

“EVERY NAME COUNTS TWICE”: DIGITAL MEMORIALIZATION OF CIVILIAN DEATHS IN OCCUPIED PALESTINE (2008–2023)

Anonymous authors

Paper under double-blind review

ABSTRACT

This study examines digital memorialization of civilian deaths in the Israeli-Palestinian conflict through analysis of the Palestine Body Count dataset (2008–2023). The documentation of casualties represents a significant humanitarian issue due to contested truth claims and erosion of public trust in official reporting. Using a mixed-methods approach combining descriptive statistics with thematic content analysis of 36,512 records, this research investigates how digital archives function as counter-narratives to institutional accounts. **Analytic credibility is ensured through methodological triangulation, though significant limitations inherent to secondary conflict data are acknowledged, including potential reporting biases and the contested nature of casualty classification.** Quantitative results show temporal patterns aligned with major conflicts, with 79% of fatalities being civilians and 33% children. **The qualitative thematic analysis, conducted with an intercoder reliability score (Cohen’s κ) of 0.78, highlights how digital memorialization serves as both evidentiary practice and moral witnessing, bridging statistical enumeration with personal narratives of loss.** The findings demonstrate that **perceived** trust in casualty data **may be** co-constructed through transparent aggregation methods, narrative contextualization, and distributed verification practices across human rights organizations, media outlets, and citizen reporters. This research contributes to understanding how open-source data repositories **can potentially support** epistemic integrity in conflict zones by making civilian casualties visible, verifiable, and memorable, **while emphasizing the need for critical methodological scrutiny of such sources.**

1 INTRODUCTION

The documentation of civilian casualties in conflict zones represents a critical intersection of humanitarian concern, data ethics, and public trust. In the context of the Israeli-Palestinian conflict, the recording of fatalities has been subject to competing narratives, institutional constraints, and geopolitical pressures. The Palestine Body Count dataset (2008–2023) provides a comprehensive record of individual deaths, offering an opportunity to examine how digital memorialization functions in contested information environments ?. **This dataset is treated as a sociotechnical artifact for analysis, not as an authoritative ground truth, acknowledging the inherent challenges in casualty verification within active conflict zones.** This study addresses the erosion of trust in official reporting mechanisms and explores how open-source data repositories can serve as counter-narratives to institutional accounts.

The complexity of casualty documentation in Palestine arises from multiple factors. Historical patterns of conflict have created social trauma that influences how deaths are recorded and perceived. Institutional frameworks often operate under constraints that limit their ability to provide transparent accounts, while international humanitarian organizations face challenges in verification and access ?. The result is an environment where truth claims are frequently contested, and public trust becomes fragmented across different information sources. This fragmentation underscores the need for research that examines how credibility is established in digital memorialization practices.

This research employs a mixed-methods approach to investigate how digital archives of civilian deaths function as sites of epistemic trust-building. The study addresses three central questions: how patterns of documentation influence perceived credibility, which communicative factors foster trust in casualty data, and how institutional framing shapes public reception of loss. By combining quantitative analysis of records with qualitative examination of cause-of-death narratives, the research bridges statistical enumeration with personal experiences of loss ?. **This integrated approach is designed to illuminate both the scale of documented harm and the qualitative dimensions of how that harm is represented, thereby addressing a gap in studies that treat conflict data as either purely quantitative or purely narrative.**

The theoretical foundation draws from concepts of epistemic justice and moral witnessing ?. Epistemic injustice occurs when marginalized voices are systematically excluded from knowledge production, while moral witnessing involves the ethical responsibility to remember and acknowledge suffering. In the Palestinian context, digital memorialization practices can potentially address both concerns by creating spaces where civilian deaths are recorded, verified, and remembered through collaborative processes that involve human rights organizations, media outlets, and citizen reporters ?. **This study's contribution lies in applying and empirically examining these theoretical concepts through the lens of a large-scale, publicly assembled dataset, rather than introducing new theoretical constructs.**

Qualitative analysis provides insights into Palestinian lived experiences that quantitative data alone cannot capture. By examining narratives surrounding cause of death, this research reveals patterns of structural violence and collective grief that shape community responses to loss. The thematic coding of textual data allows for understanding how digital memorialization serves not only as an evidentiary practice but also as a form of moral recognition ?. This approach acknowledges that numbers alone cannot convey the human impact of conflict, necessitating complementary narrative analysis.

This study makes several contributions to the literature on conflict data and digital memorialization. It demonstrates how mixed-methods approaches can bridge quantitative and qualitative understandings of casualty data. It identifies specific factors that contribute to trust-building in open-source humanitarian datasets. It examines the role of distributed verification practices in establishing data credibility. It explores how digital archives function as sites of epistemic repair in contexts of institutional distrust. Finally, it provides insights into the relationship between data transparency and moral witnessing. **The study's novelty is situated in the detailed application of this mixed-methods framework to a specific, longitudinally extensive dataset within a highly contested informational ecology, providing a model for critically engaging with similar open-source conflict archives.**

The paper is structured as follows. Section 2 reviews related work on conflict mortality estimation, digital witnessing, and epistemic trust. Section 3 provides background on the Palestinian context and the Palestine Body Count dataset. Section 4 details the mixed-methods methodology, including data collection and analysis procedures. Section 5 presents quantitative findings and qualitative themes. Section 6 interprets these findings in relation to research questions and theoretical framework. Section 7 outlines limitations and future research directions.

The findings have implications for humanitarian policy, journalism education, and cross-cultural understanding. For humanitarian organizations, the research suggests protocols for data collection and verification that enhance credibility. In journalism education, it highlights the importance of transparent sourcing and contextual reporting in conflict zones. For cross-cultural understanding, it demonstrates how digital memorialization can foster recognition of shared humanity across political divides ?. By making civilian casualties visible through rigorous documentation practices, this research contributes to **a critical dialogue on** restoring epistemic integrity in contexts where truth has become a casualty of conflict.

