

Are counter-terrorism strategies effective? The results of the Campbell systematic review on counter-terrorism evaluation research

CYNTHIA LUM*

Administration of Justice Program, George Mason University, 10900 University Blvd., MS 4F4, Manassas, VA 20110, USA

**corresponding author: E-mail: clum@gmu.edu*

LESLIE W. KENNEDY and ALISON SHERLEY

School of Criminal Justice and Center for the Study of Public Security, Rutgers University, Newark, NJ 07102, USA

Abstract. The events of September 11th have led to massive increases in personal, commercial, and governmental expenditures on anti-terrorism strategies, as well as a proliferation of programs designed to fight terrorism. These increases in spending and program development have focused attention on the most significant and central policy question related to these interventions: Are these programs effective? To explore this question, this study reports the results of a Campbell Collaboration systematic review on evaluation research of counter-terrorism strategies. Not only did we discover an almost complete absence of evaluation research on counter-terrorism interventions, but from those evaluations that we could find, it appears that some interventions either did not achieve the outcomes sought or sometimes increased the likelihood of terrorism occurring. The findings dramatically emphasize the need for government leaders, policy makers, researchers, and funding agencies to support both outcome evaluations of these programs as well as efforts to develop an infrastructure to foster counter-terrorism evaluation research.

Key words: Campbell collaboration, counter-terrorism, evidence-based, government accountability, meta-analysis, policy, September 11th, terrorism, what works

Introduction and background

Since September 11th, there has been an exponential increase in government and private spending on counter-terrorism strategies (Congressional Budget Office 2002, 2005; Guinnessy and Dawson 2002; Issues in Science and Technology 2002; Macilwain 2002; Silke 2004) generally. In 2005 the Congressional Budget Office estimated that “Appropriations for Combating Terrorism and Protecting Critical Infrastructure” since 1998 had increased from U.S. \$7.2 billion dollars to \$88.1 billion dollars (Congressional Budget Office 2005; see also Congressional Budget Office 2002). More remarkably, these estimates represent only a subset of U.S. defense spending and do not include the billions of dollars expended since September 11th on counter-terrorism measures in other sectors and around the world. The House Budget Committee has recently estimated non-Department of Defense funding for homeland security to have risen from \$9 billion in 2000 to \$32 billion in 2005.¹

These expenditures reflect both depth and scope in the funding of counter-terrorism programs. Not only has more money been spent, but there has also been a proliferation in the range of counter-terrorism policies implemented. For example, counter-terrorism strategies, tactics, and tools have included such interventions as:

airport screening	emergency preparedness	multilateral agreements
anti-terrorism home products	foreign aid	prison building/imprisonment
arrest	fortification of embassies	psychological counseling
assassination	gas masks distribution	punishment and sentencing
bilateral agreements	harsher punishment	religious interventions
blast-resistant luggage	hostage negotiation	seal/tamper-proof devices
buildings security	investigation strategies	situational crime prevention
CCTV	legislation (e.g., Patriot Act)	U.N. conventions
community and NGO efforts	medical antidotes	U.N. resolutions
weapons detection devices	media efforts/spinning	vaccinations
diplomatic efforts	metal detectors	war
educational programs	military interventions	

This proliferation has also been reflected in the shift in emphasis of government grant solicitations which have called for the development of new technologies, tools, and partnerships to facilitate these efforts.

This increase in personal, commercial, and governmental expenditures on anti-terrorism strategies, as well as the proliferation of programs designed to fight terrorism, raises the most significant policy question regarding counter-terrorism efforts: Do these programs work? Specifically, do they reduce the likelihood of, or damage from, terrorism events or discourage individuals from acquiring motivation to carry out this type of violence? How can this effectiveness be measured? And, could policies also produce harmful effects or unintended consequences?

These questions are at the core of determining the effects and effectiveness of any social intervention and reflect an evidence-based approach to policy (see Farrington and Petrosino 2001; Sherman et al. 2002). Evidence-based policies are ones that have been shown to be effective through systematic and scientific tests so to assure that the most appropriate treatments and interventions are used. Just as a medication should not be taken without knowledge of both its effectiveness as well as harmful side effects, so should social interventions be held to similar standards. Thus, evidence-based counter-terrorism policies are those policies which not only show promise in achieving outcomes sought, but at the same time do not cause harm.

Given this policy goal, what do we know about the effectiveness of counter-terrorism strategies? To assess the existing evidence on the effectiveness of counter-terrorism measures, we conducted a Campbell Collaboration² review of available evaluations of counter-terrorism interventions. Campbell reviews employ a systematic, peer-reviewed process, often using meta-analytic techniques to locate, examine, and make conclusions about evaluation studies (see Boruch et al. 2000, Farrington and Petrosino 2001).³ What follows is the analysis and findings of this systematic review.⁴ In particular, we highlight not only the research that we did find, but also what we *did not* find, and conclude with policy suggestions to government agencies and evaluation researchers regarding how to move forward.

Such an undertaking is not without problems and controversies. Most importantly, we expected that, despite the proliferation of counter-terrorism tactics (especially after September 11th), there would be few evaluations of these policies. Additionally, terrorism and its associated counter-policies, have not been easily or clearly defined, as the recent special briefing by the U.S. State Department has illustrated⁵ (see also Crenshaw 1992; Merari 1991; Wilkinson 1986). Types of strategies could potentially cover a wide range of phenomena, seek a variety of outcomes, and involve multiple points at which effectiveness could be measured. For example, counter-terrorism strategies may include prevention and alleviation of early risk factors, situational prevention of actual events, or post-event responses. How “effectiveness” should be gauged is also debatable. Furthermore, because of the rare nature of terrorism events, as well as the secrecy surrounding data associated with them, it may be difficult to determine whether strategies are effective. Finally, there may be evaluations of strategies that are shown to be effective but might violate international law or human or civil rights, challenge the values of a particular society or nation, or displace problems. While these are clearly limitations and challenges of assessing the state of counter-terrorism evaluation research, such a review is useful in providing a better understanding of what we do and do not know about program effectiveness.

In the past, a few general reviews of terrorism studies have been undertaken (see, e.g., Halkides 1995; Hoffman 1992; Miller 1988; Romano 1984; Schmid and Jongman 1988), although a systematic review of evaluation research on counter-terrorism strategies had not yet been conducted prior to this study. September 11th also has certainly necessitated the need for an updated review. Before beginning this Campbell review, we had conducted a preliminary, more general review for an in-house report for Rutgers University (see Kennedy and Lum 2003) to gain a sense of the literature up to and after September 11th. In that review, we collected all articles in published, unpublished, peer-reviewed, non-peer-reviewed, academic and non-academic sources that mentioned terms related to terrorism and political violence. We conducted this search across 17 separate literary databases.⁶ Although books, government and technical reports, online documents, and web information are included in the systematic review reported here, we limited the Rutgers study to only articles, in order to gain a basic sense of the literature. We found over 14,000 articles, most written and/or published between 1971 and 2003.

This preliminary search confirmed our initial estimations about the state of counter-terrorism evaluation research. The first unique finding compared with that of past literature reviews was that the events of September 11th have had an enormous effect on terrorism research. As Figure 1 indicates, among all the articles, as well as only those from peer-reviewed sources, over half were published in 2001 and 2002.⁷ Neither the Oklahoma City bombing in 1995 nor the World Trade Center bombing in 1993 was followed by such a large increase in research interest on terrorism.

After reading through the thousands of article abstracts from peer-reviewed sources, we also discovered that only 3–4% of them were based on studies that employed some

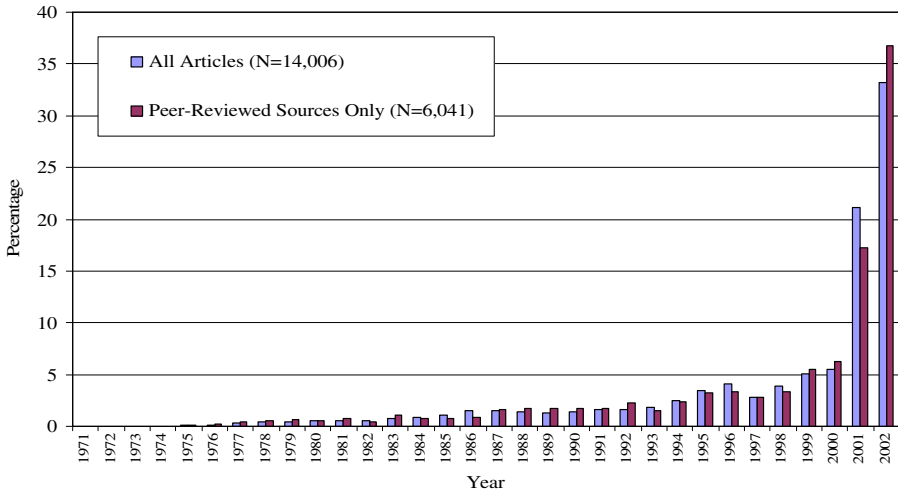


Figure 1. Yearly distribution of terrorism publications as a percentage of total publications in the Rutgers preliminary study (see Kennedy and Lum 2003).

type of empirical analysis on terrorism data or information. Thus, almost all the research on terrorism could be broadly described as thought pieces, theoretical discussions, or opinions. This was interesting in the context of the Campbell review, as we had anticipated that evaluations of counter-terrorism programs would either come from this very small subset of empirical works or be similarly scarce.

