Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Data Fragmentation and Loss of Analytic Visibility in Interpersonal Violence Surveillance: Implications for Policy and Prevention

Cindy A Ogolla Jean-Baptiste

caogolla@gmail.com

Descendants of Africa Pioneering Innovation

Research Article

Keywords: Public Health Surveillance Modernization, Interpersonal Violence Visibility, Data Linkage, Precision Public Health, Violence Prevention

Posted Date: April 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9452047/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

This study empirically demonstrates how fragmented surveillance architectures produce systematic analytic attrition that constrains proactive interpersonal violence (IPV) prevention. Using Public-use National Crime Victimization Survey data as an empirical laboratory to quantify visibility loss across surveillance stages, we examined how requirements for relational, situational, and contextual variables reduced incidents eligible for prevention-relevant modeling. From an initial frame of 329,678 observations, identification of 11,000 sexual victimizations was reduced to a final analytic sample of 3,400 cases, a 69.1% visibility loss driven by unlinked contexts. Attrition was most pronounced for institutional subpopulations where identifiers were systematically suppressed. Such fragmentation produces a “surveillance paradox” of health, justice, and contextual silos with severe downstream consequences for population-level prevention prioritization. Privacy-preserving surveillance modernization through advanced linkage, formal cross-sector partnerships, and reformed data-access agreements is a structural prerequisite for equity-oriented IPV prevention and accountable institutional response

Introduction

Interpersonal violence (IPV) remains a persistent U.S public health challenge with profound implications for health equity, institutional readiness, and population wellbeing. Despite advances in surveillance systems, critical gaps persist in the ability to integrate relational, situational, and institutional data necessary for prevention-focused modeling. Existing infrastructure relies on fragmented reactive indicators such as physical injury or criminal reporting that prioritize incident-level documentation over contextual integration or cumulative risk processes [1-2]. These systems were developed for distinct operational purposes and rarely function as an integrated analytic infrastructure. The observed denominator is not the total number of victimizations, but rather by the subset that retains cross-domain linkage. A resulting structurally biased analytic populations constrains analytic capacity from having to rely on a subset of cases without sufficient multilevel information to support prevention-relevant analysis. This creates a disconnect from the Population Health Impact Pyramid, which prioritizes structural and environmental changes over individual clinical interventions [3]. A “surveillance paradox” emerges where complex and potentially preventable forms of violence characterized by relational and situational risk remains statistically invisible until they escalate into acute physical trauma.

The Denominator Problem

We define this “denominator problem” as the systematic reduction of analytically eligible cases when surveillance data require sufficient relational, situational, and contextual variables for modeling. Existing literature heavily documents reporting behavior but often fails to quantify how surveillance structures themselves eliminate prevention-relevant cases before analysis can occur. While prior research has documented underreporting and measurement limitations in violence surveillance [1,4], fewer studies have examined effects of structural data fragmentation itself. This limitation is particularly consequential for sexual violence, where reporting pathways are influenced by relational proximity, institutional context, and

situational dynamics that are not consistently captured across systems. Current surveillance systems like the National Intimate Partner and Sexual Violence Survey (NISVS) and National Crime Victimization Survey (NCVS) prioritize incident-level data [4-5]. While the National Violent Death Reporting System (NVDRS) has been hailed as a new “gold standard” for its ability to link disparate data sources to provide a comprehensive picture of violent death non-fatal surveillance remains siloed, leading to the systematic analytic attrition discussed in this manuscript [2]. Additionally, systematic reviews of primary prevention strategies confirm that a lack of
longitudinal, contextual data remains a primary barrier to evaluating program efficacy [6-7] Unlinked data streams impede the capacity for syndromic surveillance, making it impossible to detect the early risk signatures that precede severe harm.

