Inventory of U.S. public data sources to measure the socioeconomic impact of experiencing interpersonal violence

Jordan Papp PhD, Michael Mueller-Smith PhD, Megan C. Kearns PhD, Cora Peterson PhD

PII: S2773-0654(23)00051-2
DOI: https://doi.org/10.1016/j.focus.2023.100114
Reference: FOCUS 100114

To appear in: AJPM Focus

Received date: 14 December 2022
Revised date: 26 May 2023
Accepted date: 27 May 2023

Please cite this article as: Jordan Papp PhD, Michael Mueller-Smith PhD, Megan C. Kearns PhD, Cora Peterson PhD, Inventory of U.S. public data sources to measure the socioeconomic impact of experiencing interpersonal violence, AJPM Focus (2023), doi: https://doi.org/10.1016/j.focus.2023.100114

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 Published by Elsevier Inc. on behalf of The American Journal of Preventive Medicine Board of Governors.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Highlights

- Public data exists to study the socioeconomic impact of interpersonal violence.
- Administrative data resources and linkage methods provide new opportunities.
- Linked administrative data provides cost-effective alternatives to survey data.
Inventory of U.S. public data sources to measure the socioeconomic impact of experiencing interpersonal violence

Jordan Papp¹, PhD; Michael Mueller-Smith², PhD; Megan C. Kearns³, PhD; Cora Peterson³, PhD

Affiliations: ¹ Institute for Social Research, University of Michigan; ² Department of Economics, University of Michigan; ³ National Center for Injury Prevention and Control, Centers for Disease Control and Prevention

Corresponding author: Jordan Papp, pappj@umich.edu, 426 Thompson Street Ann Arbor MI 48104

Keywords (MeSH): Violence; Child Welfare; Economics; Databases, Factual; Longitudinal Studies

Funding: JP and MMS were funded via an Intergovernmental Personnel Act agreement from the Centers for Disease Control and Prevention for this study.

Financial disclosures: The authors have no relevant financial relationships to disclose.

Conflicts of interest: The authors have no relevant conflicts of interest to disclose.

Word count (excluding abstract, references, and tables): 1,199/1,200; Pages: 11; References: 37; Exhibits: 2

Contributors: JP led and MMS, MK, and CP assisted with the study design and interpretation of results. JP analyzed data and drafted the manuscript. MMS, MK, and CP edited the manuscript. All authors approved the final manuscript as submitted.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Abstract

Introduction: There is limited recent information on the impact of interpersonal violence on individuals’ non-health related experiences and attainment including criminal activity, education, employment, family status, housing, income, quality of life, or wealth. This study aimed to identify publicly available representative data sources to measure the socioeconomic impact of experiencing interpersonal violence in the United States.

Methods: In 2022, authors reviewed data sources indexed in Data.gov, the Inter-university Consortium for Political and Social Research data archive, and the U.S. Census Bureau’s Federal Statistical Research Data Center network to identify sources that reported both nonfatal violence exposure and later socioeconomic status—or data source linking opportunities to achieve both measures—over time (i.e., longitudinal/repeated cross sections) at the individual person level. Relevant data sources were characterized in terms of data type (e.g., survey), violence measure
type (e.g., intimate partner violence), socioeconomic measure type (e.g., income), data years, and geographic coverage.

**Results:** Sixteen data sources were identified. Adverse childhood experiences, intimate partner violence, and sexual violence were the most common violence measures. Income, education, and family status were the most common socioeconomic measures. Linked administrative data had the largest breadth and depth of analytic opportunities.

**Conclusions:** Linked administrative data currently appears to offer the most comprehensive opportunities to examine the long-term impact of violence on individuals’ livelihoods. This type of data infrastructure may provide cost-effective research opportunities to better understand elements of the economic burden of violence and improve targeting of prevention strategies.