Our worries about the wide range of terrorism topics and the difficulty in finding evaluations were also confirmed when we examined the distribution of specific subject matters studied. When we tried grouping articles into common topical categories, we found 35 subject areas which could be collapsed into 17 groupings shown in Table 1. Table 1 also reports the distribution of these categorizations for studies which use some form of empirical analysis.

As Table 1 indicates, issues related to weapons of mass destruction represented the largest proportion of articles (18.9%) from peer-reviewed sources, followed by articles that focused on a specific issue, such as the Israel–Palestine conflict, the problems in Northern Ireland, Al Qaeda, or September 11th (if they could not be categorized elsewhere). Descriptions of political responses to terrorism and the sociology of terrorism (causes, motivations, explanations, definitions) also represented large proportions of terrorism research. Missing were articles which clearly indicated that an evaluation of counter-terrorism strategies had been conducted, although we anticipated that evaluations could be hidden within any of these subject matters. Even within the small subset of empirical research (in which we felt that evaluations might be found), about a quarter of those studies appeared to be conducted on victimology, focusing on the effects of terrorism on victims and their families. And, works examining any type of response involved only 6.4% of those empirical studies.

This preliminary overview of the terrorism research confirmed our concerns about coverage when we started a more systematic review. Because of all these challenges,

we initially used the broadest approach possible, making few (if any) initial assumptions about how interventions or outcomes should be defined. We also did not discriminate methodologically in the initial stages of our search, but rather first sought any study that even hinted that an evaluation had been conducted, whatever the outcome measured. Finally, in the Campbell review, we broadened our search for counter-terrorism evaluation research beyond just articles to books, government reports, web documents, and other sources of literature. We now describe the criteria and search strategy, analytic methods and results for the Campbell review.

The Campbell systematic review

Initial criteria for study consideration

We centered our initial search strategy for evaluation studies on three organizing constructs that reflected our goal of being as inclusive as possible. First, because of

Table 1. Types of subject matters examined in terrorism research.

<i>Subject matter</i>	<i>Peer-reviewed sources (N=4,458^a)</i>	<i>Empirical only (N=156^a)</i>
Weapons of mass destruction (biological, chemical, nuclear)	18.1%	10.3%
Article on a specific issue such as the Irish Republican Army (IRA), Al Qaeda or incident ^b	12.2%	5.1%
Political responses to terrorism (war, politics, international relations)	9.5%	1.9%
Causes, motivations, psychology, trends of terrorism	8.7%	18.1%
Impacts of terrorism (political, social, economic)	7.7%	5.2%
Non-political responses to terrorism (medical, social, economic)	5.5%	3.9%
Victimology, coping mechanisms, psychological effects of terrorism	5.4%	25.8%
Other (nationalism, intelligence issues, democracy and vulnerability)	5.4%	3.9%
Legal issues surrounding terrorism	5.2%	0.6%
The media and public attitudes towards terrorism	4.6%	18.7%
How to define terrorism	4.2%	1.3%
Non-conventional, cyber- and narco-terrorism	3.0%	0.6%
Religion and terrorism	2.6%	1.3%
State-sponsored terrorism	2.6%	1.3%
Law enforcement responses to terrorism (airports, police)	2.5%	0.6%
Research/science of studying terrorism	2.1%	0.6%
Domestic terrorism	0.6%	0.6%

^aExcluding book reviews and articles where not enough information was given to be categorized.

^bIf could not be placed into any other category.

continuing disagreement over the definition of terrorism, we considered any study in which the author(s) referred to the program or outcome as terrorism or terrorism-related, even if the subject matter discussed employed a unique, non-official or traditional definition of terrorism.

Secondly, we broadly included in our search any study that evaluated the effects of programs generally designed to prevent, detect, manage, and/or respond to terrorism events, related incidents and issues, risk factors, time periods, groups, and/or individuals. In other words, we approached the definition of interventions very liberally. For example, prevention strategies could include a wide range of programs designed to deter future events, such as improving embassy security, placing metal detectors in airports, or increasing penalties for crimes deemed to be terrorism. Research on detection strategies (which can also be preventative) might include evaluations of measures which improve the scanning of shipping containers, border-related strategies, immigration policy or other tactics used to detect people, places, or situations involved in terror-related activity. Examples of evaluations of management strategies might include examining the government's threat level warning system, new hospital procedures designed to address issues of health security, or programs that help people cope, adjust, or recover from the impact of a terrorist event. Evaluations studying response strategies might focus on law enforcement responses to suicide bombing or the effects of economic sanctions on terrorist-harboring states. Further, we did not limit ourselves to responses to one particular discipline since counter-terrorism strategies can include a wide variety of tactics and tools that can include political, social, legal, law enforcement, economic, preventative, reactive, or after-care responses.

Finally, the third construct we used in considering studies eligible for this review involved deciding what might be viewed as a measurable outcome of effectiveness. While the most straightforward outcome measure would be how the intervention affected (hopefully lowered) the number of terrorist incidents, we did not restrict ourselves to this requirement. Other measurable outcomes of effectiveness might include the reduction of the public's fear of terrorism, the increased ability to respond to certain events, the decline in the number of terror groups (or support for them), the reduction in the level of risk of an event occurring, or the ability to detect mechanisms of terrorism (for example, the ability to detect anthrax spores in mail). Additionally, outcomes measured could include general manifestations (i.e., "terrorism") as well as specific groupings (i.e., "skyjackings", "hostage-taking", "casualty" or "non-casualty").

Search strategy

Beginning with these general constructs, we then pursued a comprehensive search strategy to locate evaluation research using our preliminary review to guide our efforts. Since we had already read the abstracts of thousands of articles from peer-reviewed sources for the preliminary Rutgers study, we examined these studies as a

first step in locating evaluations. At this point, we included all studies that made any reference to the evaluation of a program, no matter the methodological quality of those evaluations. Additionally, because the initial collection of 14,006 terrorism studies included only studies up through January 2003 and did not include research from other media (books, government and technical reports, online documents, websites, unpublished material), we re-ran our initial search for terrorism research through December 2004 (the chosen end-date of this search) and extended our search across these additional media. We also searched the internet as well as terrorism-related organizations and data sources of terrorism research for studies.⁸ In total, we found over 20,000 documents on terrorism from which we attempted to pull evaluations. We re-ran our more specific search for evaluations using keywords (and their derivatives) such as assess, assessment, evaluate, evaluation, effect, effectiveness, empirical, intervention, policy, program, and/or works as well as forward-cited particular studies in hopes that future publications which cited evaluations were also themselves evaluations. We again more closely examined our small set of empirical studies to ensure we captured any evaluation missed in the broader search.

From this interactive process of searching through these 20,000+ works, we located 290 articles, reports, internet publications, and other published and unpublished non-book materials, as well as 64 books, that made any minimal reference that an evaluation of a counter-terrorism program had been conducted, giving us a total of 354 documents to examine more closely. It should be noted that the process of searching for studies was lengthy and involved and is only summarized here, but it is described in detail in the full Campbell systematic review.⁹

Final selection process of evaluation studies

We then proceeded to examine the 354 documents more closely. A more careful examination of the abstracts, notes, and titles of the articles of the 290 non-book studies revealed 94 works which, in the least, loosely indicated that an evaluation had been conducted or that one was discussed in the course of the study. We were able physically to locate 79 of these 94 studies by searching multiple libraries, databases, and the internet, asking for assistance from colleagues and graduate students at different universities, and contacting authors themselves for articles that could not be located. Of the 64 books initially located through our keyword search, 38 were located when we physically searched four separate university libraries and their affiliates.¹⁰ Of these 38 books, only one indicated that some evaluation of a counter-terrorism program had been undertaken (Wilkinson 1977). We now had located 80 studies that seemed, upon closer examination, even to hint that evaluation of counter-terrorism programs was of interest to the author(s).

The next stage of our selection process involved making a methodological judgment about these 80 studies to determine if they satisfied the minimum requirements to be considered an evaluation. To do this we followed a two-step process. First, we conducted an initial reading of the full text of each of the 80

studies and chose only those studies which appeared to connect an outcome or effect with a program or intervention through a minimally rigorous scientific test. For example, studies which used simple correlation statistics or even a before–after comparison were retained. This step was a conservative approach, as we felt that the nature of terrorism studies (for example, that events are rare or the difficulty in collecting information on terrorism) warranted careful consideration of any possible study to include. Additionally, if other evaluation studies were mentioned, we examined those as well.

