

Open-Source Data and the Study of Homicide

Journal of Interpersonal Violence

2017, Vol. 32(18) 2693–2723

© The Author(s) 2015

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0886260515596145

journals.sagepub.com/home/jiv



**William S. Parkin¹ and
Jeff Gruenewald²**

Abstract

To date, no discussion has taken place in the social sciences as to the appropriateness of using open-source data to augment, or replace, official data sources in homicide research. The purpose of this article is to examine whether open-source data have the potential to be used as a valid and reliable data source in testing theory and studying homicide. Official and open-source homicide data were collected as a case study in a single jurisdiction over a 1-year period. The data sets were compared to determine whether open-sources could recreate the population of homicides and variable responses collected in official data. Open-source data were able to replicate the population of homicides identified in the official data. Also, for every variable measured, the open-sources captured as much, or more, of the information presented in the official data. Also, variables not available in official data, but potentially useful for testing theory, were identified in open-sources. The results of the case study show that open-source data are potentially as effective as official data in identifying individual- and situational-level characteristics, provide access to variables not found in official homicide data, and offer geographic data that can be used to link macro-level characteristics to homicide events.

Keywords

homicide, media and violence, criminology

¹Seattle University, WA, USA

²Indiana University–Purdue University Indianapolis, IN, USA

Corresponding Author:

William S. Parkin, Criminal Justice Department, Seattle University, 901 12th Avenue, Seattle, WA 98122, USA.

Email: parkinw@seattleu.edu

Introduction

Researchers have long relied on official sources of data to study crime and advance our understanding of criminality and criminal justice responses to crime. There are many advantages of official crime statistics, including their consistency and availability over the past several decades. Based on prior research, we know quite a bit about the nature of threats to official data reliability and validity, allowing us to develop strategies for alleviating potential biases (see, for example, Coleman & Moynihan, 1996; Elliot & Ageton, 1980; Hindelang, Hirschi, & Weis, 1979; Kirk, 2006; Loftin, McDowall, & Fetzer, 2008; Maxfield, Weiler, & Widom, 2000; Wiersema, Loftin, & McDowall, 2000). Despite these advantages, studying some crime types, or certain aspects of crimes, remains challenging based on official crime data. Currently, official data on topics of increasing interest to crime scholars, including white-collar crime, terrorism, and human trafficking to name a few, are less available for study. Other common sources of crime data, such as surveys of victims and offenders, are also usually impractical options when conducting large-scale studies on the serious crimes of terrorists and human traffickers. The question then becomes, where should we turn to when studying crimes not included in official data sources?

Contrastingly, what is available for serious crimes such as these is information from publicly available open-sources that can be mined, assembled, and systematically coded into original, comprehensive databases. Open-source data come in forms such as media articles, legal transcripts, and academic writings, and can either be free or purchased by members of the public.¹ Although not nearly as prevalent as traditional sources of data, open-source data have been increasingly used in scholarly research published in peer-reviewed journals. Examples include topics such as corporate crime (Steffensmeier, Schwartz, & Roche, 2013), serial killers (Allely, Minnis, Thompson, Wilson, & Gillberg, 2014; Beard, Hunter, Kern, & Kiley, 2014), and terrorism and extremism (see, for example, Chermak & Gruenewald, 2006; Cothren, Smith, Roberts, & Damphousse, 2008; Dugan, LaFree, & Piquero, 2005; Gruenewald, Chermak, & Freilich, 2012; Gruenewald & Pridemore, 2012; LaFree, Dugan, & Korte, 2009; LaFree, Dugan, Xie, & Singh, 2012; Parkin, Freilich, & Chermak, 2014; Sageman, 2004). Considering the increasing use of these data sources in social science research, it is curious that open-source data have yet to be systematically discussed, critiqued, or evaluated as a tool for studying crime. Although many scholars rightfully remain skeptical of the viability of open-source data as a supplement or alternative to official data sources, we suggest that it is imperative to comparatively assess the reliability and validity of open-source and official crime data sources. We ask, *does*

open-source capture the same crimes, in the same amount of detail, as official data sources?

In an initial step toward answering this important question, the results of a case study of official and open-source data on homicide victimization in Seattle, Washington, are presented. Data are comparatively examined across official and open-source data for all homicides known to the police occurring in 2011 within a single jurisdiction. Official homicide victimization data come from the Federal Bureau of Investigation's (FBI) Supplementary Homicide Report (SHR), the most commonly used official data source for studying homicide. In addition, open-source data on Seattle homicide victims are collected from print, televisions, and online news media, news magazines, news blogs, and other electronic sources of data such as criminal histories, residence histories, and other available public information. We focus our case study on homicide victimization for several reasons. First, we believe that it is important to begin examining the quality of open-source data with a single form of serious crime as opposed to observing crime more generally. Although our results may apply to open-source homicide victimization data, we would hesitate to generalize them to other forms of violent and non-violent crimes. Second, homicide is a logical choice as it is one of the most reliably reported crimes in official data. In part due to its seriousness, research has also found that homicide is one of the most frequently reported crimes in public sources, such as news media (Chermak, 1995; Graber, 1980). In effect, a case study of homicide victimization provides an opportunity to establish baseline comparative findings between open-sources and official sources under ideal conditions supportive of data availability. Third, official homicide data make for a sensible comparison with alternative open-source data because, as mentioned previously, much is already known about their reliability and validity. Fourth, homicide victimization has been a reoccurring topic in prior research reliant on open-source data. For example, scholars have used open-source data to study bias homicide (Gruenewald, 2012, 2013), homicide victimization of prostitutes (Quinet, 2011), and homicide with multiple victims (Fox & Levin, 2005).

