RESOLVE
N E T W O R K

Building Secondary Source Databases on Violent Extremism
Reflections and Suggestions

Joe Whittaker

ABSTRACT

It is often argued that a limitation to studying and challenging violent extremism (VE) is a lack of data – extremists are a difficult research population to engage in primary research. However, proponents of this argument miss the wealth of rich, secondary data available. This chapter offers a reflective analysis of creating a database of violent extremists exclusively via secondary sources, including:

- Considerations relating to finding appropriate sources to collect data and the merits and limitations of different sources. While there is an abundance of rich data, discerning how and where to find it, as well as judging between conflicting sources, can be challenging;

- The difficulties in identifying robust inclusion and exclusion criteria. Parsing through what constitutes membership in VE groups, how to define VE, what constitutes specific geographic or temporal zones, etc., is made more challenging given the nebulous nature of terrorism and violent extremism.

- The development of a codebook that both draws from previous literature and creates new avenues for future research;

- Selecting and undertaking quantitative and qualitative analysis appropriate for the data at hand; and,

- Ethical issues surrounding the collection and storing of personal data, as well as potential naming of violent extremists as part of the research.

This analysis draws from the author's doctoral research, in which he created a database of Daesh actors within America. The hope is that the reflections of both the successes and failures of his own research can offer guidance for those that seek to undertake similar studies.

INTRODUCTION

It is often argued that a limitation to researching violent extremism stems from a lack of robust, primary data, leaving important questions unanswered.[1] Recently, scholars have noted access to such sources has

---

1   Andrew Silke, 'Contemporary Terrorism Studies: Issues in Research', Critical Terrorism Studies: A New Research Agenda, Editors: Richard Jackson, Marie Breen Smyth, Jeroen Gunning, 2009, 34–48; Marc Sageman, 'The Stagnation in Terrorism Research', Terrorism and Political Violence, 26.4 (2014), 565–80; Paul Gill, Emily Corner, A Thornton, and others, 'What Are the Roles of the Internet in Terrorism?', *Vox Pol*, 2015.

improved, but there is still undoubtedly a long way to go.[2] The reason for a lack of primary data often stems from the seemingly insurmountable practical and ethical issues that pertain to researching dangerous, unwilling, and vulnerable actors. Although scholars should seek to conduct primary data-driven research, it is also incumbent to look for opportunities to gain knowledge using secondary sources.

For clarity, the term "primary sources" relate first-hand accounts of data. This may be a video of an event, a law, or an autobiography. "Secondary sources" rely on interpretation of events, often by analyzing primary sources themselves. This may include journalistic articles, scholars' analysis of data, or documentaries. Often in research, primary sources are considered to be more valuable than secondary ones because they the interpreting author can have a number of vested interests and biases. However, given the historic difficulties in collecting primary data on violent extremism, secondary sources have emerged as a popular avenue of research. Fortunately for would-be researchers, violent extremists' actions are often highly "newsworthy." There is often a keen appetite in the media and among the general populace to discover as much as possible about the lives of violent extremists, particularly following an attack, resulting in a wealth of rich data.

Many scholars have taken advantage of the rich data available from secondary sources and made important contributions to the field. Sageman was one of the first to utilize secondary sources to understand violent extremists, creating a database of 172 violent extremist actors in the years leading up to 2003.[3] Bakker compiled a similar database of European actors based on secondary sources to compare with Sageman's.[4] Gill and others find that while the use of the Internet is prevalent in cases of violent extremist actors, most do not "radicalize" online; the offline domain is equally important.[5] Similarly, Horgan and others establish key differences between lone and group-based actors, both demographically and in their event behaviors.[6] Below I offer reflections from my doctoral research, in which I created a database of Da'esh[7] actors in America using a range of secondary sources. The chapter details my considerations on data collection, the importance of robust definitions and inclusion/exclusion criteria, appropriate methods of analysis, and ethical challenges. In addition, the chapter outlines many of the issues, limitations, and challenges that face scholars collecting secondary data for research on violent extremism and similar studies and offer suggestions on how to ameliorate them. These include the problems caused by the different goals and

---

2   Bart Schuurman, 'Research on Terrorism, 2007–2016: A Review of Data, Methods, and Authorship', Terrorism and Political Violence, 2018, 1–16; Lorne Dawson, 'Clarifying the Explanatory Context for Developing Theories of Radicalization: Five Basic Considerations', Journal for Deradicalization, 18 (2019), 146–84.

3   Marc Sageman, Understanding Terror Networks (Philadelphia, PA: University of Pennsylvania Press, 2004).

4   Edwin Bakker, Jihadi Terrorists in Europe: Their Characteristics and the Circumstances in Which They Joined the Jihad (The Hague, 2006).

5   Gill, Corner, Thornton, and others; Paul Gill, Emily Corner, Maura Conway, and others, 'Terrorist Use of the Internet by the Numbers: Quantifying Behaviors, Patterns, and Processes', Criminology & Public Policy, 16.1 (2017), 99-117.

6   John Horgan and others, 'Actions Speak Louder than Words: A Behavioral Analysis of 183 Individuals Convicted for Terrorist Offenses in the United States from 1995 to 2012', Journal of Forensic Sciences, 61.5 (2016), 1228–37.

7   Da'esh is an Arabic translation of the acronym for the Islamic State in the Levant, also known as ISIL (Islamic State in Iraq and the Levant) or ISIS (Islamic State in Iraq and Syria).

incentives of the original authors to the researcher; uneven data; as well as the ethical issues that scholars face.

DATA COLLECTION: SECONDARY SOURCES

There are many avenues through which researchers can collect data on violent extremists from secondary sources, each offering distinct strengths and challenges, including criminal justice documents, journalistic accounts and the media, academic literature, and pre-existing databases on violent extremist actors and incidents.

Criminal Justice Data

Criminal justice documents are an excellent place to start to gather secondary source materials. Court transcripts and criminal complaints often provide rich, granular-level detail relating to an individual's behaviors prior to and during their involvement in terrorist activities. Prosecution and defense sentencing memorandums can also paint a comprehensive picture of an individual's upbringing and demographic information. Issues associated with accessing legal documents, however, necessarily limits the utility of their use as secondary sources. The availability of legal documents in particular differs from country to country depending on privacy norms.