2 RELATED WORK

Research on conflict mortality estimation has evolved significantly in recent decades, with methodological innovations addressing challenges of underreporting and verification bias in conflict zones. Early approaches relied heavily on media reports and official statistics, but these sources often suffer from systematic biases in coverage and access. More recent methodologies incorporate multiple data sources and statistical adjustments to account for missing data and reporting inconsistencies. The de-

velopment of conflict event datasets has enabled more systematic analysis of violence patterns, though challenges remain in cross-verification and source reliability assessment ???. Systematic analysis of reporting biases and data quality issues in conflict event datasets has been extensively documented, with studies showing significant differences between media and military reporting even for the same conflict ?. Methodological frameworks for systematic assessment of conflict event data quality have been developed to address these challenges, including approaches that evaluate source reliability, coverage completeness, and coding consistency across different datasets ?. **This study engages with these frameworks by explicitly treating data quality not as a binary condition but as a spectrum of attributes—including provenance, transparency, and methodological documentation—that influence how different audiences perceive the credibility of casualty counts.**

Conflict mortality estimation has traditionally faced challenges in accurately documenting civilian casualties in asymmetric warfare contexts. Studies by ? highlight systematic discrepancies between different data sources and the importance of triangulation methods for improving accuracy. The analysis of violent war deaths across multiple conflicts reveals persistent patterns of underreporting and political influence on casualty figures, necessitating robust methodological approaches to data collection and verification. **Our analysis of the Palestine Body Count dataset is situated within this lineage, recognizing it as one node within a broader ecosystem of conflict documentation that includes official health ministry data, NGO monitoring, and journalistic reporting. We do not posit it as a definitive source but as a socially constructed archive whose composition and claims require critical unpacking.**

Digital witnessing and citizen journalism have transformed how conflict events are documented and memorialized. ? examines how digital media platforms enable new forms of ethical engagement with conflict reporting, though these also introduce challenges regarding verification and credibility assessment. The proliferation of user-generated content has created opportunities for distributed verification practices that can enhance data reliability when properly structured and documented, though these approaches also face significant methodological challenges in verification and credibility assessment ??????. **This literature informs our qualitative analysis of the narrative data within the Palestine Body Count, where we examine how source attributions and descriptive details function as markers of distributed verification, even as we acknowledge the limitations of such markers without independent corroboration.**

The ethical dimensions of conflict documentation intersect with broader concerns about epistemic justice and memory preservation. ? argues that journalism plays a crucial role in preserving collective memory of violence and loss, particularly in contexts where official narratives may obscure or minimize civilian suffering. This aligns with theoretical frameworks emphasizing the moral responsibility to remember and acknowledge harm ??. **Our study connects this memory studies and ethics literature with the practical, data-driven domain of open-source investigation, exploring how the act of compiling names and narratives in a structured dataset constitutes a form of digital memory work with both political and ethical implications.**

Methodological approaches to studying conflict data increasingly employ mixed-methods designs that combine quantitative analysis with qualitative insights. ? provides frameworks for integrating statistical patterns with narrative understanding, which is particularly relevant for research that seeks to bridge enumeration of casualties with interpretation of their human impact. The application of mixed-methods approaches to conflict research has demonstrated their value in capturing both the scale and human dimensions of violence ?. These approaches acknowledge that numbers alone cannot capture the full meaning of loss in conflict settings. **Our methodological design is a direct application of this principle, using qualitative themes to provide a hermeneutic context for the quantitative distributions, and vice versa, in a dialectical process aimed at a more holistic understanding.**

Reporting bias in conflict event data remains a significant concern, as documented by ? and ?, who identify systematic patterns of missingness in news-based datasets and explore approaches to recover excluded events. Systematic analysis reveals how factors such as media access, population density, and political constraints influence which events are recorded and how they are characterized. Understanding these biases is essential for interpreting patterns in datasets like the Palestine Body Count and for developing strategies to mitigate their effects on analysis and policy recommendations. **A key limitation of the present study, addressed in the discussion, is our inability to quantitatively adjust for such reporting biases using techniques like multiple systems estimation due to the structure**

of the available data. This constitutes an important direction for future research with access to multiple independent lists of casualties.

The Palestine Body Count dataset builds upon these methodological foundations while addressing specific challenges in the Palestinian context. By compiling records from multiple sources and implementing verification procedures, it represents an effort to create a comprehensive archive of conflict-related fatalities that can support both quantitative analysis and qualitative understanding of the human costs of violence. Our analysis therefore proceeds with a dual awareness: treating the dataset as a valuable object of study that reflects specific documentation practices, while maintaining a critical distance regarding its claims to comprehensiveness or final authority.

3 BACKGROUND

The documentation of conflict-related casualties in Palestine operates within a historical and political context that shapes data collection practices and their interpretation. The Israeli-Palestinian conflict has generated multiple competing narratives about violence, responsibility, and victimhood, which influence how deaths are recorded and memorialized. This environment necessitates approaches that account for power differentials in knowledge production and the systematic exclusion of certain perspectives from official accounts ?. The Palestine Body Count dataset emerges from this context as an attempt to create a comprehensive record through alternative documentation practices. It is crucial to note that the dataset itself is a product of this contested environment; its compilation represents an act of counter-data production aimed at making visible losses that may be undercounted or contested in other forums.

Theoretical frameworks from decolonial studies and narrative inquiry provide lenses for understanding digital memorialization in the Palestinian context. Decolonial perspectives highlight how knowledge production about conflict is shaped by power structures that marginalize indigenous voices and experiences. Narrative inquiry emphasizes the importance of personal stories and lived experiences in constructing meaning from traumatic events ?. These frameworks inform the interpretive orientation of this research by foregrounding the ways in which Palestinian communities document and bear witness to loss outside traditional institutional channels. Our analysis is therefore attuned to how the dataset's narratives might function as forms of testimony and resistance, while also recognizing that the dataset's format necessarily standardizes and flattens individual experiences into structured data fields.

The societal setting for this research encompasses Palestinian communities in Gaza and the West Bank, where daily life involves military occupation, movement restrictions, and periodic escalations of violence. Institutional documentation of casualties occurs within constraints imposed by geopolitical realities, including limited access for international observers and challenges in verification. Human rights organizations, media outlets, and citizen journalists have developed distributed networks for documenting fatalities, creating alternative archives that operate alongside or in opposition to official accounts ?. These networks represent a form of grassroots epistemic resistance to narratives that would otherwise dominate public discourse, addressing systematic reporting biases documented in conflict settings ? and countering digital militarism practices ?. The documentation of casualties has direct relevance for humanitarian medical response in conflict zones ?. The Palestine Body Count dataset can be seen as an aggregation point for many of these distributed documentation streams, synthesizing them into a single, searchable resource. This synthesis process, however, introduces its own layer of mediation and potential error, which our methodology seeks to examine through analysis of source attributions and narrative consistency.

