

USING OPEN SOURCE DATA TO TRACK WORLDWIDE TERRORISM PATTERNS*

DOI: 10.20542/2307-1494-2017-1-64-76

Keywords: Global Terrorism Database, methodology, open-source data, global trends, terrorist attacks, fatal attacks, tactics, weapons, targets, Russia, United States

Abstract: The article is the first publication in Russia providing a first-hand review and introduction into the methodology of the U.S.-based Global Terrorism Database. It also displays some key findings from the GTD data. The findings show trends in total and fatal terrorist attacks, both globally and, specifically, for Russia and the United States, as well as in terrorist tactics, weapons and targets. It concludes by assessing comparative strengths and weaknesses of open-source data collection and analysis in terrorism studies.

Ключевые слова: Глобальная база данных по терроризму, методология, данные из открытых источников, глобальные тенденции, теракты, атаки со смертельным исходом, тактика, вооружения, мишени, Россия, США

Аннотация: Данная статья представляет собой первый опубликованный в России обзор Глобальной базы данных по терроризму (США) «от первого лица» и введение в ее методологию. В ней также представлен ряд тенденций и выводов, полученных на основе анализа статистики этой базы данных, в том числе о динамике общего числа терактов и терактов с человеческими жертвами как в мире в целом, так и отдельно для России и США, а также тенденции в области используемых террористами методов, вооружений и мишеней терактов. В заключении в статье проводится сравнительный анализ сильных и слабых сторон сбора и анализа данных из открытых источников в области исследований терроризма.

I. Introduction

In recent years, the explosive growth of online media availability is ushering in a new wave of innovation in terms of the collection of worldwide open source terrorism data – that is, data collected from unclassified print and electronic media sources. Attempts to automate the collection of open source terrorism data received a major boost in the late 1960s with the rise of satellite technology and hand held audio camera equipment. Early efforts to collect data relied especially on global newspapers like the London-based “Financial Times” and “The New York Times”, but increasingly came to

* Parts of this article were presented at the U.S.–Russia Joint Inter-Academy Effort on Countering Terrorism and Violent Extremism at IMEMO, Moscow, 1–3 November 2016. The author would like to thank Bo Jiang for data support.

rely on news aggregators like the wire service Reuters and later on-line aggregators like “Lexis-Nexis”, “Factiva” and “OpenSource”. Over time, these efforts have become more sophisticated and comprehensive. This article briefly describes the Global Terrorism Database (GTD) which is currently the most comprehensive of these open source terrorism databases. The GTD is collected by a team at the University of Maryland and at present includes data on 156772 terrorist attacks worldwide from 1970 to 2015. After describing the methods used to collect the GTD, the article provides some illustrative global terrorism trends and then follows this up by looking specifically at terrorism trends for Russia and the United States.

II. Collecting the Global Terrorism Database

Before counting terrorist attacks, we must first define them. At present, there are no universally accepted, worldwide definitions of terrorism. Even the United Nations has thus far failed to come up with a universal definition. In this environment we have tried to create a definition of terrorism that is as close to being generally accepted as possible, but at the same time gives researchers and policy makers some flexibility in terms of how they define terrorism.

The GTD defines terrorism as “*the threatened or actual use of illegal force and violence by non-state actors to attain a political, economic, religious, or social goal through fear, coercion, or intimidation.*” This definition includes the assumption that terrorism may involve the threatened in addition to the actual use of violence: individuals who seize an aircraft and say they will blow it up unless their demands are met may threaten violence without actually using it. At the same time, the GTD excludes idle threats such as bomb hoaxes made by phone or threats against the life of world leaders that are never acted upon. In addition, the requirement that these events be limited to the actions of non-state actors means that the GTD does not include the considerable violence and terror that are directly carried out by governments or their militaries. And the requirement that the act have a direct political goal means that it excludes ordinary criminal violence.

In order to identify the small subset of articles that describe terrorist attacks our GTD team begins each day with a universe of more than 1.6 million articles published worldwide.¹ The team uses customized search strings to isolate an initial pool of potentially relevant articles and then relies on natural language processing methods to identify and remove duplicate source articles by measuring similarities between pairs of documents. In addition, the team has developed a machine-learning model using feedback from trained GTD staff that classifies documents identified by the initial automated processes to determine how likely they are to be relevant to terrorism. This model is continually refined using input from the research team regarding the accuracy of the classification results. To facilitate this iterative process, the team has developed a web-based interface used to provide continuous feedback to the system through the manual review of the source documents identifying both false-positives (source

documents that appear to describe terrorist attacks but do not) and false-negatives (source documents that appear to not describe terrorist attacks but actually do).