In the United States, for example, many legal and court filings are available to the public via the Public Access to Court Electronic Records (PACER). In many cases, these filings are collected and archived by researchers, such as George Washington University's Program on Extremism.[8] In addition, the U.S. Department of Justice issues press releases rich in detail on violent extremist and terrorist activity and profiles, often including criminal justice information. For example, a U.S. Department of Justice press release for Zoobia Shahnaz, a U.S. citizen who pleaded guilty to providing material support to Da'esh, contained both a summary of the allegations against her along with her actual indictment[9].

In other countries, particularly within Europe, legal documents on individuals under investigation or convicted of terrorism-related offenses are less readily available, although most countries maintain an online service with varied access to detailed documents. However, because incidents of violent extremism are considered to be of high public interest, press reports related to the incidents are often sanctioned by government agencies in the interest of transparency. These press reports often yield rich, albeit sometimes selective, data on individuals involved in terrorist incidents. After the Westminster Bridge attack in London in 2017, for example, the UK government released a detailed report on the incident which itself contained important insights useful to researchers.[10]

---

8   George Washington University Program on Extremism, ISIS in America – The Cases. Access via: https://extremism.gwu.edu/cases.

9   Department of Justice, Long Island Woman Pleads Guilty to Providing Material Support to ISIS, 26 Nov, 2018. Access via: https://www.justice.gov/usao-edny/pr/long-island-woman-pleads-guilty-providing-material-support-isis.

10   Max Hill, The Westminster Bridge Terrorist Attack, HM Government, 2018.

Journalistic Data

Journalistic data are also an invaluable source of secondary data on violent extremists, particularly in countries where court documents are less readily available. Most newspapers have courtroom beat reporters that write about the specific details of cases as they occur. For my own research, I found that local, investigative journalism offered an in-depth picture into violent extremist actors' lives and their pathways towards violent extremism. In particular, news organizations such as Minnesota Public Radio provided a wealth of information, dedicating time and resources to comprehensive research within their local community, which greatly contributed to my own data collection.[11] Online resources such as Google News and Lexis Nexus, can increase the ease of collecting journalistic resources, offering archived reporting from different sources. Users can also set up to be notified of new resources by setting up email notifications alerting them of the addition of new resources on Google News and Lexis Nexus.

Academic Literature

Another rich resource for secondary data is academic literature. Many in-depth case studies on individual violent extremists,[12] groups,[13] or specific incidents[14] exist in academic literature. These scholarly studies can provide important insights for researchers embarking on violent extremism studies using secondary data

The strength, and therefore utility, of academic resources lies in their rigor, established through peer-review processes and the incorporation of theoretical perspectives. Unfortunately, however, access to many academic studies is sometimes limited, particularly because many peer-reviewed journals exist behind paywalls, and are therefore inaccessible to those who do not have institutional or individual paid-access subscriptions. Open-access articles, those which do not require a paid subscription, are steadily becoming more commonplace and a number of think tanks and organizations provide open access to all of their resources. The sheer amount of academic articles on violent extremism, even without paywall access, makes sorting through and analyzing such articles for secondary-source material somewhat difficult. For the collection and analysis of academic resources, I recommend using file storage software, such as Mendeley, which is free and includes keyword search functionalities that enable users to search through many articles for specific information (i.e. the names of violent extremist actors or incidents).

---

11   For example: Laura Yuen, Gulad Omar: The Path to ISIS and the Story You Haven't Heard, MPR News, Dec 22, 2016, Access via: https://www.mprnews.org/story/2016/12/21/path-to-isis-minnesota-guled-omar-mpr-interviews

12   For example, for Anders Brevik: Jacob Aasland Ravndal, 'Anders Behring Breivik's Use of the Internet and Social Media', Journal EXIT-Deutschland, 2 (2013), 172–85. For Roshonara Choudhry: Elizabeth Pearson, 'The Case of Roshonara Choudhry: Implications for Theory on Online Radicalization, ISIS Women, and the Gendered Jihad', Policy and Internet, 8.1 (2016), 5–33. For Colleen LaRose: Maura Conway and Lisa McInerney, 'What's Love Got to Do with It? Framing "JihadJane" In the US Press', Media, War & Conflict, 5.1 (2012), 6–21.

13   For example on the Hofstad Group Bart Schuurman and John Horgan, 'Rationales for Terrorist Violence in Homegrown Jihadist Groups: A Case Study from the Netherlands', Aggression and Violent Behavior, 27 (2016), 55–63.

14   For example the Westgate Mall attack David Mair, '#Westgate: A Case Study: How Al-Shabaab Used Twitter during an Ongoing Attack', Studies in Conflict and Terrorism, 40.1 (2017), 24–43.

Pre-Existing Databases

Finally, pre-existing databases on violent extremist actors and incidents provide an easy-to-use and rich source of secondary data for those researching violent extremism. Many databases are free and available online, including the Global Terrorism Database (GTD), which boasts the largest collection of data on terrorist incidents and allows users to select events by date, location, and perpetrator group, using a number of different definitions.[15] Sorting by definitions is particularly useful if the research in question blurs the lines between terrorism and insurgency. The GTD offers a short description of each terrorist event and a number of corresponding quantitative variables, while also providing a sampling of journalistic resources on the event. Whereas the GTD focuses on events, databases such as the Counter-Extremism Project focus on specific actors, providing a sourced summary of their background and their path to extremism as well as information and links on known associates.[16] These databases can be particularly useful in the initial phase of identifying violent extremists for inclusion in research or the construction of a new database.

Collection Challenges and Limitations

Individually, these four sources of secondary data are likely to prove insufficient in gaining a clear picture of to create a database for analysis. However, they can complement each other effectively. For example, criminal justice information can offer a detailed and granular account of the events directly leading up to violent extremist attacks or arrests, such as what the actor posted on social media, whether they travelled abroad, or if they conducted training. Data from criminal justice and legal materials can be supplemented by data from investigative journalism, which often provides information from interviews with friends and family of individual perpetrators of violent extremist activities who can attest to that individual's upbringing or any recent changes in behavior. Academic materials can also offer a layer of expert analysis, incorporating theoretical perspectives, quantitative and qualitative analysis, and placing the actor within a larger movement. Finally, databases can be an excellent place both to initially identify actors or incidents to be analyzed, as well as a resource from which to triangulate data collected from other sources.