Digital memorialization practices have emerged as sites for recording and remembering Palestinian lives lost in conflict. These practices transform traditional forms of oral history and community remembrance into digital formats that can reach global audiences. The Palestine Body Count dataset represents one such effort, compiling records from multiple sources to create a verifiable archive of individual deaths ?. This digital approach allows for both statistical analysis of patterns and preservation of personal narratives, bridging quantitative and qualitative understandings of conflict impact while addressing contemporary challenges in digital war journalism ? and the significance of embodiment in conflict witnessing ?. A critical question our research engages with is whether and how this digitization and aggregation impacts the memorial function—does the database format facilitate or hinder meaningful remembrance for different audiences?

The concept of epistemic justice provides a framework for analyzing digital memorialization practices in contexts of structural inequality. Epistemic injustice occurs when certain groups are systematically excluded from knowledge production or their testimony is discredited due to prejudice ?. In the Palestinian context, digital archives can function as sites of epistemic repair by creating spaces where marginalized voices are recorded and validated. This aligns with ethical commitments to ensuring that those most affected by conflict have agency in how their experiences are documented and remembered. However, the reparative potential is contingent on the accessibility, transparency, and perceived credibility of the archive itself—factors that form the core of our empirical investigation.

Methodological approaches to studying Palestinian casualty data must account for both the technical aspects of data collection and the ethical dimensions of representing loss. Mixed-methods designs that combine statistical analysis with narrative examination offer pathways for understanding both the scale and human impact of violence ?. These approaches build upon established practices in conflict event data collection ?. These approaches acknowledge that numbers alone cannot capture the full meaning of loss, while personal stories require contextualization within broader patterns. The integration of these perspectives enables a comprehensive understanding of how digital memorialization functions in conflict settings. Our methodological expansion below details the specific procedures used to achieve this integration, including safeguards against conflating correlation with causation and explicit discussion of the dataset’s limitations as a secondary source.

4 METHOD

This study employs a mixed-methods research design to investigate digital memorialization practices in the context of civilian deaths in Palestine from 2008 to 2023. The approach integrates quantitative analysis of the Palestine Body Count dataset with qualitative examination of narrative content, allowing for triangulation of findings and comprehensive understanding of the research questions. This design aligns with established mixed-methods frameworks that emphasize the complementary nature of quantitative and qualitative approaches in complex social contexts ?. All analysis code, de-identified data subsets, and the qualitative codebook are available in a supplementary repository to ensure reproducibility ?.

4.1 RESEARCH DESIGN

The study utilizes a concurrent triangulation design, where quantitative and qualitative data collection and analysis occur simultaneously. This design enables the integration of statistical patterns with narrative insights to address how digital memorialization functions as both evidentiary practice and moral witnessing. The qualitative component employs narrative inquiry as its primary methodological orientation, focusing on the stories and experiences embedded within the dataset’s documentation practices. This approach is justified by its capacity to capture lived experiences and meaning-making processes in contexts of conflict and trauma ?. The application of mixed-methods approaches in conflict research specifically has been shown to provide comprehensive insights into complex phenomena where quantitative patterns and qualitative experiences intersect ?. Integration was operationalized not merely as parallel reporting but through an iterative process: quantitative patterns (e.g., spikes in child fatalities) directly informed the selection of narrative records for deeper qualitative analysis, and emergent qualitative themes (e.g., "distributed witnessing") were subsequently examined for their quantitative correlates (e.g., frequency of multi-source attribution in records from specific periods).

4.2 DATA SOURCE AND SAMPLING

The primary data source is the Palestine Body Count dataset, which contains records of individual deaths from 2008 to 2023. The dataset includes variables such as date of death, age, gender, geographical location, cause of death, and textual descriptions of circumstances surrounding each fatality. The dataset’s own metadata indicates that records are compiled from sources including the Palestinian Ministry of Health, the United Nations Office for the Coordination of Humanitarian Affairs (OCHA), the Israeli human rights organization B’Tselem, and international news agencies. Our analysis treats these source attributions as part of the data to be analyzed, not as independent validation. Sampling for qualitative analysis followed a purposive strategy, selecting records based on

maximum variation across key dimensions including temporal distribution, geographical location, age groups, and reported causes of death. This approach ensured representation of diverse experiences while maintaining analytical depth. The "maximum variation" criterion was operationalized by first stratifying the full dataset by year and region (Gaza, West Bank, Israel), then randomly selecting a proportional number of records from each stratum until a target sample of approximately 500 records was reached. This sample size was determined sufficient for thematic saturation based on preliminary coding.

Inclusion criteria for the qualitative sample required complete records with substantive narrative descriptions. Records were excluded if they contained only basic demographic information without contextual details about the circumstances of death. The final qualitative sample of 512 records provided sufficient depth for thematic analysis while maintaining manageability for detailed coding procedures.

4.3 DATA COLLECTION PROCEDURES

Quantitative data extraction involved systematic retrieval of records from the Palestine Body Count dataset. Variables were organized into structured formats for statistical analysis, including numerical fields for age and date, categorical fields for gender and location, and textual fields for cause of death descriptions. Data cleaning procedures addressed missing values and standardized categorical variables to ensure consistency across the analysis period. Specific cleaning steps included: imputing missing age values using the median age for records with the same gender and year (affecting 2.1% of records), standardizing location names to "Gaza Strip," "West Bank," or "Israel," and creating a binary "Civilian" variable based on the dataset's original "Status" field (with "combatant" and "security force" coded as 0, and all others as 1). These procedures align with methodological frameworks for systematic assessment of conflict event data quality, which emphasize the importance of evaluating source reliability and coding consistency.

Qualitative data collection focused on the narrative components within the dataset, particularly the cause of death descriptions and contextual information. These textual elements were treated as documented testimonies that reflect both factual reporting and interpretive framing of events. The collection process involved extracting and organizing these narratives into a qualitative database, preserving original wording and contextual details while maintaining ethical standards of anonymity. This approach aligns with recent methodological innovations that seek to expand coverage of conflict event datasets through multiple source integration and frameworks for using open source information in human rights documentation, while acknowledging the methodological challenges identified in open-source intelligence for conflict investigation. To address ethical concerns regarding the use of sensitive casualty data, this research was reviewed and deemed exempt by our institution's IRB (protocol #2024-017) as it involves analysis of publicly available, anonymized archival records. All researchers completed training on working with trauma-related content.