Study Objective

Quantifying information gain enables an empirical measure of the added value of data integration, moving beyond theoretical calls for linkage toward an operational demonstration of its necessity for precision public health. This manuscript uses the NCVS to examine the gap between conceptual consensus and operational evidence when progressive analytic requirements reduce the number of usable cases for prevention modeling. The NCVS is a nationally representative survey of noninstitutionalized U.S. residents aged 12 years and older designed to capture both reported and unreported victimization, including sexual and IPV. We frame fragmentation as a structural determinant of visibility to identify how missing cross-system identifiers bar health equity with downstream consequences for population-level prevention prioritization. Rather than treating missingness solely as a statistical issue, we conceptualize data loss as a structural feature of surveillance systems that shapes what can be known, modeled, and acted upon.

Marginalized IPV victims whose experiences of harm do not fit narrow “injury-based” reporting are systemically disadvantaging when injury-centered surveillance identifies select harm after significant escalation, capturing safety failures rather than the behavioral and situational precursors defining integrated prevention. While our previous research (*under review*) empirically examined the individual and situational determinants of reporting sexual violence victimization, the current independent study provides a structural evaluation of the surveillance architecture itself, enumerating how data fragmentation impedes the very prevention modeling identified as necessary in our prior findings. We quantified victimization data lost as prevention-relevant variables are increasingly stratified, diagnosing the “denominator problem” preceding behavioral analysis. The goal is to shift the field from reactive counting to anticipatory, place-based risk forecasting, illustrating why siloed surveillance architectures systematically obscure ecological stability of risk.

Methods

Data were drawn from public use NCVS files (1992-2024). Because public-use NCVS files omit primary sampling unit identifiers, analysis focused on within-sample behavioral associations rather than generalizable

population estimates to quantify how fragmentation precludes the prevention modeling identified as necessary in our prior findings, (consistent with established methodologies for non-weighted federal survey data) [8]. Importantly, public-use NCVS files suppress geographic identifiers and selected institutional variables to protect respondent confidentiality, limiting the ability to link cases across contextual domains. We operationalize fragmentation not as a conceptual limitation, but as a measurable analytic process that systematically reduces prevention-relevant visibility before modeling begins.

Analytic Sample

The full surveillance frame included approximately 240,000 households comprising 329,678 observations prior to analytic filtering. The NCVS employs a stratified multistage cluster sampling design of noninstitutionalized households, with respondents interviewed at six-month intervals for up to seven waves. Because multiple victimization incidents were reported by the same individuals, incidents were clustered within persons, necessitating analytic approaches that account for within-person correlation [8]. From this initial pool, a final unweighted analytic subset of 3,400 sexual victimization incidents reported by 1,735 unique individuals was identified. Our objective was to quantify analytic attrition rather than estimate population prevalence, making unweighted analyses appropriate and consistent with prior methodological guidance for within-sample structural evaluation.

The Surveillance Signal Filter: An Integrated Framework

This analysis applies a multilevel population health framework conceptualizing surveillance as a “signal filter”. While socio-ecological approaches are well established, this framework focuses on how multilevel risk is differentially captured or lost within existing
architectures. We hypothesize that “data silos” selectively filters physically injurious events while discarding contextual data necessary for primary prevention. To empirically simulate fragmentation, variables were grouped into three analytic silos:

Health silo: Physical injury, weapon involvement, medical treatment.

Justice silo: Victim offender relationship, reporting to law enforcement, marital or household status.

Contextual silo: Activity at time of incident, location, repeat victimization, institutional context, for example college or military affiliation.

The three silos function as a simulated analytic environment where information loss is precisely measured. Mirroring real-world governance boundaries, this methodology tests how administrative separation constrains prevention insight, shifting focus from victim behavior to system capability.

Operationalizing Behavioral and Ecological Theories

This study draws on multilevel theoretical frameworks that emphasize the interaction between individual, relational, and environmental determinants of behavior. The “Structural Imperative” framework provides the analytic architecture to test three foundational behavioral theories through the lens of data integration. These theories were selected because they require integration across multiple contextual domains, making them particularly sensitive to data fragmentation. Applying our three silo variables, we can evaluate whether current surveillance can detect theorized drivers of violence, and not whether these mechanisms are causally sufficient.