We identified 21 of the 80 studies that satisfied this minimum criterion. Common reasons for excluding the other 59 studies were that a study:

- was not an evaluation of a program or policy designed to detect, prevent, respond to, or manage terrorism
- described the process of a program, but did not evaluate it
- made claims that a program was effective without any empirical test of that claim
- was a news article reporting on individuals claiming effectiveness
- was a review of non-evaluation literature
- advocated that an evaluation should be done but did not conduct one
- surveyed individuals about their feelings as to a program’s effectiveness
- surveyed individuals about how prepared they felt for another attack
- determined the effects of terrorism, rather than the effectiveness of a counter-terrorism program¹¹
- examined attrition for participation in a program but not the effects of the program
- made suggestions for the treatment of terrorism-related injuries and/or
- described criteria as to what effective policy should look like but did not evaluate the policy

The second step in selecting the final studies for analysis was to only include studies in which authors used at least a moderately rigorous methodological design. Not all evaluations are created equal; they vary in both internal and external validity and quality which lends to the believability of their findings (Cook and Campbell 1979; Farrington 2003; Shadish et al. 2002). The minimal requirements of the first stage only indicated that a weak test was conducted connecting the program with the desired outcome. For example, a program designed to treat post-traumatic stress disorder (PTSD) given to victims of terrorist acts may determine that the program worked because scores for indications of PTSD reduced over time after an intervention. However, this finding could be spurious; those individuals who were not given the program may have improved similarly in the same amount of time.

On the other hand, only including randomized controlled experiments in our review, which some Campbell-sponsored systematic reviews do, would be impracticable in this specific case. Not only would we have nothing to report, but the use of experimentation might have been a criterion too strict to use on outcomes that are rare, or where, in some cases, it would be truly unethical or impossible to allocate treatment randomly. Thus, as we initially outlined in our protocol, we were guided by what Sherman et al. (1997) describe as a moderate or mid-level scientifically rigorous design. Sherman et al. devised a five point “Scientific Methods Scale” (SMS) to score the methodological quality of evaluation research.¹²

For them, the highest quality evaluation method, the randomized controlled experiment, was ranked “5”, while simple correlational studies were scored as “1” or “2”. We recognize there have been critiques of this scale, and we used it only as a guide to assist with our selection process.

We used the middle SMS score of “3” as a general guide to exclude those studies which were not at least moderately rigorous, but we warn the reader that moderately rigorous designs are less believable than more rigorous ones. Sherman et al. (1997) describe an SMS score of 3 as “a comparison between two or more units of analysis, one with and one without the program.” Throughout Sherman et al. (1997) there are also elaborations on the meaning of a “3” by different authors; a methodological score of 3 can point to studies with multiple units who receive or do not receive the program and attempt to control for other factors or time series approaches which analyze time series data.¹³ Shadish et al. (2002) suggest that methods like autoregressive integrated moving average (ARIMA) that analyze time series data can be powerful research designs and equivalent to quasi-experimental studies.

Finally, we decided to exclude studies from the medical field, such as those which tested smallpox vaccinations, therapy for nerve agents, or anthrax vaccines. Those subjects have been more thoroughly examined outside the context of terrorism, and it would therefore be misleading to use the studies we found specifically discussing counter-terrorism as representing the research of these topics in the medical field.¹⁴ Furthermore, while we generally understood the methodological strategies used by these researchers, the context by which the findings were discussed was much removed from our expertise. Additionally, the focus of the Campbell Collaboration is social, rather than medical, interventions.¹⁵

Using these criteria, seven of the 20,000 studies satisfied our final requirement for inclusion of being a moderately rigorous evaluation study of a counter-terrorism program. At minimum, these studies:

- evaluated two or more units of analysis, comparing some with and without the counter-terrorism intervention, or
- made some attempt to provide for controls within a statistical analysis, or
- conducted an interrupted time series or intervention analysis to indicate some temporal ordering of effects, and
- were not medical studies

Summaries of the final seven studies chosen

Landes (1978)

Landes’ study on skyjackings is the earliest evaluation study of counter-terrorism strategies that we could locate and is often referred to in subsequent evaluations. Landes examined the effects of changes in laws and security measures through the increased probability of apprehension, incarceration, longer sentences, and being killed (by authorities) on the quarterly rate of domestic hijackings and the number of days and flights between successive hijackings for the period between 1961 and

1976. Landes also controlled for other variables, including the number of air flight operations per quarter, civilian unemployment rates, population rates, and personal consumption. He used a variety of ordinary least-squares regression techniques to ascertain relationships between these variables.

Cauley and Im (1988)

Cauley and Im offered an interrupted time series analysis (which they and others sometimes refer to as intervention analysis) of terrorism incidents occurring between 1968 and 1979. In their study they examine the effectiveness of increased airport security screening measures that occurred in 1973, increased security at embassies and other diplomatic missions in 1976, and a United Nations convention on preventing crimes against diplomatic personnel enacted in 1977. Cauley and Im not only analyzed multiple interventions but also the effects of these interventions on different outcomes, including skyjackings and non-skyjacking incidents, such as hostage taking, barricades, and attacks on diplomats.

Enders, Sandler and Cauley (1990)

This evaluation of multiple interventions and outcomes used an interrupted time series approach for events between 1968 and 1988. Enders et al. examined the effects of metal detectors in airports in 1973, the United Nations Convention on the Prevention and Punishment of Crimes Against Internationally Protected Persons Including Diplomatic Agents in 1977, United Nations resolutions against General Assembly and Security Council hostage taking (1985), United Nations resolutions against aerial hijacking (1969–1970), as well as the U.S. retaliatory raid on Libya in 1986.

Enders and Sandler (1993)

Enders and Sandler improved upon their previous statistical techniques in this 1993 evaluation of the substitution effects that policies may have between different types of terrorism. Like Enders et al. (1990), they examined the effects of metal detectors and resolutions during the period of 1968–1988, but they also examined security fortification measures taken on U.S. embassies. However, this study differs from their 1990 paper in that they analyzed interactions and substitution/displacement effects of different interventions across different types of terrorism.

Brophy-Baermann and Conybeare (1994)

Brophy-Baermann and Conybeare also employed an interrupted time series/intervention analysis to determine the effectiveness of six Israeli military-led retaliation attacks on reducing terrorism from the PLO and Lebanon. These retaliations began in September 1972 in response to the killings of Israeli athletes at the Munich Olympic Games, from which five more retaliations followed through 1988.

Enders and Sandler (2000)

In 2000, Enders and Sandler expanded their analysis to terrorism events between 1970 and 1996. They studied the effectiveness of metal detectors, embassy fortification, and the Libyan raid in 1986, as well as the reduction in totalitarian

governments that occurred after the end of the Cold War. Additionally, rather than study the effects of these interventions on different types of terrorism (or their substitutions) as they did before, they used outcomes that measured the type of person-based destruction. These outcomes included death, wounded, and non-casualties per quarter.

Barros (2003)

Barros employed intervention analysis by examining information collected by Abadie and Gardeazabal (2001) on the Spanish terrorist group ETA. The data included yearly counts of assassinations and kidnappings conducted by the ETA between 1968 and 2000. Barros studied the effects of different political ideologies in power, police and military expenditures as well as increases in foreign investment on incidents of kidnappings and assassinations conducted by the Spanish ETA using a vector autoregression time series framework.

Seven studies, 86 findings: data extraction and illustrative meta-analysis

In our Campbell protocol (the protocol is a peer-reviewed proposal that is submitted prior to conducting the actual review), we proposed to use meta-analytic techniques (see Lipsey and Wilson 2001) to make generalizations and to determine patterns from evaluation studies. However, the empirical studies that we uncovered presented unique challenges to using meta-analyses for this review. Specifically, each of the studies, except that of Landes, used intervention analysis with interrupted time series methods (Landes used a more modest, yet related, approach). Meta-analyses on time series data are uncommon, and reducing information provided by time series could over-simplify the initial study's findings or knowledge gained.¹⁶ Not only were different time periods analyzed across each of the studies, but, in two cases, a different unit of analysis was used. For example, five of the seven studies used quarterly time periods (3-month intervals) as their unit of analysis. However, Barros (2003) used years as units, while Cauley and Im (1988) chose months. Standardizing this unit would not necessarily solve the problem of generalization. And, the concept of time (and therefore outcomes associated with it) is qualitatively different from other more static units of analysis. Thus, we use meta-analytic techniques in our review for only illustrative purposes, and we strongly emphasize the limitations of these techniques in this particular situation. For this reason, we show each of our findings in the following figures, rather than providing only a summary meta-analytic statistic (for example, a mean effect size) for groups of findings.

To begin, we first determined whether multiple findings were reported in the same study. As our summaries indicate, each of these studies (with the exception of that of Brophy-Baermann and Conybear 1994) comprised several findings for different interventions, time periods, and/or outcomes. Many articles reported the evaluation of multiple interventions within the same study. Some studies also evaluated the effects of various interventions on different outcomes, for example, examining the effects of metal detectors on skyjackings and on embassy attacks.

Additionally, because the studies used interrupted time series analysis, we also had findings across different time periods for the same study (for example, “short-term” and “long-term”). Thus, in total, we discovered there were actually 86 findings within the seven studies relevant to this review.