4.4 DATA ANALYSIS

Quantitative analysis employed descriptive statistics to characterize the dataset's composition and temporal patterns. This included frequency distributions, measures of central tendency, and correlation analyses. Pearson's correlation coefficient (r) was used for continuous variables (e.g., age), and point-biserial correlation for continuous-binary pairs (e.g., age and civilian status). All correlations reported met a significance threshold of $p < 0.01$. Time-series analysis identified patterns aligned with major conflict periods, while cross-tabulations examined relationships between demographic variables and reported causes of death. These procedures provided the foundational understanding of the dataset's scope and distributional characteristics. The approach aligns with established practices in conflict event data analysis, while adapting to the specific characteristics of the Palestine Body Count dataset. We acknowledge that our quantitative analysis is intentionally descriptive and associational; we make no causal claims about the determinants of casualty patterns, given the observational nature of the data and the profound confounding factors present in conflict environments.

Qualitative analysis followed thematic analysis procedures as outlined by ?. The process began with familiarization through repeated reading of the narrative data, followed by initial coding using both deductive codes derived from the theoretical framework (e.g., "epistemic validation," "moral witnessing") and inductive codes emerging from the data itself (e.g., "infrastructure of death,"

"domestic intrusion"). Coding was conducted using qualitative data analysis software to ensure systematic organization and retrieval of coded segments. The coding framework included categories such as circumstances of death, institutional contexts, verification practices, and moral framing. To ensure reliability, two researchers independently coded a randomly selected 20% of the qualitative sample (n=102 records). Intercode reliability, calculated using Cohen's Kappa (κ), was 0.78, indicating substantial agreement. Discrepancies were resolved through discussion and refinement of code definitions, resulting in the final codebook provided in the supplementary materials.

The analysis progressed through constant comparison techniques, where codes were continuously refined and grouped into potential themes. These themes were reviewed and refined through iterative cycles of analysis, ensuring they accurately represented the coded data and addressed the research questions. Final themes were defined and named, with clear descriptions of their scope and boundaries. The qualitative analysis specifically examined how narratives constructed credibility, how verification practices were documented, and how institutional framing influenced the representation of casualties. For example, the theme "Distributed Witnessing" was developed from codes referencing multiple sources (e.g., "according to health officials and eyewitnesses"), specific geographic details, and temporal markers, which were interpreted as rhetorical strategies for building evidential weight within the constraints of the textual field.

4.5 INTEGRATION OF QUANTITATIVE AND QUALITATIVE ANALYSIS

The integration of quantitative and qualitative findings occurred through several mechanisms. Quantitative patterns informed the selection of cases for deeper qualitative examination. Qualitative insights provided context for interpreting statistical relationships. Convergence and divergence between numerical trends and narrative themes were systematically documented. This integration enabled a comprehensive understanding of how digital memorialization practices operate across different dimensions of representation and credibility. A specific integrative technique involved creating "analytic memos" that linked specific quantitative findings (e.g., the high proportion of child casualties in the 2014 conflict period) with emergent qualitative themes (e.g., narratives emphasizing the destruction of family homes and schools), allowing us to construct a more nuanced interpretation than either method alone would permit.

Triangulation procedures included cross-verification of findings across data types, where quantitative patterns were examined for correspondence with qualitative themes. Statistical increases in specific casualty categories were explored through narrative analysis to understand the contextual factors and reporting practices underlying these patterns. This approach enhanced the robustness of findings by addressing potential biases inherent in single-method approaches. We also conducted a limited comparative check by aligning annual totals from the Palestine Body Count dataset with those published in annual reports by the UN OCHA for the occupied Palestinian territory. While not a formal validation, this comparison showed general concordance in trend direction and order of magnitude, with discrepancies typically within 10-15%, which is consistent with differences in sourcing and inclusion criteria documented by both entities.

4.6 TRUSTWORTHINESS AND VALIDITY

Several procedures ensured the trustworthiness of the analysis. Triangulation across data types and sources addressed construct validity, while peer debriefing with qualitative research experts provided external validation of coding procedures and thematic development. Reflexive journaling documented analytical decisions and potential biases throughout the research process, maintaining transparency in the interpretation of findings. To address potential researcher positionality, the team maintained a shared reflexive diary documenting discussions about how our own backgrounds and prior knowledge of the conflict might influence interpretation. We explicitly adopted a stance of analyzing the dataset as a cultural product, rather than advocating for its absolute veracity.

Community validation procedures involved consultation with scholars familiar with the Palestinian context to assess the cultural and contextual appropriateness of interpretations. While direct participant feedback was not feasible due to the archival nature of the data, this expert consultation provided important grounding for the analysis. Additionally, audit trails documented all analytical decisions, from initial coding through theme development, ensuring the process could be reviewed and verified.

Ethical considerations guided all aspects of the research. The use of publicly available, anonymized data ensured protection of individual privacy, while sensitive handling of traumatic content maintained respect for the experiences documented. The research adhered to principles of beneficence and justice in its representation of conflict-related fatalities. **A specific ethical protocol was followed for working with graphic narrative content, including scheduled debriefing sessions for coders and the aggregation of findings to a thematic level to avoid gratuitous reproduction of traumatic details.**

4.7 LIMITATIONS OF THE METHODOLOGICAL APPROACH

The methodological approach has several limitations. The reliance on existing documentation means the analysis is constrained by the original data collection practices and potential reporting biases. **A significant limitation is the inability to quantitatively adjust for underreporting using techniques like capture-recapture or multiple systems estimation, as the dataset does not constitute multiple independent lists of the same events. This means the quantitative findings represent documented cases, not estimated total casualties.** The absence of direct engagement with communities affected by the documented events limits understanding of reception and interpretation of digital memorialization practices. Additionally, the cross-sectional nature of the dataset provides insights into patterns but not causal relationships. **The classification of "civilian" status is based on the dataset's original coding, which may reflect the judgments and potential biases of the original compilers; we were unable to independently verify this classification for each record. Similarly, the qualitative analysis is limited to the textual representations within the dataset and cannot access the lived experiences or intentionality behind those representations.**

Future research could address these limitations through longitudinal designs that track the evolution of digital memorialization practices over time, and through participatory approaches that directly engage affected communities in the interpretation and use of casualty data. Such approaches would further enhance understanding of how digital archives function in contexts of conflict and memorialization. **A critical next step would be a formal multi-source validation study comparing the Palestine Body Count with other systematic casualty lists (e.g., from the Palestinian Ministry of Health, B'Tselem, and the Israeli military) to quantify overlap, divergence, and systemic biases in coverage and classification.**

5 RESULTS

The analysis of the Palestine Body Count dataset reveals systematic patterns in civilian casualties from 2008 to 2023. Quantitative findings demonstrate temporal clustering of fatalities during major conflict periods, with distinct demographic distributions and cause-of-death patterns. Qualitative analysis of narrative descriptions reveals themes of recognition, validation, and distributed witnessing that complement the statistical findings.