The GTD team reviews all the source documents that have been classified as relevant by the machine learning model and generates database entries for individual attacks that satisfy the GTD inclusion criteria. GTD analysts use script analysis tools to facilitate this process by clustering similar documents together based on key identifying features of the text. In addition, as the set of identified incidents expands, they use this information to supply coders with details of already created events or related sources that are potential matches for a given attack under review. The GTD team applies automated tools, including Boolean filtering, Natural Language Processing, Named Entity Recognition, and Machine Learning models to the source documents to begin the event definition process in order to identify those events most likely to qualify as terrorist attacks. At present, approximately 115000 articles are manually reviewed to identify attacks for each month of data collection.

To this point in time most open source terrorism data collection efforts have relied on experts for different regions of the world, like Latin America or Southeast Asia.² The GTD instead relies on domain-specific research teams organized to collect data on specific characteristics of attacks, including location, perpetrators, targets, weapons, tactics, casualties, and consequences. Each domain-specific team records information according to the ever-evolving specifications of the GTD Codebook.³ In short, the GTD team uses automated tools to process millions of documents a day but human coders to digest the information and ensure the quality of the resulting data.

Based on these processes the GTD team reduces the stream of data to about 50000 articles per week – small enough that we can process them with our staff of about 25 researchers and students. Only about 7% of these cases will eventually be included in the database. We collect about 120 individual pieces of information on each attack. The next several sections provide an overview of worldwide terrorism characteristics based on an analysis of the GTD.

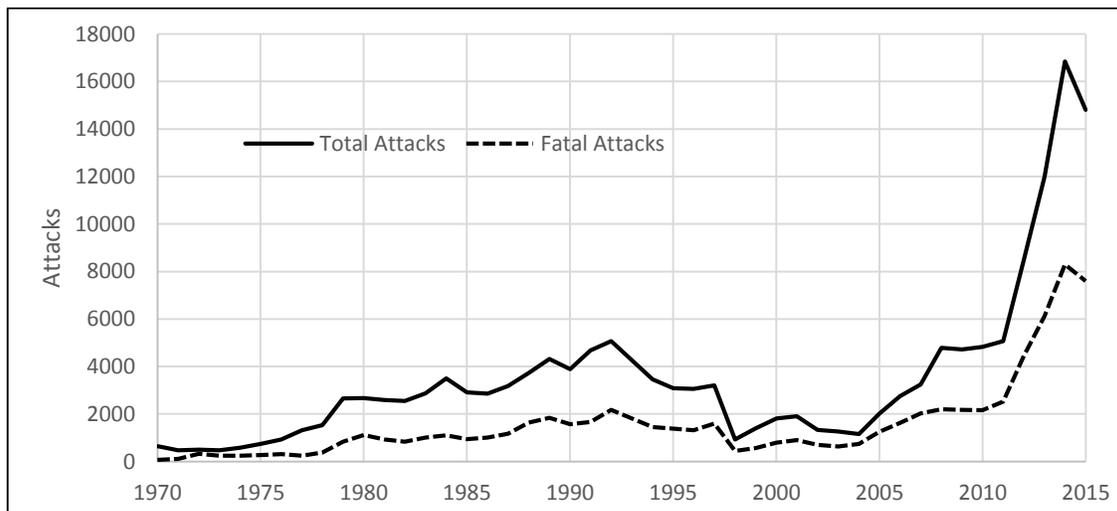
III. Worldwide trends in total and fatal attacks

Figure 1 shows worldwide terrorism trends in total attacks and fatal attacks from 1970 to 2015. According to Figure 1, attacks were relatively infrequent during the early 1970s, with fewer than 1000 incidents each year until 1977. However, we see steady increases throughout the decade: between 1970 and 1979, the number of attacks increased by more than 300%, from 651 to 2661. This rise is associated especially with high levels of activity during the 1970s in Western Europe and the United States. For example, 47% of all 1970s attacks in the GTD were from Western Europe. Leading countries here included Northern Ireland (for convenience treated here as a country; 32%), Italy (22%), and Spain (19%). At the same time, 14% of all 1970s terrorist attacks in the GTD were from the United States.