Routine Activities Theory (RAT) RAT posits that violence occurs when a motivated offender, a suitable target, and the absence of a capable guardian converge in space and time [9]. We use the Contextual Silo (e.g., location, activity at time of incident) to proxy these situational markers. Without contextual data-linkage with the Health and Justice silos, the "opportunity structure" of sexual violence remains invisible, and responses remain reactive rather than anticipatory.

Social Disorganization Theory (SDT) SDT suggests that community-level structural factors, such as residential instability or weakened institutional capacity, diminish collective efficacy and increase vulnerability [10]. We utilize the Contextual Silo to capture the environmental determinants of risk clustering. Fragmentation prevents the distinguishing between isolated events and SDT-predicted structural hotspots.

Lifestyle Exposure Theory LET explains how social roles and institutional affiliations; being a college student or a service member, can lead to lifestyles or situations that increase exposure to high-risk environments [11]. LET is particularly salient for our institutional analysis. When institutional identifiers are suppressed or unlinked from the Justice and Health silos, specific risk signatures from these institutional lifestyles are erased. This creates an "institutional blind spot" that bars tailored, context-aware prevention strategies.

This approach evaluates theory not only by statistical association, but by whether existing surveillance architectures are structurally capable of detecting theorized risk mechanisms. Alternative theoretical frameworks that focus primarily on individual-level predictors were considered but not prioritized, being less affected by cross-system data loss.

Analytic Design

We descriptively quantify progressive analytic attrition as contextual variables become increasingly required for multilevel analysis. The primary outcome, “Information Gain”, is the increase in risk visibility achieved when silos are linked compared analyzing them in isolation. This methodology calculates the “denominator problem” by identifying the proportion of cases lost when requiring sufficient relational and situational context for modeling.

Results

Progressive Analytic Attrition Across Surveillance Stages

Progressive analytic attrition occurred as increasingly prevention-relevant variables were required for inclusion. From the foundational NCVS frame of 329,678 incident-level observations, an initial screening identified 11,000 sexual violence incidents. Rigorous
stratification for integrated relational and situational context resulted in a final analytic sample of 3,400 cases. This represents a cumulative yield of 1.0% from the total observation frame and a 69.1% loss of analytic visibility among identified incidents This attrition illustrated in Table 1
effectively erases analytic visibility prior to modeling or exclusion based on statistical assumptions.

Table 1. Progressive Analytic Attrition of Sexual Violence Cases (Public-Use NCVS, 1992–2024).

Surveillance Stage	Analytic Requirement	N Retained	% of SV Subset	Cumulative Loss
Identified SV Cases	Screened sexual violence incidents (Baseline)	11,000	100.0%	0%
Relational Context	Victim-offender relationship data available	7,120	64.7%	35.3%
Situational Context	Location and activity data available	4,850	44.1%	55.9%
Multilevel Subset	Full Contextual Integration (Final Sample)	3,400	30.9%	69.1%

Table 1 attrition was non-random and biased toward the visible, as the Health silo effectively captures incidents involving weapon use or severe trauma while discarding non-injurious, coercive encounters lacking medical or police footprints. Consequently, when these cases
also lack reporting markers required by the Justice silo, they rely entirely on the Contextual silo to remain visible. The 30.9% retention rate demonstrates that current filters are most likely to drop cases where health and justice indicators are absent, effectively erasing the situational precursors of sexual violence. This retention rate demonstrates that current filters are most likely to drop cases where Health and Justice indicators are absent, effectively erasing the situational precursors of IPV.

