Mapping Political Violence in Turkey: When Terrorism and Civil War Converge

Betul Ozturan[1]

Abstract

This study introduces the Political Violence in Türkiye Event Dataset (POLVITED), a novel resource designed to provide a more comprehensive estimation of political violence incidents and fatalities in Türkiye. I argue that reporting biases, along with variations in the scope and definitions of existing event datasets, significantly affect the quantitative study of conflict and violence. To address these challenges, I employ the Matching Event Data by Location, Time, and Type (MELTT) methodology to integrate data from the Global Terrorism Database (GTD) and the Uppsala Conflict Data Programme Georeferenced Event Dataset (UCDP-GED). The findings reveal that UCDP-GED captures 60% of political violence incidents in Türkiye, GTD records 40%, and only 15% of incidents are documented in both datasets. By eliminating duplicate records and mitigating reporting bias, POLVITED offers a more comprehensive representation of political violence in Turkey. Additionally, a comparison with official reports demonstrates that while reporting bias cannot be entirely eradicated, data integration significantly enhances dataset coverage and reliability. This new data will help researchers better understand the nature of violence and the relation between groups.

Keywords

Political Violence, Terrorism, Turkey

Introduction

Research on terrorism and civil war has made significant progress in uncovering the drivers and mechanisms behind political violence, as well as its far-reaching consequences. Scholars have developed sophisticated theoretical frameworks and quantitative methods to analyze these phenomena, deepening our understanding of conflict dynamics. However, an important dimension remains underexplored: the interconnected nature of different forms of political violence. Current research often isolates terrorism from civil war, leaving unresolved questions about how these forms overlap and influence each other. Additionally, the lack of universally accepted definitions for various types of political violence has created challenges in distinguishing between them, introducing ambiguity into both conceptual and empirical analyses. This uncertainty is further reflected in the datasets used to study political violence, where issues related to validity, accuracy, and reliability often arise. In response to these challenges, this study aims to advance the field by addressing both the conceptual and methodological gaps that hinder a comprehensive understanding of political violence.

To address these gaps, this study emphasizes the importance of comprehensive datasets to better understand the complexities of political violence. Definitions play a crucial role in shaping the scope and focus of research on terrorism and civil war, yet their variability across studies has significant implications for empirical analyses. By examining how definitions influence the categorization and measurement of political violence, this paper contributes to ongoing discussions about the reliability and comparability of conflict datasets. Moreover, concerns about reporting biases, a longstanding challenge in the field, call for innovative approaches to mitigate their impact and improve the validity of data used in quantitative analyses.

This study leverages advancements in data integration methodologies to address these challenges. Specifically, I apply the Matching Event Data by Location, Time, and Type (**MELTT**) method (Donnay et al. 2019) to integrate two of the most widely used datasets on political violence: the Global Terrorism Database (**GTD**) (LaFree and Dugan 2007) and the Uppsala Conflict Data Programme Georeferenced Event Dataset (**UCDP-GED**) (Sundberg and Melander 2013). The MELTT method facilitates the systematic identification and removal of duplicate records across these datasets, enabling the creation of a novel resource that contains uniquely coded events. This approach not only eliminates redundancies but also provides a more comprehensive and accurate representation of political violence in Türkiye, addressing key methodological and empirical gaps in the literature.

In this study, I use Türkiye as a case study to estimate the total number of fatalities resulting from political violence and to demonstrate how and when terrorism is used as a tactic by rebel groups within the context of a civil war. This estimation provides a clearer understanding and comparison of the human cost of political violence as reported in various datasets. Türkiye, which has experienced numerous terrorist attacks, offers an ideal context for examining the reporting of violent political incidents. Three key

[1]Boston College, USA

Corresponding author:
Betul Ozturan
Email: ozturan@bc.edu

Prepared using sagej.cls *[Version: 2017/01/17 v1.20]*

factors make it particularly suitable. First, the infrastructure for news reporting has largely remained intact, ensuring relatively consistent coverage. Second, the fluctuating levels of political violence in Türkiye help minimize the risk of reporting fatigue for specific types of events. Third, although not entirely consistent or comprehensive, official reports on political violence are accessible to researchers. These factors collectively make Türkiye a relevant and informative case for studying the dynamics and reporting of political violence, as well as understanding the role of terrorism within civil conflict.

Beyond the methodological suitability of Türkiye as a case study, there is a clear need for a comprehensive dataset to thoroughly investigate political violence within the country. While the Committee on Human Rights Inquiry of the Grand National Assembly of Türkiye provides official data on terror-related fatalities up until 2012 (TBMM 2013), accurate and systematic reporting on the human cost of political violence has been lacking since then, particularly for terrorist organizations other than the PKK. This study addresses this gap by emphasizing the importance of reliable and accessible data in a context where information is often fragmented and difficult to obtain.

The contributions of this paper are threefold. First, it establishes a theoretical link between terrorism and civil war, emphasizing the overlapping concepts that introduce ambiguities in data collection. Second, it demonstrates how this theoretical framework can inform and guide empirical research. Third, it provides statistical estimates of political violence through the Political Violence in Türkiye Event Dataset (**POLVITED**).

The structure of the paper is as follows: First, I examine the importance of event datasets in conflict studies and address the methodological concerns associated with their use. Next, I discuss how integrating different datasets can help mitigate these biases. Third, I describe the data and methods employed in this study. In the results section, I present the **POLVITED** dataset and discuss its significance. Finally, I compare the officially published statistics on fatalities from political violence with the **POLVITED** dataset to assess its coverage. The paper concludes by exploring potential directions for future research and the policy implications of this novel dataset.

Theory

This section explores current debates in political science regarding the use of event datasets in quantitative conflict studies. After outlining the data collection methods employed in event datasets, I discuss key statistical and methodological concerns associated with their use. Specifically, I focus on underreporting bias and definitional ambiguities in the classification of political violence. These challenges include the absence of a universally accepted definition of terrorism, the methodological implications of this gap, the varying criteria used to define civil wars, and how these definitional differences impact research outcomes. The section concludes by outlining how this paper aims to address these issues, contributing to the development of more reliable and comprehensive approaches to studying political violence.

Methodological Concerns of Data Collection on Political Violence

Two primary biases are commonly discussed in the political science literature as critical challenges to the reliability of empirical studies using event datasets. The first is the variation in scope and definitions employed by different datasets, and the second is the reporting bias introduced by news agencies, governmental archives, or non-governmental organizations. Both biases significantly impact the accuracy and reliability of conflict research.

This section evaluates these two biases, focusing on their implications for empirical studies of political violence. It also explores potential approaches to address and mitigate these challenges, contributing to the development of more robust methodologies in event dataset research.

Differences in Definitions and Scope of the Datasets

Each dataset relies on its own methodology and coding rules to determine how events are classified and recorded. A critical first step in creating any dataset is defining the phenomena it aims to capture. In a field as contested as conflict studies, reaching universally accepted definitions for key concepts is particularly challenging. Sambanis highlights how definitional differences among datasets can significantly impact analyses. By applying the same probit model to ten datasets—COW 1994, COW 2000, Collier and Hoeffler 2001 (Collier and Hoeffler 2004), Licklider 1995 (Licklider 1995), Gleditsch et al. 2001 (Gleditsch et al. 2001), Fearon and Laitin 2003 (Fearon and Laitin 2003), Leitenberg 1991 (Leitenberg 2001), Regan 1996 (Regan 1996), Doyle and Sambanis 2000 (Doyle and Sambanis 2000), and Sambanis 2004 (Sambanis 2004)—Sambanis demonstrates how these differences can lead to overestimation or underestimation of correlations between variables, or to variations in statistical significance.

The discrepancies among these datasets often arise from differences in coding rules, such as those governing the onset and termination of civil wars, or the thresholds for casualties required to classify a conflict as a civil war. These inconsistencies pose significant challenges for researchers attempting to compare findings across studies. To address this issue, Sambanis advocates for a broader investigation of political violence as a general phenomenon. By examining when and how different forms of political violence emerge, researchers can move beyond isolating civil war as a distinct category and develop a more comprehensive understanding of conflict dynamics (Sambanis 2004).

The definition and scope of political violence remain contested within the field of political science. Political violence is a broad concept encompassing various forms of activity. One fundamental challenge in categorizing an event as violence lies in defining what constitutes violence itself (Mars 1975). Does violence refer exclusively to acts involving physical force, or does it also include threats of violence? Moreover, identifying an act as political violence requires understanding its purpose and motivations—an inherently complex and often subjective task (Mars 1975).

