

LESSONS LEARNED ON THE METHODOLOGICAL CHALLENGES IN STUDYING RARE VIOLENT INCIDENTS

BY **BASIA E. LOPEZ, DANIELLE M. CRIMMINS, AISHA J. QURESHI, AND NADINE FREDERIQUE**

To increase knowledge and aid prevention efforts, the research community must develop a strategy to source, code, check, and analyze the data surrounding rare violent incidents.



Violent incidents such as public mass shootings and terrorist events are rare but have widespread, catastrophic impacts on society. These devastating and often high-profile events raise questions about their causes and how best to prevent them. Yet their infrequent occurrences make them hard to predict.

Finding answers to questions that can guide the prevention of such events is complicated. Naturally, one may turn to “what the research shows” to inform these discussions, but research on rare — albeit high-impact — events is incredibly difficult to carry out and even more challenging to generalize. Research designs and methods are continuously evolving; however, rare incidents remain an exceptionally difficult phenomenon to study.

For example, public mass shootings are both infrequent and context-dependent, meaning that the situation, background, or location differs in each case. As a result, it is challenging for researchers to quantify their impact.¹ Further, the rarity of public mass shootings makes it difficult to develop and test theoretical models, owing to the dearth of good-quality data.

Research must move beyond these limitations to advance the criminal justice field’s capacity to prevent rare violent incidents. For example, we need reliable and valid information on which factors may lead an individual to commit a mass shooting. Determining these factors requires rigorous data collection methods and analysis. But this is just the starting point. We can also capitalize on methods used in other disciplines that study rare events to potentially help forecast violent incidents and suggest appropriate mitigation strategies.²

Research on rare — albeit high-impact — events is incredibly difficult to carry out and even more challenging to generalize.

As the scientific research, development, and evaluation arm of the U.S. Department of Justice, the National Institute of Justice (NIJ) has played a crucial role in improving the knowledge and understanding of rare violent incidents, such as public mass shootings and terrorist events. The Institute has funded various studies over the years that seek to provide rigorous data, a better understanding of what the data convey, improved connectivity between data sources, and consistent definitions so that the field has more information about who perpetrates mass violence, their motivations, and how they plan and carry out their attack.

In late 2019, NIJ convened a meeting with current and former grantees and research scholars focusing on rare mass murder incidents.³ (For a list of meeting participants, see sidebar “NIJ Topical Meeting on Rare Incidents Data Collection Models To Advance Research on Mass Violence.”) The participants discussed methodologies, the nuances of collecting and analyzing data on these incidents, and challenges related to validity and accuracy. This article briefly summarizes these discussions, along with NIJ’s efforts on studying rare violent incidents. It closes with implications for the research community and the criminal justice field to consider.

Defining Key Terms

A mass shooting falls under the broader category of mass murder, which is defined as the willful (nonnegligent) killing of at least four human beings by another person by whatever means (for example, bomb, knife/machete, firearm, or use of a vehicle).⁴ A mass shooting is a mass murder that involves at least one firearm and is typically carried out at a single point in time and in one location.⁵

Mass shootings have additional subcategories, which include:

- Public mass shootings (school shootings, workplace shootings, and shootings in other publicly accessible places, such as cinemas, restaurants, bars, places of worship, or outdoor events in open spaces).
- Domestic mass shootings (familicide with a firearm).
- Shootings connected to crimes, such as robberies and gang-related shootings (a form of organized crime and “turf wars”).

Another example of mass murder is a terrorist attack during which multiple victims are intentionally murdered. Unlike in mass shootings (which may or may not be driven by an ideological motive), terrorist attacks are always driven by ideological objectives and are often carried out by means other than a firearm.

NIJ’s History of Research on Rare Incidents

Over the years, NIJ has been committed to supporting research on rare incidents such as terrorism and mass shootings, with a particular emphasis on school shootings. Recognizing that there are also knowledge gaps around mass shootings outside of the school setting, the Institute began investing in research on domestic radicalization to violent extremism and terrorism in 2012 and on other public mass shootings in 2018. These rare violent incidents are all linked by a lack of data, which has prompted researchers to use open-source databases to better understand them.

