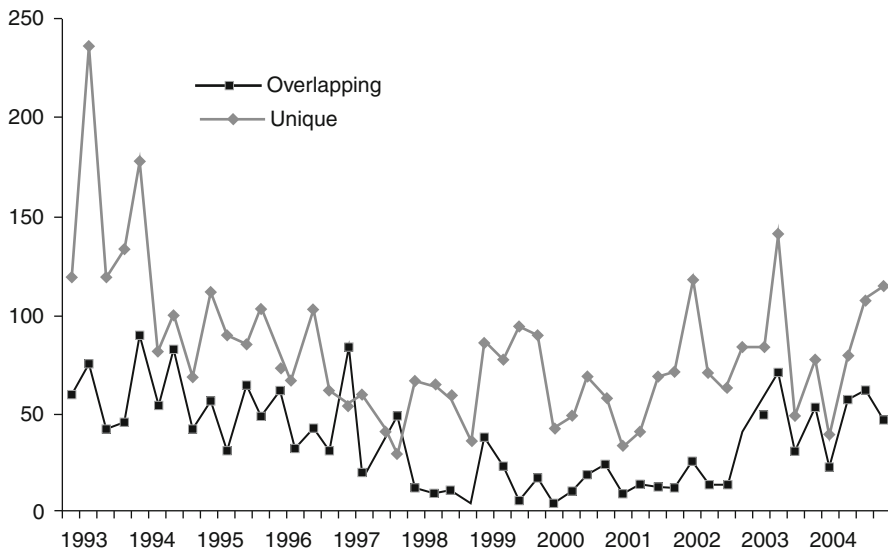


Fig. 2.3 ITERATE vs. RAND-MIPT (1993–2004): overlapping and unique international terrorist incidents. *ITERATE* International Terrorism: Attributes of Terrorist Events; *RAND-MIPT* RAND-Memorial Institute for Prevention of Terrorism Database

of all incidents. Overlapping terrorist incidents are the ones the databases agree on. In general they are more likely to be reported and more likely to be deadly (Sheehan, 2007). However, they constitute a relatively small proportion of all incidents. This may be because the databases have selection biases in terms of sources, because they use different operational definitions of terrorism, because they have different inclusion or exclusion rules, or simply because the enormity of capturing so much data in real time is so great that each database is only able to cover a segment of terrorism events. Unique events, however, constitute larger portions of each dataset and they are responsible for the variability in terrorism datasets.

The point is that from a practical viewpoint to assess the quality and comparability of information across terrorism databases, researchers need some benchmarks or normative criteria.

Theoretical Reasons

There are also good theoretical reasons for developing best practices criteria for terrorism data. As Alex Schmid observes, scholars seek out terrorism data for a number of reasons. They may want to use the data to uncover underlying patterns or

trends in terrorism, to compare terrorist campaigns cross-nationally and over time, to predict future events, to examine the causes, concomitants, or consequences of terrorism in relation to other phenomena, or to evaluate the success of counterterrorism (Schmid, 2004). To meet these different requirements, terrorism databases should ideally fulfill a number of criteria. In the best of worlds, it should be relevant and transparent. The definition of terrorism should be consistent over time and across regions. In addition, the data should be replicable and reliable and the validity and integrity of observations should be such that they can be checked (Drakos, 2009).

Unfortunately, terrorism data like much of what Colin Robson (2002) calls “real world” data rarely lives up to these expectations. However, to the extent that a terrorism database considers and at least tries to approach some norms and standards, we can have greater confidence in its credibility. The existence of norms and standards, moreover, can be a guide to particular researchers in their search for the best database for a particular project.