An alternative approach in empirical studies is to categorize political violence into distinct types, such as state repression, civil war, domestic violence, terrorism, and electoral violence. However, these categories are not

universally agreed upon, and their boundaries often remain ambiguous. In this study, I focus on civil war and terrorism to illustrate how different categories of political violence can overlap, highlighting the challenges these overlaps pose for empirical research and data collection.

Firstly, civil war can be defined as the use of force between two organized parties to achieve political objectives. The general prerequisites for an event to be classified as a civil war include being perpetrated by organized groups, involving reciprocal violence (as opposed to unilateral acts such as pogroms or genocide), and challenging sovereign authority (Sambanis 2004). However, empirical definitions of civil war can vary significantly. For example, the Uppsala Conflict Data Program (UCDP) defines armed conflict as "a contested incompatibility that concerns government or territory, over which the use of armed force between the military forces of two parties, at least one of which is the government of a state, has resulted in at least 25 battle-related deaths each year" (Gleditsch et al. 2002). In contrast, the Correlates of War (COW) project defines civil war as "any armed conflict that involved: (1) military action internal to the metropole of the state system member; (2) the active participation of the national government; (3) effective resistance by both sides; and (4) a total of at least 1,000 battle-deaths during each year of the war" (Sarkees and Wayman 2010).

The main empirical disputes over the definition of civil war center around three key issues:

- **Start and end dates:** How should researchers code the beginning and end of a civil war? Should the start be marked by the first casualty, or only after a certain threshold of violence is reached?
- **Thresholds:** What should the threshold for violence be? Should it account for population differences between countries (e.g., violence per capita), or should it be an absolute number of casualties?
- **Casualty classification:** Should researchers include civilian casualties in the count, or only consider battle-related casualties?

These debates illustrate the complexity of operationalizing civil war as a research concept (Sambanis 2004).

The answers to these definitional questions significantly influence whether a country is classified as experiencing a civil war and the specific years attributed to that conflict. These variations impact the quantitative analysis of civil wars, which often relies on numerical data and estimates derived from existing datasets. In the country-year format commonly used in such datasets, these differences are almost inevitable, leaving researchers with limited control over the decisions made in data collection and aggregation. In contrast, event datasets provide researchers with greater autonomy to define and control variables in their studies, offering more flexibility compared to the rigid structure of the country-year format.

Another form of political violence is terrorism, the definition of which is even more debated and complex than that of civil war (Ganor 2002; Weinberg et al. 2004; Marsden and Schmid 2011; Schmid 2011a; Shanahan 2016). A broad definition of terrorism involves the use or threat of violence to instill fear and terror in the public to achieve political objectives. However, as with civil war, identifying and classifying terrorist events is far from straightforward. What constitutes a terrorist event, a terrorist individual, or a terrorist group is a highly contested area of study.

Weinberg et al. analyze existing definitions of terrorism in the literature, examining the frequency of specific elements within these definitions (Weinberg et al. 2004). The emphasis on particular definitional elements often varies depending on the author's field of study, region, or the period in which they published their work. For example, while some definitions highlight the threat of violence as a key characteristic of terrorism, others exclude this element entirely. This diversity in definitions underscores the lack of consensus in the field.

For instance, Ganor draws a clear distinction between terrorists and rebels, arguing that attacking civilians is the defining feature of terrorism. According to Ganor, any group that targets civilians qualifies as a terrorist organization, whereas guerrilla fighters exclusively target military personnel (Ganor 2002). However, empirical studies suggest that guerrilla groups frequently adopt terrorist tactics, blurring the line between these categories (Stanton 2013; Fortna et al. 2020; Polo and Gleditsch 2016)..

The answers to these definitional questions significantly influence whether a country is classified as experiencing a civil war and the specific years attributed to that conflict. These variations impact the quantitative analysis of civil wars, which often relies on numerical data and estimates derived from existing datasets. In the country-year format commonly used in datasets, such differences are almost inevitable, leaving researchers with limited control over the decisions made in data collection and aggregation. In contrast, event datasets provide researchers with greater autonomy to define and control variables, offering more flexibility compared to the rigid country-year format.

The lack of consensus among scholars on the definition of terrorism has significant implications for terrorism studies. For example, Kaczkowski et al. compare the Global Terrorism Database (**GTD**) with U.S. governmental reports and identify substantial discrepancies in the reported location, type, perpetrators, and targets of attacks (Kaczkowski et al. 2019). These differences stem from the varying definitions of terrorism adopted by the **GTD** and the U.S. government, illustrating how definitional ambiguities can lead to inconsistencies in empirical findings.

Beyond the definitional ambiguities of political violence concepts themselves, the distinctions between different types of violence are also unclear. Both terrorism and civil war share the objective of achieving political goals and often rely on similar means. Several studies highlight that terrorism is frequently employed as a tactic in civil wars by rebel organizations (Fortna et al. 2020; Stanton 2013; Polo and Gleditsch 2016). These findings suggest that the boundary between terrorism and armed warfare is far from clear-cut; instead, the two phenomena are deeply intertwined.

The theoretical and empirical challenges of defining political violence concepts have significant repercussions for our understanding of these phenomena. Ambiguities in definitions and overlaps between categories hinder the development of precise theoretical frameworks and complicate the interpretation of empirical findings, underscoring the need for more rigorous conceptualization in conflict studies.

Reporting and Underreporting Bias A second major bias that can influence the results of analyses based on conflict data arises from the nature of the data collection process. Conflict data are typically compiled from primary or secondary sources, rather than objective observations of the phenomena. Instead, these datasets rely on reports aggregated from external sources such as media outlets, governmental or non-governmental organizations, and other textual resources. Reporting bias, defined as the systematic overreporting or underreporting of certain events, is especially prevalent in media sources and has significant implications for conflict research. This bias can directly affect the dependent variable of the analysis, leading to errors in the researcher's conclusions.

Reporting bias can be broadly categorized into two types: selection bias and description bias. Selection bias refers to systematic differences in decisions to report specific events based on their characteristics, such as location or consequences. When the decision to report an event is non-random and correlates with the event features included in the analysis, it can skew the findings, leading to biased conclusions.

Description bias, on the other hand, arises from variations in the way the same event is reported. Factors such as audience costs or the political orientation of the reporting organization can lead to differing interpretations or emphases. For example, one news outlet may portray an event as highly consequential, while another might downplay its significance or omit critical details. These variations in content can propagate through event datasets, which often rely on textual sources for data collection, ultimately impacting the reliability and validity of the analysis.

The causes of reporting bias in conflict data can often be traced to the interplay of demand and supply in news production (Weidmann 2016), as well as the political context in which a media outlet operates. Certain events receive disproportionately high media coverage because they are perceived as more engaging to the public. Events that are surprising, dramatic, or culturally and geographically relevant are more likely to become newsworthy (Galtung and Ruge 1965). Additionally, urban centers tend to attract more media attention than rural areas (Davenport and Ball 2002). This discrepancy may stem from the relative ease of access to urban locations for news agencies or the perception that urban events hold greater interest for audiences.

Cubukcu and Forst compare official reports of terrorism events in Türkiye with the Global Terrorism Database (**GTD**) and uncover significant discrepancies in reporting. They argue that these differences are not random but rather influenced by "incident characteristics (victim/target, offender, and incident types, temporal and spatial factors) and rational factors (especially newsworthiness)" (Cubukcu and Forst 2018).

Another significant source of reporting bias is the political context of the media outlet. Democracies and non-democracies differ substantially in how their press systems operate. In non-democratic states, high levels of public ownership of media often curtail political freedom (Djankov et al. 2003). Direct government ownership provides a mechanism for controlling which news is broadcast or published. Furthermore, states can exercise indirect control through licensing requirements, taxation, or legal restrictions on reporting (Whitten-Woodring and James 2012). This political environment often creates norms among media owners and journalists, leading to self-censorship to safeguard revenues or job security (Whitten-Woodring and James 2012).