However, as previously noted, a number of problems can arise from collecting secondary data. A key fact to keep in mind when collecting any type of data – primary or secondary – is that data itself is likely to include a number of biases, inaccuracies, limited data collection priorities, and coding differentials. When using secondary sources in particular, researchers must be mindful of the fact that the original data-collectors may not—and almost always do not—have the same goals as the secondary collector. Law enforcement and public prosecutors write criminal complaints to secure prosecutions, journalists write (and editors edit) articles to sell newspapers or generate clicks. Even academic works or databases are likely to have different purposes when used as data the second time around. This is problematic because the original authors may not think to include information that is vital for the secondary author, or may have used different exclusion or inclusion criteria that imbue a certain level of bias in the data they present.

---

15   National Consortium for the Study of Terrorism and Responses to Terrorism (START), Global Terrorism Database, (2018). Access via: https://www.start.umd.edu/gtd

16   Counter-Extremism Project, Terrorist and Extremist Database, Access via: https://www.counterextremism.com/extremists.

Consequently, using a diversity of types of sources is vital, as is attempting to triangulate data to ensure the greatest validity possible.

Similarly, some events and actors are deemed more newsworthy than others, making available data on different topics and actors unequal. Incidents of terrorism, particularly jihadist, are the subject of disproportionate news coverage,[17] as are female perpetrators.[18] I found this to be the case in my research, wherein certain types of actors and incidents had far more data available than others. For example, successful attackers received far more coverage than those that travelled to Iraq or Syria, white converts to Islam were covered more than those born as Muslims, and women involved in violent extremism were covered more than men. This runs the risk of certain demographics being artificially over-represented, which could skew analyses. There is no easy way to correct for this. To address this in my research, I removed cases that had insufficient evidence or available data after establishing the key variables that I was investigating—antecedent and event-focused behaviors.

SELECTING DEFINITIONS AND INCLUSION/EXCLUSION CRITERIA

The next important aspect of developing a research project or database is adopting accurate and representative definitions for the data included and making them transparent. Definitions are difficult to establish, notably because concepts around "violent extremism" are not universally defined or adopted. Nevertheless, establishing clear and consistent definitions that guide the research process and database construction is necessary in ensuring that the ultimate analysis and findings are accurately represented.

Defining Broad Concepts

The most fundamental decision any researcher faces, and particularly those within the ill-defined fields of terrorism and violent extremism, is to define the phenomenon that will be researched. I cannot overstate the importance of opting against the "you know extremism when you see it" approach. The long-lasting inability to define or reach consensus on "terrorism" has long marred the field. This lack of a coherent definition stems, in part, from political issues,[19] and it is not clear terms such as "extremism", "violent extremism", or "radicalization" are any less nebulous or politically loaded.[20] It is prudent, therefore, for scholars to be clear about exactly what they are researching. In many instances, it can be helpful to use

---

17   Zachary S. Mitnik, Joshua D. Freilich and Steven M. Chermak, 'Post-9/11 Coverage of Terrorism in the New York Times', Justice Quarterly, 2018, 1–25.

18   Conway and McInerney.

19   Boaz Ganor, 'Defining Terrorism: Is One Man's Terrorist Another Man's Freedom Fighter?', Police Practice and Research, 29.3 (2002), 123–33; Alex P. Schmid, 'Terrorism - The Definitional Problem', Case Western Reserve Journal of International Law, 36.2 (2004), 375–419.

20   Alex P. Schmid, Radicalisation, De-Radicalisation, Counter-Radicalisation: A Conceptual Discussion and Literature Review, International Centre for Counter Terrorism.

a pre-existing definition. For example, the aforementioned GTD has three different criteria for users to choose between when searching data:

- The act must be towards a political, economic, religious, or social goal.

- There must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims.

- The action must be outside the context of legitimate warfare activities, i.e. the act must be outside the parameters permitted by international humanitarian law (particularly the admonition against deliberately targeting civilians or non-combatants).[21]

These criteria offer researchers different tools, depending on what they are studying. For example, if one is studying violent extremism defined as more closely resembling an insurgency, then the first definition will be more applicable, but if the topic is on a more traditional conception of terrorism, the second or third would be more appropriate. In other cases, it may be more suitable to use a governmental definition, particularly if the research is focused on a specific country. However, it may also be beneficial to deliberately eschew a governmental decision if the researcher feels that it is too value laden or problematic. For example, the British government defines extremism as "vocal or active opposition to *fundamental British values*."[22, 23] Ultimately, there are a number of ways in which researchers can define their phenomenon—what is important, is that what the researcher is addressing is defined clearly.

One way to bypass politically charged definitional debates is to adopt definitions based on group membership, only selecting cases in which the actor is part of a particular violent extremist organization. For example, if researching violent extremist Boko Haram actors, one would not need to distinguish between acts of terrorism and acts of insurgency, but rather include events or actors associated with the group. However, this brings about another set of problems.

Designating and Defining Individuals

Group membership is often more nebulous than presented, as for example, Da'esh acts of terrorism in the West in recent years. Although there are instances of "official" group members conducting attacks, the majority of cases represent smaller cells or lone actors who may have varied links to the core organization. Europol observes that "jihadist actors can be both directed by [Da'esh] or merely inspired by [Da'esh] ideology and rhetoric."[24] This means that, if a researcher were to select cases on the basis of formal group membership, it would not capture the reality of contemporary violent extremism. To ameliorate this, in my

---

21   Global Terrorism Database.

22   Emphasis added.

23   Home Office, Revised Prevent Duty Guidance for England and Wales (London, 2015), 22.

24   Europol, Terrorism Situation and Trend Report (TE-SAT) 2017, 2017, p. 5

work, I followed the lead of the Study of Terrorism and Responses to Terrorism's "Profiles of Individual Radicalization in the United States" (PIRUS) codebook:

We define "member" broadly. This includes official members, individuals [that governments] claimed were members…and includes credible media sources link to that group (but not based on pure speculation). It also includes individuals who claim membership…even if the group itself does not acknowledge membership.[25]

This criterion allows for self-identification to a violent extremist organization. Next, I established that there are two primary ways in which one can be judged to self-identify:

- Outward: Words or actions that display support for a group such as publicly stating it (online or offline) or attempting to recruit others or;

- Inward: Activities such as downloading, reading, or listening to the group's or sympathizers' media content.