Trends and locations of terrorism changed considerably in the 1980s. The annual frequency continued to increase throughout the 1980s until the 1992 peak (5078 attacks), with smaller peaks in 1984 (3494 attacks) and 1989 (4322 attacks). This steady rise in attacks was due in large part to a surge of attacks in Latin America. More than 55% of all terrorist attacks in the 1980s took place in South America (31%) and Central America and the Caribbean (24%). Groups that were especially important during this period include “Sendero Luminoso” in Peru, the FARC in Colombia and the FMLN in El Salvador. After 1992, the number of terrorist attacks dropped dramatically to a twenty-year low in 1998. Declines in attacks before and after 1990 were especially pronounced in El Salvador (where total attacks dropped by 82% from the 1980s to the 1990s) and Guatemala (where attacks dropped by 71% during the same period).

Figure 1 shows another major transition in the 2000s. Total attacks in 2000 (1813) were just a few hundred more than the corresponding figure for 1978 (1526). However, total attacks rose again sharply in the aftermath of the United States and its allies invading Iraq in 2003. By 2011, total attacks (5065) were barely less than the record level experienced in 1992. Since 2011, total attacks have shattered all previous records. In 2013, total attacks stood at 11952, or 135% higher than the peak in 1992. Interestingly, both total attacks and worldwide attacks dropped in 2015 – for the first time in more than a decade. Taken together, this ebb and flow results in a pronounced U-shape pattern in total terrorist attacks from 1992 to 2014, with a decline for the first time in 2015.

Figure 1. Terrorist attacks worldwide, 1970–2015



Source: the Global Terrorism Database

Less than half of all attacks result in fatalities (1575 fatal attacks per year compared to 3483 total attacks per year worldwide on average). Despite this difference, the two time series are highly correlated (correlation coefficient = 0.99). Until 1979, the

GTD recorded less than 400 fatal terrorist attacks per year. Between 1978 and 1979, fatal attacks more than doubled (from 374 to 836). Throughout most of the 1980s, fatal attacks hovered close to 1000 each year. The trend shifted again in 1988, after which fatal attacks rose to a 1992 peak of 2173 fatal terrorist attacks. Fatal attacks then declined over time, bottoming out in 1998 with 451 fatal attacks. This was followed by another rise, especially after 2003, with the 1992 peak eventually surpassed for the first time in 2011 (2521). From 2011 to 2014, fatal attacks rose precipitously by 229%. However, Figure 1 shows that fatal attacks, like total attacks, showed a moderate 8.44% decline in the final year of the series.