However, multiple findings within the same study could lead to the possible complication of non-independence of the units of analysis. In meta-analysis, analyzed findings should be independent of each other to ensure that assumptions of statistical analyses are not violated. In our data, dependence of findings could be an issue. For example, short-run and long-run effects may not be considered independent, but effects of interventions on different types of terrorism might be. In some cases (Cauley and Im 1988, Enders and Sandler 1993) substitution or displacement effects were measured from one outcome to another, suggesting non-independence. Some researchers reported the effects of the same intervention on the same *general* outcome (e.g., “terrorism”) but also specified outcomes (e.g., “skyjackings”, “assassinations”, “barricade and hostage situations”). To gain the most from our data, we chose to code each finding (a strategy suggested by Lipsey and Wilson 2001), although we acknowledge the problem of possible non-independence of our data. Again, we caution the reader that we used meta-analytic techniques only as an illustrative tool.

To compare these multiple findings we extracted the effect size from each. When a meta-analysis is conducted, the size of the effect must often be standardized across study findings because of the different ways in which effect sizes are reported. Although results from time series present many challenges, one positive aspect of these seven studies was that all the studies reported their findings in the same way because they used similar methods—as the change from the natural rate of events per time period. Thus, we did not convert the findings to a standardized effect size. In other words, we treated these time series statistics like an unstandardized mean gain of a pre–post contrast (see Lipsey and Wilson 2001, p. 42), which normally compares the effect of an intervention at two time periods (before and after a treatment is administered).

Many of the findings also provided enough information, such as a *t* statistic or standard error of the change, for us to calculate a 95% confidence interval around the effect size. In many of the cases that reported a long-run effect in addition to a short-run effect, there was not enough information provided in the long-run effects for us to calculate a standard error. However, we retained the effect size in our visual displays, despite not including them in our meta-analyses. Finally, to calculate an average effect size across study findings, we employed a weighted mean effect size and its associated standard error and confidence intervals.¹⁷

We also conducted a homogeneity analysis to test the assumption that there was an underlying population mean that all effect sizes were estimating. Across all findings in which a weighted effect size could be calculated, we rejected the assumption of homogeneity (homogeneity *Q* statistic = 840.25, *df* = 59). This was expected, given that the studies measured a wide range of outcomes at different time periods. Even within the category groupings of specific interventions (discussed

below), homogeneity was rejected. Thus, when calculating an average effect size for a group of studies, we employed a random, rather than a fixed, effects model.

Given these caveats, for each of the 86 findings, we extracted and/or derived the following information:

- the complete citation
- the specific intervention evaluated (e.g., metal detectors or a military raid)
- the outcome measured as specified by the author (e.g., “skyjacking”, “assassination”, “deaths”)¹⁸
- the data source the authors used to evaluate interventions
- specific information about the method used by the researcher(s)
- the unit of analysis that was employed (specifically, the length of the time interval)
- the sample size (i.e., the number of time intervals used in the time series)
- the “effect size” for each finding (i.e., how much the intervention increased or decreased the natural rate of events per time period)
- whether the effect was a short-run or long-run effect (or whether this was not indicated)
- whether the finding was statistically significant
- the standard error or *t* statistic for the change
- the inverse variance weight (to calculate the weighted mean effect size)
- the 95% confidence interval around each effect size

Analyses and findings

We present our analysis of these 86 findings in two general ways—through visual displays and by reporting the meta-analytic statistic of the weighted mean effect size. We do so by first grouping findings into six general intervention categories and then conducting separate meta-analyses for each category when appropriate. We felt that grouping findings into categories of similar interventions was a more meaningful approach to discerning what is known about the effectiveness of different types of counter-terrorism programs than to provide just an overall grand weighted mean effect size. For example, in the case of our 86 findings, the meta-analytic weighted mean effect size across findings in which a standard error could be calculated equaled 0.30 events, with a standard error of 0.656 and a 95% confidence interval of $\{-0.98, 1.59\}$. In other words, overall, there appears to be no statistically significant evidence that interventions had an effect on terrorism. However, it is unclear from this statistic whether this general finding holds true for specific categories of findings.

Thus, we grouped the 86 findings into six categories of interventions. Two categories could be described as measures which attempted to harden targets, three categories focused on deterrence of offenders, and the final grouping was related to changes in political ideology or regimes:

TARGET-HARDENING STRATEGIES.

1. Interventions which increased detection at airports, including installing metal detectors and increasing security screening more generally

2. Interventions which protected embassies and/or diplomats through increased security measures, screening, and the fortification of buildings

DETERRENCE STRATEGIES.

3. Interventions which increased the length and/or severity of punishment for those apprehended and convicted of terrorism
4. United Nations resolutions against terrorism
5. Military interventions and/or retaliations, specifically, the Israeli retaliation attacks on the Palestine Liberation Organization (PLO) and Lebanon in the 1970s and 1980s and the United States of America's attack on Libya in 1986

POLITICAL CHANGES.

6. Changes in political governance, such as a different political party/ideology in power or the end of the Cold War

Figures 3 through 7 illustrate our findings for each of these categories. Each graph has the following attributes:

- The title indicates the category of intervention in which findings were grouped.
- The citations for each finding are listed on the y-axis. Only 56 (of 86) citations are listed, because, in the case of some findings, there are both short-run and long-run findings. The long-run indicator, if given (see below), is visually displayed directly above its short-run counterpart for the same citation.
- The size of the effect is indicated on the horizontal axis. Effects greater than 0 indicate that terrorism events increased after the intervention, and negative effects indicate that terrorism decreased after the intervention.
- A circle with a bar represents the effect size and the 95% confidence intervals for the size of the effect when a standard error was available. When the confidence interval crosses 0, this indicates a non-significant effect. If a bar is not given, then no information was available for us to calculate the confidence interval.
- A stand-alone circle with a dot in the center is sometimes given directly above the 95% confidence interval of a finding which represents a long-run finding, if provided by a study. In these cases, there was no information in which a confidence interval could be calculated.
- In some cases the specific outcomes measured are also indicated next to a finding or group of findings.

Metal detectors and security screening

We first grouped all findings related to interventions that increased detection of potential terrorism through target hardening at airports, one of the most common interventions analyzed by researchers. Specifically, these findings focused on increasing security of airports in the early 1970s, including the installation of metal detectors and the more general increase in security screening of passengers.

Figure 2 displays the array of effects related to this intervention and highlights an interesting nuance. Metal detectors in airports have often been widely recognized as effective in reducing hijackings, and the findings here support this assertion. The weighted mean effect of airport security on hijacking events was a statistically significant reduction of 6.25 events with a 95% confidence interval (CI) of $-8.79, -3.14$. Put another way, the placing of metal detectors in airports in the 1970s after a series of hijackings led to a statistically significant reduction of about six hijackings per unit of time.

However, Figure 2 shows an important qualification to the widely held belief of the success of metal detectors and airport screening. For findings where effect sizes were less than zero, where the intervention led to a decline in terrorism events, all outcomes measured were hijacking events. For those findings indicating a harmful effect (an increase in terrorism events), the outcomes measured were non-skyjacking events. As Cauley and Im (1988), Enders et al. (1990) and Enders and Sandler (1993) have pointed out, this could indicate a substitution or displacement effect of airport security on other types of terrorism. In other words, these researchers have

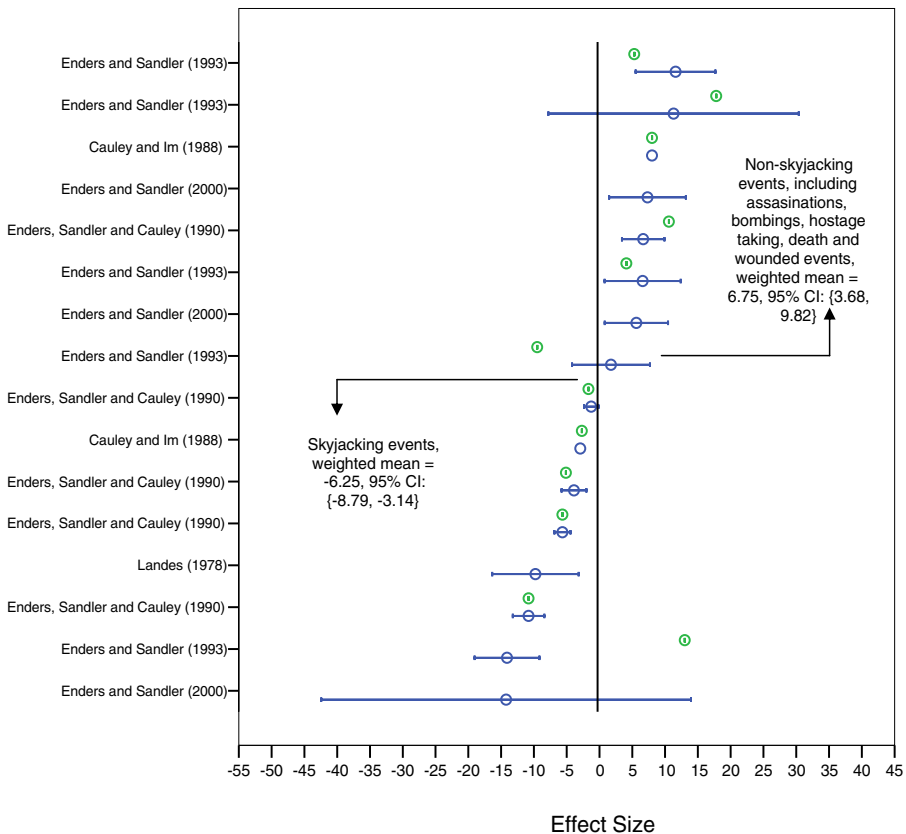


Figure 2. Increased detection: metal detectors and security screening.

suggested that, while airport security may decrease airplane hijacking, other types of terrorism may have increased during the same time, such as miscellaneous bombings, armed attacks, hostage taking, and events which included death or wounded individuals (as opposed to non-casualty incidents) in both the short and long run. The weighted mean effect for these non-hijacking events was a statistically significant *increase* of 6.8 events (CI 3.68, 9.82). When the overall weighted mean effect size is calculated for all the findings examining the effectiveness of metal detectors, the positive and harmful effects cancel each other out. The weighted mean (-0.96) is not statistically significant (CI -2.91, 0.998).