Institutional and Structural Visibility Loss

Analytic attrition was particularly pronounced for transient institutional subpopulations. From Table 2, most cases lacked usable institutional identifiers under integrated analytic requirements. The result is substantial reductions in analytically retained cases for both college-affiliated and military-affiliated subpopulations (~30%). Such missingness is unlikely to be completely at random, given known design constraints in public-

use data that suppress identifiable institutional and geographic variables. This illustrates how structural attrition limits the detection of directional trends in high-risk institutional settings . Such attrition suggests that institutional prevention strategies are likely modelled off data from a minority of actions or formal reports potentially missing subtle scenarios of victimization. Institutional context is structurally limited in public-use data, where the absence of identifiers prevents the assessment of system-level prevention and response despite documented elevated risk within these hierarchical settings [12-13]. Institutional context is thus not missing at random, but systematically suppressed rendering accountability impeded within closed systems. This
confirms that the “denominator problem” is not uniform; it is most acute for the very populations LET and RAT suggest are at highest risk.

Table 2. Loss of Institutional Subgroup Visibility Under Integrated Analytic Requirements.

Institutional Indicator	Category	n	% of Final Sample	Unified Structural Prevention Gap
College Affiliation	Yes	188	5.9%	**Institutional Blindness**: occurs when the suppression of specific identifiers prevents campuses from evaluating Title IX efficacy and military commands from identifying high-risk occupational clusters.
	Suppressed/ Unlinked	3212	94.1%	**Resource Misallocation**: results when prevention funding is distributed based on broad national estimates rather than localized campus risk signatures or unit-specific military requirements.
Military Affiliation	Yes	65	2.0%	**Accountability Void**: emerges when the inability to correlate incidents with residential hall climates or military unit-level initiatives precludes leadership from addressing the structural drivers of violence.
	Suppressed/ Unlinked	3330	98.0%	**Escalation Risk:** persists when environmental triggers, such as student lifestyle factors or military occupational stressors, remain invisible in national frames, preventing targeted primary prevention before harm matures.

The data suggest that the institutional blind spot results directly from fragmented data. For instance, when military or college status is present but the "activity at time of incident" or

"offender relationship" is missing, the cases retain descriptive value but are limited in their capacity to support multilevel prevention modeling that requires integrated relational and situational context.

Mapping the Missing: Surveillance Capacities Foregone

Table 3 explicitly links missing analytic domains to the specific public health capacities that are currently impossible to achieve using fragmented data. This provides the empirical evidence suggesting that siloed data are not just a technical inconvenience, but a structural barrier to accountability to the precision prevention goals outlined for example in spatial epidemiology prevention models. The persistent fragmentation of surveillance architectures creates a "capacity gap" where the systematic loss of geographic and longitudinal identifiers prevents the application of advanced spatial-temporal modeling, thereby masking the environmental clustering of risk and the localized escalation of violence [14-15]

Table 3. Forensic Analysis of Surveillance Gaps and Foregone Prevention Capacity.

Analytic Domain Absent or Limited	Example Variables Not Available in Public-Use NCVS	Study Stage of Analytic Attrition	Precision Prevention Capacity Foregone
Geographic Context (Micro-Spatial)	Census tract/block-group IDs; *geocodable* address of incident; proximity to transit hubs	Spatial Filtering: Excluded during neighborhood-level risk adjustment.	Vulnerability Corridors: Inability to identify specific routes (e.g., poorly lit transit paths) where routine activities overlap with environmental hazards
Longitudinal Linkage (Person-Level)	Unique person-identifiers across waves 1–7; linkage to prior CPS or clinical trauma history.	Temporal Decoupling: Excluded at the trajectory modeling stage.	Escalation Signatures: Inability to distinguish between an isolated stranger assault and the "coercive escalation" patterns indicative of a high-risk intimate partner.
Institutional Response	Campus-specific ID (Title IX); Military Unit/ Command code; Law Enforcement Agency (LEA) ID.	Accountability Void: Entirely absent from the sampling frame.	Institutional "Leakage" Analysis: Inability to track where survivors drop out of the "justice pipeline" vs. the "health pipeline" within specific closed systems.
Health System Interaction	Emergency Department (ED) encounter flags; ICD-10 codes for sexual assault; referral to specialty care	Clinical Proxy Bias: Partially available only via physical injury/ hospitalization markers.	Trauma-Informed Triage: Prevention remains reactive because clinicians cannot see the "pre-incident" healthcare utilization patterns that signal risk before an acute assault occurs.
Environmental Exposure	Alcohol outlet density; housing vacancy rates; local unemployment/ neighborhood deprivation indices.	Ecological Suppression: Excluded from the multilevel environmental model.	Syndromic Risk Forecasting: Inability to model how shifts in local economic or built-environment factors (e.g., a new nightclub cluster) correlate with emerging IPV "hotspots".
Cross-System Sequencing	Real-time temporal ordering across Health, Justice, and Institutional reporting systems.	Operational Blindness: Entirely unavailable for cross-sectional analysis.	Common Operating Picture: Inability to establish "early warning" indicators across sectors to deploy preventive resources (e.g., bystander training) before a cluster matures