Importantly, one must establish that this self-identification plays a significant role in the actor's behaviors.

Researchers must also take spatial and temporal criteria into account; that is to say, where the research is focused and when it occurred. The violent extremist movement of choice may dictate these decisions. For example, researching The Troubles in Northern Ireland already predefines a broad area and time period of focus. However, if either the area or time is too expansive to collect all of the relevant data, researchers should consider narrowing their focus to be more specific. In this case, for example, it may be more realistic to gather data on actors in Derry, Northern Ireland between 1970 and 1974. Of course, data need not always be representative, particularly when conducting qualitative analysis. However, if the focus of research is to make generalizable claims, then it must be representative. For my own research, my choice of group made establishing the date easy; I simply began from the first established incidents of Daesh in America, which dated back to around 2012. However, deciding what constituted "being in America" is more problematic as there are a number of different ways this could be interpreted. I decided that to be included, the actor must:

- Have been charged in the U.S., or

- Be a U.S. citizen, or

- Be a U.S. permanent resident, or

---

25   START, 'Profiles of Individual Radicalization in the United States (PIRUS) Codebook', 2018, p. 6

- Have resided in the U.S. at the time of the event.

Violent extremist movements can be complex, transnational, and nebulous. If I were to include every actor that had any relation to the U.S., then the list would be unmanageable and would likely not end up not saying much about the country of focus. Ultimately, there is no correct way to define or include/exclude data; researchers must establish their parameters for themselves using sound reasoning, stick to them, and clearly articulate them to external audiences.

ANALYZING DATABASES: QUANTITATIVE, QUALITATIVE, AND MIXED-METHODS

After collecting data and establishing inclusion and exclusion criteria, the next step is deciding the most appropriate way to analyze the database using a quantitative, qualitative, or mixed-method approach. Although there are few limits on what can be analyzed, researchers should stick to the strengths of secondary data and avoid its weaknesses. Most analyses focus on the behaviors of violent extremists, rather than, for example, cognitive factors, as rich data on the former are easier to obtain. Secondary data is not well-suited for establishing when and how actors' beliefs began to become more extreme and when they started to justify violence. On the other hand, secondary sources are more helpful in determining whether actors displayed specific behaviors, such as interacting with co-ideologues or scouting out potential targets in advance. One must be careful, however, not to over-interpret these sources given the limitations outlined above are still present. Being overly ambitious in terms of trying to establish different phases and turning points of "radicalization," is unlikely to yield robust findings if the researcher relies solely on second-hand evidence. One potential exception to this is if the researcher has access to closed-source police data which often outlines the details of the case in much greater detail.[26]

Quantitative methods

The majority of secondary database studies use quantitative measures of analysis.[27] For this method, a codebook must be created; a method of classifying content to be analyzed. In many instances, this takes the form of demographic data, such as age, gender, country of birth, employment, and criminal record,[28] often to establish whether there are any common demographic factors shared amongst these actors. Furthermore, some studies code for event-specific behaviors, such as if and how the actor used the Internet,[29] how actors plan and prepare for attacks,[30] or how far actors travel to conduct attacks.[31] While researchers

---

26   Paul Gill, The Online Behaviours of British Terrorists, Vox Pol, 2016.

27   For example, see: Gill, Corner, Thornton, and others. Gill, Corner, Conway, and others; Horgan and others.

28   Bakker; Sageman, Understanding Terror Networks.

29   Gill, Corner, Conway, and others.

30   Bart Schuurman and others, 'Lone Actor Terrorist Attack Planning and Preparation: A Data-Driven Analysis', Journal of Forensic Sciences, 63.4 (2018).

31   Zoe Marchment, Noémie Bouhana and Paul Gill, 'Lone Actor Terrorists : A Residence-to-Crime Approach', Terrorism and Political Violence, 2018, 1–26

should have a codebook in place before analysis, the process can be iterative, adding new coding variables as they emerge from the data.[32]

One challenge that must be considered is how to code the data in cases with insufficient evidence. For example, when coding for whether an actor chose their attack target online, open source data may not mention how such operational decision-makings were made. In some instances, scholars provide three data points – *Yes*, *No*, or *Don't Know* [33] - and remove "Don't Knows" from the analysis. Others, including myself, code dichotomously, using two data points: *Yes* or *Insufficient Evidence for Yes.*[34]

Neither method is without flaws; the former is likely to undercount both *Yes* and *No* responses, while the latter is likely to undercount *Yes* responses. I use the two-point coding system because, as explained by Horgan et al., "it is universally true in open-source coding that the number of true 'Yes's' will likely be a truer representation than the number of 'Nos'… [and] it may be especially true for behavioral factors associated with non-illegal or terrorist activities that occurred abroad." [35] This, again, is playing to the strengths of the data available; it is very difficult to prove a negative in open source-reporting. For example, a journalist will not report on many things that *did not* happen because most of them are irrelevant to the piece they are writing but is more likely to report things that *did* happen as long the variables are not too obscure.

There are a number of different ways to analyze data quantitatively once it is coded. Most simply, researchers can use descriptive statistics and frequencies that yield a basic overview of the sample: the average age; what percentage used the Internet to engage with co-ideologues; how many co-offenders the actor worked with; or what percentage are male or female.