Fortifying embassies and protecting diplomats

Figure 3 reports findings regarding the fortification of embassies and the protection of diplomats through increased security measures and target hardening. As the confidence intervals indicate, many findings were non-significant and close to 0, indicating that there is no existing scientific evidence which points to the effectiveness of these interventions. Indeed, the weighted mean effect size for these findings was not statistically significant (weighted mean effect size = - 0.45, CI -2.17, 1.27). In total, these findings do not indicate that the fortification of

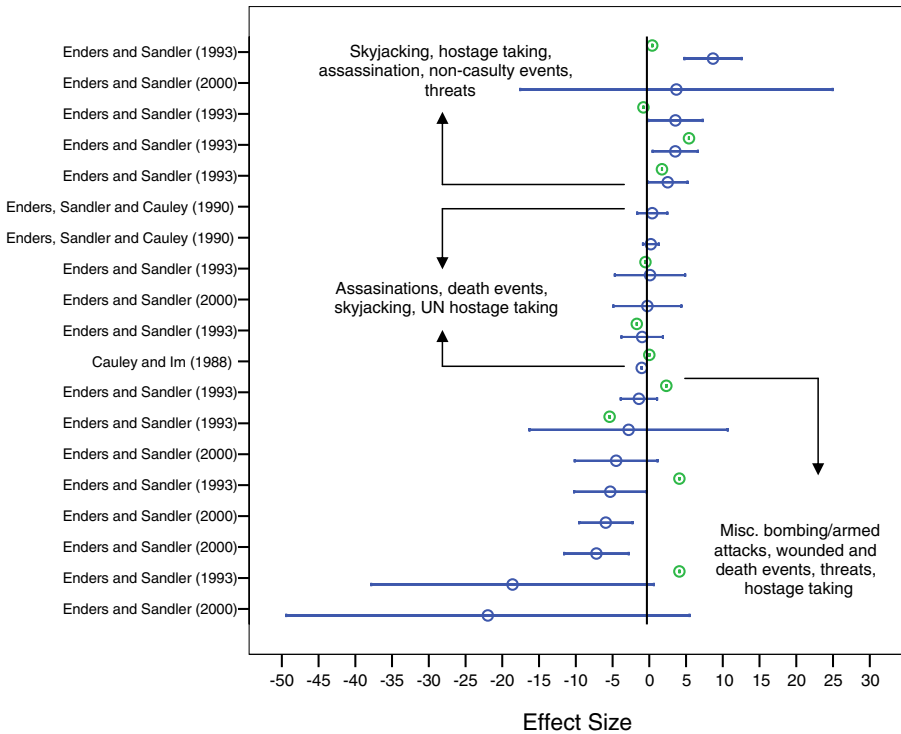


Figure 3. Increased protection: fortifying embassies and protecting diplomats.

embassies and efforts to protect diplomats have been effective in reducing terrorist attacks on these targets. Nor does there appear to be any logical grouping of the types of outcomes in which harmful or beneficial effects might occur.

Increasing the severity of punishment

Landes (1978) provided the only findings concerning increasing the severity of punishment for hijackers who were apprehended, reporting two of the 86 findings (making an illustrative figure or meta-analysis unnecessary). Both findings were non-significant—Landes' work does not show that increasing the severity of punishment had a statistically discernible effect on reducing skyjacking incidents. Again, this does not mean that these strategies “don't work,” but only that when Landes conducted this relatively weak study on data available to him at the time, he did not find any discernible effects. In other words there is no current evidence to show either harmful or beneficial effects of harsher punishment and laws against individuals who carry out acts of terror.

United Nations resolutions against terrorism

Yet another type of intervention found in the evaluation literature concerned the use of United Nations resolutions against terrorism. Although these resolutions are more general in nature, they may provide a deterrent effect on terrorism by establishing international norms which may then strengthen national policies against terrorism or reduce the demand for terrorism (see Telhami 2002 for a discussion of terrorism demand). As Figure 4 shows, Enders et al. (1990) discovered that only a United Nations resolution against aerial hijackings that also supported the use of metal detectors in airports appeared effective in reducing the number of skyjacking events in both the short and long term (the finding at the bottom of the graph). However, the second finding illustrated in the middle of Figure 4 indicates that resolutions without the implementation of metal detectors were not useful in reducing terrorism. Further, resolutions intended to “prevent and punish crimes against internationally protected persons” did not seem to have a statistically discernible effect.

Military retaliations

One often-researched counter-terrorism strategy has been the United States of America's 1986 attack on Libya after Libya's involvement in the bombing of the LaBelle Discotheque in West Berlin. While some have incorrectly reported the effects of the raid as reducing terrorism (see Prunckun and Mohr 1997), it is generally believed that this raid increased terrorist attacks, at least in the short run (see Silke 2005). Brophy-Baermann and Conybeare (1994) also reported on findings regarding military strikes by Israel on PLO targets.

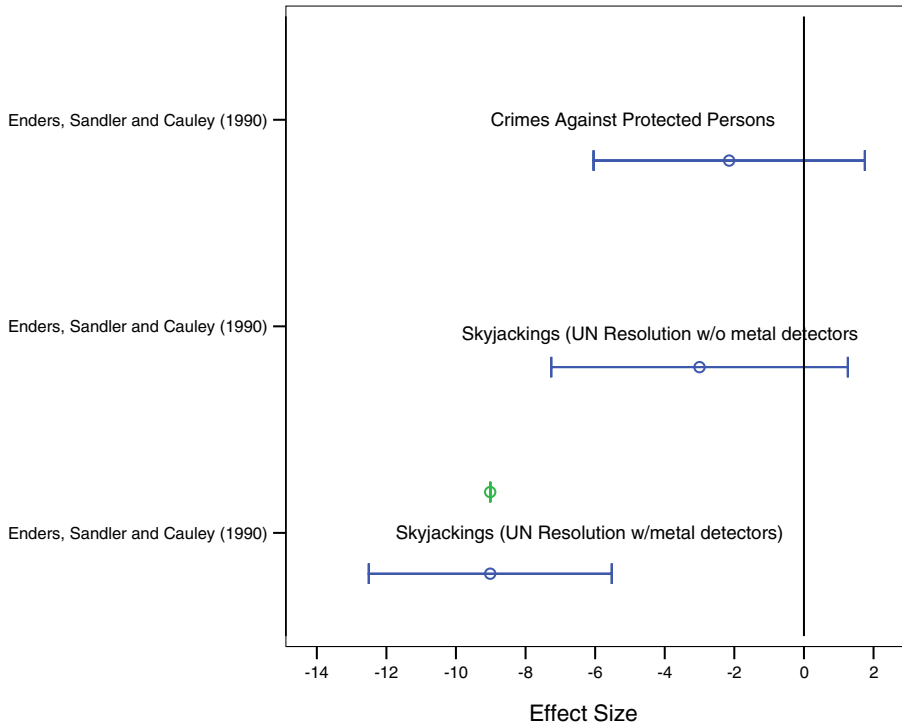


Figure 4. Resolution interventions: United Nations resolutions against terrorism.

The findings in Figure 5 point to the short-run effects of the attack on Libya (the short-run distinction is suggested by the authors themselves, as unlike the use of metal detectors, the attacks on Libya were not sustained over the time series) as well as Israeli retaliatory strikes on Palestinians. Figure 5 generally shows that the attack on Libya resulted in a statistically significant *increase* in the number of terrorist attacks in the short run, with a weighted effect size of 15.33 events, with a 95% confidence interval of 3.46, 27.20. However, Figure 5 also shows that the Libyan attack affected non-casualty events, threats and miscellaneous bombings more so than “resource-utilizing” (Enders et al. 1990) attacks such as hijackings, hostage events, and events which led to death or wounded individuals. Additionally, the specific retaliations researched seemed to increase attacks on the United States of America, the United Kingdom, and Israel. The harmful effects of these interventions were not sustained over the long run, and no statistically significant long-run effects of military interventions were found in these studies.