Terrorism

Following the September 11 terrorist attacks, NIJ began funding research focused on developing terrorism databases, improving responses to terrorism, and studying the composition of terrorist organizations. In 2012, NIJ received new congressional direction to fund research specifically focused on developing a better understanding of the domestic radicalization process and advancing

evidence-based strategies to effectively intervene and prevent radicalization in the United States.⁶

While this shift in focus emphasized the study of the radicalization process, data collection efforts on terrorist attacks and related incidents, including factors surrounding radicalization, continued. This NIJ portfolio dedicated resources to developing and supporting open-source databases to understand rare violent incidents. For example, NIJ has helped fund the Terrorism and Extremist Violence in the United States (TEVUS) database and portal, which “compiles behavioral, geographic, and temporal characteristics of extremist violence in the United States dating back to 1970.”⁷ Similarly, NIJ has helped fund the Profiles of Individual Radicalization in the United States (PIRUS), a database that “contains deidentified individual-level information on the backgrounds, attributes, and radicalization processes of over 2,200 violent and non-violent extremists who adhere to far right, far left, Islamist, or single-issue ideologies in the United States covering 1948-2018.”⁸

These efforts have highlighted that researchers collecting data on terrorist attacks face similar methodological problems as researchers collecting data on school and other public mass shootings.

School Shootings

One of the first NIJ-funded studies on school shootings was the Safe School Initiative (1999-2004). This project, carried out by the U.S. Department of Education and the U.S. Secret Service, focused on threat assessment. The project’s goal was to identify information that could be obtainable, or “knowable,” prior to an attack.⁹ It used administrative source data, including investigative, school, court, and mental health records, as well as data derived from interviews. However, the researchers had limited access to administrative sources, and they often had to obtain information through publicly available means.

From 2014 to 2017, NIJ funded several other research projects related to school shootings under the Comprehensive School Safety Initiative. One of these projects, The American School Shooting

Study (TASSS), developed a database modeled after terrorism studies that used open-source data to examine both school shootings and mass shootings.¹⁰ TASSS is a national, open-source database that includes all publicly known shootings that resulted in at least one injury on K-12 school grounds in the United States between January 1, 1990, and December 31, 2016.

Other Public Mass Shootings

In 2015, NIJ collaborated with the Office for Victims of Crime to fund a project examining the impact of school and other mass shootings on communities and individuals.¹¹ Since then, NIJ has prioritized research on public mass shootings, funding four projects in 2018 and one in 2019. Three of these projects aimed to build or expand databases on public mass shootings using publicly available sources to answer specific research questions and ensure the data were rigorously collected.¹²

Collecting Data on Rare Incidents

When studying rare incidents, researchers must often build a database from scratch. These databases are typically based on research questions and the variables needed to answer those questions, thus requiring a well-developed process. During the NIJ meeting on rare incidents data collection, participants described the following steps in building a database (see exhibit 1):

1. Construct the codebook.
2. Identify cases.
3. Manage sources.
4. Recruit and train a coding team.
5. Code cases.
6. Perform quality control and data testing.
7. Disseminate.

Participants also discussed key challenges that often emerge during this process and, where applicable, identified possible solutions. They noted, however, that in some cases, there are no clear solutions. The

Exhibit 1. Steps To Build a Database



overall goal is to create a nuanced and comprehensive database that is reliable and usable to help the criminal justice system prevent violent crime.

Constructing the Codebook

A database must be functional and reliable in order to be valid and useful. The research team should construct a “codebook” based on an analysis of existing literature and databases, their critique of the important issues, the questions they want to answer, and the theories they want to examine.

Publicly available data sources include a wide range of variables, which results in both complexity and missing information. When analyzed, this may produce invalid or even erroneous findings. Thus, the meeting participants stated that, instead of including all possible variables, the research team should pick the “right” variables (considering the sources available). The participants suggested using broad definitions and allowing future users to filter the data to maximize functionality. They also stressed the importance of using the right measurement scheme when

constructing the codebook — for example, entering the person’s exact age, not an age range.

Participants noted that it is imperative to use several test cases to see what the codebook captures and misses and what is too subjective — and then make necessary changes. If the research team wants to expand the codebook, they must focus on new research questions and a new set of variables and then operationalize them accordingly.