**keywords:** violence, socioeconomic outcomes, data resources, data linkage

Research has identified relationships between interpersonal violence and a range of negative long-term health and other consequences.\(^1\)\(^-\)\(^5\) Previous studies have also estimated the substantial cost of violence,\(^6\)\(^-\)\(^12\) providing important information for understanding the cost-effectiveness of violence prevention strategies.\(^13\)\(^-\)\(^15\) Recent cost of violence studies have focused on the health impact of violence using mathematical modelling (combined data from separate studies) of observational outcomes from cross-sectional sample survey data.\(^9\)\(^-\)\(^11\) For example, to estimate the long-term cost of intimate partner violence, researchers combined the lifetime cost of medical care for depression with the attributable prevalence of depression among survivors of intimate partner violence.\(^11\) Using observational cross-sectional data in this way is relevant given the data landscape but complicates the interpretation of violence as a cause of long-term health-related conditions and costs.

Further, there is limited recent information on the impact of interpersonal violence on individuals’ non-health related experiences and attainment including criminal activity, education, employment, family (including marital) status, housing, income, quality of life, or wealth. For example, one frequently cited study compared arrest records among adults up to age 40 years old with or without substantiated reports of child maltreatment processed by U.S. courts during 1967–71; that study quantified the way in which childhood abuse and neglect were associated with worse educational performance, worse mental health, lower achievement, and higher incidence of arrests in adulthood.\(^16\) New original direct estimates of the socioeconomic impact of
interpersonal violence using individual-level longitudinal data and econometric methods to strengthen the causal interpretation of the impact of violence would improve estimates of the economic burden of violence and the cost-effectiveness of prevention strategies. This study aimed to identify opportunities in current publicly available representative data sources to measure the impact of interpersonal violence on socioeconomic outcomes in the United States.

Methods

This study used publicly available data and no human subjects. In 2022, authors searched three archives of publicly available data: Data.gov, the Inter-university Consortium for Political and Social Research (ICPSR) data archive, and the U.S. Census Bureau’s Federal Statistical Research Data Center (FSRDC) network. Data.gov operationalizes the 2018 OPEN Government Data Act requiring government data to be available in machine-readable formats while ensuring privacy and security. ICPSR maintains an extensive archive of data sources for social and behavioral sciences research. FSRDC provides qualified researchers access to restricted-use anonymized microdata (such as person-level records) with opportunities for linking to networked data sources using a securely assigned person-level Personal Identification Key.

Interpersonal violence was defined as adverse childhood experiences (ACE; abuse—physical, emotional, or sexual—or household member substance misuse, incarceration, mental illness, parental divorce, or intimate partner violence), bullying, community violence, elder abuse, intimate partner violence, sexual violence, or youth violence (Table 1). Socioeconomic status measures were defined as criminal activity, education, employment, family status, housing, income, quality of life (i.e., formal scale such as the Short Form Health Survey), or wealth. Data source search inclusion criteria were: ≥1 nonfatal interpersonal violence and temporally later socioeconomic status measure (directly reported or via data source linking), representative (not a convenience sample), individual person reporting unit, and longitudinal record type (multiple records over time for the same person or repeated cross-sections that measure violence exposure and are linkable to outcomes measured at a later point in time, but no requirement for the amount of time between measurements).

A keyword search was applied in Data.gov and ICPSR and the current FSRDC networked data source list was reviewed (Table 1). Keywords were searched in available meta data to identify
an initial pool of data resources that might include analytic measures that matched our definitions of violence and socioeconomic outcomes. Then, available data documentation was reviewed to identify data resources that included analytic measures of concepts matching our definitions of violence and socioeconomic outcomes. Qualifying data sources were characterized in terms of data type (e.g., survey), violence measure type (e.g., intimate partner violence), socioeconomic measure type (e.g., income; via direct report or data source linking), data years, and geographic coverage. Qualifying data sources were reviewed by two of the study authors independently and analyzed by author agreement.