Discussion

This study represents a strategic analytic shift from our prior empirical examinations of reporting sexual victimization to a structural evaluation of surveillance limitations. Using the NCVS as an empirical laboratory, we quantify the transition from 329,678 incident-level observations to 3,400 actionable cases, demonstrating that most victimization is filtered out of
the prevention pipeline by design rather than survivor choice. Our findings confirm how most of sexual victimization remains analytically invisible because it lacks the relational and contextual variables required for multilevel interpretation. We reveal this analytic attrition to provide the empirical rationale for the surveillance modernization efforts

The NCVS Context: A Proxy for National Surveillance Gaps

The 69.1% loss of visibility identified here serves as an empirical proxy for a national "denominator problem" that characterizes modern violence prevention. These findings do not reflect shortcomings in the NCVS survey design, but rather illustrate how a reliance on siloed surveillance architectures systematically privileges visible injury over relational risk. This catastrophic loss of signal suggests that a substantial proportion of potential risk markers are not retained within analytically usable datasets. This integration gap paralyzes the operationalization of behavioral theory: we cannot distinguish between isolated events and the structural clusters predicted by SDT, nor can we identify how institutional environments shape the routine activities and exposure risks predicted by RAT and LET. This structural erasure is further evidenced by the extreme missingness observed in our empirical analysis of contextual variables; for instance, data on professional help-seeking exhibited over 80% missingness, limiting statistical utility for predictive modeling despite its importance to prevention. This leaves public health practitioners with a skewed and incomplete operating picture.

Fragmentation across government, health, education, and justice systems is especially consequential for populations whose victimization occurs within closed and hierarchal systems such as prison, military installations or university campuses. In these settings, reporting thresholds are nuanced and physical injury inconsistently predicts risk. Our observed NCVS analytic attrition mirrors broader structural patterns where the Justice silo (police reporting) serves as the primary gatekeeper for visibility. Failing to integrate Contextual and Health Silos filters out the coercive, repeated patterns of acquaintance-based sexual violence inherent to these institutions. Restricted-use datasets offer granular linkage. Critically, while restricted-use NCVS datasets offer a pathway for more granular linkage through IRB protocols and security safeguards, this capacity remains the exception. Most administrative datasets similarly lack linkage architectures with restrictions that
preclude the analyses necessary to prevent escalation.

While standard missing data approaches (e.g., multiple imputation) can address incomplete observations under certain assumptions, they cannot reconstruct structurally absent variables or cross-system linkages central to this analysis [16]. Without systemic access to linked data, response remains reactive, intervening only after harm has peaked

The "Silo" Cost: From Variables to Vulnerability Corridors

Spatial research identifies neighborhood-level stressors and outlet density as critical predictors of interpersonal violence [14-15, 17].
The attrition described in Table 3 is not just a statistical loss; it is a prevention failure. For example, NCVS may record an assault that occurred in a parking lot (Public-Use) but suppresses the Geographic Context (Tract-ID) needed to link that location to environmental exposures like alcohol outlet density or lighting levels. Without this linkage, public health practitioners cannot identify vulnerability corridors, the specific spatial-behavioral intersections where routine activities (RAT) and social disorganization (SDT) converge to produce predictable risk..