Identifying Cases

Next, the research team must identify cases to include in the database. The team must be clear from the start about inclusion criteria and set benchmarks.

The meeting participants suggested using multiple methods to help build a representative sample, including:

- Conducting searches from lists generated by news aggregators and Boolean search strings to select cases based on specified inclusion criteria.¹³
- Using customized news alerts.
- Following external sources and social media accounts for cases that are not high profile.

The research team should use protocols to search for and collect open-source documents to identify incidents. They should conduct a systematic review of the literature to locate sources that could have pertinent information, including databases from other academics and organizations. They can verify leads from internet sources against official, publicly available, or requested records of specific incidents.

According to the participants, to ensure they have enough cases for critical analysis, researchers should not limit the number of incidents included in the database. Nevertheless, inclusion criteria should be based strictly on the study’s goals or research questions and the associated operational definitions. Participants added that, instead of limiting data to events only, researchers should also collect data on individuals, their lives, and their characteristics, as well as any additional variables required to inform

prevention. Accessing and capturing life history data can be difficult; however, the participants noted that this information is often available through open sources and can be triangulated (that is, use multiple data sources to ensure the information is accurate).

Managing Sources

The meeting participants stressed that databases should include a wide range of sources to limit poor source validity. The research team should consider all possible data sources even though they likely will not use all of them.

For example, the research team should discuss whether they will use court records and have a budget line item to cover potential costs. Most state and local cases are not available on the internet, which limits access to those court documents. As such, the researchers will have to contact a given agency to obtain information, sometimes through relevant state or local government's freedom of information laws.¹⁴ If unsuccessful, the fallback is often news sources. Participants stated that researchers must use multiple sources to corroborate news data and use caution when entering these data, warning that an author might edit or update information already obtained by the research team.

The participants added that researchers can also access primary sources and historical archive data. For example, the FBI's Supplementary Homicide Report includes voluntarily entered data.

Research teams often develop a ranking system for the reliability of source documents, with court documents at the top and personal opinions expressed in blogs or editorials at the bottom. Meeting participants suggested developing a ranking system that includes the number of sources used for corroboration. Thus, the research team would include the source with higher-ranking information (for example, a court case file) in the database, especially when corroborated with another source. The team can exclude the lowest-ranked information (for example, internet blogs), even when corroborated by another, higher-ranking source.

The research team can use an online relational database to manage sources, which offers efficiency at a reasonable cost. This approach allows team members to link sources to individuals and single documents to multiple people. Thus, people internal to the project, as well as external users, can view the same pool of sources collated in the same place.

Recruiting and Training a Coding Team

The first step in the coding process is to recruit and train a coding team. When recruiting coders, the research team must be transparent about the tedious nature of coding to ensure that project staff are fully aware of their assigned tasks and the larger, overarching goals of these ambitious data collection efforts.

It is crucial to allow adequate time for training, including a test period to work out technical issues and to build a team that understands what it means to be a data collector or coder. The coders should undergo extensive training to review the codebook, practice cases, and perform partner coding based on a rigorous training protocol and established inclusion and exclusion criteria. Untrained coders may misidentify or miss information, resulting in data errors and poor validity of the results.

After initial training, the coders can perform multiple rounds of independent searching and critiques. Their training should emphasize how to think like an investigator and track down leads. For instance, they should be intimately familiar with case files so that when they come across buried information (for example, a prior conviction), they can use it as a lead to get more information. The coders should also meet frequently to discuss questions and have regular meetings with the research team to address any issues that arise.

The meeting participants acknowledged the time and resources needed to conduct this type of research and suggested training a large team, if possible. Depending on an organization's resources, they noted that the coding team could consist of interns, research assistants, graduate students, or contract

staff. Another possible solution is using technology — such as machine learning, web crawlers, and textual analysis programs — to identify incidents. The participants discussed bringing computer scientists into the conversation and merging social science and computer science to make the process more effective. The field of computational social science offers promise for applying new methods of analyzing complex social science problems in dynamic social systems and complex organizations. These can include the dynamics of epidemics or social movements, among others.¹⁵

Coding Cases

Coding must be conducted effectively, reliably, and accurately. The meeting participants said that to overcome subjective judgment in the coding process, the research team must establish clear codebook guidelines with thorough instructions. Setting a goal (for example, number of cases per a specific period) is also helpful.