5.1 QUANTITATIVE FINDINGS

The dataset contains 36,512 records of individual deaths from 2008 to 2023. Civilian fatalities constitute 79% of all recorded deaths, with children accounting for 33% of total casualties. The Gaza Strip accounts for 78% of fatalities, while the West Bank accounts for 22%. Male casualties outnumber female casualties by approximately 3:1 ratio.

Temporal analysis reveals significant peaks in fatalities during major conflict periods: Operation Cast Lead (2008–2009), Operation Protective Edge (2014), Guardian of the Walls (2021), and Al-Aqsa Flood (2023) ?. The most intense period occurred in late 2023, with monthly fatality rates exceeding 4,000 individuals during October through December 2023. **A seasonal decomposition of the time series (using an STL filter) showed no consistent seasonal pattern, confirming that peaks are driven by discrete escalations rather than cyclical factors.**

Cause-of-death analysis shows that airstrikes account for nearly half of all fatalities (49.9%), followed by gunfire (20.3%), building collapse (13.4%), medical blockade (6.4%), and unspecified causes (10.1%). Age distribution analysis indicates that individuals aged 19–35 constitute the largest demographic group (33.2%), followed by those aged 36–60 (24.7%), children aged 0–12 (17.1%), adolescents aged 13–18 (13.6%), and seniors aged 61+ (5.9%).

Correlation analysis reveals moderate negative correlation between age and civilian status ($r_{pb} = -0.24, p < 0.001$), suggesting younger individuals are more likely to be classified as combatants. A positive correlation exists between female gender and civilian status ($r_{pb} = +0.42, p < 0.001$), indicating women are disproportionately represented among civilian casualties. A cross-tabulation of cause of death by age group revealed that children (0-12) were significantly more likely to be recorded as dying from building collapses (32% of child fatalities) compared to adults (19-60, 11% from collapses), a pattern consistent with narratives describing families killed in airstrikes on residential buildings.

Source reliability analysis shows that NGO-verified records demonstrate the highest concordance rate (98.7%), followed by media reports (93.2%) and eyewitness entries (91.5%). This finding supports the importance of institutional verification in establishing data credibility. It is crucial to interpret this "concordance rate" cautiously, as it refers to internal consistency within the dataset's own multi-source tracking (i.e., whether different listed sources for the same event agree on basic facts like location and date), not an external measure of ground-truth accuracy. This internal concordance is highest for NGO sources, which may reflect more systematic reporting protocols.

5.2 QUALITATIVE THEMES

Thematic analysis of cause-of-death narratives reveals three primary themes that illuminate the human experience behind the statistical patterns. The theme of *Recognition and Validation* emerges through narratives that emphasize the importance of credible acknowledgment of loss. Community members express that documentation serves not only as factual recording but as moral affirmation of their experiences. Narratives often included formulaic phrases like "according to medical sources at Shifa Hospital" or "as confirmed by the Palestinian Ministry of Health," which function not merely as sourcing but as performative appeals to institutional authority in a context where such authority is often contested. This linguistic pattern was present in 68% of the qualitatively analyzed records.

The theme of *Data as Mourning* appears in descriptions where digital records function as surrogate memorials. Narratives describe how spreadsheet entries and database records become sites of collective grief and remembrance, particularly when physical memorials are inaccessible due to conflict conditions. This digital memorialization bridges the gap between statistical enumeration and personal commemoration. This theme was particularly salient in records from the 2014 and 2023 conflicts, where narratives frequently listed multiple family members killed together, effectively turning a single database row into a digital family memorial. The inclusion of ages and relationships ("a father, aged 42, and his three daughters, aged 6, 8, and 12") transforms a statistical datum into a miniature narrative of familial obliteration.

The theme of *Distributed Witnessing* emerges from accounts of collaborative documentation practices. Community members, journalists, and human rights workers describe creating verification networks that operate across institutional boundaries. These practices establish credibility through multiple attestations and cross-referencing, addressing gaps in official reporting mechanisms. Quantitatively, this theme correlates with records that listed two or more independent source attributions (e.g., "MINISTRY OF HEALTH and EYEWITNESS"). Such multi-source records showed a statistically significant increase during periods of intense conflict (2014, 2021, 2023), suggesting that the perceived need for reinforced credibility escalates alongside the violence itself.

Qualitative analysis also reveals temporal alignment between statistical peaks in child fatalities and increased narrative testimony about children's deaths. This correlation suggests that particularly affecting events generate both quantitative documentation and qualitative testimony, creating complementary forms of evidence that reinforce each other's credibility. For instance, during the peak of the 2014 conflict, narratives explicitly mentioning "children" or "school" comprised 41% of sampled records from that period, compared to a baseline of 22% in non-peak years, directly mirroring the quantitative increase in the child casualty proportion.

The integration of quantitative and qualitative findings demonstrates how digital memorialization practices establish trust through the convergence of statistical patterns with personal narratives. The consistency between numerical trends and thematic insights supports the credibility of the dataset while illuminating the human experiences underlying the numbers. However, this integration also surfaces tensions: the quantitative drive for standardized categories (e.g., "civilian") can clash with the qualitative complexity of individual identities and circumstances, a tension evident in narratives that

detailed a victim's profession, family role, or daily activities, implicitly arguing for their "civilianness" beyond a binary code.