Results

Sixteen data sources were identified that provide an opportunity to study the longitudinal socioeconomic impact of violence at the individual person level (Table 2). Adverse childhood experiences, intimate partner violence, and sexual violence were the most common analyzable violence measures (Table 1). Income, education, and family status were the most common analyzable socioeconomic measures. Self-reported information on violence victimization and socioeconomic status from longitudinal surveys was the source of the information in nearly all (14 of 16) of the analyzed data sources but two data sources were based on administrative records (e.g., substantiated child maltreatment or criminal justice records) (Table 2). The two administrative data sources were unique among the analyzed data sources in terms of being linkable at the record level to external longitudinal data sources reporting socioeconomic measures (Table 2). Nearly all (15 of 16) data sources had national geographic coverage and most had extensive temporal coverage (the longest was from 1970 to present and the shortest was 2016 to 2018) (Table 2).

Discussion

This study generated an inventory of publicly available representative data sources to classify current opportunities for expanding research evidence on the socioeconomic impact of experiencing interpersonal violence. Past research in this area has faced methodological limitations which have inhibited our understanding of the causal link between experiencing interpersonal violence and subsequent outcomes. This study’s primary contribution is a comprehensive and concise description of the current data landscape to investigate a complex
and understudied question of critical public health importance. Most opportunities identified here
to study the person-level socioeconomic impact of violence were based on self-reported
longitudinal survey data. A major benefit of this type of data is that it can capture a wider variety
of violence exposure than administrative data sources based on, for example, law enforcement
involvement. Another major benefit is that surveys typically include richer measures of various
forms of violence (as compared to administrative records). This is important for studying the
implications of experiencing different types of violence across the life course. However,
longitudinal surveys are resource intensive and for sensitive topics such as violence exposure and
socioeconomic status, survey respondents’ willingness to report and recall bias might reduce the
value of the information that is gathered.

This study identified a limited number of administrative data sources that offer applicable
analytic opportunities—primarily based on being linkable to other data sources. Recent
advancements in data linkage infrastructure, data and network security, and research surrounding
methodologies for preventing inadvertent disclosure of personal information has created new
opportunities to conduct research using linked administrative records. For example, the FSRDC
network facilitates data linkage across disparate data resources through the use of personal
identification keys that are generated using probabilistic matching algorithms. These personal
identification keys are an anonymized identifier which facilitate data linkage while maintaining
individual anonymity.

It should be noted that administrative records of violence exposure are necessarily a subset of all
violence that occurs, which limits the generalizability of related research findings. Making
generalizations to populations based on administrative records carries the risk of perpetuating the
biases inherent in the systems that generate administrative records. Therefore, caution should be
taken when considering the external validity of findings from studies using administrative
records. However, the major benefit of administrative records as highlighted here is that data
linking infrastructure can enable expansive opportunities for longitudinal analysis of victims’
and perpetrators’ subsequent socioeconomic experiences. It should also be noted that surveys
share similar limitations in terms of external validity. For example, a survey that aims to be
nationally representative may miss certain hard-to-reach groups of the population. Victims of
violence are a notable example of this because the consequences of violence exposure could
make these individuals less likely to be represented in surveys that are not specifically designed to measure this group, more susceptible to social desirability bias, and more difficult to measure at multiple points in time.

**Limitations**

This data inventory addressed only data sources that are publicly available and archived in Data.gov, ICPSR, or available in the FSRDC network. Data.gov has been criticized for a non-systematic approach.\textsuperscript{20} Classification of data sources was based on publicly available data documentation. Some data sources facilitate investigation of household, rather than individual, socioeconomic outcomes. Authors did not systematically examine data source quality issues, such as data completeness for key measures. For example, the Criminal Justice Administrative Records System (CJARS) had opportunities for the largest depth (records spanning 1970 to present) and breadth of socioeconomic status measures (criminal activity, education, employment, income, housing, family status, wealth) but identification of violence exposure in that data source is based on criminal justice system involvement (arrest, criminal court case filings, terms of incarceration, probation, and parole records). There are also a number of other aspects of data quality that were not assessed, such as survey participation rates over time, sample generalizability, concerns of accuracy of self-reported measures, changes in measurement over time, representativeness of administrative records, and limitations in data linkage. Because data quality assessment was outside of the scope of this data inventory, future researchers wishing to use the data resources identified here will need to evaluate data quality against the specific research questions they plan to answer and methodologies they plan to implement. This study primarily identified opportunities to study socioeconomic outcomes in terms of income, education, and family status but a wider variety of socioeconomic and quality of life measures would be valuable for public health research and practice.