Researchers can also use bivariate tests such as Person's chi-square and Fisher's exact tests, which test the relationship between two categorical variables by comparing the frequencies in the categories against the frequencies that one might expect given a random distribution. If there is a statistical significance (the p value is less than 0.05) then the null hypothesis (that the variables' relationship is randomly distributed) can be rejected.[36]

---

32   Gill, Corner, Conway, and others.
33   Horgan and others.
34   Gill, Corner, Conway, and others.
35   Horgan and others, p. 1236.
36   Andy Field, Discovering Statistics Using IBM SPSS Statistics, 5th edn (London: Sage Publications, 2018). This can also be used to calculate the "Odds Ratio", which calculates the ratio of two categorical events occurring in the sample. For example, if I collect data on whether actors used social media and whether they divulged their plans to others, the odds ratio is calculated by multiplying all the instances in which actors used neither social media nor did they divulge (a) by the instances in which they did both (b), before dividing by the number of instances in which only one of the two occurred (b) (c). This is summarised in a 2 x 2 contingency table (a x d)/ (b x c). This can be used to assess how much more likely or unlikely one outcome is than another Field, 2018). For example, Gill, Corner, Conway, and others find that those that plot against a government target are 4.50 times more likely to use the Internet to learn about their attack than those that plot against civilian targets.

For example, Gill and others found statistical significance between those that engaged with co-ideologues online and those that engaged in an offline network, which suggests that although Internet usage is prevalent within the sample, the offline domain also remains important.[37] A number of other analyses can be utilized too, such as conditional logistic regression[38] or binary logistic regression [39] to predictively model behaviors.[40]

Qualitative methods

It may, however, be more appropriate to conduct a qualitative analysis. Rather than establishing a codebook prior to analysis, researchers can employ qualitative approaches, including thematic grouping based on patterns that emerge in the data. While qualitative analyses of datasets are not as representative as quantitative, the former can add valuable insight into the subject matter of focus by taking a deeper, more nuanced dive into the cases beyond that provided in a binary quantitative codebook. Various studies on violent extremism and foreign fighters have adopted this approach. For example, in their research on U.S. foreign fighters in Iraq and Syria, Meleagrou-Hitchens, Hughes, and Clifford identified three non-exhaustive themes of individuals traveling to participate in violent extremism from their data—Pioneers, Networked Travelers, and Loners—offering descriptive accounts of their lives and travel.[41] Similarly, in a study on female violent extremists, Alexander used qualitative grouping to identify three categories of female involvement in violent extremism—Plotters, Supporters, and Travelers—in analyzing the diversity of roles for women.[42] Finally, in a report on French jihadist actors, Hecker used qualitative analysis to identify individual pathways to extremism that appeared within his data sample, drawing together distinct patterns with important insights for future research.[43]

Mixed Methods

For my research on violent extremists' use of the Internet, I opted for a mixed-method approach, incorporating a number of the qualitative and quantitative techniques outlined above. Mixed methods ensure that the analysis carried out benefits from both the extrapolatable and the highly nuanced findings inherent in quantitative and qualitative methods respectively. More specifically, my database research included quantitative variables and testing, as well as qualitative sorting and theory building.

---

37   Gill, Corner, Conway, and others.

38   Zoe Marchment and Paul Gill, 'Modelling the Spatial Decision Making of Terrorists : The Discrete Choice Approach', Applied Geography, 104. (2019), 21–31.

39   Gill, Corner, Conway, and others

40   A logistic regression is a model for predicting categorical outcomes from categorical or continuous variables. For example, Marchment and Gill used a dataset of Irish Republican Army attacks to establish which characteristics increase the likelihood that an area would be a target, finding that factors such as distance from the actor's home decreased the likelihood, and proximity to main roads and police and military bases increased the chances.

41   Alexander Meleagrou-Hitchens, Seamus Hughes and Bennett Clifford, 'The Travelers: American Jihadists in Syria and Iraq', Program on Extremism, 2018.

42   Audrey Alexander, 'Cruel Intentions: Female Jihadists in America', Program on Extremism, 2016.

43   Marc Hecker, '137 Shades of Terrorism: French Jihadists Before the Courts', Security Studies Center, 2018.

Building off of the work of other studies, most pertinently Gill et al.'s research on the use of the Internet among convicted British terrorists, I analyze my dataset quantitatively by coding for online and offline behaviors related to individual interactions with violent extremist networks and behaviors indicative of learning or planning related to violent extremist activities.[44] I replicated the key variables from these studies to use as a basis of comparison against my own. I also conducted an iterative stage of coding, in which I included variables not already present - such as the use of end-to-end encryption, the type of social media platform used, and the presence of undercover law enforcement agents in individual cases. Like Gill et al., I performed descriptive, bivariate, and multivariate statistical tests to analyze whether online and offline variables were independently randomly distributed. This allowed me to assess, for example, the degree to which individuals that learned the techniques necessary and/or planned violent extremist attacks online were likely to have also done so offline.

Qualitatively, I decided to analyze my data using Grounded Theory Methodology (GTM). In contrast to deductive methods of coding and analyzing against a pre-existing codebook, GTM is an inductive method aimed at creating new theories grounded in data.[45] While the majority of academic research tests theories or hypotheses in some way – looking at economic market data to establish whether rational choice theory explains a certain phenomenon, for example - GTM is undertaken with as few preconceptions as possible. Rather, GTM aims to generate hypotheses from the data itself which can be tested against future data. GTM, in essence, lets the data speak for themselves, developing a much wider and more imaginative set of research insights and questions than hypothesis testing alone.[46]

The specific grounded theory method I utilized included three rounds of coding. First, I used descriptive *open coding* to assess and code data line-by-line for the presence of online behavior. Second, using *selective coding,* I grouped the descriptive codes from the first round into larger categories. For example, I coded for each instance of an actor financing their plot during *open coding*, and then analyzed these and grouped them using *selective coding* into categories such as: cash; goods; FinTech; Cryptocurrency. Finally, I used *theoretical coding* to consider the categories developed in the second phase of coding in relationship to each other, which allowed me to construct theories from observation of the data itself.[47] Using the financing example, it allowed me to generate the hypothesis that terrorists in this sample were *not*, by and large, exploiting cryptocurrencies to fund their plots and that offline cash transactions are still most common. The inclusion of qualitative coding enabled more in-depth investigation of the dataset facilitating the establishment themes and theories otherwise beyond the scope of a purely quantitative codebook. For example, qualitative coding allowed me to investigate the role of the Internet in financing plots, methods that actors took to evade law enforcement online, and gender dynamics in the online domain in more depth than quantitative methods alone. There are a number of software packages that can be used to facilitate the

---

44   Gill, Corner, Thornton, and others; Paul Gill and Emily Corner, 'Lone Actor Terrorist Use of the Internet and Behavioural Correlates', in Terrorism Online: Politics Law and Technology, Edited by Lee Jarvis, Stuart Macdonald, and Thomas M. Chen (Abingdon, Oxon: Routledge, 2015), pp. 35–53; Gill, Corner, Conway, and others.