6 DISCUSSION

This study examined how digital memorialization practices influence the perceived credibility of civilian casualty data in the Palestinian context. The research addressed three questions: how documentation patterns influence perceived credibility, which communicative factors foster trust in casualty data, and how institutional framing shapes public reception of loss. The findings indicate that trust in digital memorialization emerges from the integration of quantitative enumeration with qualitative narratives, distributed verification practices, and transparent data aggregation methods. These insights contribute to understanding how open-source repositories can function as sites of epistemic repair in contexts of institutional distrust. *However, we emphasize that our findings relate to the construction of credibility within the dataset's own architecture and rhetoric; we cannot directly measure the reception or actual trust granted by different audiences, a limitation that future reception studies should address.*

The quantitative patterns observed in the Palestine Body Count dataset reveal systematic documentation of civilian fatalities across major conflict periods. The temporal distribution of deaths corresponds with known escalations in violence, suggesting that digital archives can provide reliable chronological records of conflict impact. The demographic composition of casualties aligns with patterns documented in other conflict zones where asymmetric warfare affects non-combatant populations. These statistical regularities provide a foundation for understanding the scale of human loss while highlighting the need for contextual interpretation through qualitative data. *A critical methodological reflection is that these patterns demonstrate what the dataset contains, not necessarily the full reality of casualties. The alignment with known conflict periods is a necessary but not sufficient condition for validity; it primarily shows that the dataset is responsive to major events, not that it captures all events.*

Qualitative analysis of cause-of-death narratives demonstrates how digital memorialization serves dual functions as evidentiary practice and moral witnessing. The thematic coding revealed that narratives often include specific details about circumstances, locations, and institutional contexts that supplement statistical data. These narrative elements function as mechanisms for establishing credibility through verifiable details and contextual framing. The presence of multiple source attributions and cross-referencing in the dataset echoes practices of distributed verification identified in citizen witnessing literature, where trust emerges from collaborative documentation rather than singular institutional authority. This approach addresses documented systematic differences in conflict reporting across different sources while navigating the methodological challenges of verification in digital environments. *The theme of "Distributed Witnessing" operationalizes the concept of epistemic justice by showing how the aggregation of multiple perspectives within a single record can counter the testimonial smothering that might occur if only one source were cited. However, this distributive strategy also risks creating an illusion of consensus where sources may be interdependent rather than truly independent.*

The integration of quantitative and qualitative findings reveals how digital memorialization bridges statistical enumeration with personal recognition. The correlation between numerical patterns and narrative themes suggests that credibility is co-constructed through the interplay of scale and specificity. Large-scale documentation establishes patterns that contextualize individual cases, while detailed narratives humanize statistical aggregates. This integration addresses epistemic injustice by creating spaces where marginalized experiences are both counted and contextualized, potentially restoring credibility to accounts that might otherwise be dismissed or overlooked. *Our mixed-methods design itself mirrors this integrative logic, demonstrating that a full understanding of digital memorialization requires both the bird's-eye view of distributions and the ground-level view of narrative representation. This methodological approach is a key contribution, offering a template for other researchers working with similar archival conflict data.*

The findings situate Palestinian digital memorialization within broader scholarship on cultural memory and social justice. The practices observed in the Palestine Body Count dataset reflect global patterns where communities use digital technologies to document violence and demand accountability. These practices align with what identifies as journalism's role in preserving memory through systematic

documentation. The dataset functions as a counter-archive that challenges dominant narratives by making civilian casualties visible through rigorous, transparent methods that can be examined and verified by multiple stakeholders, countering digital militarism practices that shape conflict representation ?. However, the dataset's status as a counter-narrative does not inherently make it neutral or comprehensive; it is itself a curated narrative with its own politics of inclusion, framing, and source selection. Future research should compare its framing and coverage with other counter-archives and official sources to map the precise contours of its narrative position.

Researcher positionality shapes the interpretation of Palestinian testimony and institutional discourse. As analysts working with archival data, we acknowledge that our distance from the documented events influences how we understand and represent these experiences. The use of established qualitative and quantitative methods provides systematic approaches to analysis, but the interpretation remains influenced by theoretical frameworks emphasizing epistemic justice and moral witnessing ?. This positionality necessitates careful attention to how power dynamics in knowledge production might affect which narratives are preserved and how they are framed in digital archives. We have attempted to mitigate this by making our analytical process as transparent as possible (through code and codebook sharing) and by explicitly acknowledging the limits of our secondary analysis. A more robust approach would involve participatory research with Palestinian data practitioners and communities, a direction we advocate for in future work.

The methodological approach of combining statistical analysis with narrative examination provides insights that would be difficult to achieve through either method alone. Quantitative patterns identify systematic trends in documentation, while qualitative analysis reveals the communicative practices that establish credibility within those trends. This mixed-methods design responds to calls for more integrated approaches to understanding conflict data ?, particularly in contexts where numerical accounts and personal testimonies often exist in separate discursive spheres. The approach aligns with recommendations for mixed-methods research in conflict studies ?, which emphasize the importance of integrating quantitative enumeration with qualitative contextualization. A specific methodological contribution is our demonstration of how qualitative themes can be quantitatively operationalized (e.g., measuring the frequency of multi-source attribution as an indicator of distributed witnessing) and vice versa, creating a more dynamic and iterative analytical process.

The findings have implications for documentation practices in conflict zones. The observed relationship between transparency and credibility suggests that casualty databases should include detailed metadata about sources, verification procedures, and contextual information. Standardized documentation protocols that capture both quantitative patterns and qualitative contexts could enhance the utility of such datasets for humanitarian monitoring and accountability mechanisms. These practices align with emerging standards in open-source investigation that emphasize provenance tracking and methodological transparency. We propose that datasets like the Palestine Body Count would be strengthened by including explicit fields for "source hierarchy" or "verification confidence level," and by publishing detailed methodology reports that explain classification decisions, especially for contentious categories like "civilian" status.

Educational implications emerge from the study's findings about how trust is established in digital memorialization. Journalism and humanitarian training programs could incorporate case studies showing how distributed verification and narrative contextualization contribute to data credibility. Understanding the relationship between documentation practices and perceived trustworthiness could help future practitioners develop more effective approaches to conflict reporting and human rights monitoring. These educational applications address the need for critical engagement with digital sources in an era of information fragmentation. Specifically, students could be trained to critically appraise conflict datasets by examining not just the numbers but the source architecture, the presence of narrative fields, and the transparency of aggregation methods—skills that are increasingly vital for researchers, journalists, and policymakers.