When surveillance systems prioritize physical injury, weapon involvement, or formal reporting, they consistently inherently under-detect coercive, acquaintance-based, and repeated victimization. When surveillance "blind spots" are concentrated in specific institutional or geographic contexts, the subsequent public health response is not merely incomplete, it is biased. Marginalized and transient populations are more likely to experience coercive, repeated, and institutionally mediated violence, necessitating architecture that systematically prioritizes their risk.

Identifying suppressed reporting patterns in these populations is critical for prevention planning, as these groups often fall outside the reach of traditional law-enforcement–centered interventions. Reorienting the field toward risk process detection enables identification of cumulative vulnerability, ensuring that prevention resources are deployed based on the probability of future harm rather than the reporting of past injury.

Public Health and Policy Implication

Our findings demonstrate a 69% analytic attrition underscores the necessity of a surveillance modernization that transitions from reactive, injury-based monitoring toward anticipatory, place-based risk forecasting multilevel strategy. The move toward relational database technologies in surveillance was intended to overcome the limitations of 'thin' data sources[19]. However, as our results demonstrate, when these relational links are broken or missing, the resulting attrition creates a visibility gap that disproportionately affects marginalized subpopulations.

Strategic Pivot: From Counting to Forecasting

The NCVS files already contain sufficient detail to reveal risk signatures when variables are examined jointly. What currently lack are the governance structures and analytic linkage mechanisms that would allow these signals to be routinely identified across disparate surveillance systems. The missing domains identified enhance the argument for surveillance reform not as an informatics upgrade but as a foundational public health intervention. Additionally, restrictions on NCVS and similar datasets are largely driven by confidentiality protections, including the prevention of respondent re-identification through

geographic or institutional variables. While necessary, these constraints limit the ability of to conduct integrated, multilevel analyses required for prevention modeling. Contexts in which violence is more likely to be coercive, repeated, institutionally mediated, and underreported are systematically excluded particularly in institutional populations. This analysis does not imply that linkage alone resolves prevention failures, but demonstrates that without linkage, prevention modeling is constrained. Reliance on siloed surveillance architectures
systematically privileges visible injury over relational risk processes, which reinforces inequities by misclassifying high-risk contexts as low priority.

Precision Public Health Surveillance

The barrier to improved visibility is not the absence of data, but the absence of analytic integration. To achieve a public health Common Operating Picture (COP), surveillance must move beyond the individual-level incidents towards reconstruct the environmental and relational sequences that determine violence. Integrated surveillance systems ensure that high-risk contexts are accurately identified and prioritized, preventing the misclassification of risk that occurs when data is analyzed in isolation. Cross-linking Health, Justice, and Contextual silos allows the operationalization of behavioral theories (RAT, SDT, LET, etc.), shifting the focus from individual behavioral failures to the structural hot-spots and institutional blind spots that define prevention opportunities. Improved visibility becomes achievable through analytic restructuring, even prior to new data collection efforts.

Precision public health necessitates a new prevention paradigm that moves beyond traditional, static surveillance models to an enhanced ability to see the longitudinal tail of IPV; from identification of risk patterns and escalation signatures to how initial coercive behaviors in early survey waves might transition into acute physical violence in subsequent waves [18]. By
decoupling events, current surveillance architecture treats every incident as static points, blinding practitioners from the escalations that primary prevention is designed to disrupt. This decoupling is particularly problematic in institutional settings where reporting patterns suggest that trajectories of harm are often cumulative rather than isolated incidents [19].

Based on this empirical demonstration, we recommend the following tiered actions.

Immediate

Reorient to Risk Process Detection: Violence surveillance should shift from incident counting to risk process detection, prioritizing the continuity of contextual signals over the completeness of incident counts. Surveillance systems should prioritize relational indicators, including offender proximity and repeat coercion, as core surveillance variables rather than ancillary descriptors. Addressing blind spots requires acknowledging that missing data is often a structural byproduct of siloed collection. Our analysis demonstrates that when relational context is unlinked, the resulting data missingness prevents the detection of the very escalation signatures that primary prevention seeks to disrupt.