45   Barney G Glaser and Anselm L Strauss, The Discovery of Grounded Theory: Strategies for Qualitative Research (New Brunswick, NJ: AldineTransaction, 1967).

46   Glaser and Strauss

47   Cathy Urquhart, Grounded Theory for Qualitative Research: A Practical Guide (London: Sage Publications, 2013).

qualitative process; I used NVivo, which offers line-by-line coding and searching. Although NVivo is not free, many educational and research institutions have it included in their set of software.

ETHICAL CONSIDERATIONS

One might assume that there are minimal ethical issues involved in conducting open-source, secondary research. After all, as there is no direct contact with research subjects, many of the traditional safeguarding issues seem as if they do not apply. Similarly, if all of the data are in the public domain, privacy issues may seem less pertinent. However, while ethical issues may seem minimal on the surface, there are a number of considerations that researchers using open-source and secondary sources in studies on violent extremism must take into account.

Storing and Processing Personal Data

First, when building and analyzing a database, the researcher is storing and processing personal data without the consent of the research subjects, which can be unlawful in some contexts, particularly within the jurisdiction of the European Union General Data Protection Regulation (GDPR). The GDPR does, however, offer an exemption: Article 6 (1) (e) permits processing of personal data if it "is necessary for the performance of a task carried out in the public interest".[48] Violent extremism is considered a top priority by most governments,[49] making satisfying this threshold attainable for researchers conducting studies in these territories. The principles embodied in this exemption serve as an ethical norm for researching secondary sources without consent beyond the GDPR. It is not enough to simply argue "the data are in the public domain, therefore anything goes." Particularly when dealing with personal information on topics as sensitive as violent extremism, researchers should strive to adopt adhere to strict ethical guidelines, whether mandated by a governing body or institution or not. Researchers should develop and/or adopt a set of defined ethical guiding principles regarding how to store, process, and publish personal information, even if derived from open-access, already publicly available sources.

Identifying Individuals as "Violent Extremists"

One particular ethical issue that researchers should be cognizant of is how and whether to name individuals as "violent extremists" and the implications of doing so. Many legal cases involving individuals accused of terrorism-related charges that may populate a secondary-source database on violent extremism are ongoing. What is more, many individuals accused of terrorism-related charges do not ever face trial (i.e. many are deceased or allocated in a foreign jurisdiction). In addition, once-extremist actors may pay their debt to society, rehabilitate, and wish to forget their involvement; publicly naming those individuals may have consequences for the rest of their life. In any case, the tension between wanting to conduct

---

48   European Parliament, General Data Protection Regulation, 2018, Access via: https://gdpr-info.eu/art-6-gdpr/.

49   For example, see: Federal Bureau of Investigation, What We Investigate – Terrorism, Access via: https://www.fbi.gov/investigate/terrorism. Home Office, Revised Prevent Duty Guidance for England and Wales (London, 2015).

rigorous research that includes every potential violent extremist subject identified during data collection versus the harm that erroneously naming someone as a violent extremist may cause presents a unique ethical challenge. To include only people that have been convicted or plead guilty to terrorism-related charges would result in a skewed sample, representative of only those found guilty in a court of law, undercounting others suspected to have participated in violent extremist activity but never convicted for one reason or another. At the same time, everyone is entitled to the presumption of innocence. To label an individual not found guilty by a court of law as a violent extremist contains many potentially harmful ramifications for not only the individual, but also their entire social circle.

In purely quantitative research, where results are reported numerically, this is not necessarily an issue so long as personal information on subjects is not identified in reporting on the statistics. However, in qualitative research, this issue is far more prominent. Typical privacy techniques such as anonymization and pseudonymization are not necessarily possible in qualitative studies, which require a higher degree of referencing to establish academic rigor. To protect oneself and one's research subjects, researchers undertaking qualitative studies should make a common practice of noting which criminal cases under study are ongoing or unresolved during the coding process, and always reference that the state of the case identified when discussing findings – i.e. "x is *alleged* to have left for Iraq in June 2014." Similarly, if actors included in the database are later found acquitted of charges, researchers should at the very least reference this within their findings and database, and also seriously consider removing the individual or case from the database entirely. In my research, I removed Noor Salman – the wife of Omar Mateen, the shooter in the June 12, 2016 Pulse Nightclub attack in Orlando, Florida – from my sample after she was acquitted at trial.

Researcher Self-Care

A final ethical issue that researchers should consider when compiling databases with secondary sources is that of researcher self-care, both psychological and legal. In this field of study, there is a great deal of gruesome, distressing, and extremely violent content. Even research like mine, which does not focus on notably violent and distressing propaganda from violent extremists like Da'esh, carries with it the possibility of a significant psychological toll for researchers. At times, court documents may contain disturbing images of injuries sustained due to violent extremist attacks or violent still images from propaganda videos. Those undertaking this research – even those not focused on researching the most disturbing violent extremist content – may not be mentally prepared for the content that they may ultimately encounter. It is, therefore, incumbent on researchers and their affiliated institutions to find effective mechanisms to deal with this problem, both proactively and as it emerges. While access to sustained psychological care for researchers encountering disturbing images and violent extremist content is necessary, it is not always economically viable, particularly for smaller research teams or institutions. In lieu of this, institutions should consider creating environments wherein researchers can discuss distressing data without fear of burdening others. King, for one, has argued that building resilience in researchers of violent extremism is vital, suggesting the establishment of formal interventions (i.e. psychological counseling and debriefs) and also less formal

interventions (i.e. monthly research team discussions in a safe and open environment).[50] Most importantly, researchers should avoid isolating themselves and researchers and institutions should work to identify options should researchers need psychological support.