Changes in political governance

Finally, we examined findings that we generally label as “political changes”, specifically Barros’ (2003) discussion of the election of the Socialist Party in Spain

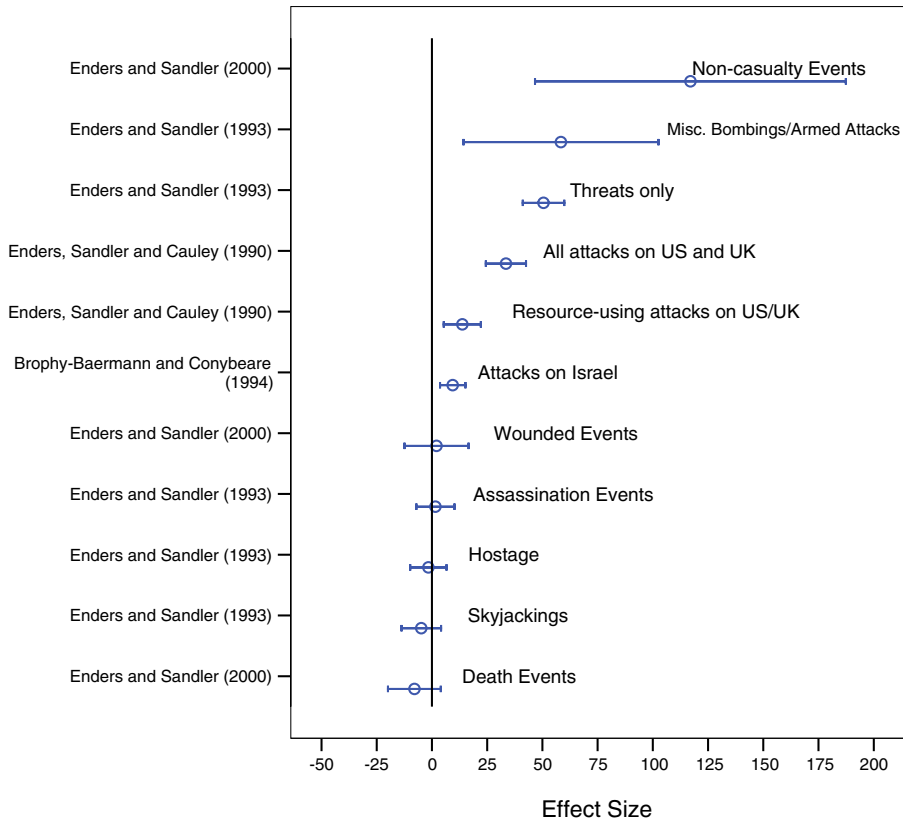


Figure 5. Military retaliation by states: Israeli military-led retaliation attacks on the PLO and Lebanon and United States retaliatory raids on Libya.

(which he describes as more intolerant and harsh against rogue political groups like ETA) and Enders and Sandler’s (2000) discussion of the effects that the end of the Cold War had on patterns of terrorism. Whether these might be considered interventions are debatable; a political party change or a shift in political ideologies may or may not be due to the public’s attempt to stop terrorism, and, therefore, could be described as a deliberate or non-deliberate attempt at a counter-terrorism strategy. More obviously, the end of the Cold War was not specifically sought to reduce terrorism. Yet, these changes are interesting because these findings illustrate the broader political nature of terrorism and its responses.

The combination of findings in Figure 6 indicates an uncertainty about the effects of political change on terrorism. The weighted mean effect size of these findings was 6.16, indicating that these “strategies” led to an increase in terrorism, although the confidence interval indicated a null effect (CI -0.46, 12.79). However, the homogeneity test did not indicate a strongly significant chi-squared statistic, and when a fixed-effects model of this category was run, it revealed a statistically significant harmful effect of these changes on terrorism (weighted

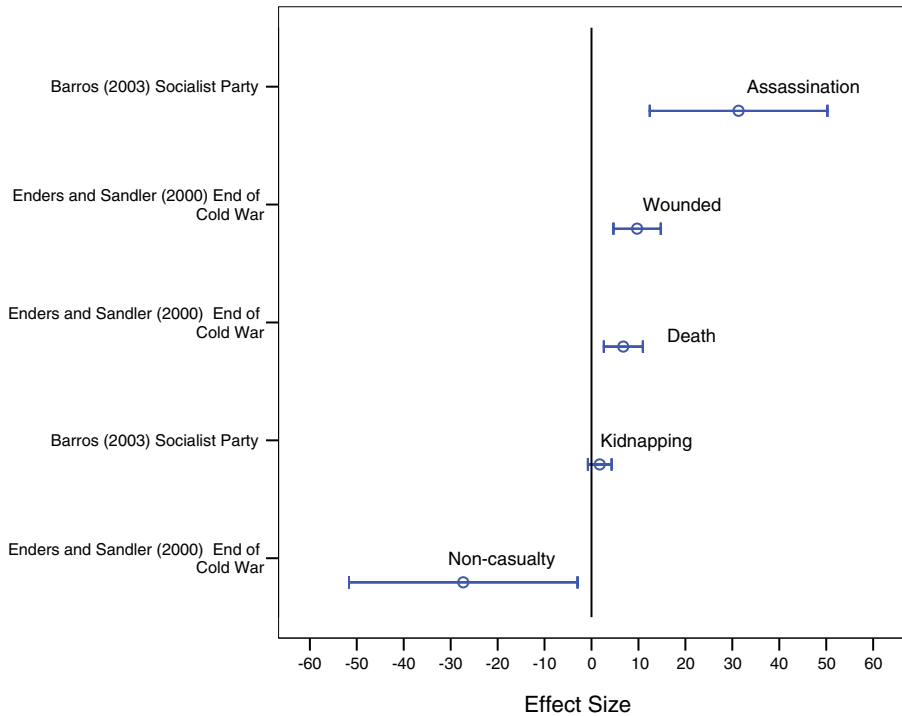


Figure 6. Political governance: socialist party in power or after the Cold War.

mean=4.23, CI 2.56, 6.20). Also interesting when the individual findings were examined was that the harmful effect of both an intolerant party as well as the end of the Cold War was reflected in more dangerous outcomes (assassinations and events which led to individuals becoming wounded or killed), while these aspects of political governance reduced the likelihood of less serious, non-casualty events.

Conclusions and recommendations

Two main points are emphasized from these findings. First, many of these interventions did not have a statistically discernible effect on reducing terrorism across time and, in some cases, led to increases in terrorism. Secondly, there is more uncertainty than certainty about the effectiveness of counter-terrorism programs, mostly because only a small subset of strategies have actually been evaluated, and some evaluations are of modest quality (although Enders and Sandler have advanced this shortcoming). Also, the Campbell process tends to focus on quantitative evaluations, and, as some reviewers of this paper have remarked, useful qualitative evaluations may exist.¹⁹ Indeed, the aforementioned limitations of this review highlight how the scarcity of evaluation research leaves us with limited ability to draw strong conclusions about the effectiveness of counter-terrorism strategies.

However, there are specific lessons within each intervention category that cannot be ignored. While metal detectors are effective in reducing airplane hijackings, Cauley and Im (1988) and Enders and his colleagues (1990, 1993, 2000) have found that there may be displacement or substitution effects leading to increases in other types of terrorism not involving aircraft. This type of target hardening remains one of the more researched counter-terrorism tools and, indeed, suggests that recent attempts to improve airport security and the establishment of special agencies like the Transportation Security Administration (TSA) may be fruitful endeavors. Furthermore, these findings suggest that the increasing of security at other types of transportation hubs such as train, subway, and bus stations may also prove useful. However, as the evidence indicates, individuals can also change their targets, their mechanism of violence, or the extent of damage they wish to inflict.

The fortification of embassies and the protection of diplomats have not been shown to be particularly effective strategies and, in specific situations, may increase terrorism against these targets. This is somewhat surprising given the results of airport security which are also target-hardening measures. The question of interest for current counter-terrorism efforts may be why these target-hardening strategies do not work as well as other target-hardening strategies such as airport security. The difference may lie in the possibility that airports are much more controlled environments than embassies, buildings, or exposed diplomats. The combination of natural controls of the environment may help facilitate target-hardening strategies in such a way that makes them more effective.²⁰ A similar finding was discovered when Welsh and Farrington (2006) examined the effects of CCTV on crime. They found that CCTV worked much better in car parks than in other settings, such as public housing complexes.

In terms of attempting to deter would-be offenders wielding terrorism, we do not know much about whether increasing the severity or certainty of punishment is an appropriate deterrent. The weak evidence indicates ineffectiveness of these strategies, which seems logical given that the motivations for terrorism can be much stronger than the threat of apprehension, punishment, or even death. Additionally, deterrent effects may also be mediated by the nature of the target, the offenders, and the political motivation. However, because of the equivocal nature of these evaluations, it is irresponsible to draw conclusions from them. More is needed, for example, on gauging the deterrent effects of the U.S. Patriot Act, in order to discern the effects of such laws on individuals motivated to use terrorism.

Regarding United Nations resolutions and conventions, it is clear that to stop terrorism or other international crimes requires much more than these international legal resolutions. As Telhami (2002) suggests, it requires reducing both the supply of terrorism and the demand for terrorism in addition to de-legitimizing it through resolutions or conventions. The de-legitimization of the Irish Republican Army (IRA) in the 1990s is an example of how the reduction in citizen demand for terrorism, as represented by direct, indirect, or rhetorical support, was probably significantly responsible (as compared with government action alone) in reducing IRA terrorism. However, government agencies also acted during this time of de-

legitimization by implementing major reforms in the police service through the Patten Commission (see Independent Commission on Policing in Northern Ireland 1999). Thus, both symbolic gestures (like resolutions and conventions) and real actions (for example, conventions combined with metal detectors) are needed to counter terrorism, not simply one or the other.