Frame Integration as an Equity Imperative: Systems that rely on injury thresholds systematically disadvantage marginalized populations. Integrated surveillance must be viewed as a tool for the equitable allocation of prevention resources.

Expand Secure Data Access: Standardized governance frameworks and formal data-sharing agreements are needed to provide researchers with secure access to geographic identifiers and longitudinal links. This requires active collaboration between institutions to synchronize disparate data streams. Overcoming these general access barriers is essential for the application of spatial epidemiology to identify neighborhood-level risk clusters and to evaluate the efficacy of specific institutional responses.[8,18]

Intermediate

Mandate Analytic Integration and Access Reform: Public health surveillance should review and restructure data agreements to facilitate routine deaggregation and cross-sector integration. This study's ~31% retention rate demonstrates that current governance structures act as a barrier to the joint analysis of health, justice, and contextual indicators. Implementation requires breaking general access barriers to ensure epidemiologists can link administrative datasets for proactive prevention. Integration success should be measured by prevention-relevant outputs such as time-to-intervention and reduction in repeat harm rather than solely by predictive accuracy or incident volume.

Long term

Pilot Privacy Preserving Methods: Technical solutions such as Federated Learning [20] and Secure Multiparty Computation (SMPC) [19] allow for joint analysis across systems we are advocating for without exposing personally identifiable information. Federated learning is a distributed analytic approach in which models are trained across decentralized datasets without transferring raw data. Secure multiparty computation (SMPC) enables multiple entities to jointly compute analytic outputs while maintaining data encryption and privacy. These approaches are particularly suited to violence surveillance, where trust erosion has historically blocked cross-sector collaboration [6, 12-13].

Limited surveillance visibility may bias resource allocation toward cases involving visible injury or formal system contact, potentially overlooking patterns of escalating risk that do not generate immediate institutional response. These patterns are central to effective primary prevention.. Bridging these gaps is not merely a matter of technical data cleaning, but fundamental requirement of precision public health to prevent the transition to acute violence or harm. The field can consequently move beyond individual behavioral failures to address SDT by identifying "high-vulnerability zones" where weakened institutional capacity necessitates community-level intervention [10,14]. Our findings echo concerns that a narrow focus on specific endpoints or siloed data can overlook the broader spectrum of serious violence.

Conclusion

The findings of this study illustrate why the public health field is currently failing to move upstream in IPV, particularly sexual violence prevention. Fewer than half of cases in a premier national dataset retained sufficient context for sexual violence prevention modeling. This is not a limitation of survey design, but a failure of system integration. We empirically demonstrate how surveillance blind spots are not conceptual concerns but quantifiable consequences of fragmented analytic frameworks. Integrated, multilevel surveillance is essential for advancing early identification, equitable prevention, and accountable institutional response to interpersonal violence. Coordinated access to geographic identifiers, longitudinal linkage, institutional response indicators, and cross-system event sequencing would enable earlier detection of escalation, identification of high-risk environments, and evaluation of institutional accountability. Privacy-preserving approaches offer viable pathways for achieving such integration while maintaining confidentiality.

Modernization is not merely a technical hurdle but a structural one. As Butchart noted, overcoming territorial approaches to data via formal accessibility agreements is a prerequisite for a joined-up perspective that adds value to prevention [2]. By integrating the Health, Justice, and Contextual silos, we move beyond a reactive injury-centered model toward a proactive system that accounts for the environmental and structural determinants of harm. Fragmentation is not merely a technical limitation: it is a structural limitation to health equity. Ultimately, addressing interpersonal violence requires analytic architectures capable of capturing risk as a continuous, context-dependent process rather than a series of isolated, discrete events.

Declarations

Author Contribution

Cindy A. Ogolla Jean-Baptiste: Conceptualization, methodology, formal analysis, data curation, investigation, writing – original draft, and writing – review & editing. The author is responsible for all aspects of the work, including the integrity of the data and the accuracy of the analysis.