Related to psychological self-care is ensuring that the research being conducted is lawful, as different countries have different legal codes dealing with violent extremist content. For example, while my research focused on Da'esh in the United States, I conducted my research in the United Kingdom (UK). This proved somewhat problematic as the two countries have vastly different interpretations with regards to the legality of violent extremist propaganda. In the UK, an individual can receive a prison sentence of up to fifteen years for accessing terrorist propaganda online,[51] whereas in the U.S., accessing terrorist propaganda is almost universally protected under the First Amendment because it does not meet the "direct incitement" criterion of *Brandenburg v. Ohio 1969*.[52] The implications of this for my research meant that a number of resources from which I gathered data contain links to content that was illegal in the UK, but not in the U.S.

To make matters even more complicated with relation to legality, criminal justice documents and sources sometimes included photos of violent extremist propaganda (otherwise deemed illegal to access in certain jurisdictions, including the UK) without any warning that it was included in the document. I encountered this a few times throughout my research. Researchers must be sure to be cognizant of local law and legal codes and how those might affect their work. While some loopholes to legal codes around violent extremist content and individuals exist for those engaged in research,[53] law enforcement may not look kindly upon propaganda that has been accessed online and downloaded to an external hard drive or printed out into hard copies. For researchers living and working in a country which proscribes violent extremist propaganda that may be accessed as a part of their study, it is best to notify law enforcement prior to beginning research, rather than deal with legal issues after accessing the content.

CONCLUDING THOUGHTS

Although it is possible that scholars of violent extremism will be able to continue making more use of primary sources in the future, it is unlikely that the data from secondary sources will diminish in importance. Indeed, secondary sources can and do offer rich and detailed data that produce excellent, nuanced research in areas where primary data may not be available or reliable. This chapter offered a number of reflections on using open, secondary sources to build and analyze a database of violent extremist actors, including: possible avenues for data collection and problems inherent in each, the importance of using robust defi-

---

50   Peter King, 'Building Resilience for Terrorism Researchers', Vox Pol, 2018.

51   Counter-Terrorism and Border Security Act 2019, Available at: https://services.parliament.uk/bills/2017-19/counterterrorismandborder-security.html.

52   Ofer Raban, 'Observations on the First Amendment and the War on Terror', Tulsa Law Review, 53.2 (2018), 141–57.

53   In the United Kingdom, for example, the Counter-Terrorism and Border Security Act 2019 lists "academic research" as a reasonable excuse for possessing such data, and will therefore not be prosecuted.

nitions and clearly articulated inclusion and exclusion criteria in database construction, different methods for data analysis, and ethical challenges for consideration.

There are a number of practical suggestions that researchers seeking to use secondary sources in database construction and/or research on violent extremism should take into account:

First, software can provide invaluable assistance to scholars and it is often available at no cost or through research institutions. Whether used for literature storage (i.e. Mendeley); news aggregators (such as Google News or LexisNexus) statistical analysis, (i.e. SPSS); or qualitative coding (i.e. NVivo), technological tools for research vastly augment the ease of conducting otherwise difficult research tasks.

Second, existing literature serves as an important source of insights on good practice and ideas that can be adopted for new research endeavors. It is well worth the time invested to read as many academic studies that have used similar methodologies within existing research. Doing so both reduces redundancies in existing research and provides important opportunities to build off of and supplement previous study. There is no need to reinvent the wheel.

Finally, it is important that researchers continually engage with others in the research community. Engaging with existing literature provides one avenue for doing this, as does attendance at research conferences and informal discussions with other researchers. Scholars are often keen to share their experiences to prevent others from making similar mistakes in the future.

The problem of a lack of primary data, whilst improving, is unlikely to go away soon and secondary sources can, despite a number of limitations, have been, and will continue to be a vitally important type of data for understanding violent extremism.

SOURCES

Alexander, Audrey, 'Cruel Intentions: Female Jihadists in America', Program on Extremism, 2016

Bakker, Edwin, Jihadi Terrorists in Europe: Their Characteristics and the Circumstances in Which They Joined the Jihad (The Hague, 2006)

Conway, Maura, and Lisa McInerney, 'What's Love Got to Do with It? Framing "JihadJane" In the US Press', Media, War & Conflict, 5 (2012), 6–21 <http://dx.doi.org/10.1177/1750635211434373>

Counter-Terrorism and Border Security Act 2019, Available at: https://services.parliament.uk/bills/2017-19/counterterrorismandbordersecurity.html

Dawson, Lorne, 'Clarifying the Explanatory Context for Developing Theories of Radicalization: Five Basic Considerations', Journal for Deradicalization, 18 (2019), 146–84

Department of Justice, Long Island Woman Pleads Guilty to Providing Material Support to ISIS, 26 Nov, 2018. Access via: https://www.justice.gov/usao-edny/pr/long-island-woman-pleads-guilty-providing-material-support-isis

European Parliament, General Data Protection Regulation, 2018, Access via: https://gdpr-info.eu/art-6-gdpr/

Europol, Terrorism Situation and Trend Report (TE-SAT) 2017, 2017

Federal Bureau of Investigation, What We Investigate – Terrorism, Access via: https://www.fbi.gov/investigate/terrorism

Field, Andy Discovering Statistics Using IBM SPSS Statistics, 5th edn (London: Sage Publications, 2018)

Ganor, Boaz, 'Defining Terrorism: Is One Man's Terrorist Another Man's Freedom Fighter?', Police Practice and Research, 29 (2002), 123–33 <http://dx.doi.org/10.1080/1561426022000032060>

Gill, Paul, 'Online Behaviours of Convicted Terrorists', Vox Pol, 2016

Gill, Paul, and Emily Corner, 'Lone Actor Terrorist Use of the Internet and Behavioural Correlates', in Terrorism Online: Politics Law and Technology, ed. by Lee Jarvis, Stuart Macdonald, and Thomas M. Chen (Abingdon, Oxon: Routledge, 2015), pp. 35–53

Gill, Paul, and Emily Corner, 'There and Back Again There and Back Again: The Study of Mental Disorder and Terrorist Involvement', American Psychologist, 2017, 1–35 <http://discovery.ucl.ac.uk/1521071/>