Drakos and Gofas argue that press freedom in democracies inflates the number of recorded terrorism incidents, making democracies appear more prone to terrorist attacks. In contrast, underreporting in autocracies distorts the relationship between regime type and terrorism activities (Drakos and Gofas 2006). Similarly, Baum and Zhukov examine how political regimes influence reporting during the Libyan Civil War (Baum and Zhukov 2015). They find that media in non-democratic states tend to be biased toward the status quo, while democratic media often favor revisionist narratives. This divergence reflects the domestic political interests of the regimes in question. For authoritarian regimes, the desire to delegitimize domestic opposition influences how international events are covered, often resulting in biased reporting that aligns with regime objectives.

Although the literature extensively addresses bias in media reporting, government and non-governmental organizations' (NGO) reports are not exempt from reporting bias. Government documents, particularly archival records, are among the most commonly used sources of information. However, archives are subject to three distinct forms of bias: survival bias, transfer bias, and source selection bias (Lee 2022).

Survival bias refers to the destruction—whether intentional or accidental—of archival documents. For example, the air raid on Potsdam in 1945 destroyed a significant portion of Germany's archives, leaving researchers with limited documentation on Germany's World War II period. Transfer bias arises because archival records are often physical documents that have not been digitized, making them susceptible to loss or damage. This bias is particularly prevalent during transfers between buildings, administrations, or custodians. Although these biases may seem random, they are often systematic; during crises, documents deemed important are more likely to be preserved.

The most pervasive archival bias is source selection bias, which reflects the tendency of governments to document information aligned with their own interests rather than providing a comprehensive summary of societal events. Government records are frequently bureaucratic and narrowly focused. For instance, various branches of the Government of Türkiye collect information on military personnel, civilians, and alleged terrorists killed due to terrorism. However, only one document summarizing the overall cost of terrorism was released to the public in 2013 during the Peace Process (TBMM 2013). Although intelligence agencies, police, military, and other government bodies maintain extensive records on terrorism, access to these documents is often limited and contingent on specific political contexts, such as peace negotiations.

The Peace Process (2013–2015) aimed to resolve the Kurdish question during the early 2010s, with the Justice and Development Party (AKP) and Abdullah Öcalan, the imprisoned leader of the PKK, serving as key

NGO reports are also prone to bias, similar to government reports. NGOs typically document issues that align with their organizational missions. For instance, an NGO focused on humanitarian aid for refugees will prioritize statistics and reports related to refugees, while another focused on children's welfare will highlight issues concerning children. Ideally, a diverse array of NGOs operating within a country would provide researchers with a statistically accurate and comprehensive picture. However, in war-torn countries, the number of active NGOs is often insufficient. Furthermore, NGOs may exaggerate the severity of situations to attract international attention and funding. As such, researchers must approach NGO reports with caution, recognizing the potential for selective or exaggerated reporting.

Reporting bias has been a central topic of debate in the political science literature for decades, and potential solutions to this issue have been actively discussed. Kreutz examines ambiguous cases in the Uppsala Conflict Data Program (UCDP) dataset to assess the scope of this problem. These unclear cases involve events with incomplete or disputed information, making them challenging to code with certainty. Kreutz finds that including these ambiguous cases reveals a higher number of active conflicts in the 1980s than previously estimated. This finding challenges the prevailing notion that the bipolar world system was more stable than the multipolar system (Kreutz 2015). This highlights how the data sources and compilation methods used in conflict studies can significantly influence the conclusions drawn.

Addressing Underreporting Bias One approach to mitigating underreporting bias is the mark and recapture estimation method (Hendrix and Salehyan 2015). Originally developed in ecology to estimate animal populations, this technique involves capturing, tagging, and releasing a sample of animals back into their habitat. A second sample is then recaptured, and the proportion of tagged to untagged animals is used to estimate the total population (Hendrix and Salehyan 2015). Hendrix and Salehyan adapt this method to the Social Conflict in Africa Database (SCAD) to estimate underreporting bias, which can arise either from the choice of data sources or during the aggregation process.

Addressing Description Bias The second type of reporting bias, description bias, occurs when the same event is reported differently or when certain details are obscured due to varying interpretations. Since event datasets rely on these sources, they inherit the biases present in the original reports, leading to inconsistencies or systematic distortions in the coded data. For instance, Davenport and Ball analyze reports on Guatemalan state terror and compare the reporting patterns of media outlets, government documents, and NGO reports. They find that these sources emphasize different aspects of the same events, leading to significant discrepancies in how the events are represented (Davenport and Ball 2002).

One proposed solution to address description bias in event datasets is to record data at the report level (Cook and Weidmann 2019; Weidmann and Rød 2015). This approach involves documenting individual reports rather than aggregating them into single events during the initial coding stage. While this method requires additional effort and does not fully eliminate aggregation bias—since researchers must eventually combine multiple reports into single events for statistical analysis—it provides greater transparency. By maintaining report-level data, researchers can make informed decisions about how to aggregate information and remain aware of the aggregation process, thereby reducing potential distortions.

Integration as a Solution

To address the challenges of reporting and description biases in event datasets, researchers have increasingly turned to integrating multiple datasets as a methodological solution. This approach can enhance conflict studies by mitigating the limitations inherent in individual datasets and providing more comprehensive insights into political violence (Donnay et al. 2019).

For instance, Polo and Gleditsch examine the use of terrorist tactics in civil wars by linking actors from the Uppsala Conflict Data Program (UCDP) to the Global Terrorism Database (GTD) (Polo and Gleditsch 2016). Similarly, Fortna et al. developed the Terrorism in Armed Conflict dataset, which investigates the use of terrorism by rebel groups by integrating the UCDP sample of rebel organizations with START's Global Terrorism Database (Fortna et al. 2020). However, Fortna et al. focus specifically on the names of civil war actors recorded in UCDP and analyze their actions as recorded in the GTD. While this effort provides valuable insights into one dimension of political violence—terrorism perpetrated by rebel groups—it does not capture the broader dynamics of political violence.

In another example, Stanton explores the relationship between the use of terrorism in civil wars and the regime type of the country (Stanton 2013). Such studies demonstrate the potential of integrated datasets to reveal patterns and relationships that remain obscured when relying on single-source data.

These developments in the quantitative study of conflict underscore the necessity for new integration methods. Effective integration allows researchers to compare different types of political violence, observe the evolution of terrorist or rebel groups, and analyze processes leading to the escalation of civil wars. No individual dataset is currently sufficient to address these complex research questions. The integration of multiple data sources, therefore, represents a critical step forward in overcoming the limitations of existing datasets and advancing the study of political violence.

In addition to enabling the comparison of different forms of events, integration also helps detect reporting bias (Donnay et al. 2019). Different datasets may omit variable values in distinct ways. By integrating datasets that cover the same or similar phenomena, researchers can compare the discrepancies and gain insights into the nature of reporting bias. Furthermore, access to information on civil wars is inherently challenging, often leading to missing critical data. For instance, the Uppsala Conflict Data Program (UCDP) requires at least one known fatality and the involvement

actors. The process involved three main steps: Gradual withdrawal of PKK elements from Türkiye's territory, Implementation of democratic reforms by the government, Integration of PKK members into political and civil life following disarmament (Köse 2017).

of a known organization to record an event. However, the complex nature of civil wars can make it difficult to obtain such information, resulting in some events going unrecorded. Conversely, the Global Terrorism Database (GTD) records events even when the perpetrator is unknown and in cases where the event results in zero fatalities. Integrating these datasets provides a more comprehensive and accurate representation of political violence.

This paper demonstrates the integration of different datasets covering diverse forms of political violence as a means to address biases stemming from definitional differences and reduce reporting bias. To achieve this, I apply the Matching Event Data by Location, Time, and Type (MELTT) method, a novel approach for automated integration and disambiguation of event data (Donnay et al. 2019). By combining data on terrorism and organized violence in Türkiye, I create a unique resource for studying political violence: the Political Violence in Türkiye Event Dataset (POLVITED).

Method

This study employs the Matching Event Data by Location, Time, and Type (MELTT) method, an automated approach for integrating and disambiguating event data, to create a comprehensive dataset of political violence in Türkiye. The dataset is designed to estimate the number of fatalities resulting from political violence and to examine the distinctions and overlaps between the Global Terrorism Database (GTD) and the Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP GED). By using Türkiye as a case study, this research also explores the implications of dataset differences for quantitative conflict studies. This section provides a detailed explanation of the datasets used (UCDP GED and GTD) and the MELTT method.