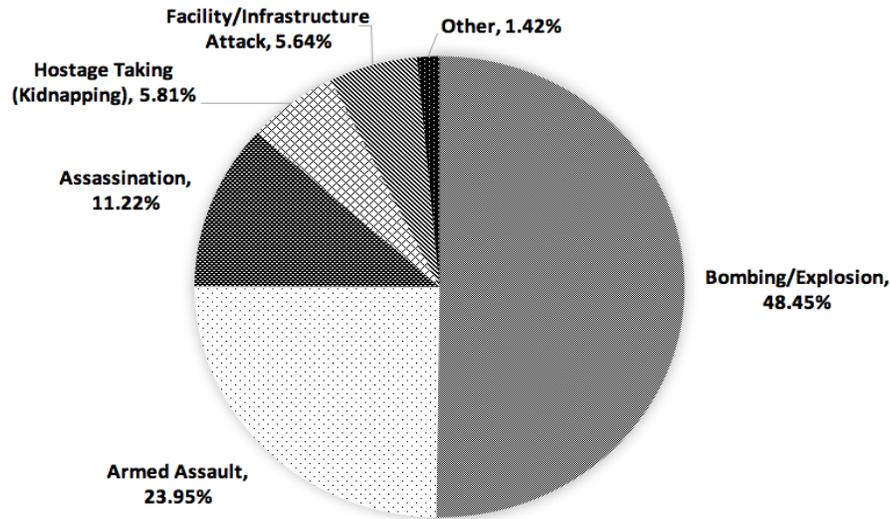
IV. Tactics used by terrorists

Figure 2 shows the major tactics used in terrorist attacks from 1970 to 2015. The most common tactic accounting for more than half of the attacks is bombings. All attacks that use explosive devices or multiple explosive devices, including bombs detonated manually or by remote timer, including suicide bombings, are classified as bombings. Next most common, accounting for a quarter of all attacks, are armed assaults. Armed assaults include assaults on specific persons or facilities by any means other than explosives. For example, the November 2008 coordinated attacks in Mumbai would be identified as armed assaults because the perpetrators relied on firearms and not explosives. In contrast, the perpetrators of the 2004 Madrid and 2005 London attacks planted explosives, and hence these attacks were classified as bombings. Note that if bombings and armed assaults are combined, these two tactics account for three-quarters of all the attacks in the database.

The next most common tactic is assassination, accounting for 12% of all attacks. Assassinations are attacks that kill or attempt to kill specific high profile or prominent figures. Assassinations may use explosive devices, but even if explosive devices are used in assassinations, the latter are identified as assassinations and not bombings. For example, the May 1991 suicide attack by a woman who took the life of Indian Prime Minister Rajiv Gandhi by detonating concealed explosives as she knelt before him is classified as an assassination.

All of the other tactics summarized in Figure 2 are responsible for less than 10% of total attacks. Facility attacks, about 6% of the total, are those whose primary objective is to cause damage to non-human targets, such as buildings or monuments. Another 6% of attacks are kidnappings; abductions of persons or group of people in which ransom is not the primary objective (which would make it a crime rather than a terrorist incident). Finally, a small number of attacks (under 2%) is included in an “other” category that includes hostage taking where the perpetrators barricade themselves into some type of structure and unarmed assaults.

FIGURE 2. TACTICS USED IN ATTACKS, 1970-2015

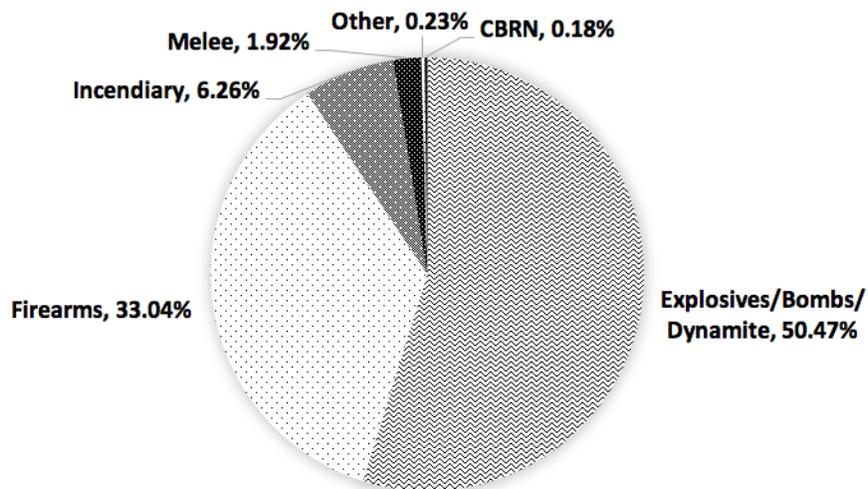


Source: the Global Terrorism Database

V. Terrorist weapons

Because of high profile cases like 9/11 and also the ubiquitous treatment of terrorism by the film and media industry, there is a tendency to think that most terrorist strikes are complex, carefully orchestrated and rely heavily on sophisticated weaponry. Figure 3 shows the distribution of weapons for the thousands of cases included in the GTD. Contrary to the view of terrorism that one commonly gets from the media, the vast majority of terrorist attacks rely on readily accessible weapons.

FIGURE 3. DISTRIBUTION OF WEAPONS, 1970-2015



Source: the Global Terrorism Database

According to Figure 3, the most common weapons in the GTD were explosives and firearms. These two categories account for about 90% of all attacks. Moreover, the explosives used were for the most part readily available, especially dynamite, grenades, mortars and improvised devices placed inside vehicles (“car bombs”). Similarly, the most common firearms were also widely available, including especially automatic weapons, shot guns, and pistols. After explosives and firearms, incendiaries account for about 7% of the incidents. Melee attacks, which are attacks where the perpetrator comes into direct contact with the target using low technology weapons such as stones, fists, or knives, account for about 2% of all attacks.