**Conclusions**

Linked administrative data currently appears to offer the most comprehensive opportunities to examine the long-term impact of violence on individuals’ livelihoods. Despite the limitations of administrative data to comprehensively capture violence exposure, administrative data linking
infrastructure may provide cost-effective research opportunities to better understand elements of the economic burden of violence and improve targeting of prevention strategies.

References

### Table 1. Data source search

<table>
<thead>
<tr>
<th>Search term</th>
<th>Data.gov</th>
<th>ICPSR</th>
<th>FSRDC</th>
<th>Data sources analyzed, n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interpersonal violence (a)(b)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse childhood experiences</td>
<td>54</td>
<td>9</td>
<td>167</td>
<td>10</td>
</tr>
<tr>
<td>Bullying</td>
<td>42</td>
<td>56</td>
<td>167</td>
<td>3</td>
</tr>
<tr>
<td>Child abuse and neglect</td>
<td>55</td>
<td>35</td>
<td>167</td>
<td>9</td>
</tr>
<tr>
<td>Community violence</td>
<td>146</td>
<td>10</td>
<td>167</td>
<td>6</td>
</tr>
<tr>
<td>Elder abuse</td>
<td>41</td>
<td>21</td>
<td>167</td>
<td>2</td>
</tr>
<tr>
<td>Intimate partner violence</td>
<td>113</td>
<td>33</td>
<td>167</td>
<td>10</td>
</tr>
<tr>
<td>Sexual violence</td>
<td>123</td>
<td>31</td>
<td>167</td>
<td>10</td>
</tr>
<tr>
<td>Youth violence</td>
<td>82</td>
<td>32</td>
<td>167</td>
<td>2</td>
</tr>
<tr>
<td><strong>Socioeconomic status (c)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criminal activity</td>
<td>147</td>
<td>50</td>
<td>167</td>
<td>8</td>
</tr>
<tr>
<td>Education</td>
<td>12,094</td>
<td>282</td>
<td>167</td>
<td>14</td>
</tr>
<tr>
<td>Employment</td>
<td>3,276</td>
<td>266</td>
<td>167</td>
<td>12</td>
</tr>
<tr>
<td>Family status</td>
<td>1,853</td>
<td>28</td>
<td>167</td>
<td>14</td>
</tr>
<tr>
<td>Housing</td>
<td>5,367</td>
<td>261</td>
<td>167</td>
<td>8</td>
</tr>
<tr>
<td>Income</td>
<td>2,356</td>
<td>247</td>
<td>167</td>
<td>14</td>
</tr>
<tr>
<td>Quality of life</td>
<td>1,626</td>
<td>90</td>
<td>167</td>
<td>4</td>
</tr>
<tr>
<td>Wealth</td>
<td>375</td>
<td>101</td>
<td>167</td>
<td>3</td>
</tr>
</tbody>
</table>

ICPSR= Inter-university Consortium for Political and Social Research; FSRDC=U.S. Census Bureau’s Federal Statistical Research Data Center.

Notes. Searches conducted August 2022. Data.gov and ICPSR “considered” counts are the number of results returned for each keyword search (row titles in this table). Authors applied Data.gov’s sort by relevance function and ended review after the ranked list became no longer relevant for study aims. FSRDC counts are the number of indexed data sources in that research environment at the time of review. Analyzed counts match data source descriptions in Table 2.