Türkiye, which has experienced attacks by more than 25 different terrorist organizations in the past quarter-century and has been engaged in a prolonged civil conflict with the PKK, provides an ideal case for investigating political violence. Countries experiencing political violence often lack the infrastructure to adequately report incidents, resulting in significant underreporting of violent events, such as due to limited communication capacity. However, Türkiye offers relatively accessible reporting on violent events, making it a feasible context for quantitative research.

Although researchers can access some official reports on political violence (TBMM 2013), these efforts are not systematic enough to provide reliable data for the international research community. Integrating datasets on political violence in Türkiye is therefore essential to regularly inform the international community about estimates of political violence while also assessing the accuracy of open-source datasets. Considering the availability of data and the pressing need for more comprehensive information, this study focuses on Türkiye as a case to demonstrate the potential of integrating event datasets to address reporting challenges and enhance our understanding of political violence.

Data

An essential step in integrating datasets is understanding in detail what these datasets record and how they do so. This section explains the coding rules and scope of the Global Terrorism Database (GTD) and the Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP-GED). Additionally, it highlights their methodological differences and areas of overlap, providing a foundation for their integration.

Global Terrorism Database As noted earlier, a universally accepted definition of terrorism is absent in the field of political science. To address this methodological challenge, the GTD employs a set of criteria that reflect key characteristics of terrorism while allowing researchers to apply their own criteria as filters (LaFree and Dugan 2007). For an incident to be included in the GTD, it must satisfy the following three conditions:

1. **The incident must be intentional** – it must result from a deliberate calculation by the perpetrator.
2. **The incident must entail some level of violence or immediate threat of violence** – this includes both violence against people and property damage.
3. **The perpetrators of the incident must be sub-national actors** – acts of state terrorism are excluded from the dataset.

These criteria aim to capture the defining elements of terrorism while maintaining flexibility for researchers to refine their analyses based on specific research objectives (LaFree and Dugan 2007).

Furthermore, for an incident to be recorded in the GTD, it must meet at least two of the following three additional criteria:

1. **Criterion 1:** The act must be aimed at attaining a political, economic, religious, or social goal. For economic goals, the exclusive pursuit of profit does not satisfy this criterion; the goal must involve the pursuit of more systemic economic change (LaFree and Dugan 2007).
2. **Criterion 2:** There must be evidence of an intention to coerce, intimidate, or convey a message to a broader audience beyond the immediate victims. The act is evaluated as a whole, irrespective of whether all individuals involved were aware of this intention. As long as any planners or decision-makers intended to coerce, intimidate, or publicize, this criterion is satisfied (LaFree and Dugan 2007).
3. **Criterion 3:** The action must fall outside the context of legitimate warfare activities as defined by international humanitarian law. Specifically, it must target non-combatants to be considered terrorism (LaFree and Dugan 2007).

These criteria enable the GTD to distinguish acts of terrorism from other forms of violence while allowing flexibility for researchers to apply additional filters based on their specific research needs.

The following table presents the number of events recorded in the GTD for Türkiye between 1989 and 2019,

categorized by the inclusion criteria they satisfy. The results indicate that 28% of the recorded events do not meet the third criterion, which requires actions to fall outside the scope of legitimate warfare. This suggests that approximately 28% of the GTD events in Türkiye could potentially overlap with incidents recorded in the UCDP-GED as part of armed conflict events.

Table 1. Number of events in GTD based on inclusion criteria (Türkiye, 1989–2019).

Inclusion Criteria	Number of Events
Criteria 1 and 2	799
Criteria 2 and 3	28
Criteria 1 and 3	19
Criteria 1, 2 and 3	2801

Uppsala Conflict Data Programme Georeferenced Event Dataset The second dataset utilized in this analysis is the Uppsala Conflict Data Programme Georeferenced Event Dataset (UCDP-GED), which records incidents of organized violence at the subnational level (Sundberg and Melander 2013). Unlike the country-year UCDP dataset, UCDP-GED enables subnational analyses, providing a more granular understanding of organized violence. The dataset adheres to UCDP's coding procedures and is compiled by human coders who systematically extract information from a wide range of sources, including news reports, NGO documents, case studies, truth commission reports, and historical archives (Eck and Hultman 2007; Sundberg et al. 2012; Sundberg and Melander 2013).

UCDP-GED records individual incidents of armed violence involving organized groups against other organized actors or civilians. Each incident must result in at least one known fatality to be included. This criterion aligns with UCDP's definition of active armed conflicts, which requires at least 25 battle-related deaths in a calendar year. By adopting this threshold, UCDP-GED ensures compatibility with UCDP's broader conflict definitions (Sundberg et al. 2012). Importantly, UCDP-GED serves as a complement to UCDP, focusing on dyads and actors identified by UCDP rather than incorporating entirely new data sources. Coders track these dyads during both active years (those meeting the threshold of 25 fatalities) and non-active years (Sundberg and Melander 2013).

Another essential criterion for inclusion in UCDP-GED is that incidents must involve organized violence. Methodologically, this means the events must be perpetrated by organized actors targeting other organized actors or civilians. Consequently, riots, protests, and terrorist attacks by non-organized or semi-organized actors are excluded from the dataset.

Table 2. Total number of events in UCDP-GED by type of violence (Türkiye, 1989-2019)

Type of Violence	Number of Events
State-based Violence	4861
Non-State Violence	14
One-sided Violence	442

UCDP-GED categorizes incidents of organized violence in Türkiye into three main types: state-based violence, non-state violence, and one-sided violence. The total number of recorded events for each category is summarized in the table 2. The majority of organized violence events in Türkiye between 1989 and 2019 are classified as state-based violence.

MELTT Method

Matching Event Data by Location, Time, and Type (MELTT) is an automated methodology for integrating event datasets, providing researchers with a transparent and reproducible framework to combine datasets based on their research interests and criteria (Donnay et al. 2019). The method employs at least three taxonomies for integration: geolocation, date, and a third taxonomy chosen by the researcher. By leveraging these taxonomies, MELTT identifies potential matched entries across the input datasets. Researchers retain control over how the algorithm applies the taxonomies, allowing for customization based on specific requirements. Furthermore, MELTT can integrate an unlimited number of datasets, making it a versatile tool for event data analysis.

The first taxonomy employed by MELTT is geolocation variables. Both the Global Terrorism Database (GTD) and Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP-GED) include longitude and latitude data for event locations. Longitude specifies how far east or west an event is from the prime meridian at Greenwich, while latitude indicates how far north or south it is from the Equator. These coordinates are widely used in spatial analysis to determine the precise location of events.

For this study, the geolocation interval for the MELTT analysis was set at 100 kilometers. This distance corresponds to the approximate range from a city center to its administrative border in southeastern Türkiye, where the majority of events occurred. Setting this interval ensures that events coded at slightly different but nearby locations are recognized as potential matches. At the same time, it minimizes the risk of incorrectly matching events occurring in different cities.

The second variable used by MELTT for integration is the date. In the GTD, the date variable is already provided, whereas the UCDP-GED records both the start and end dates of events. For MELTT integration, I utilized the start date from the UCDP-GED. Approximately 90% of UCDP events are recorded with precise dates, as indicated by the *date_prec* variable. Meanwhile, 97% of GTD events are reported to last less than 24 hours. Based on this information, I set a temporal window of 1 day for MELTT integration. This ensures that only events recorded on the same day are matched, minimizing the risk of misclassifying events that occurred on consecutive days as duplicates.

The most critical step in the MELTT analysis involves creating actor taxonomies, which serve as the foundation for dataset integration. Different datasets often use varying names, abbreviations, and formats for the actors involved in political violence. For example, UCDP refers to the *PKK*, while GTD uses *Kurdistan Workers' Party (PKK)*.

Identifying and reconciling these differences is essential for MELTT integration. Actors that appear in only one dataset retain their original coding within that dataset.