Note that more sophisticated weapons, chemical, biological, and radiological and nuclear are all exceedingly rare. Among the more sophisticated weapon types were 231 attacks using chemical agents, 35 attacks involving biological agents, and 13 attacks involving radiological agents. Chemical agents range from letters containing rat poison to tainted water supplies. Ten of the 35 biological weapons cases were the U.S. anthrax attacks of 2001 in which 7 people died. Other biological agents include salmonella and ricin. Radiological agents were only used in two tenths of 1% of the attacks. Most of the radiological weapons were monazite, while plutonium and iodine were each used once. Fortunately, to this point in time the GTD has never recorded a terrorist attack using a nuclear weapon. Other weapons account for 0.23% of the total cases and include equipment sabotage where attackers attempt to disrupt the functioning of an existing system (e. g., removing bolts to dismantle vehicles or cutting cables) and attacks where a vehicle is used as a weapon.

VI. Terrorist targets

Table 1 shows the targets of terrorist attacks in the GTD. Overall, there is considerable variation in target types, with the top five targets for the entire series ranking as private citizens and property, military, police, government and businesses. Together, these five targets account for more than three-quarters of all attacks. Perhaps unsurprisingly, private citizens and their property are the number one target of terrorists. It should be noted that many of the other categories could also be classified as attacks on private citizens, including attacks on businesses, transportation, educational institutions and airports. While the GTD does not include violence by state actors, it does include attacks on the military and the police if they are perpetrated by non-state actors and if they fit the other parts of the terrorism definition given above.

Following the top five target types, in order, are transportation, utilities, educational institutions, religious figures and institutions, unknown targets, government-diplomatic, journalists and the media, other terrorist groups, and violent political parties. While attacks on airports and aircraft raise a good deal of public concern, they account for less than 1% of total targets. In addition, attacks on telecommunications, nongovernmental organizations (NGOs), other, tourists, maritime, food or water supply, and abortion providers are all less than 1% of attacks.

Table 1. Targets of Terrorism, 1970–2015

Type of Primary Target	Number of Attacks	Percent of Total Attacks
Private Citizens & Property	35877	22.88%
Military	22924	14.62%
Police	21241	13.55%
Government (General)	19251	12.28%
Business	18882	12.04%
Transportation	6419	4.09%
Utilities	5504	3.51%
Educational Institution	3947	2.52%
Religious Figures/Institutions	3891	2.48%
Unknown	3805	2.43%
Government (Diplomatic)	3333	2.13%
Journalists & Media	2658	1.70%
Terrorists/Non-State Militia	2540	1.62%
Violent Political Party	1637	1.04%
Airports & Aircraft	1303	0.83%
Telecommunication	936	0.60%
NGO	864	0.55%
Other	464	0.30%
Tourists	421	0.27%
Maritime	319	0.20%
Food or Water Supply	294	0.19%
Abortion Related	262	0.17%

Source: the Global Terrorism Database

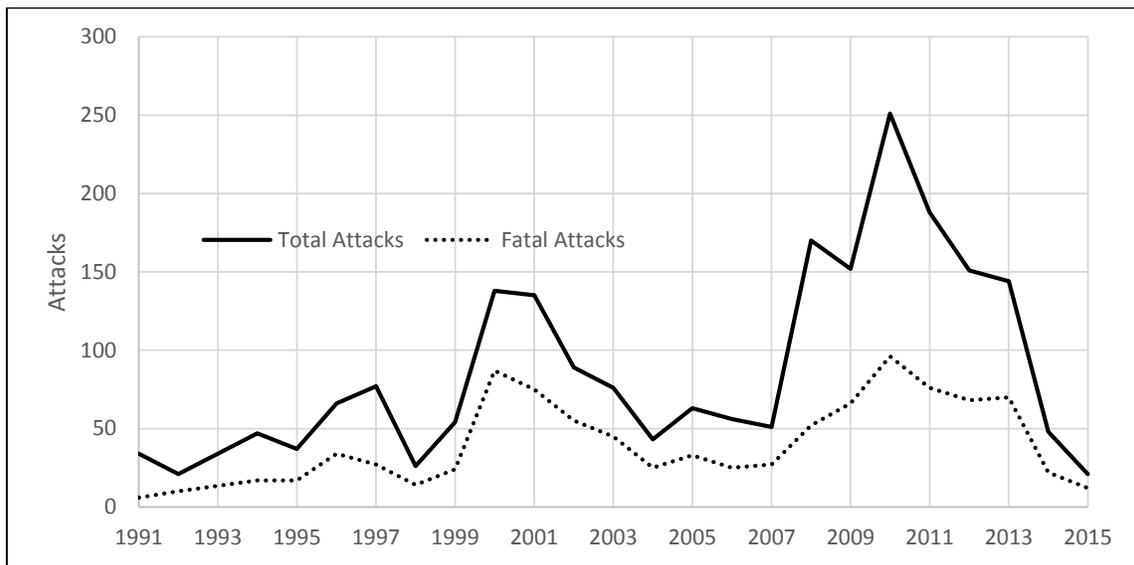
The GTD gives us the capacity to examine not only worldwide trends, but also trends for regions, countries and cities. In the next sections, trends in total and fatal attacks for Russia and the United States are examined.