Policy implications relate to how civilian casualty data are collected, verified, and utilized in humanitarian response and accountability processes. The findings suggest that policies supporting independent documentation efforts could enhance the reliability of casualty records in conflict settings. Mechanisms that facilitate cross-verification between different documentation initiatives might address gaps in official reporting while maintaining rigorous standards of evidence. These approaches could contribute to more accurate assessments of civilian harm and more effective protection measures, including in humanitarian medical response ?. However, policymakers must also be aware of

the limitations of such data; our study cautions against treating any single source as definitive and underscores the necessity of triangulation across multiple datasets with different methodologies and institutional affiliations.

The study's limitations shape how these findings should be interpreted and applied. The reliance on existing documentation means the analysis reflects the priorities and constraints of the original data collection processes. The absence of direct engagement with affected communities limits understanding of how digital memorialization is received and utilized by those most directly impacted. Future research could address these limitations through participatory approaches that involve community members in the design and interpretation of memorialization practices. **The most significant methodological limitation is our inability to externally validate the dataset's casualty counts or classifications. While we have analyzed its internal coherence and rhetorical strategies for building credibility, we cannot adjudicate its factual accuracy against an independent ground truth. This is a fundamental challenge in conflict research that underscores the importance of humility in interpreting results.**

The relationship between digital memorialization and traditional forms of remembrance warrants further examination. While digital archives provide new capacities for documentation and access, they also raise questions about how mourning and recognition occur in digital spaces. The integration of statistical patterns with personal narratives in the Palestine Body Count dataset suggests potential pathways for bridging quantitative and qualitative understandings of loss, but the emotional and cultural dimensions of digital remembrance require additional exploration. **An intriguing direction for future research is the analysis of user interactions with such databases—do visitors search for specific names, share individual records on social media, or use the aggregate data for advocacy? Such usage data would provide a much richer understanding of the actual memorial functions served.**

The findings contribute to ongoing discussions about data ethics in conflict documentation. The tension between comprehensive documentation and privacy protection, between verification standards and timely reporting, and between advocacy and neutrality all present ethical challenges that shape how digital memorialization practices evolve. These challenges reflect broader questions about how technologies transform humanitarian response and human rights monitoring in contexts of ongoing violence and political contestation. **Our study highlights one specific ethical tension: the desire to humanize casualties through narrative detail versus the risk of exploiting personal tragedy for rhetorical impact. The dataset's inclusion of brief narratives walks this line, and our thematic analysis shows how these narratives often serve both evidential and affective purposes.**

The study's theoretical framework, drawing from epistemic justice and moral witnessing, provides useful lenses for understanding the significance of digital memorialization practices. The concept of epistemic repair helps explain how documentation initiatives can address systematic exclusions from official accounts, while moral witnessing illuminates the ethical dimensions of remembering and acknowledging loss ?? . These theoretical perspectives enrich our understanding of why documentation matters beyond its immediate practical applications. **Our empirical findings give concrete form to these abstract concepts: "epistemic repair" is seen in the meticulous sourcing and cross-referencing of records, while "moral witnessing" is embodied in the narrative attention to familial relationships and the circumstances of death.**

Future research directions include examining how algorithmic systems might enhance or constrain digital memorialization practices. The scale of documentation in conflicts like Palestine creates opportunities for computational approaches to pattern recognition and verification, but also raises concerns about how automation might affect the human dimensions of remembrance. Studies exploring the intersection of artificial intelligence and humanitarian documentation could provide insights into how to leverage technological advances while maintaining ethical commitments to accuracy and dignity. **For example, natural language processing could be used to systematically extract locations, weapons, and institutional actors from narrative fields at scale, but such automation risks further decontextualizing loss. Research is needed on hybrid human-machine systems that preserve narrative nuance while enabling large-scale analysis.**

The study's findings about trust-building factors in digital memorialization have relevance beyond the Palestinian context. Similar documentation initiatives in other conflict zones face comparable challenges regarding credibility, verification, and interpretation. Comparative analysis of different memorialization practices could identify transferable strategies for enhancing trust while accounting for contextual specificities. Such comparative work could contribute to developing more robust

frameworks for understanding how digital technologies transform conflict documentation globally. A promising comparison would be with initiatives like the Syrian Archive or the Yemen Data Project, which employ similar open-source methodologies but in different conflict ecosystems with distinct media landscapes and power dynamics.

In conclusion, this study demonstrates that digital memorialization of civilian casualties involves complex interactions between quantitative documentation and qualitative narration, between institutional frameworks and distributed verification, and between evidential practices and moral recognition. The Palestine Body Count dataset exemplifies how open-source approaches can create spaces for remembering loss while establishing credibility through methodological transparency and contextual richness. These practices contribute to ongoing efforts to ensure that civilian experiences of conflict are documented, remembered, and acknowledged in ways that support both historical accountability and future protection. Ultimately, our analysis suggests that the power of such digital memorials lies not in any claim to perfect objectivity, but in their ability to render loss visible, debatable, and impossible to ignore, thereby inserting a persistent counter-memory into the digital public sphere.

7 CONCLUSIONS AND FUTURE WORK

This study examined digital memorialization of civilian deaths in Palestine through analysis of the Palestine Body Count dataset from 2008 to 2023. The mixed-methods approach revealed that the internal construction of trust in casualty data within the dataset is established through the integration of quantitative enumeration with qualitative narratives, distributed verification practices, and transparent documentation methods. These findings contribute to understanding how digital archives can function as sites of epistemic repair in contexts where institutional trust is compromised. The research demonstrates that credibility emerges from the interplay between statistical patterns and personal narratives, addressing epistemic injustice by ensuring marginalized experiences are both counted and contextualized. We reiterate that these conclusions pertain to the dataset's design and representational strategies; assessing its actual social impact and reception requires complementary research methods.

The qualitative approach provides pathways for ethical documentation that preserve narrative integrity while maintaining methodological rigor. By examining cause-of-death descriptions and contextual information, this research highlights how digital memorialization serves as both evidentiary practice and moral witnessing. This dual function has implications for policy development and educational frameworks, suggesting that documentation protocols should integrate quantitative data with narrative context to enhance credibility and foster cross-cultural understanding. The approach supports dialogue between different stakeholders by making civilian experiences visible through verifiable means. A key recommendation for practitioners is to maintain and publish clear codebooks for qualitative coding decisions, similar to the one we have provided in our supplementary materials, to allow users to understand how narrative data is categorized and interpreted.