However, for actors appearing in both datasets, I conducted a thorough review of terrorist groups in Türkiye and their histories to create a unified taxonomy. This taxonomy facilitates accurate matching of actors across datasets and ensures consistency in the integrated dataset. The resulting taxonomy for actors coded in both datasets is outlined below:

Table 3. The names of actors

UCDP-GED	GTD
al-Qaida	Al-Qaida
IS	Islamic State of Iraq and the Levant (ISIL)
MKP	Maoist Communist Party (MKP)
PKK	Kurdistan Workers' Party (PKK)
SDF	Syrian Democratic Forces (SDF)
Yurtta Sulh Konseyi	Peace at Home Council
DHKP-C	Devrimici Halk Kurtulus Cephesi (DHKP/C)
Kurdish extremists	Kurdish extremists
TAK	Kurdistan Freedom Hawks (TAK)
Muslim extremists	Muslim extremists

The final step in the MELTT analysis involves the examination of duplicates to verify that the datasets have been correctly matched. This step ensures that each matched entry is accurate and that no mismatches have occurred. Verification can be conducted in two ways: by manually reviewing matched entries one by one or by checking variables such as locations, the number of total casualties, and attack types to confirm consistency across datasets.

Table 4. List of variables in UCDP-GED and GTD with definitions [*]

UCDP-GED	GTD	Definition
X	X	Row number in the original dataset
id	eventid	Event ID in the original dataset
type_of_violence	terrorism	Type of violence [†]
date	date	Start date (UCDP-GED) / Event date (GTD)
longitude	longitude	Geolocation (longitude)
latitude	latitude	Geolocation (latitude)
actor_tax	actor_tax	Name of the actor involved in the event [‡]
death_a+death_b +death_civilian	nkill	Number of fatalities
conflict_name	–	Name of the conflict
Dyad_name	–	Name of the conflict dyad
Side_a	–	Side A
Side_b	–	Side B [§]
Active_year	–	Active conflict indicator
–	attacktyp1	Type of attack
–	tartype1	General type of the target
–	tarsubtype1	Detailed type of the target

Upon completing the MELTT analysis, the output identifies matched events across the datasets but does not automatically merge them beyond the variables designated as taxonomies. In this study, MELTT produced a deduplicated index that included the dataset name, event code, date, longitude, latitude, and actor taxonomy. Using the event code and dataset name, I manually integrated the Global Terrorism Database (GTD) and Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP-GED) into a single unified dataset, referred to as the Political Violence in Turkey Event Dataset (POLVITED). The variables included in POLVITED are listed in Table 4.

For the integration process, UCDP-GED served as the base dataset. When an event was matched across datasets, its information was primarily sourced from UCDP-GED. For events unique to a single dataset, the relevant data were retained from their respective original sources. This approach ensured comprehensive coverage while maintaining the consistency and accuracy of the integrated dataset.

Results

This study aims to create a comprehensive dataset encompassing events of political violence in Türkiye to estimate the number of fatalities resulting from such violence. To achieve this objective, the MELTT method was applied to the Global Terrorism Database (GTD) and the Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP-GED), as described in the previous section.

This section is organized as follows: First, I will present the output of the MELTT analysis, including the identification of duplicated events in GTD and UCDP-GED, and discuss how these duplicates reveal differences in the definitions of terrorism and organized violence. Next, I will introduce the Political Violence in Turkey Event Dataset (POLVITED), a product of the integration process, and compare its scope and coverage to the original GTD and UCDP-GED datasets. After evaluating the merged dataset and the results of the MELTT analysis, I will provide an estimate of the extent of political violence in Türkiye. Finally, I will compare the findings from POLVITED with officially released data to assess the comprehensiveness and accuracy of the merged dataset.

MELTT Output

Using the specified temporal and spatial parameters, as well as the actor taxonomies developed in this study, the MELTT algorithm identified 895 events recorded in both datasets as matches. Additionally, the analysis revealed that 2,752 entries in the Global Terrorism Database (GTD) and 4,422 [†]entries in the Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP-GED) were unique to their respective datasets (see Figure 2).

[*]For more information, please consult the codebooks of GTD or UCDP-GED

[†]Indicates the actor's involvement in the incident; it does not necessarily mean the actor is the perpetrator (UCDP-GED).

[‡]In the merged dataset, GTD events are coded as "terrorism," while the dataset distinguishes four types of violence: state-based violence, one-sided violence, non-state violence, and terrorism.

[§]In UCDP-GED, side_a refers to the government in events involving a government and an organized armed group. The actor_tax is coded as side_b unless side_b is "civilians." In such cases, actor_tax refers to side_a.

[†]During the integration process, UCDP was used as the base dataset. Figure 2 displays 5,317 unique entries, comprising 4,422 unique events and 895

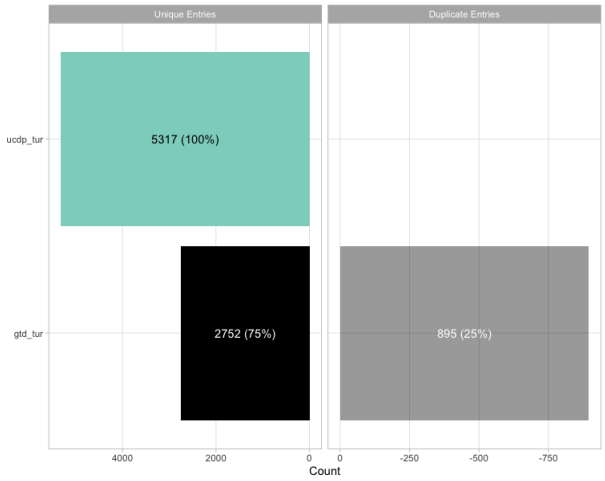

Figure 1. The Output of MELTT analysis

Figure 3 illustrates the temporal distribution of events. The upper panel displays the distribution of unique events over time, while the lower panel shows the distribution of duplicated events identified by the MELTT analysis. The first graph represents the distribution of events on a weekly basis, while the second graph aggregates the data annually. The highest overlap between UCDP-GED and GTD is observed in the years 1992, 1994, 2015, and 2016.‡

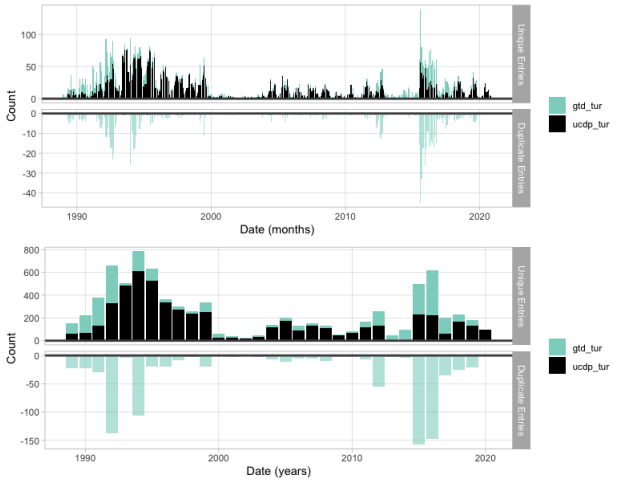

Figure 2. The Timeline of MELTT analysis

Matched Events

This section analyzes the duplicate events identified by the MELTT algorithm to understand why both the Global Terrorism Database (GTD), which records terrorism, and the Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP-GED), which records organized violence, included these events in their datasets.

A comparison of the types of violence, as shown in Figure 3, reveals that matched events primarily fall under the category of state-based conflict paired with terrorism. This finding aligns with the fact that 91% of the organized violence recorded in Türkiye consists of state-based conflict.

However, this result also underscores the misconception that terrorism operates independently of civil wars or armed conflicts involving a government. Instead, the analysis demonstrates that a significant proportion of terrorist activity occurs within the broader context of such conflicts.

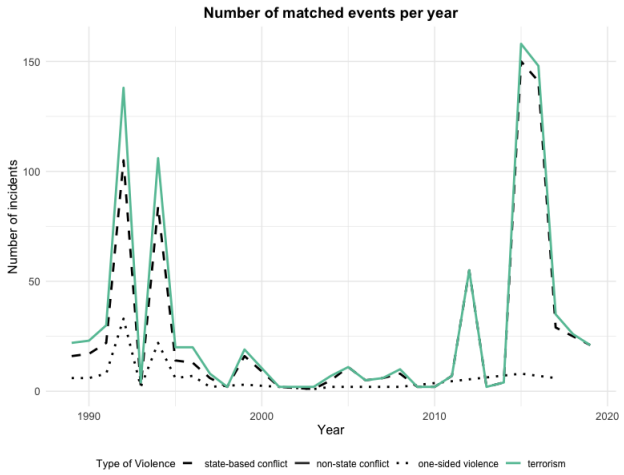

Figure 3. The comparison of type of violence in matched entries

§

POLVITED allows for a deeper understanding of how and when rebel groups incorporate terrorist tactics within their broader strategic repertoire. By identifying patterns in the use of terrorism, we can assess whether such tactics emerge as a tool for signaling strength or legitimacy, or as a means of coercion in the absence of territorial control. Table 5 highlights the specific conflicts in which terrorism has been employed. The prevalence of terrorism in the Turkey: Kurdistan conflict, for instance, underscores its role as a key strategy within the broader dynamics of organized violence. Similarly, the inclusion of incidents targeting civilians or government entities points to the dual utility of terrorism in both direct engagement with adversaries and psychological warfare.