VII. Russia: total and fatal attacks

Figure 4 shows trends for total and fatal terrorist attacks in Russia from 1991 through 2015. Total attacks in Russia were relatively low during the early 1990s but then began to increase somewhat in the late 1990s, exceeding 100 per year for the first time in 2000. From 2001 to 2004, total attacks again declined, but then began to increase again in 2007, reaching their highest point in the series with 250 attacks occurring in 2010. From 2010 forward attacks again declined so that there were the same number of total attacks in 2015 as in 1992 (n = 21).

Total fatal attacks surpass 50 during only two periods in the series from 1996 to 1997 and from 2008 through 2013. Total and fatal attack trends diverge especially during the period from about 2007 to 2015 where total attacks rose more rapidly than fatal attacks. Despite obvious differences total and fatal attacks are highly correlated (correlation coefficient = 0.93). According to the GTD, the majority of fatal terrorist attacks on Russian soil during this period can be attributed to Chechen and other North Caucasian rebels. This includes the devastating Beslan school siege in 2004 where the GTD records 344 fatalities.

Figure 4. Trends in total and fatal attacks in Russia, 1991–2015



Source: the Global Terrorism Database

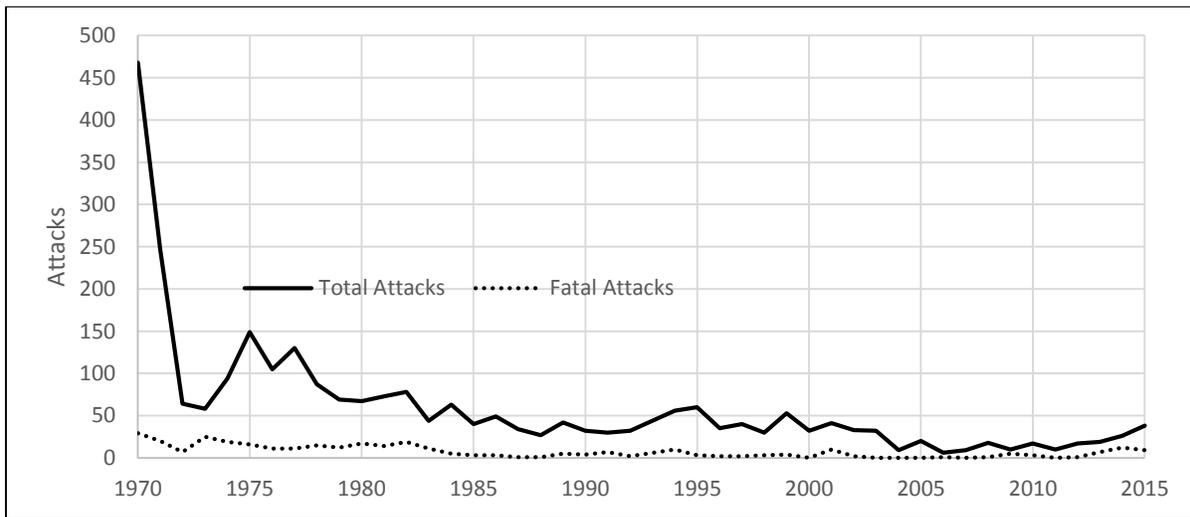
VIII. United States: total and fatal attacks