Developing Standards for Terrorism Data

The Case for a “Criteria Framework”

Users of large- n datasets usually look for tests of validity and reliability to evaluate a given dataset. Unfortunately, traditional tests of validity and reliability are problematic for terrorism data. Validity tests typically rely on a gold standard, but there is no gold standard for terrorism data. There is not even a universally accepted definition of terrorism. By one count there are as many as 109 definitions of terrorism (Schmid & Jongman, 1988). Reliability tests typically depend on consistency, but terrorism databases are often viewed as “living databases” (Wigle, 2010) that can be changed, even retrospectively, as new information emerges, as perspectives change or as the operational definition of terrorism is revised (Paull, 1982; Schmid, 1983; Wilkinson, 1986; Reid, 1997).² This means that inter-coder reliability tests, even if conducted at the time of collection of a discrete item of data (e.g., a terrorist event from a report in the media), are rendered meaningless and the conditions for test-retest may not exist. To complicate matters, terrorism databases have relied almost

²In the 1970s and 1980s, it was not at all uncommon for terrorism databases to redefine and reclassify terrorist incidents for political reasons (Paull, 1982, p. 46; Schmid, 1983, p. 260). This practice led to wide disparities in the annual Patterns of International Terrorism report across the Ford, Carter, and Reagan administrations. For example, although only eight types of incidents were classified as terrorism under Ford, as many as 17 were classified as terrorism under Reagan. Reclassification, moreover, was applied retrospectively. As a result the 1980 Patterns of Global Terrorism report estimated that the number of worldwide terrorism incidents for the period 1968–1980 was 6,714 although only a year earlier it estimated that terrorism incidents for approximately the same time period 1968–1979 were half that number or 3,336 (Wilkinson, 1986, p.44, cited in Reid, 1997).

exclusively on reports in the news media (Fowler, 1981) and all too often such reports have been accepted unquestioningly despite their known biases and unreliability (Wardlaw, 1982; Wilkinson, 1986; Herman & Chomsky, 1988).

Alternative criteria, developed for small-*n* research may be well suited to evaluating terrorism data with its naturalistic roots in “real world” inquiry and its emergent qualities (openness to adaptation and change as new information emerges). The mindset that informs these criteria is also well suited to terrorism data. Although large-*n* research is typically guided by a positivist belief that a real world of objects apart from people exists out there and that researchers can accurately describe it and compare their descriptions with this objective reality, this is not the case with small-*n* research. Small-*n* researchers, coming from post-positivist, constructionist, and interpretivist traditions, are more likely to take the position that researchers can only know the world from their own perspective of it.³ One implication of this view is that researchers’ and informants’ own preconceptions, values, and biases are relevant to all phases of research and are critical to its credibility. Another implication of this view is that knowledge (the truth) can only be approximated, never fully known, and uncertainty needs to be acknowledged. This mindset may be particularly helpful in evaluating terrorism data where, because of the clandestine nature of the phenomenon, objective information is not always available.

At the same time, small-*n* researchers, mindful of the importance of validity, have developed criteria such as credibility, transferability, generalizability, and dependability that parallel the concepts of validity and reliability (see Guba, 1981). This situation has led to increasing recognition that shared norms for designing and evaluating research can be built across the two approaches (Brady & Collier, 2004).

The concept of a criterial framework was introduced and developed by Gerring (2001, 2010) as a means of helping social science researchers bridge small-*n*/large-*n* and other divides (qualitative/quantitative, positivist/interpretivist) to find common ground in designing research. It has since been extended to evaluating large-*n* political data (Lieberman, 2010). One of the advantages of applying such a framework large-*n* data is that it has the potential to bring rich detailed descriptions that can serve as an alternative to traditional validity and reliability testing when such testing is not possible.

Proposed Criteria

In this section, I propose six criteria derived from small-*n* research that could be used to evaluate and compare terrorism datasets. To make them easier to remember, I have chosen words that all start with the letter C. They include Conceptual clarity, Context and immediacy of observation, Citation transparency, Coding

³ It should be noted that these approaches are not uniform. Post-positivists are more likely to accept the “objective” nature of reality than interpretivists and constructionists.