Gill, Paul, Emily Corner, Maura Conway, Amy Thornton, Mia Bloom, and John Horgan, 'Terrorist Use of the Internet by the Numbers: Quantifying Behaviors, Patterns, and Processes', Criminology & Public Policy, 16.1 (2017), 99-117' <http://dx.doi.org/10.1111/1745-9133.12249>

Gill, Paul, Emily Corner, A Thornton, and Maura Conway, 'What Are the Roles of the Internet in Terrorism?', Vox Pol, 2015

Glaser, Barney G, and Anselm L Strauss, The Discovery of Grounded Theory: Strategies for Qualitative Research (New Brunswick, NJ: AldineTransaction, 1967) <http://dx.doi.org/10.2307/2575405>

Hecker, Marc, '137 Shades of Terrorism: French Jihadists Before the Courts', Security Studies Center, 2018

Hill, Max, The Westminster Bridge Terrorist Attack, HM Government, 2018

Home Office, Revised Prevent Duty Guidance for England and Wales (London, 2015) <http://dx.doi.org/ISBN: 978-1-78246-7933-9>

Horgan, John, Neil Shortland, Suzzette Abbasciano, and Shaun Walsh, 'Actions Speak Louder than Words: A Behavioral Analysis of 183 Individuals Convicted for Terrorist Offenses in the United States from 1995 to 2012', Journal of Forensic Sciences, 61 (2016), 1228–37 <http://dx.doi.org/10.1111/1556-4029.13115>

King, Peter 'Building Resilience for Terrorism Researchers', Vox Pol, 2018

Mair, David, '#Westgate: A Case Study: How Al-Shabaab Used Twitter during an Ongoing Attack', Studies in Conflict and Terrorism, 40 (2017), 24–43 <http://dx.doi.org/10.1080/1057610X.2016.1157404>

Marchment, Zoe, Noémie Bouhana and Paul Gill, 'Lone Actor Terrorists : A Residence-to-Crime Approach', Terrorism and Political Violence, 2018, 1–26

Marchment, Zoe, and Paul Gill, 'Modelling the Spatial Decision Making of Terrorists : The Discrete Choice Approach', Applied Geography, 104 (2019), 21–31 <http://dx.doi.org/10.1016/j.apgeog.2019.01.009>

Meleagrou-Hitchens, Alexander, Seamus Hughes, and Bennett Clifford, 'The Travelers: American Jihadists in Syria and Iraq', Program on Extremism, George Washington University, 2018

National Consortium for the Study of Terrorism and Responses to Terrorism (START), Global Terrorism Database, (2018). Access via: https://www.start.umd.edu/gtd

Zachary S. Mitnik, Joshua D. Freilich and Steven M. Chermak, 'Post-9/11 Coverage of Terrorism in the New York Times', Justice Quarterly, 2018, 1–25

Pearson, Elizabeth, 'The Case of Roshonara Choudhry: Implications for Theory on Online Radicalization, ISIS Women, and the Gendered Jihad', Policy and Internet, 8 (2016), 5–33 <http://dx.doi.org/10.1002/poi3.101>

START, 'Profiles of Individual Radicalization in the United States (PIRUS) Codebook', 2018

Ravndal, Jacob Aasland, 'Anders Behring Breivik's Use of the Internet and Social Media', Journal EXIT-Deutschland, 2 (2013), 172–85

Sageman, Marc, 'The Stagnation in Terrorism Research', Terrorism and Political Violence, 26 (2014), 565–80 <http://dx.doi.org/10.1080/09546553.2014.89564 9>

Sageman, Marc, Understanding Terror Networks (Philadelphia, PA: University of Pennsylvania Press, 2004)

Schmid, Alex P., Radicalisation, De-Radicalisation, Counter-Radicalisation: A Conceptual Discussion and Literature Review, International Centre for Count- er-Terrorism (2013) <http://www.icct.nl/download/file/ICCT-Schmid-Radicalisation-De-Radicalisation-Counter-Radicalisation-March-2013.pdf>

Schmid, Alex P, 'Terrorism - The Definitional Problem', Case Western Reserve Journal of International Law, 36 (2004), 375–419 <http://libproxy.northgeorgia. edu/login?url=http://search.proquest.com/docview/211184999?accountid=12774>

Schuurman, Bart, 'Research on Terrorism, 2007–2016: A Review of Data, Methods, and Authorship', Terrorism and Political Violence, 2018, 1–16 <http://dx. doi.org/10.1080/09546553.2018.1439023>

Schuurman, Bart, and John Horgan, 'Rationales for Terrorist Violence in Homegrown Jihadist Groups: A Case Study from the Netherlands', Aggression and Violent Behavior, 27 (2016), 55–63 <http://dx.doi.org/10.1016/j.avb.2016.02.005>

Silke, Andrew, 'Contemporary Terrorism Studies: Issues in Research', Critical Terrorism Studies: A New Research Agenda, 2009, 34–48 <http://dx.doi. org/10.4324/9780203880227>

Cathy Urquhart, Grounded Theory for Qualitative Research: A Practical Guide (London: Sage Publications, 2013)

Yuen, L. (2016) Gulad Omar: The Path to ISIS and the Story You Haven't Heard, MPR News, Dec 22, Available at: https://www.mprnews.org/story/2016/12/21/ path-to-isis-minnesota-guled-omar-mpr-interviews.

RESOLVE /
NETWORK

INSIGHT INTO VIOLENT EXTREMISM AROUND THE WORLD

The RESOLVE Network is a global consortium of researchers and research organizations committed to delivering fresh insight into violent extremism around the world. The Network provides access to open-source data, tools, and curated research to ensure policy responses to violent extremism are evidence based. Members of the Network work in parts of Africa, Asia, Europe, and the Middle East to promote empirically driven, locally defined responses to conflict and to support grassroots research leadership on violent extremism.

Our partners operate in more than 25 countries where challenges with conflict are an everyday reality. We are passionate about amplifying credible local voices in the fight to mitigate the destabilizing risks of social polarization and political violence. The RESOLVE Network Secretariat is housed at the U.S. Institute of Peace, building upon the Institute's decades-long legacy of deep engagement in conflict-affected communities.

To learn more about the RESOLVE Network, our partners and how to get involved visit our website, www.resolvenet.org, and follow us on Twitter: @resolvenet.