Figure 5 shows total terrorist attacks and fatal attack for the United States from 1970 to 2015. Terrorism trends in the United States look decidedly different than they do in Russia. The most obvious conclusion is that trends for both total and fatal attacks in the United States are sharply down over more than four decades included in the GTD. By far, the largest number of total attacks recorded in the United States for a single year (468) happened in 1970, the first year of the series, and the year with the lowest number of total attacks (6) was 2006 – seven years before the end of the series (there were 38 attacks in 2015). Following a large drop after 1970, total attacks hit a much lower peak in the mid-1970s with about 120 attacks per year. Total attacks continued to decline throughout the period spanned by the data, dropping below 50 for the first time in 1983 (44) and below 30 for the first time in 1988 (27).

According to Figure 5, total attacks have been about eight times more common than fatal attacks in the United States. As with total attacks, the highest number of fatal

attacks happened in the early 1970s and then fell off sharply during the late 1970s and 1980s. In fact, there were no fatal attacks recorded in the GTD for the United States during the year 2000 and the years 2003 to 2007. The single fatal attack in 2012 occurred on August 5 when a gunman attacked a Sikh temple in Oak Creek, near Milwaukee, Wisconsin. Six people were killed and four others were wounded before the gunman was killed by a police officer called to the scene. No group claimed responsibility for the incident, but the perpetrator was a member of a white supremacist organization.

Figure 5. Trends in total and fatal attacks in the United States, 1970–2015



Source: the Global Terrorism Database

Perpetrator groups responsible for these terrorist attacks on the U.S. homeland were extremely diverse. Of the U.S. attacks where the GTD includes specific information about the perpetrator (81.8% of attacks in the United States), ten organizations responsible for the largest number of attacks since 1970 are: Fuerzas Armadas de Liberacion Nacional (FALN; 119 attacks); New World Liberation Front (NWLF; 86 attacks); Jewish Defense League (JDL; 74 attacks); Animal Liberation Front (ALF; 73 attacks); Earth Liberation Front (ELF; 65 attacks); “Omega 7” (54 attacks), “Weather Underground” (45 attacks); “Macheteros” (37 attacks), Black Liberation Army (36 attacks), and Chicano Liberation Front (31 attacks). Many more terrorist attacks in the United States are attributed to general categories of perpetrators, rather than formal organizations. In fact, unidentified far-left militants were responsible for 169 attacks, all of which took place in the early 1970s. Unidentified anti-abortion extremists were responsible for 168 attacks. While the majority of these attacks took place in the 1980s and 1990s, they have not entirely stopped. Finally, an additional 155 attacks in the United States were carried out by unaffiliated individuals, including Theodore Kaczynski

(the “Unabomber”), Timothy McVeigh (the Oklahoma City bombing), and Nidal Hasan (the Fort Hood shooting spree).

IX. Open source data on terrorism: strengths and weaknesses

It should be acknowledged that terrorism event databases have limitations. The media may report inaccuracies and lies. There may be conflicting information or false, multiple or no claims of responsibility. Government censorship and disinformation may also affect results. When closed societies like North Korea, Sudan, or Myanmar report extremely low terrorism rates, one cannot say for sure whether it is because of actual low reporting or the ability of these societies to minimize coverage by the print or electronic media.

On balance, three weaknesses stand out. First, because there is no universally accepted definition of terrorism, individual event databases all rely on different operational definitions. Second, event databases also face a variety of general biases, many of which have not been empirically explored. Because they rely on news sources, it is usually impossible to know the extent to which reported events reflect real outcomes or the freedom of the press in a particular country or region. For example, Drakos and Gofas⁴ show that the country-level distribution of press freedom strongly depends on level of democratization, especially for extreme values of the latter. In other words, for the vast majority of strongly autocratic states, the press is not free, while for strongly democratic states, it is essentially free. The authors caution that research linking the frequency of terrorist attacks to democratization levels may be biased by underreporting of terrorist attacks in countries with low press freedom (that also tend to be more autocratic countries). Even in countries with high levels of press freedom it seems incontrovertible that news sources will be more likely to report more serious than less serious attacks.