(a) Violence types that intersect are separately identified. Example: A data source that reported child abuse and neglect exposure is counted in both “Adverse childhood experiences” and “Child abuse and neglect.”

(b) Data sources retained for analysis based on Centers for Disease Control and Prevention definitions (https://www.cdc.gov/violenceprevention): Adverse childhood experiences—physical, emotional, or sexual abuse, or household member substance misuse, incarceration, mental illness, parental divorce, or intimate partner violence; Bullying—any unwanted aggressive behavior(s) by another youth or group of youths, who are not siblings or current dating partners, that involves an observed or perceived power imbalance, and is repeated multiple times or is highly likely to be repeated; Child abuse and neglect—any act or series of acts of commission or omission by a parent, caregiver, or another person in custodial role that results in harm, potential harm, or threat of harm to a child; Community violence—happens between unrelated individuals, who may or may not know each other, generally outside the home. Examples include assaults or fights among groups and shootings in public places, such as schools and on the streets; Elder abuse—act or failure to act that causes or creates a risk of harm to an older adult; Intimate partner violence—abuse or aggression that occurs in a romantic relationship; Sexual violence—sexual activity when consent in not obtained or not freely given; Youth violence—intentional use of physical force or power to threaten or harm others by young people ages 10-24.

(c) Data sources retained for analysis based on: Criminal activity—conduct prohibited by criminal laws; Education—highest credential/attainment; Employment—having paid work; Family status—familial structure (e.g., marital status); Housing—one’s living circumstances (e.g., renting versus owning); Income—money received from paid work, investments, etc.; Quality of life—measure of health, comfort, and happiness experienced by an individual; Wealth—total value of one’s assets.
<table>
<thead>
<tr>
<th>Data source</th>
<th>Data source type</th>
<th>Violence type</th>
<th>Data years</th>
<th>Geography</th>
<th>Longitudinal socioeconomic measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add Health 22 (e) (f)</td>
<td>Survey</td>
<td>ACE, CAN, CV, IPV, SV, YV (S)</td>
<td>1994+</td>
<td>National</td>
<td>Direct report Criminal activity, education, employment, family status, housing, income, QOL, wealth</td>
</tr>
</tbody>
</table>
| Criminal Justice Administrative Records System 23 (a) (g) | Administrative records | ACE, CAN, IPV, SV (A) | 1970+      | 29 states    | Direct report Criminal activity  
  Data source linking  
  • ACS (2005+): Education, employment, income, housing, family status  
  • AHS (2005+): Employment, housing, income, criminal activity  
  • CPS (1962+): Education, employment, income, housing, family status  
  • Decennial Census (2000+): Education, income, housing, family status  
  • HPS (2020+): Education, employment, family status, housing  
  • IPUMS (1850+): Education, employment, income, housing, family status  
  • IRS W2 and 1040 (1969+): Income, employment  
  • Moving to opportunity (1994-2010, 2018): Employment, housing, income, education, criminal activity  
  • MSIS (2000-2016): Income  
  • PBI-TRA (1995+): Income, housing  
  • PBI-IC (2000-2016): Income  
  • PSID (1968+): Income, wealth, education, employment, family status  
  • SAIPE (1989+): Income  
  • SIPP (1937-2017): Income, employment, education, family status  
  • SNAP(s) (2004+): Income  
  • SSA-MBR (2015+): Income  
  • SSA-PHUS (2019+): Income  
  • SSI (2010+): Income  
  • TANF(s) (2004+): Income  
  • T-MSSIS (2014-2017): Income  
  • TRAC (1994-2014): Income, housing  
  • Unemployment insurance (2009-2017): Income, employment  
  • WIC (e) (2004+): Income  
  • WIC (e) (2004+): Income |
| Education Longitudinal Study Series 24 (e) (f) | Survey              | Bullying, CV, YV | 2002-2012 | National    | Direct report Education, employment, family status, income |
| Head Start Impact Study Series 25 (e) (f) | Survey              | ACE, CAN, CV | 2002-2008 | National    | Direct report Family status, income, housing |
| Interpersonal Conflict and         | Survey              | ACE, CAN, IPV, SV | 2016-2018 | National    | Direct report Education, employment, family status, housing, income |
|--------------------------|--------|-----|-----------|----------|---------------|---------------------------------------------|
| Midlife in the United States Series (e) (f) | Survey | ACE, CAN, elder abuse, SV | 1995+ | National | Direct report | Education, employment, family status, housing, income |
| Monitoring the Future Panel Data (f) | Survey | Bullying | 1975+ | National | Direct report | Criminal activity, education, employment, family status, income |
| National Child Abuse and Neglect Data System (e) (f) | Administrative records | ACE, CAN, SV | 1992+ | National | Direct report | Family status, income |
| National Crime Victimization Survey (a) (f) (g) | Survey | CV, IPV, SV | 1973+ | National | Direct report | Criminal activity, education, income |
| National Survey of Child and Adolescent Well-Being (e) (f) | Survey | ACE, CAN, CY, IPV, SV | 1997+ | National | Direct report | Education, employment, family status, housing, income, QOL, wealth |
| National Social Life, Health, and Aging Project (e) (f) | Survey | Elder abuse | 2005-2013 | National | Direct report | Criminal activity, education, employment, family status, housing, income, QOL |
| National Youth Survey (c) (f) | Survey | CV, IPV, SV | 1976-1987 | National | Direct report | Education, employment, income, family status, criminal activity |