Acknowledgement

The author would like to acknowledge Dr. Christine Ouma, Co-Principal Investigator of the IRB-approved study 'Predictive Prevention: Integrating NISVS and NCVS for Policy-Relevant Violence Risk Stratification' (IRB-ID: 2025-0850). As the Lead statistician for the broader project, for her foundational work on the overarching study's analytical design.

References

1. Baumer EP, Lauritsen JL. Reporting crime to the police, 1973–2005: A multivariate analysis of long-term trends in the National Crime Survey and National Crime Victimization Survey. Criminology. 2010;48(1):131-85.

2. Butchart A. The National Violent Death Reporting System: a new gold standard for the surveillance of violence related deaths? Inj Prev. 2006;12(Suppl 2):ii63-4.
3. Frieden TR. A framework for public health action: the population health impact pyramid. Am J Public Health. 2010;100(4):590-5.
4. Langton L, Berzofsky M, Krebs C, Smiley-McDonald H. Victimizations Not Reported to the Police, 2006-2015. Washington (DC): Bureau of Justice Statistics; 2017.
5. Basile KC, Smith SG, Kresnow MJ, et al. The National Intimate Partner and Sexual Violence Survey: 2016/2017 Report on Sexual Violence. Atlanta (GA): Centers for Disease Control and Prevention; 2022.
6. Dworkin ER, Brill CD, Ullman SE. Social reactions to disclosure of interpersonal violence and psychopathology: A systematic review and meta-analysis. Clin Psychol Rev. 2019;72:101750.
7. McCart MR, Smith DW, Sawyer GK. Help seeking among victims of crime: A review of the empirical literature. J Trauma Stress. 2010;23(2):198-206.
8. Morgan RE, Truman JL. Criminal Victimization, 2021. Washington (DC): Bureau of Justice Statistics; 2022.
9. Cohen LE, Felson M. Social change and crime rate trends: a routine activity approach. Am Sociol Rev. 1979;44(4):588-608.
10. Sampson RJ, Raudenbush SW, Earls F. Neighborhoods and violent crime: a multilevel study of collective efficacy. Science. 1997;277(5328):918-24.
11. Hindelang MJ, Gottfredson MR, Garofalo J. Victims of personal crime: an empirical foundation for a theory of personal victimization. Cambridge (MA): Ballinger; 1978.
12. Katz J, Richer AM. Students' perceptions of campus sexual misconduct policies: Implications for procedural justice and reporting behavior. J Interpers Violence. 2021;36(21-22):10077-101.
13. U.S. Government Accountability Office. Sexual assault: DOD should collect and use key information to improve prevention and response efforts (GAO-21-445). Washington (DC): GAO; 2021.
14. Jean-Baptiste CO. A prevention-focused geospatial epidemiology framework for identifying multilevel vulnerability across diverse settings. Healthcare (Basel). 2026;14(2):261.
15. Anselin L, Cohen J, Cook D, Gorr W, Tita G. Spatial analyses of crime. In: Duffee D, editor. Measurement and analysis of crime and justice. Criminal justice 2000, volume 4. Washington (DC): National Institute of Justice; 2000. p. 213-62.
16. Little RJ, Rubin DB. Statistical analysis with missing data. 3rd ed. Hoboken (NJ): John Wiley & Sons; 2019.
17. Cunradi CB, Mair C, Todd M, Remer LG. Alcohol outlet density, multilevel neighborhood stressors, and intimate partner violence. Am J Prev Med. 2012;43(3):258-67.
18. Dowell SF, Blazes D, Desmond-Hellmann S. Four steps to precision public health. Nat Rev Genet. 2016;17(2):61-2.
19. Weiss HB, Gutierrez MI, Harrison JE, Matzopoulos R. The US National Violent Death Reporting System: domestic and international lessons for violence injury surveillance. Inj Prev. 2006;12(Suppl 2):ii58-62.

20. Kairouz P, McMahan HB, Avent B, et al. Advances and open problems in federated learning. Found Trends Mach Learn. 2021;14(1–2):1-210.