Setting a range of possible data values in the data collection tool can help reduce data entry errors. Practicing version control (that is, having only one master file) also preserves data accuracy. The participants said that quality control and data tests are necessary and reiterated that researchers can use multiple sources to corroborate and triangulate information. The participants also suggested double- or triple-coding at least 25% of the cases to check intercoder reliability. This will allow the team to reconcile disagreements and modify the codebook if needed.

Another important aspect to consider when coding cases is to account for the time that has elapsed between the incident and coding. The workflow may not be in “real time,” so monitoring any updates to the case — including corrections made to initial codes based on new information — is crucial. Data must go through layers of fact-checking and independent coding to ensure accuracy.

Quality Control and Data Testing

The research team should incorporate quality control and data testing from the start — for example,

double coding, preliminary analyses to look for logical impossibilities, and having multiple people check for errors.

The meeting participants underscored the issue of missing data, as there is a difference between missing variables and no information found. For example, some variables have a higher likelihood of leaving a paper trail (such as military records) compared with variables that are harder to determine (such as substance use). To adequately address this challenge, the participants said that the research team should create a plan for dealing with missing data and follow it consistently.

The research team must collaborate closely with the coding team to work through difficult cases. After the coders review a file and conduct targeted searches to fill in missing values, the research team should assess the type and number of documents reviewed in the search file. Nevertheless, some values will always be missing. A different coder must then confirm values and conduct a final targeted search to fill in missing data, while also flagging any reliability concerns. As new information comes out (for example, in a book written years after the event), the research team must ask coders to reevaluate the data and make any necessary changes.

Dissemination

The participants discussed ways to communicate findings to multiple audiences, including academics, practitioners (such as law enforcement), and policymakers — something the participants agreed has become increasingly more difficult in recent years. They added that data dissemination often receives inadequate attention until too late, yet it is key to avoiding data misuse and misinterpretation.

The media is a primary (although imperfect) vehicle to get messages to the public. The participants acknowledged that there is no quick way to explain the process of creating a database. However, they recommended highlighting the three or four most important data pieces, saying this strategy can help prevent misinterpretation.

Several participants also suggested that when working with the media, researchers should:

- Clearly explain how the data can and cannot be used and the limitations of the data.
- Make the data available through multiple outlets.
- Provide a user guide and codebook.
- Write a frequently asked questions section for publicly accessible databases.
- Use technology tools to visualize the data and make the data interactive and accessible (for example, create infographics).

They also recommended using language that everyone can digest (for example, writing op-eds).

Implications for the Field

Rare incident research on terrorism and public mass shootings has the potential to affect policy and aid prevention efforts. Although researchers must consider many methodological steps, challenges, and limitations when investigating rare incidents, establishing strong data collection and data entry procedures is imperative to producing rigorous research. Researchers must understand the nuances and procedures essential for studying these rare phenomena in order to accurately translate the findings to the field.

Researchers must adequately describe their methods, any known limitations, and the operational definitions so that those outside the project team and outside the research community can understand the findings. Policymakers and practitioners must have all the necessary information to interpret the research and inform their decisions. Specifically, they should review the data sources and how the project team created the database and defined the variables. Researchers should present the findings to practitioners and policymakers to help remove skepticism regarding open-source data from sources such as media reports.

NIJ Topical Meeting on Rare Incidents Data Collection Models To Advance Research on Mass Violence

San Antonio, Texas, September 24-25, 2019

Participants:

- Steven Chermak, Michigan State University
- Nadine M. Connell, Griffith University, Queensland (Australia)
- James Densley, Metropolitan State University (Minnesota)
- Grant Duwe, Minnesota Department of Corrections
- James Alan Fox, Northeastern University (Massachusetts)
- Joshua D. Freilich, John Jay College of Criminal Justice (New York)
- Michael Jensen, University of Maryland
- Hannah Laqueur, University of California, Davis
- Jillian Peterson, Hamline University (Minnesota)
- Travis C. Pratt, University of Cincinnati Corrections Institute
- Michael Rocque, Bates College (Maine)
- Jillian J. Turanovic, Florida State University
- Basia Lopez, NIJ
- Danielle Crimmins, NIJ
- Nadine P. Frederique (former NIJ social science analyst)
- Mark Morgan (former NIJ policy advisor)
- David B. Muhlhausen (former NIJ director)
- Phelan Wyrick, Office of Justice Programs/ Office of the Assistant Attorney General
- Notetaker: Mary Beth de Ribeaux, CSR, Incorporated