Military retaliations can increase terrorism in the short run and, over the long run, may not affect terrorism at all. This finding has recently been supported in preliminary work by Sherman (2005) on how the United States-led offensive in Iraq has led to increases in suicide bombings. While military retaliations may be seen as justified for reasons other than the prevention and deterrence of future terrorism (for example as punishment or detection of offenders), the costs of such interventions may not only be monetary but actually lead to a short-term increase in these events. At the same time, military interventions are an example of how some strategies might vary dramatically in terms of short-run versus long-run effects.

The findings on changes in governance provide an interesting nuance to counter-terrorism efforts. Like crime, broader social, political and economic factors can also affect trends in terrorism. Here, it appears that, while the end of the Cold War led to a decline in non-casualty terrorism incidents, there may have been an increase in more violent ones. Of course, this does not suggest that democratization should then be avoided, because it could lead to more violent conflicts between groups, nations, or between groups and nations. Nonetheless, these findings remind us that counter-terrorism efforts, sometimes more so than ordinary crimes, occur within a political context that cannot be ignored.

Perhaps what is equally (if not more) interesting is what we did not find from our review. Only a small subset of interventions was analyzed across the seven studies, using one primary method (time series) and one general outcome type (incidents). However, when we examined the 94 studies which seemed remotely connected to some evaluation, a number of interventions were discussed and are mentioned in the list at the beginning of this article. We suspect that this list only partially represents the different types of counter-terrorism interventions that can be evaluated, and, if anything, our review suggests we know very little about the effects and effectiveness of counter-terrorism tactics, strategies, and tools.

The findings confirm our initial hypotheses about the state of counter-terrorism strategies that were derived from the general overview of terrorism research. There has been a proliferation of anti-terrorism programs and policies as well as massive increases in expenditures toward combating terrorism. Yet, we currently know almost nothing about the effectiveness of any of these programs. From the over 20,000 pieces of literature found on terrorism, there have been only seven which have evaluated the effectiveness of counter-terrorism programs in a moderately rigorous way. This is a compelling finding—it says that we actually do know some things about terrorism generally, but we hardly know anything about whether our efforts to counter it are effective.

The call for more evaluation research and funding for evaluation research on counter-terrorism interventions, however, is more easily said than done. The small

amount of evaluation research in the area of terrorism may be due to a number of reasons. First, practical problems in either the qualitative or quantitative study of terrorism are often thought to be significant obstacles to overcome. Not only are terrorism events rare, but other units of analysis (individuals or groups who employ terror violence) may be difficult to locate and study. Unlike criminological research that contains large numbers of data from frequently occurring events, terrorism research relies on much scarcer occurrences, which makes analysis more challenging (in terms of identifying patterns). Common challenges with using crime data are also accentuated in terrorism prevention programs. These include the difficulty in detecting intervention effects of major programs where treatment effects may be unclear, or where alternative explanations are difficult to control for. For terrorism, threats are often not clearly defined, so the impacts of the programs are almost impossible to assess.

Compounding this problem is the nature of the counter-terrorism enterprise itself, which, unlike much of current local policing, is shrouded in secrecy. Because of this, information about incidents and outcomes is difficult to collect for scientific purposes. There is also a real problem in establishing, for the data that are available, clear standards of accuracy and reliability, problems that have been exhaustively addressed in the reform of police statistical standards through Uniform Crime Reports (UCRs) and other standardization procedures. Further, in terms of evaluation, there is likely to be less willingness in this area to engage in experimental programs that might test differential program effects (a reluctance that was, to a degree, overcome in police evaluations, such as those performed by Sherman et al. (1992) in the area of domestic violence interventions). The evaluations, then, would have to be conducted *ex post facto*, as illustrated in the case studies that we report above. While these studies reveal some interesting and important effects, their real impact is diminished through the inability to adjust program effects or increase the sensitivity of measurement in an ongoing assessment process.

It is certainly the case that researchers can extend the types of evaluations that would use existing data sets to test the effects of major interventions, such as the Patriot Act. The availability of the ITERATE data base (by Edward Mickolus and colleagues)²¹, the files available from MIPT, and the new data collected by Gary LaFree and his colleagues at the Department of Homeland Security START Research Center at the University of Maryland²² have and will encourage more detailed, albeit retrospective, analyses of terrorism. As we move forward, it is important that these data are improved through the addition of important contextual information that will help uncover the multitude of factors that influence these events. And, there needs to be a new initiative that evolves from the intelligence community that would mirror the increased collaborations between researchers and police over the past few decades.

Addressing these concerns requires building and improving the infrastructure of terrorism research. Briefly,²³ this endeavor includes many suggestions related to crime prevention evaluation research more generally. These include:

- changing the focus of government-sponsored research evaluations towards more methodologically rigorous outcome evaluations of multiple types of counter-terrorism programs

- exploring alternative methods for evaluation, especially for rare events
- creating access routes to both classified and unclassified data for researchers
- improving and increasing dialog between researchers, policy makers and practitioners, including overcoming myths, fears, and uncertainties about the intentions and capabilities of all three groups
- building into public policies and laws requirements for evaluation of those policies
- developing mechanisms of both delivery and interpretation of research evaluation results to practitioners and policy makers

Evidence-based counter-terrorism policy should be lawful, rational, effective, and should cause as little harm as possible. This means that evaluation researchers must provide policy makers with rigorous outcome evaluations, attempt to match evaluations with the breadth of interventions and information that exists, and consider the goals of decision makers. At the same time, those who create policies need to pay attention to available scientific evidence or push for the generation of more evidence if there is not enough (as is the problem in counter-terrorism policy). These goals can be facilitated by developing an infrastructure for counter-terrorism evaluation research. This Campbell systematic review clearly indicates that more is needed.

Acknowledgements

We would like to acknowledge with gratitude the financial support of the Campbell Collaboration and the Jerry Lee Center of Criminology at the University of Pennsylvania for supporting the preparation of the protocol and review. The original funding for the bibliographical review was provided by the Office of the Provost, Rutgers Newark and the Rutgers School of Criminal Justice to the Center for the Study of Public Security at Rutgers Newark. In particular, we are indebted to David Wilson for his methodological assistance and comments. Especially helpful comments were provided by Peter Grabosky, Phil Davies, and the anonymous reviewers. Finally, we thank Louise Stanton, Michelle Grillo, Lauren Giordano, Deborah Vegh and Sue-Ming Yang for their research assistance.

Notes

- 1 See <http://www.house.gov/budget/>.
- 2 See <http://www.aic.gov.au/campbellcj/>.
- 3 See also <http://www.campbellcollaboration.org/Fraguidelines.html>.
- 4 For the full report, see http://www.campbellcollaboration.org/doc-pdf/Lum_Terrorism_Review.pdf.
- 5 The briefing was in reference to the 2005 Country Reports on Terrorism and was held on April 28, 2006. The briefing and full report can be found at <http://www.state.gov/s/ct/rls/rm2/65482.htm> and <http://www.state.gov/s/ct/rls/crt/c17689.htm>, respectively.
- 6 Academic Search Premier, ArticleFirst (OCLC), Contemporary Women's Issues, Criminal Justice Abstracts, EbscoHost, EconLit, Educational Abstracts,

- Electronic Collections Online, ERIC(OCLC), GEOBASE, Humanities Abstracts, Ingenta, ISI Web of Science, MEDLINE, National Criminal Justice Reference Service, PAIS International Articles Only, PUBMEDLINE, Social Science Abstracts, Sociological Abstracts.
- 7 This initial review was initiated by the authors in 2003; hence, only literature through the end of 2002 is represented.
 - 8 The complete list of websites and available online databases we researched are listed in Appendices A and B of the main report to the Campbell Collaboration, located at: http://www.campbellcollaboration.org/doc-pdf/Lum_Terrorism_Review.pdf.
 - 9 The full systematic review can be accessed at http://www.campbellcollaboration.org/doc-pdf/Lum_Terrorism_Review.pdf.
 - 10 Graduate assistants helped search for books at Harvard University, Northeastern University, Rutgers University and the University of Maryland at College Park.
 - 11 While we were completing the full report, one anonymous reviewer questioned why we did not consider studies examining the effectiveness of terrorism itself. We believed that this literature was generally inappropriate for this review unless the study authors framed their interest in the effectiveness of terrorism in terms of evaluating how terrorism might thwart a counter-terrorism strategy.
 - 12 The Scientific Methods Scale can be found in chapter 2 of the Report and can be found at <http://www.ncjrs.org/works/chapter2.htm>.
 - 13 This is mentioned in Sherman et al. (1997) on pages 2–19, note 7.
 - 14 For example, a Cochrane review related to anthrax (Jefferson et al., 1998) already exists. There has also been a large literature on the treatment of post-traumatic stress disorder (Bisson and Andrew 2005; Rose et al. 2002; Stein et al. 2000), and for us to include one article without including others simply because one referenced terrorism victims would be unfair to the literature.
 - 15 See <http://www.cochrane.org/>.
 - 16 David Wilson (George Mason University) in personal correspondence to the authors on June 16, 2005.
 - 17 See Lipsey and Wilson (2001, pp. 113–114) for computation formalis. Also, see worksheets provided by David Wilson at <http://mason.gmu.edu/~dwilsonb/downloads/overview.ppt>.
 - 18 We recognize that incidents such as skyjackings and the number of fatalities are two different types of “events”. In a skyjacking, there could be multiple deaths. We will specifically note this on visual displays of information as well as on analysis that uses both these types of measures.
 - 19 To address this limitation the lead author is undertaking an additional study examining whether there exist comparable qualitative evaluations of these counter-terrorism programs.
 - 20 This suggestion was pointed out to the authors by Dr. Christopher Koper (University of Pennsylvania) in personal correspondence.
 - 21 The ITERATE data is housed at ICPSR (see <http://www.icpsr.umich.edu/>).
 - 22 See <http://www.start.umd.edu/>.
 - 23 The creation of an infrastructure for terrorism research is discussed in detail in a forthcoming article by Lum and Kennedy.