The majority of incidents are associated with the conflict labeled ”Turkey: Kurdistan”, which accounts for 741 events and underscores the prolonged and intense nature of violence between the Turkish state and the Kurdistan Workers’ Party (PKK). The second largest category, ”PKK - Civilians” (108 incidents), highlights the frequent targeting of non-combatants, demonstrating how insurgent groups often employ terrorist tactics within broader conflict campaigns. Other notable categories, such as ”Turkey: Government” (18 incidents) and ”Turkey: Islamic State” (13 incidents), reveal the overlapping nature of state-based violence and terrorism, particularly in cases where international terrorist organizations, like ISIS, contribute to localized

duplicates. This discrepancy arises from a limitation of the visualization integrated into the mellt package

§Note that data for 1993 from GTD were lost during the transition. I used an additional data source Acosta and Ramos (2016) to account for this loss; however, the data have a suprisingly low coverage for Turkiye in 1993. Please keep this in mind when using the dataset

violence.These results emphasize the relation between terrorism and organized violence, with terrorism often operating within the broader context of state-based or armed conflicts rather than independently.

Table 5. Matched events categorized by conflict name (source: UCDP), with percentages representing the proportion of the total incidents

Conflict Name	Groups	Incidents
Turkey: Kurdistan	PKK	741
PKK - Civilians	PKK	108
Turkey: Government	DHKP-C	8
	TAK	7
Turkey: Islamic State	IS	13
IS - Civilians	IS	9
TAK - Civilians	TAK	3
SNA - SDF	SNA & SDF	1
al-Qaida - Civilians	al-Qaida	1
USA:Government	al-Qaida	1

This next table examines the distribution of attack types, weapon types, and target types in matched incidents. Attack types reflect the general methods used by perpetrators, as defined in the Global Terrorism Database (GTD) codebook. Armed assaults (416 incidents) and bombings/explosions (325 incidents) emerge as the most prevalent forms of attack. Other methods, such as assassinations (65 incidents) and hostage-taking through kidnappings (24 incidents) or barricade incidents (2 incidents), illustrate targeted and coercive strategies. Facility or infrastructure attacks (12 incidents) and incidents with unknown methods (50 incidents) further highlight the diversity of tactics employed.

Weapon types similarly reveal the strategic preferences of perpetrators. Firearms (434 incidents) and explosives (375 incidents) dominate, reflecting their accessibility and utility in achieving various operational objectives. Less common weapon types include incendiaries (16 incidents), which cause fires, and melee weapons (2 incidents), used in close-contact assaults.

Target type indicate the primary victims of violence. Military forces (392 incidents) and police (199 incidents) are the most frequently targeted, reflecting the intersection of terrorism with state-based conflicts. Private citizens and property (108 incidents) also feature prominently, demonstrating the psychological toll on the public.

Political Violence in Turkey Event Dataset (POLVITED)

This section details the merged dataset, POLVITED, and evaluates the human cost of political violence in Türkiye by analyzing the combined records of terrorism and organized violence. By integrating two datasets with differing values for the number of attacks and fatalities per year, POLVITED provides a more comprehensive understanding of political violence. Using Türkiye as a case study, this section demonstrates the importance of a merged dataset in accurately capturing the dynamics and impact of political violence.

Between 1989 and 2019, the Global Terrorism Database (GTD) recorded 3,647 terrorism events in Türkiye. During the same period, the Uppsala Conflict Data Program

Table 6. Distribution of attack types, weapon types, and target types of matched events in Türkiye between 1989 and 2019

Variable	Category	Incidents
Attack Type	Armed Assault	416
	Bombing/Explosion	325
	Assassination	65
	Unknown	50
	Hostage Taking (Kidnapping)	24
	Facility/Infrastructure Attack	12
	Hostage Taking(Barricade Incident)	2
	Hijacking	1
Weapon Type	Firearms	434
	Explosives	375
	Unknown	68
	Incendiary	16
	Melee	2
Target Type	Military	392
	Police	199
	Private Citizens or Property	108
	Business	44
	Government (General)	35
	Transportation	34
	Educational Institution	20
	Terrorists/Non-State Militia	18
	Utilities	17
	Unknown	9
	Journalists or Media	4
	Violent Political Party	4
	Tourists	4
	Food or Water Supply	3
	Airports or Aircraft	2
	Government (Diplomatic)	1
	Religious Figures/Institutions	1

Georeferenced Event Dataset (UCDP-GED) recorded 5,317 incidents of organized violence, extending its coverage to 2020. In contrast, POLVITED, the merged dataset, comprises 8,069 incidents of political violence, reflecting the integration of both datasets while accounting for matched and unique events.

The analysis reveals that 25% of terrorism incidents recorded in GTD are also identified as organized violence in UCDP-GED, meaning that one in every four terrorism events exhibits characteristics of organized violence as defined in the literature. Conversely, 15% of organized violence events are simultaneously recorded as terrorism. These findings underscore the interconnectedness of terrorism and organized violence and highlight the value of integrating datasets to capture the full scope of political violence.

Furthermore, the findings indicate that UCDP-GED captures 60% of the total number of political violence incidents, while GTD records 40% of these incidents. This suggests that relying on only one of these datasets to analyze violence associated with political aims leads to inaccurate conclusions. For instance, a researcher investigating the total number of fatalities in Türkiye resulting from political violence would underestimate the fatalities if solely using UCDP-GED. Conversely, combining UCDP-GED and GTD without addressing duplicate entries would overestimate the fatalities, as illustrated in Figure 7. The figure demonstrates that incorporating GTD increases the estimated number of fatalities compared to using UCDP-GED alone. However,

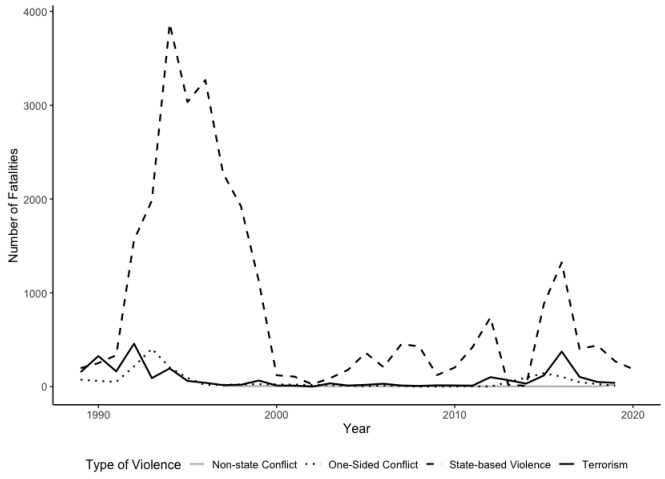

Figure 4. Number of fatalities per year in POLVITED

failing to account for duplicates results in an overestimation of the human cost of violence.

Figure 4 shows that the most lethal type of violence is state_based violence. On the other hand, the state_based violence and terrorism are the most common type of political violence.

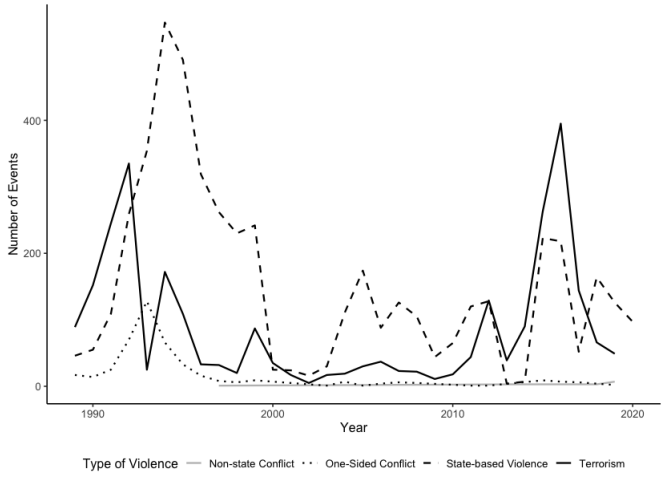

Figure 5. Number of political violence incidents per year in POLVITED

Terrorist groups operating in Türkiye often function through various branches, complicating their analysis. For instance, one of the most recent lethal attacks in central Istanbul was carried out by the Kurdistan Freedom Hawks (TAK), a breakaway faction of the Kurdistan Workers' Party (PKK). The attack, which occurred on Saturday, December 10, 2016, killed 44 people, including 36 police officers and 8 civilians (Damon et al. 2016). TAK claimed the attack as a response to recent clashes between the Turkish Armed Forces and the PKK.