It is also important for policymakers to understand what research questions the project team considered, because results will vary based on the questions and operational definitions. For example, if the researchers defined a mass shooting as three fatalities versus four fatalities, the trend line will differ. When these factors are not considered, misinformation spreads.

Those who work in fields related to rare violent incidents — whether in policy, research, or practice — are often asked what is being done to prevent these tragedies from occurring. But it is not always possible to identify the number of incidents that were prevented. Foiling plots and implementing timely interventions are critical; however, given the developing nature of this research compared with other areas in criminology, one of our best opportunities to identify patterns and trends, answers to questions, and functional tools for law enforcement is to bolster databases and data collection efforts. Developing a meticulous strategy to source, code, check, and analyze the data surrounding rare violent incidents remains paramount. We must consider the lessons learned from the creation and expansion of other pioneering databases, especially as the nature of terrorist and mass shooting threats in the United States continues to evolve. To get ahead of — or even keep up with — the threat, a strong foundation of knowledge will remain key to prevention and intervention efforts.

About the Authors

Basia E. Lopez, MPA, is a social science analyst in NIJ's Office of Research, Evaluation, and Technology, where she leads the research and evaluation portfolio on firearms violence and mass shootings, forensic intelligence, and community violence intervention and prevention. **Danielle M. Crimmins**, Ph.D., is a social science research analyst in NIJ's Office of Research, Evaluation, and Technology. **Aisha J. Qureshi**, M.A., is a social science analyst in NIJ's Office of Research, Evaluation, and Technology, where she leads the research and evaluation portfolio on domestic radicalization and violent extremism.

Nadine Frederique, Ph.D., is a former social science

analyst at NIJ and currently a social science analyst at the U.S. Department of Homeland Security's Science & Technology Directorate, where she leads research efforts on terrorism, targeted violence, and radicalization.

This article discusses the following awards:

- "Understanding the Causes of School Violence Using Open Source Data," award number 2016-CK-BX-0013
- "Examination of the Short- and Long-Term Impact of School Shootings," award number 2015-VF-GX-K113
- "A Comprehensive Assessment of Deadly Mass Shootings, 1980-2018," award number 2018-75-CX-0024
- "Mass Shooter Database," award number 2018-75-CX-0023
- "The Nature, Trends, Correlates and Prevention of Mass Public Shootings in America, 1976-2018," award number 2018-75-CX-0025
- "Firearm Purchase Behavior and Subsequent Adverse Events," award number 2018-75-CX-0026
- "Improving the Understanding of Mass Shooting Plots," award number 2019-R2-CX-0003
- "Empirical Assessment of Domestic Radicalization," award number 2012-ZA-BX-0005

Notes

1. Shelby M. Scott and Louis J. Gross, "The Mathematics Underlying Gun Violence," *SIAM News*, Philadelphia, PA: Society for Industrial and Applied Mathematics, <https://sinews.siam.org/Details-Page/the-mathematics-underlying-gun-violence>.
2. Grant Duwe et al., "Forecasting the Severity of Mass Public Shootings in the United States," *Journal of Quantitative Criminology* 38 (2021): 385-423, <https://doi.org/10.1007/s10940-021-09499-5>.
3. NIJ Topical Meeting on Rare Incidents Data Collection Models, San Antonio, TX, September 24-25, 2019.
4. This is the most commonly used definition. However, it hinders examination of incidents where there was a premeditated intent to kill multiple victims but the plan was either not carried out or resulted in fewer or no fatalities and injuries. For the development of prevention strategies, there are benefits to including both averted attacks (e.g., due to official or bystander intervention) and less deadly attacks (e.g., due to near misses, weapon malfunction, rapid medical response, or fewer potential victims at the crime scene).