ACE=Adverse childhood experience; ACS=American Community Survey; AF-CARS= Adoption and Foster Care Analysis and Reporting System; Add Health=National Longitudinal Study of Adolescent Health; AHS=American Housing Survey; AN=Child abuse and neglect; CBSA=Core based statistical area; CPS=Current Population Survey; CV=Community violence; FSRDC=Federal Statistical Research Data Centers; IPV=intimate partner violence; IRS=Internal Revenue Service; MSA=Metropolitan Statistical Area; MSIS=Medicaid Statistical Information System; PIH IC=Public Indian Housing Information Center; QOL=Quality of life (i.e., formal scale such as the Short Form Health Survey, or SF-36); PIH TRA=Public Indian Housing and Tenant Rental Assistance Longitudinal Information; PSID=The Panel Study of Income Dynamics; SAIPE=Small Area Income and Poverty Estimates; SIPP=Survey of Income and Program Participation; SNAP=Supplemental Nutrition Assistance Program; SSA-MBR= Social Security Administration Master Beneficiary Record; SSA-PHUS=Social Security Administration Payment History Update System, SSI=Supplemental Security Income; SV=Sexual violence; TANF=Temporary Assistance for Needy Families; T-MESIS=Transformed Medicaid Statistical Information System; TRAC=Tenant Rental Assistance Certification Center; WIC=Special Supplemental Nutrition Assistance Program for Women, Infants and Children; YV=youth violence.

(a) Available in U.S. Census Bureau FSRDC.
(b) Arrest, criminal court case filings, terms of incarceration, probation, and parole records.
(c) Not all states.
(d) Potential for same as CJARS (data source available in FSRDC but not currently linkable to other networked data sources).
(e) Identified in Data.gov.
(f) Identified in ICPSR.
(g) Identified in FSRDC network.

Credit Author Statement

**Jordan Papp**: conceptualization, methodology, formal analysis, investigation, data curation, writing – original draft

**Michael Mueller-Smith**: conceptualization, methodology, supervision, writing – review and editing, funding acquisition

**Megan Kearns**: conceptualization, methodology, writing – review and editing

**Cora Peterson**: conceptualization, methodology, project administration, formal analysis, writing – review and editing

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

*Michael Mueller-Smith reports financial support was provided by Centers for Disease Control and Prevention.*