This incident is recorded in both the Global Terrorism Database (GTD) and the Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP-GED), but with notable differences. GTD lists the event as two separate bombings targeting businesses and police forces in Istanbul, each resulting in 24 fatalities (48 in total). In contrast, UCDP-GED records it as a single event categorized as state-based violence between the Government of Türkiye and TAK, with 44 fatalities recorded (36 from the Government of Türkiye, 7 civilians, and 1 unknown).

This example highlights the importance of tracking the activities of different branches within a terrorist organization for accurate analysis. Not only do rebel groups engage in terrorist tactics directly, but their sub-branches also operate independently to support the main group's objectives. The discrepancies between GTD and UCDP-GED in recording this event underscore the need for a comprehensive dataset to evaluate and reconcile differences, ensuring a more accurate representation of political violence.

Comparison of POLVITED with Official Reports

Although various terrorist groups have carried out attacks in Türkiye, the Kurdistan Workers' Party (PKK) remains the most active and impactful, making it the primary security concern. Formed by Abdullah Öcalan in the late 1970s, the PKK initially sought to establish an independent Kurdish state in southeastern Türkiye. Since 1984, the PKK has engaged in periodic confrontations with the Turkish Armed Forces while also employing terrorist tactics, such as targeting civilians, to achieve its political objectives (Stanton 2013).

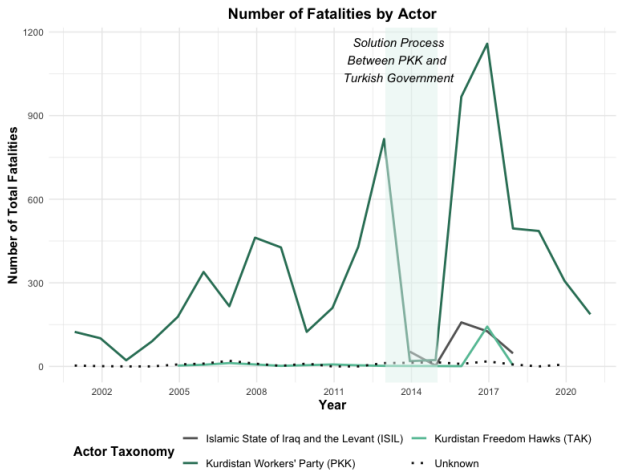

Figure 6. Number of fatalities of actors

The PKK's strategy illustrates how a rebel group can simultaneously function as a terrorist organization, employing both guerrilla warfare and terrorism as tools to pressure the government into making concessions. This duality highlights the relevance of analyzing the PKK to demonstrate how a single actor can be categorized differently across datasets, such as those focused on organized violence and terrorism. Furthermore, the complexity of the PKK's activities underscores the importance of integrating datasets to capture the full scope of its operations. Additionally, the availability of detailed statistics on PKK activities provides an opportunity to validate the integration results, ensuring the accuracy and comprehensiveness of the merged dataset.

I compare the records of violence involving clashes between Turkish security forces and the Kurdistan Workers' Party (PKK) with data published by the Committee on Human Rights Inquiry of the Grand National Assembly of

Türkiye in 2013. This report, covering the period from 1984 to 2012, provides a detailed account of violations of the right to life associated with acts of terror and violence (TBMM 2013).

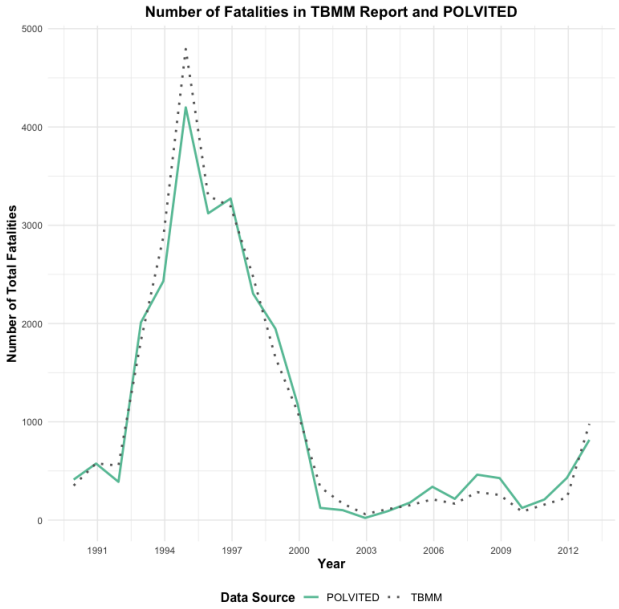

Figure 7. Number of Fatalities in TBMM Report ¶

A comparison of the official report by the Committee on Human Rights Inquiry of the Grand National Assembly of Türkiye (TBMM) with our merged dataset reveals that the official report records a higher number of fatalities (see Figures 9 and 10). This discrepancy arises because UCDP-GED and GTD primarily rely on media reports, which are susceptible to reporting bias, as discussed earlier. In contrast, TBMM has access to local and national intelligence documents, enabling it to capture killings or disappearances that may not be reported by or escape the attention of the media. As a result, such incidents are more likely to appear in the official report while being missed in the UCDP-GED and GTD datasets.

This finding aligns with the study by Cubukcu and Forst, which compared confidential reports from the Turkish Police Department with GTD and found significant underreporting in GTD, particularly for events deemed less newsworthy (Cubukcu and Forst 2018). Although discrepancies between official and internationally recorded datasets are inevitable, the overall patterns of political violence observed in the merged dataset are consistent with those in the TBMM report.

It is also important to note that the TBMM report only covers a specific period, and other official accounts of terrorist events remain inaccessible to the public. This limitation underscores the importance of internationally recorded datasets, such as UCDP-GED and GTD, for tracking political violence comprehensively over extended periods.

Conclusion

This paper aimed to create a novel dataset on political violence to estimate the number of incidents and fatalities in Türkiye. In the process of constructing the dataset, the study evaluated matched entries from the Global Terrorism Database (GTD) and the Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP-GED) to theoretically compare the definitions of terrorism and organized violence. The analysis revealed that the distinction between terrorism and organized violence becomes blurred in cases where rebel groups, actively engaged in conflict with a government, use terrorism as a strategic tool.

The findings suggest that UCDP-GED captures 60% of the total incidents of political violence, while GTD records 40%. These results indicate that relying on a single dataset leads to incomplete conclusions about political violence. Integrating these datasets significantly increases the coverage, providing a more comprehensive picture. However, 15% of political violence incidents are recorded in both datasets, necessitating a deduplication algorithm to address double recordings during integration.

As a result of this integration, the Political Violence in Turkey Event Dataset (POLVITED) was created, encompassing incidents of terrorism and organized violence in Türkiye from 1989 to 2020. POLVITED offers a unique advantage by including detailed information on over 50 different actors at the event level, enhancing the dataset's utility for in-depth analyses.

A critical challenge in collecting data on political violence is the nature of the actions being studied. These acts are often illegal and occur in contexts where media access is restricted, leading to under-reporting. For example, a major issue in Türkiye is the phenomenon of murders committed by unknown perpetrators (TBMM 2013). UCDP's coding rules exclude incidents involving unknown actors, which significantly limits the recorded data in conflict zones. By integrating GTD data, POLVITED addresses this gap, including information on incidents perpetrated by unknown actors. This makes the dataset a valuable resource for analyzing both specific actors and events involving unidentified perpetrators.

POLVITED opens new avenues for researchers to explore questions that existing datasets cannot address. For instance, it allows for an investigation into when and why rebel groups employ terrorism as a strategy or transition from terrorism to civil war. Additionally, researchers can examine whether terrorist groups abandon civil wars after achieving their objectives through terrorism. By including incidents with zero fatalities and removing arbitrary thresholds for start dates, POLVITED also enables the tracking of the evolution of terrorist groups and their tactics over time.