Table 2.1 Criterial framework for evaluating terrorism databases

Criterion	Rationale	Assessment
Conceptual clarity	The relevance of a database depends on the definition(s) of terrorism used	Is the definition of terrorism used in construction of the database specified?
Context and immediacy of observations	Data collected from primary and secondary sources are often more valued than those collected from tertiary sources	Do the authors report the context and immediacy of the observations?
Citation transparency	The replicability of a dataset depends on citation transparency	Are the actual sources of the data described? Are clear references to original data provided?
Coding consistency	The reliability of a dataset depends on coding consistency over time and across raters	Do the authors provide a codebook? Do they discuss how they resolve coding conflicts and make decisions in ambiguous cases?
Certainty	The validity of observations depends on certainty. Contradictions and ambiguity in data should be reported	Are contradictory facts reported?
Conflict of interest issues	The integrity of observations can be compromised in the presence of competing interests	Are funding sources and other potential conflicts of interest reported?
Convenience/ functionality	Differences in data sources can be uncovered more easily when those sources are easily accessible and the data can be disaggregated	How accessible is the database? Can it be downloaded? Can it be disaggregated for fine analyses?

consistency, Certainty, and Conflict of interest. I have added an additional criterion, Convenience and functionality, since many users of terrorism data want data that they can find easily and that will fulfill different functions. My hope is that the criteria proposed here will help researchers be better able to evaluate the quality of terrorism data and be in a better position to choose between terrorism data sources (Table 2.1).

Conceptual Clarity

How well do the authors of the database communicate the underlying concept of terrorism they use to choose individual cases? How well do they define it? As Lazarsfeld and Barfeld (1951, p. 155) once wrote,

Before we can investigate the presence or absence of some attribute... or before we can rank objects or measure them in terms of some variables we must form the concept of that variable.

Concepts, as Goertz (2005, p 5) observes, are ontological: “they are theories about the fundamental constitutive elements of a phenomenon.” Traditionally, concepts implied necessary and sufficient conditions. In the last few decades, however, what Collier and Mahon (1993) call “family resemblance” approaches have often been used as a substitute for necessary and sufficient conditions. In these approaches, one condition substitutes for another. Such substitutions can lead to “conceptual stretching” or “traveling” and more permissive inclusion of cases (Collier & Mahon, 1993, p. 845).

While conceptual stretching and traveling may have benefits (ambiguous cases will not as easily be lost), they pose problems for researchers who want to generalize from a set of data. This was an issue when the U.S. State Department released its 2003 *Patterns of Global Terrorism Report*. This report specified terrorism as “premeditated, politically motivated violence perpetrated against noncombatant targets by subnational groups or clandestine agents, usually intended to influence an audience” (U.S. State Department, *Patterns of Global Terrorism 2003*, p. vii). It further specified that an international terrorist attack was an act committed by sub-state actors from one nation against citizens or property of another and that an incident was “judged significant if it results in loss of life or serious injury to persons, major property damage, and/or is an act or attempted act that could reasonably be expected to create the conditions noted” (U.S. State Department, *Patterns of Global Terrorism 2003*, Appendix A). The key problem here was the use of the word “significant.” As Krueger and Laitin (2004a, 2004b) quickly pointed out, almost no information was provided about how the government authors distinguished significant from nonsignificant events. In the end, a reanalysis of the data with better specification of what the government authors meant by the word “significant” produced a very different set of data and one which embarrassingly contradicted previous findings that terrorism events had decreased that year. In fact, the new evidence indicated that terrorist events had increased.

Conceptual clarity has implications for inclusion rules. Different concepts of what constitutes an *event*, for example, can lead to different inclusion rules and widely varying estimates. This was the case when in its original database RAND counted 40 bombings by one group in one city as one event when ITERATE counted 40 separate events (Jenkins & Johnson, 1975). Similarly, different concepts of what is *international* or *transnational* can lead to other discrepancies. As discussed earlier in this chapter, RAND but not ITERATE appears to have treated incidents involving Palestinians and Israelis as transnational with the result that its estimates of total transnational terrorist incidents were much higher than ITERATE’s at some intervals in the last 12 years.