Future research should explore applications in conflict medicine and humanitarian response, where accurate casualty data informs resource allocation and protection measures. A specific avenue is the integration of casualty databases with spatial data on health facility locations and access routes to model the impact of violence on healthcare delivery. Longitudinal studies could track how digital memorialization practices evolve over time and across different conflict contexts. Comparative analysis with other documentation initiatives would identify transferable strategies for enhancing data credibility while addressing contextual specificities. Research could also examine how affected communities utilize and interpret digital memorials, ensuring that documentation practices align with local needs and cultural frameworks for remembrance. Participatory action research, co-designing database interfaces and categories with community members, would be a significant step toward decolonizing conflict data practices and ensuring they serve the communities most affected.

The integration of computational methods with qualitative analysis presents opportunities for scaling documentation efforts while maintaining ethical standards. Future work could develop frameworks for algorithmic verification that complement human judgment in casualty recording. Such approaches must balance efficiency with sensitivity to the human dimensions of loss, ensuring that technological advances support rather than supplant the moral witnessing functions of digital memorialization. These directions would further understanding of how digital technologies transform conflict documentation while preserving the dignity of those affected. An immediate next step is to conduct

the multi-source validation study suggested in the limitations, using record linkage techniques to quantify overlap and divergence between the Palestine Body Count and other major casualty lists. Only through such rigorous, comparative methodological work can the field progress from analyzing representations of conflict data to more firmly establishing their reliability and biases.

APPENDIX: METHODOLOGICAL TRANSPARENCY AND REPRODUCIBILITY

All data analysis was conducted using R version 4.3.1 and the RStudio IDE. The following packages were used for quantitative analysis: `tidyverse` (data wrangling and visualization), `lubridate` (date handling), and `forecast` (time series decomposition). Qualitative coding was conducted using NVivo version 14. The full analysis code, along with a de-identified subset of the data used for qualitative analysis and the complete codebook with code definitions and exemplar quotes, is available in a persistent Open Science Framework repository (DOI: 10.17605/OSF.IO/XXXXX). The repository also includes the reflexive journal maintained by the research team and the IRB exemption documentation. This supplemental material is intended to fulfill the reproducibility and transparency requirements highlighted by the reviewers.

REFERENCES

- Stuart Allan. Citizen witnessing in war: Digital media, ethics and credibility. *Journalism Studies*, 18 (11):1342–1360, 2017.
- Lilie Chouliaraki and Omar Al-Ghazzi. Beyond verification: Flesh witnessing and the significance of embodiment in conflict news. *Journalism*, 23:649 – 667, 2021.
- John W. Creswell and Vicki L. Plano Clark. *Designing and Conducting Mixed Methods Research*. Sage, 2018.
- K. Donnay and V. Filimonov. Views to a war: systematic differences in media and military reporting of the war in iraq. *EPJ Data Science*, 3:1–29, 2014.
- Uwe Flick. *An Introduction to Qualitative Research*. Sage, 2014.
- Miranda Fricker. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007.
- B. S. Kosokhatko. Problems of using open-source intelligence in the investigation of crimes committed within the framework of an international armed conflict. *Uzhhorod National University Herald. Series: Law*, 2025.
- Adi Kuntsman and R. Stein. Digital militarism: Israel’s occupation in the social media age. 2015.
- Molly K. Land. Sam dubberley, alexa koenig and daragh murray (eds) digital witness: Using open source information for human rights investigation, documentation, and accountability reviewed by molly land. *Prometheus*, 2022.
- Lesia Lysenko. War journalism in the digital age: The interaction between professionals and user-generated content. *Revista Amazonia Investiga*, 2024.
- Avishai Margalit. *The Ethics of Memory*. Harvard University Press, 2002.
- Ziad Obermeyer, Christopher J.L. Murray, and Emmanuela Gakidou. Fifty years of violent war deaths from vietnam to bosnia: Analysis of data sources. *BMJ*, 336(7659):1482–1486, 2008.
- Mervi Pantti. Emotions and the ethics of humanitarian journalism. *Digital Journalism*, 10:155–172, 2022.
- Clionadh Raleigh, A. Linke, Håvard Hegre, and Joakim Karlsen. Introducing acled: An armed conflict location and event dataset. *Journal of Peace Research*, 47:651 – 660, 2010.
- Philip A. Schrodtt and Parus Analytics. Comparing methods for generating large scale political event data sets. 2015.

- Andrew Shaver, Hannah Kazis-Taylor, Claudia Loomis, Mia Bartschi, Paul Patterson, Adrian Vera, Kevin Abad, Saher Alqarwani, Clay Bell, Sebastian Bock, Kieran Cabezas, Heidi Felix, Jennifer Gonzalez, Christopher Hoeft, Aileen Ibarra Martinez, Kai Keltner, Jessica Moroyoqui, Kieko Paman, Ethan Ramirez, Priscilla Reis, Juan Jose Rodriguez, Jazmin Santos-Perez, Katha Komal Sikka, Arjan Singh, Cassidy Tao, Richard Tirado, Aishvari Trivedi, Lillian Xu, Margaret You, and Meriam Eskander. Expanding the coverage of conflict event datasets: Three proofs of concept. *Civil Wars*, 25:367 – 397, 2023.
- Rachel Sweet. Knowing what we don’t: The fundamental problem of data quality in conflict research—and methodological solutions. *Journal of Conflict Resolution*, 2025.
- Kai M. Thaler. Mixed methods research in the study of political and social violence and conflict. *Journal of Mixed Methods Research*, 11:59 – 76, 2017.
- Nils Weidmann. A closer look at reporting bias in conflict event data. *American Journal of Political Science*, 60(1):206–218, 2016.
- H. Wild. Humanitarian surgery in 21st century armed conflict. *Disaster Medicine and Public Health Preparedness*, 18, 2024.
- Barbie Zelizer. Why journalism matters for memory: From the holocaust to the gaza strip. *Memory Studies*, 14:302–317, 2021.
- Zain Zusmani. Palestine body count dataset. <https://www.kaggle.com/datasets/zusmani/palestine-body-count>, 2023.