In summary, POLVITED provides a robust, comprehensive dataset for analyzing political violence in Türkiye, bridging the gaps left by existing data collection efforts. This integrated dataset not only enhances our understanding of political violence but also offers a valuable tool for answering complex questions about the dynamics of terrorism and organized violence.

¶The official report includes only the number of fatalities of PKK terrorists and civilians on yearly basis, excludes military personnel and police.

References

Acosta B and Ramos K (2016) Introducing the 1993 terrorism and political violence dataset. *Studies in Conflict & Terrorism* 40(3): 232–247. DOI:10.1080/1057610X.2016.1184061. URL https://doi.org/10.1080/1057610X.2016.1184061.

Baum MA and Zhukov YM (2015) Filtering revolution: Reporting bias in international newspaper coverage of the Libyan civil war. *Journal of Peace Research* 52(3): 384–400. DOI: 10.1177/0022343314554791.

Collier P and Hoeffler A (2004) Greed and grievance in civil war. *Oxford Economic Papers* 56(4): 563–595.

Cook SJ and Weidmann NB (2019) Lost in Aggregation: Improving Event Analysis with Report-Level Data. *American Journal of Political Science* 63(1): 250–264. DOI:10.1111/ajps.12398.

Cubukcu S and Forst B (2018) Measuring Terrorism. *Homicide Studies* 22(1): 94–116. DOI:10.1177/1088767917737808.

Damon A, Narayan C and Blau M (2016) Istanbul explosions: 44 killed, 155 wounded in twin blasts. *BBC* URL https://edition.cnn.com/2016/12/11/europe/istanbul-explosions/index.html.

Davenport C and Ball P (2002) Views to a kill: Exploring the implications of source selection in the case of Guatemalan state terror, 1977-1995. *Journal of Conflict Resolution* 46(3): 427–450. DOI:10.1177/0022002702046003005.

Djankov S, McLiesh C, Nenova T and Shleifer A (2003) Who owns the media? *Journal of Law and Economics* 46(2): 341–381.

Donnay K, Dunford ET, McGrath EC, Backer D and Cunningham DE (2019) Integrating Conflict Event Data. *Journal of Conflict Resolution* 63(5): 1337–1364.

Doyle M and Sambanis N (2000) International peacebuilding: A theoretical and quantitative analysis. *American Political Science Review* 94(4): 779–801.

Drakos K and Gofas A (2006) The devil you know but are afraid to face: Underreporting bias and its distorting effects on the study of terrorism. *Journal of Conflict Resolution* 50(5): 714–735. DOI:10.1177/0022002706291051.

Eck K and Hultman L (2007) One-sided violence against civilians in war: Insights from new fatality data. *Journal of Peace Research* 44(2): 233–246. DOI:10.1177/0022343307075124.

Fearon JD and Laitin DD (2003) Ethnicity, Insurgency, and Civil War. *American Political Science Review* 97(1): 75–90.

Fortna VP, Lotito NJ and Rubin MA (2020) Terrorism in armed conflict: new data attributing terrorism to rebel organizations. *Conflict Management and Peace Science* : 1–23DOI:10.1177/0738894220972996.

Galtung J and Ruge MH (1965) The structure of foreign news: The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers. *Journal of Peace Research* 2(1): 64–90. DOI:10.1177/002234336500200104.

Ganor B (2002) Defining Terrorism: Is One Man's Terrorist another Man's Freedom Fighter? *Police Practice and Research* 3(4): 287–304. DOI:10.1080/1561426022000032060.

Gleditsch NP, Wallensteen P, Eriksson M, Sollenberg M and Strand H (2001) Armed conflict 1946-1999: A new dataset. *Unpublished Paper* .

Gleditsch NP, Wallensteen P, Eriksson M, Sollenberg M and Strand H (2002) Armed conflict 1946-2001: A new dataset. *Journal of Peace Research* 39(5): 615–637.

Hendrix CS and Salehyan I (2015) No News Is Good News: Mark and Recapture for Event Data When Reporting Probabilities Are Less Than One. *International Interactions* 41(2): 392–406. DOI:10.1080/03050629.2015.982117.

Kaczkowski W, Lokmanoglu A and Winkler C (2019) Definitions matter: a comparison of the global terrorism database and the U.S. governmental reports of terrorist incidents in Western Europe, 2002-2016. *Cambridge Review of International Affairs* 0(0): 1–18. DOI:10.1080/09557571.2019.1705246.

Köse T (2017) Çözüm Sürecinin Yükseliş ve Düşüşü. *Türkiye Ortadoğu Çalışmaları Dergisi* 4(1): 13–13. DOI:10.26513/tocd.315152.

Kreutz J (2015) The war that wasn't there: Managing unclear cases in conflict data. *Journal of Peace Research* 52(1): 120–124. DOI:10.1177/0022343314541845.

LaFree G and Dugan L (2007) Introducing the global terrorism database. *Terrorism and Political Violence* 19(2): 181–204. DOI:10.1080/09546550701246817.

Lee A (2022) The library of babel: How (and how not) to use archival sources in political science. *Journal of Historical Political Economy* 2(3): 499–526. DOI:10.1561/115.00000038. URL http://dx.doi.org/10.1561/115.00000038.

Leitenberg M (2001) Deaths in wars and conflicts between 1945 and 2000. *Paper prepared for the Conference on Data Collection in Armed Conflict, June 8-9, Uppsala, Sweden* .

Licklider R (1995) The consequences of negotiated settlements in civil wars 1945-1993. *American Political Science Review* 89(3): 681–90.

Mars P (1975) The Nature of Political Science. *Social and Economic Studies* 24(2): 221–238.

Marsden SV and Schmid AP (2011) Typologies of terrorism and political violence. In: Schmid (2011b), pp. 158–201.

Polo SM and Gleditsch KS (2016) Twisting arms and sending messages: Terrorist tactics in civil war. *Journal of Peace Research* 53(6): 815–829. DOI:10.1177/0022343316667999.

Regan P (1996) Conditions for successful third party interventions. *Journal of Conflict Resolution* 40(1): 336–59.

Sambanis N (2004) What is civil war? Conceptual and empirical complexities of an operational definition. *Journal of Conflict Resolution* 48(6): 814–858. DOI:10.1177/0022002704269355.

Sarkees MR and Wayman F (2010) *Resort to War: 1816 – 2007*. Washington DC: CQ Press.

Schmid AP (2011a) The definition of terrorism. In: Schmid (2011b), pp. 39–99.

Schmid AP (2011b) *The Routledge Handbook of Terrorism Research*. Taylor and Francis.

Shanahan T (2016) The definition of terrorism. In: *Routledge Handbook of Critical Terrorism Studies*. Taylor and Francis. ISBN 9781317801610, pp. 223–246.

Stanton JA (2013) Terrorism in the Context of Civil War. *The Journal of Politics* 75(4): 1009–1022. DOI:10.1017/S0022381613000984.

Sundberg R, Lindgren M and Padskocimaite A (2012) UCDP Georeferenced Event Dataset (GED) Codebook (March): 1–34. URL www.ucdp.uu.se.

Sundberg R and Melander E (2013) Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research* 50(4): 523–532. DOI:10.1177/0022343313484347.

TBMM (2013) Terör ve şiddet olayları kapsamında yaşam hakkı ihlallerini inceleme raporu : 1–327.

Weidmann NB (2016) A Closer Look at Reporting Bias in Conflict Event Data. *American Journal of Political Science* 60(1): 206–218. DOI:10.1111/ajps.12196.

Weidmann NB and Rød EG (2015) Making uncertainty explicit: Separating reports and events in the coding of violence and contention. *Journal of Peace Research* 52(1): 125–128. DOI: 10.1177/0022343314523807.

Weinberg L, Pedahzur A and Hirsch-Hoefler S (2004) The challenges of conceptualizing terrorism. *Terrorism and Political Violence* 16(4): 777–794. DOI:10.1080/095465590899768.

Whitten-Woodring J and James P (2012) Fourth estate or mouthpiece? a formal model of media, protest, and government repression. *Political Communication* 29(2): 113–136. DOI: 10.1080/10584609.2012.671232.