5. William J. Krouse and Daniel J. Richardson, "Mass Murder with Firearms: Incidents and Victims, 1999-2013," Washington, DC: Congressional Research Service, July 2015, <https://sgp.fas.org/crs/misc/R44126.pdf>, 2.
6. See related article in this issue, "What NIJ Research Tells Us About Domestic Terrorism."
7. "Terrorism and Extremist Violence in the United States (TEVUS) Portal," National Consortium for the Study of Terrorism and Responses to Terrorism, <https://www.start.umd.edu/tevus-portal>.
8. "Profiles of Individual Radicalization in the United States (PIRUS)," National Consortium for the Study of Terrorism and Responses to Terrorism, <https://www.start.umd.edu/data-tools/profiles-individual-radicalization-united-states-pirus>.
9. Bryan Vossekuil et al., "The Final Report and Findings of the *Safe School Initiative*: Implications for the Prevention of School Attacks in the United States," Washington, DC: U.S. Secret Service and U.S. Department of Education, July 2004, <https://www2.ed.gov/admins/lead/safety/preventingattacksreport.pdf>.
10. Joshua D. Freilich et al., "Understanding the Causes of School Violence Using Open Source Data," Final report to the National Institute of Justice, award number 2016-CK-BX-0013, August 2021, NCJ 301665, <https://nij.ojp.gov/library/publications/understanding-causes-school-violence-using-open-source-data>.
11. National Institute of Justice funding award description, "Examination of the Short- and Long-Term Impact of School Shootings," at the Regents of the University of California, Los Angeles, award number 2015-VF-GX-K113, <https://nij.ojp.gov/funding/awards/2015-vf-gx-k113>.
12. See National Institute of Justice funding award description, "A Comprehensive Assessment of Deadly Mass Shootings, 1980-2018," at Florida State University, award number 2018-75-CX-0024, <https://nij.ojp.gov/funding/awards/2018-75-cx-0024>; National Institute of Justice funding award description, "Mass Shooter Database," at The Hamline University of Minnesota, award number 2018-75-CX-0023, <https://nij.ojp.gov/funding/awards/2018-75-cx-0023>; National Institute of Justice funding award description, "The Nature, Trends, Correlates and Prevention of Mass Public Shootings in America, 1976-2018," at Northeastern University, award number 2018-75-CX-0025, <https://nij.ojp.gov/funding/awards/2018-75-cx-0025>; National Institute of Justice funding award description, "Firearm Purchase Behavior and Subsequent Adverse Events," at the University of California, Davis, award number 2018-75-CX-0026, <https://nij.ojp.gov/funding/awards/2018-75-cx-0026>; and National Institute of Justice funding award description, "Improving the Understanding of Mass Shooting Plots," at The Rand Corporation, award number 2019-R2-CX-0003, <https://nij.ojp.gov/funding/awards/2019-r2-cx-0003>.
13. This is a type of search that allows users to combine keywords with operators (or modifiers) such as AND, NOT, and OR to produce more relevant results.
14. The federal Freedom of Information Act (FOIA) requires federal agencies to provide access to certain types of records, subject to exceptions as set out in the statute (5 U.S.C. § 552 (2) (A)-(E)). It applies only to federal agencies as defined in the statute and does not apply to state agencies or local governments. FOIA does not create a right of access to records held by Congress, the courts, or by state or local government agencies. Although FOIA doesn't directly apply to local governments, many states and some local governments have passed their own freedom of information laws. Requests for information from a state or local government must be made under that jurisdiction's freedom of information legislation.
15. National Academy of Engineering, *Frontiers of Engineering: Reports on Leading-Edge Engineering From the 2013 Symposium*, Washington, DC: The National Academies Press, 2014, <https://doi.org/10.17226/18558>.

Image source: David Gyung/Shutterstock.

NCJ 306125

Cite this article as: Basia E. Lopez et al., "Lessons Learned on the Methodological Challenges in Studying Rare Violent Incidents," *NIJ Journal* 285, June 2024, <https://nij.ojp.gov/topics/articles/lessons-learned-methodological-challenges-studying-rare-violent-incidents>.