The extent to which countries are covered by the international press also varies by region and over time. For example, Crenshaw and LaFree⁵ show that the percentage of terrorist attacks in the Global Terrorism Database where responsibility for attacks cannot be attributed to a specific group varies widely across regions of the world. While perpetrators responsible for attacks can be identified in more than 60% of the attacks from South America and Western Europe, responsible perpetrators are identified in less than 20% of attacks occurring in Russia and the states of the former Soviet Union, Eastern Europe and Central Asia. Similarly, Fariss argues that the quality of media reporting on issues like respect for human rights may be changing over time.⁶ To the extent that media sources are more likely to identify terrorist attacks over time, open source reporting showing increases in terrorism may be picking up media change rather than increases in actual terrorist attacks.

Finally, data collection efforts to this point have also been strongly biased toward coverage of English language sources. GTD and other major open source datasets all endeavor to monitor non-English sources but in each case resources limit the extent to

which this is possible. Moreover, for all terrorism data collection efforts, the extensiveness of non-English coverage varies over time. Beyond these general problems there are more subtle biases related to the media itself. For example, there is a well-known tendency for news sources to fit individual stories into particular news frames so that compared to other events, preselected themes (e. g., “improvised explosive devices” or “suicide attacks”) may be more likely to receive coverage.⁷ Also, even with reliable media sources unintentional inaccuracy and intentional misinformation are constant concerns.

Still, despite limitations, compared to more traditional data options or even compared to crime data in general, event databases have some important advantages. In particular, because of the compelling interest that terrorist groups have in media attention, open source information may be uniquely useful in the study of terrorism. Many terrorists, unlike most common criminals, actively seek media attention. Terrorism expert Brian Jenkins⁸ once famously declared that “terrorism is theatre” and explained how terrorist attacks are often carefully choreographed to attract the attention of the electronic media and the international press. The fact that terrorists are specifically seeking attention through the media means that compared to coverage of more common crimes, coverage of terrorism is likely to be more complete. Thus, while few researchers would suggest tracking burglary or fraud rates by studying electronic and print media, it seems more defensible to track terrorist attacks in this way. For example, it is hard to imagine that it is possible today for an aerial hijacking or high level politically motivated assassination – even in remote parts of the world – to elude attention of the media.

Event databases on terrorism also have another important advantage. One of the most serious limitations of cross-national crime research is that it has been focused overwhelmingly on a small number of highly industrialized western-style democracies. For example, reviews of cross-national research on homicide⁹ show that most prior research had been based on fewer than 40 of the world’s countries. These countries are of course not a random sample of the nations of the world, but strongly over-represent Europe and North America while almost entirely excluding countries of Africa, the Middle East and Asia. By contrast, open source terrorism databases offer at least some coverage for all countries. While it is the case that the media in highly industrialized countries may under-report news stemming from industrializing countries or highly autocratic states, the salience of terrorism as a phenomenon today makes it more likely than ever that media will report these incidents.

In sum, open source event databases have important strengths as well as weaknesses. The methods for collecting these data have grown more sophisticated over time and there is no reason to assume that this will not continue into the future. Terrorism event databases are likely to persist for now because despite their drawbacks, there is no obvious alternative to them for those interested in tracking worldwide terrorism.

ENDNOTES

¹ For a full description of the methods used to collect the GTD, see LaFree G., Dugan L., Miller E. *Putting Terrorism in Context: Lessons Learned from the Global Terrorism Database*. – New York: Routledge, 2015.

² *Ibid.* P. 15–18.

³ See GTD Codebook: URL: <<http://www.start.umd.edu/gtd/downloads/Codebook.pdf>>.

⁴ Drakos K., Gofas A. The devil you know but are afraid to face underreporting bias and its distorting effects on the study of terrorism // *Journal of Conflict Resolution*. 2006. V. 50. № 5. P. 714–735.

⁵ Crenshaw M., LaFree G. *Countering Terrorism: No Simple Solutions*. – Washington, DC: Brookings, 2017.

⁶ Fariss C.J. Respect for human rights has improved over time: modeling the changing standard of accountability // *American Political Science Review*. 2014. V. 108. № 2. P. 297–318.

⁷ Fishman M. *Manufacturing the News*. – Austin, TX: University of Texas Press, 1980.

⁸ Jenkins B.M. *Will Terrorists Go Nuclear?* RAND Report – Santa Monica, CA: RAND Corporation, 1975.

⁹ Nivette A.E. Cross-national predictors of homicide: A meta-analysis // *Homicide Studies*. 2011. V. 15. № 2. P. 103–131.