References

- Abadie, A. & Gardeazabal, J. (2001). *The economic costs of conflict: A case control study for the Basque Country*. NEBR Working Paper 8478. Cambridge, MA.
- Barros, C. P. (2003). An intervention analysis of terrorism: The Spanish Eta case. *Defence and Peace Economics* 14(6), 401–412.
- Bisson, J. & Andrew, M. (2005). Psychological treatment of post-traumatic stress disorder (PTSD). *The Cochrane Database of Systematic Reviews*, Issue 3. Art. No.: CD003388. pub2. doi: [10.1002/14651858.CD003388.pub2](https://doi.org/10.1002/14651858.CD003388.pub2).
- Boruch, Robert F., Snyder, B. & DeMoya, D. (2000). The importance of randomized field trials. *Crime and Delinquency* 46, 156–180.
- Brophy-Baermann, B. & Conybeare, J. A. (1994). Retaliating against terrorism: Rational expectations and the optimality of rules versus discretion. *American Journal of Political Science* 38(1), 196–210.
- Cauley, J. & Im, E. (1988). Intervention policy analysis of skyjackings and other terrorist incidents. *The American Economic Review* 78(2), 27–31.
- Congressional Budget Office. (2002). *The budget and economic outlook: Fiscal years 2003 to 2012*. Washington, DC: Congress of the United States.
- Congressional Budget Office. (2005). *The budget and economic outlook: Fiscal years 2006 to 2015*. Washington, DC: Congress of the United States.
- Cook, T. & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Crenshaw, M. (1992). Current research on terrorism: The academic perspective. *Studies in Conflict and Terrorism* 15(1), 1–11.
- 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., & Cauley, J. (1990). UN conventions, terrorism, and retaliation in the fight against terrorism: An econometric evaluation. *Terrorism and Political Violence* 2 (1), 83.
- Farrington, D. (2003). Methodological quality standards for evaluation research. *The Annals of the American Academy of Political and Social Science* 587(1), 49–68.
- Farrington, D. & Petrosino, A. (2001). The Campbell Collaboration Crime and Justice Group. *Annals of the American Academy of Political and Social Science* 58, 35–49.
- Guinnessy, J. & Dawson, P. (2002). Terrorism drives Bush R and D money to defense and NIJ; The science funding flat in fiscal 2003. *Physics Today* 55, 30.
- Halkides, M. (1995). How not to study terrorism. *Peace Review* 7, 253–260.
- Hoffman, B. (1992). Current research on terrorism and low-intensity conflict. *Studies in Conflict and Terrorism* 15, 25–37.
- Independent Commission on Policing in Northern Ireland (Patten Commission). 1999. *A New Beginning: Policing in Northern Ireland*.
- Issues in Science and Technology (Report). (2002). Federal R and D in FY 2002 will have biggest percentage gain in 20 years. *Issues in Science and Technology* 18, 23.
- Jefferson T., Demicheli, V., Deeks J., Graves, P., Pratt, M. & Rivetti, D. (1998). Vaccines for preventing anthrax. *The Cochrane Database of Systematic Reviews*, Issue 1. Art. No.: CD000975. doi: [10.1002/14651858.CD000975](https://doi.org/10.1002/14651858.CD000975).

- Kennedy, L. & Lum, C. (2003). *Developing a foundation for policy relevant terrorism research in criminology*. Prepared for the Center for the Study of Public Security, Rutgers University, Newark, New Jersey.
- Landes, W. M. (1978). An economic study of U.S. aircraft hijackings, 1961–1976. *Journal of Law and Economics* 21, 1–31.
- Lipsey, M. & Wilson, D. (2001). *Practical meta-analysis. Applied social research methods series 49*. Thousand Oaks: Sage.
- Macilwain, C. (2002). Bush goes to war as budget boosts R and D. *Nature* 415, 564.
- Merari, A. (1991). Academic research and government policy on terrorism. In C. McCauley (Ed.), *Terrorism research and public policy*. London: Frank Cass.
- Miller, R. (1988). The literature of terrorism. *Terrorism* 11, 63–87.
- Prunckun, H. & Mohr, P. (1997). Military deterrence of international terrorism: An evaluation of Operation El Dorado Canyon. *Studies in Conflict and Terrorism* 20, 267–280.
- Romano, T. (1984). *Terrorism: An analysis of the literature*. Dissertation, Fordham University, Department of Sociology, Criminology and Penology.
- Rose, S., Bisson, J., Churchill, R. & Wessely, S. (2002). Psychological debriefing for preventing post traumatic stress disorder (PTSD). *The Cochrane Database of Systematic Reviews*, Issue 2. Art. No.: CD000560. doi: [10.1002/14651858.CD000560](https://doi.org/10.1002/14651858.CD000560).
- Schmid, A. P. & Jongman, A. J. (1988). *Political terrorism: A new guide to actors, authors, concepts, databases, theories and literature*. Amsterdam: North-Holland Publishing Company.
- Shadish, W., Cook, T. & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inferences*. Boston: Houghton-Mifflin.
- Sherman, L. W. (2005). Enlightened Justice: Reducing Human Misery through Empirical Consequentialism. Presidential Panel at the 14th World Congress of Criminology. Philadelphia, PA.
- Sherman, L. W., Schmidt, J. D. & Rogan, D. P. (1992). *Policing domestic violence: Experiments and dilemmas*. New York: Free Press.
- Sherman, L. W., Gottfredson, D., MacKenzie, D. L., Eck, J., Reuter, P. & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising: A report to the United States Congress*. Washington, DC: National Institute of Justice.
- Sherman, L. W., Farrington, D. P., Welsh, B. C. & MacKenzie, D. L. (Eds.). (2002). *Evidence based crime prevention*. London: Routledge.
- Silke, A. (ed.). (2004). *Research on terrorism: Trends, achievements and failures*. New York: Taylor and Francis.
- Silke, A. (2005). Fire of Iolau: The role of state counter-measures in causing terrorism and what needs to be done. In T. Bjorgo (Ed.), *Root causes of terrorism: Myths, reality and ways forward*. Oxford, UK: Routledge.
- Stein, D. J., Zungu-Dirwayi, N., van der Linden, G. J. H. & Seedat, S. (2000). Pharmacotherapy for post traumatic stress disorder (PTSD). *The Cochrane Database of Systematic Reviews*, Issue 4. Art. No.: CD002795. doi: [10.1002/14651858.CD002795](https://doi.org/10.1002/14651858.CD002795).
- Telhami, S. (2002). *The Stakes: American in the Middle East - The Consequences of Power and the Choice for Peace*. Boulder, CO: Westview Press.
- Welsh, B. & Farrington, D. (2006). Closed-Circuit Television Surveillance. In B. Welsh & D. Farrington (Eds.), *Preventing crime: What works for children, offenders, victims, and places*. Dordrecht, Netherlands: Springer.
- Wilkinson, P. (1977). *Terrorism and the liberal state*. New York: John Wiley and Sons, Inc.
- Wilkinson, P. (1986). *Terrorism and the Liberal State*, rev. ed. London: Macmillan.

About the authors

Dr. Cynthia Lum is an Assistant Professor in the Administration of Justice Program at George Mason University. Her current projects include exploring the link between counter-terrorism and crime prevention, measuring democratic policing across thirty-one transitioning nations, developing police information and crime analysis tools, and training police commanders from developing nations through the U.S. State Department.

Dr. Leslie W. Kennedy Dr. Leslie W. Kennedy is Professor and Dean at the Rutgers School of Criminal Justice. He has published extensively in the areas of fear of crime, victimology, and violence. Among his published works, he is the co-author of *The Criminal Event*, in which he advocates a holistic approach to the study crime in social context. Dr. Kennedy's current approach to risk balance in public security draws on a multi-disciplinary background including law, health sciences, and criminology drawing on the point of view of the event analysis.

Dr. Alison J. Sherley Dr. Alison J. Sherley received her PhD from Rutgers School of Criminal Justice in 2005. Her interests include vulnerability and risk, terrorism, research methods and statistics.