Context and Immediacy of Observation

What is the context of the data? How was it collected? How close were the authors to the source? Terrorism data may be generated from primary sources, from secondary sources, tertiary ones, and from experts. Academic scholars often put a premium on

data collected through direct observation because of its immediacy. Interview data, such as that collected by Merari (2010) in his interviews with suicide terrorists, is especially valued because of its immediacy. The value of such data, however, has to be weighed against the potential for selection bias. After direct observation, scholars usually value primary sources such as newspaper accounts at the time of an event over secondary sources (e.g., journal articles) and tertiary ones (e.g., textbooks, other datasets).

Citation Transparency

Do the authors provide the sources for the data they report? Terrorism data is derived from a wide variety of sources. The sources may include United States and foreign news and wire services. They may also include information from interviews, books, memoirs, and interviews with principals (Mickolus, 2002). Knowing the sources of a particular event matters since coders often have to make decisions. They may have to choose, for example, between conflicting estimates from different sources on the number of fatalities associated with an event. Confidence in data is heightened when the actual source of the data is provided. Citation transparency is also critical to correcting and replicating datasets.

Coding and Consistency

Do the authors use consistent rules to code data? Do they use a codebook? And have they institutionalized systems to ensure coding consistency across raters (inter-rater reliability) and over time (test–retest reliability)? If so, how do these systems work? In cases where there are conflicting reports for example, about claims of responsibility for an event, or the number of casualties, how are final decisions made? By fiat? By consensus, majority rule or some other way? Finally, is there an “audit trail” that users can follow if the operational definition of terrorism or inclusion rules is altered?

Certainty of Record

Do the authors report the presence of uncertainty in the coding of a particular variable or attribute of a terrorist event? Conflicting accounts, as discussed above, may lead to uncertainty about whether an event was actually a terrorist event or something else. Or, an attribute of an event may not quite fit within a given coding scheme. The scheme, for example, may allow coding of one or two targets, but the terrorist attack had multiple targets. How do the authors resolve such quandaries and do they highlight them?

Conflict of Interest

Do the creators and maintainers of the database have conflicts of interest and do they report them? This is tricky. Since terrorism is a national security concern and since the creation and maintenance of terrorism databases is costly, most of the databases we cover here have received significant funding from a government body and one, WITS, is a direct output of a government agency, the NCTC. The question is really the extent to which sponsors, funders, or others influence the content and the extent to which operators of the database are upfront about such real or potential influences. In election years, in particular, political pressure may be exerted to show that terrorism rates have diminished. How do the authors handle such potential conflicts of interest? For example, do they disclose funding? Do they discuss the extent to which the sponsors have control over the data?

Convenience/Accessibility/Functions

How convenient is the database to use? How accessible is it? Is it available online? Is it searchable with key words? How functional is it? Can users perform online searches with key words? Can they use the data to create graphs and tables online? Is the dataset fully downloadable so that researchers can conduct their own analyses on it? Can the data be disaggregated for fine analyses? And finally, is there a fee to use it?

Application of the Framework

Assessment of Existing Terrorism Databases

Below I review and assess a range of existing terrorism databases and try to apply the criteria developed above as a basis for comparison. The databases I cover include five terrorism events databases.

Terrorism events databases are systematic numeric records, usually derived from newspapers, wire, and other media, of the occurrence of individual terrorist events and the events' characteristics (e.g., date, location, name or type of perpetrator group when it can be identified, type of attack, and number of casualties). Terrorism events data can be linked in turn with other data to study the causes and consequences of terrorism. Events databases allow researchers to examine trends and patterns in terrorism over long periods of time and geographical space. They can be used in conjunction with other data (e.g., political or economic indicators) in analyses of the causes and consequences of terrorism and can contribute information to analyses of how, when and why and terrorism events and campaigns decline or end (Schmid, 2004).