Prior Research

Those studying political and other forms of ideologically motivated violence have turned to the use of open-source data when traditional sources of crime data are unavailable (LaFree & Dugan, 2007). For example, large data collection efforts, such as the American Terrorism Study (ATS; Smith & Damphousse, 2007), the Global Terrorism Database (LaFree & Dugan, 2007), and the Extremist Crime Database (Freilich, Chermak, Belli, Gruenewald, & Parkin,

2014), rely on open-sources to identify and code attributes of criminal activity by domestic terrorists and extremists. Importantly, several of these databases have relied on open-source data to advance the application of criminological theory to terrorism. Based on data from the Extremist Crime Database, for example, Parkin and Freilich (2015) applied victimization theories to the study of far-right homicide, while Chermak and Gruenewald (2015) relied on Extremist Crime Database data to test several structural theories of crime to violent offending by far-rightists, eco-terrorists, and Islamic extremists. Dugan et al. (2005) combined data from the Global Terrorism Database and other open-sources to determine whether government policies have deterred airline hijackings based on tenets of rational choice theory. Smith and Damphousse (1996, 1998) applied Hagan's (1989) structural-contextual theory to explain variations in the sentencing of domestic terrorists. In addition to theory testing, the aforementioned databases and open-sources have also been used to study the characteristics and behaviors of terrorist organizations (Chermak, Freilich, & Suttmoeller, 2013; Miller, 2012), counter-terrorism policies (Sproat, 2010), spatial patterns in the use of improvised explosive devices (Barker, 2011), dissident violence in Ireland (Horgan & Morrison, 2011), and issues of radicalization (Porter & Kebbell, 2011).

Terrorism research is not the only topic in which open-source data have been used, though not all research articles have used the term *open-source*.² For example, researchers studying gender and corporate crime augmented their primary data set with open-source information that came from financial reports, industry publications, government press releases, Department of Correction's data, and web searches (Steffensmeier et al., 2013). Other examples include studies on product counterfeiting (Heinonen & Wilson, 2012), human trafficking (Kangaspunta, 2003; Wilson & Dalton, 2008), and other forms of financial crimes (Kankaanranta & Mutttilainen, 2010). Research on other "specialized" types of homicide victimization have also benefitted from open-source materials (see, for example, Fox & Levin, 2005; Quinet, 2011). Gruenewald and colleagues, for example, have focused on fatal violence committed by suspects belonging to extremist and hate movements in the United States, integrating official homicide data and open-source data to comparatively examine "extremist" and "routine" forms of homicide in the United States (Gruenewald, 2011; Gruenewald & Pridemore, 2012). They have also explored suspect, victim, and incident-level characteristics of two forms of bias (or "hate") homicide, including homicides targeting homeless persons (Gruenewald, 2013) and members of the lesbian, gay, bisexual, and transgender community (Gruenewald, 2012; Gruenewald & Kelley, 2014; Kelley & Gruenewald, 2015). This research has demonstrated that while there remain obstacles to studying bias homicide at the individual and event

levels based only on official data sources, integrating official data with open-source data allows scholars to address types of homicide largely ignored in previous research.

To date, open-source data have primarily been used to study aspects of the criminal justice system that have little to no data available from victimization surveys, self-report surveys, or official data. We suggest that open-source data can be used in a much broader context in social science research, supplementing or, in specific cases, even replacing traditional data sources to test theory and inform policy. Homicide victimization specifically, which primarily relies on official data, can benefit from the information provided in open-sources. To demonstrate the effectiveness of open-sources as an additional source of data beyond official statistics, a case study on homicide victimization was conducted. Empirical data from official and open-sources are presented and compared at the individual and situational levels of analysis. Finally, the strengths and weaknesses of both data types are discussed, followed by suggestions for future research to move forward using open-sources to test theory and evaluate the validity and reliability of such data.

Method

Seattle has a population of more than 600,000 and is the largest city in the Pacific Northwest. Its metropolitan area, which includes surrounding King County, has a population of approximately 2,000,000 (U.S. Census Bureau, 2013). In 2011, Seattle had 20 homicide victims, a rate much lower than many cities its size (McNerthney, 2012). The utilization of Seattle and homicide victims for this case study allows for the opportunity to present open-source data collection in the best light possible—in a single jurisdiction where homicide victimization should routinely be covered by the media because of its rarity and severity. It is acknowledged that without further research the results of this case study are not generalizable to varying jurisdiction and time periods. The results, however, should be viewed as a proof of concept, allowing future researchers to develop methods for assessing the reliability and validity of open-source data across different crime types occurring in different types of places and conducted for various purposes (e.g., theory testing).

Official Data

When studying non-fatal victimization, researchers have relied on survey data such as the National Crime Victimization Survey or self-report data. For fatal victimization, official government data are most often utilized. One of these sources comes from the FBI's Uniform Crime Report, which releases data on

homicide incidents in the United States referred to as the SHR. Officers at local or state agencies submit information about the homicide either to a state-level agency in charge of data collection or to the FBI directly. The form elicits from law enforcement information on characteristics of the incident, suspect, and victim. While homicide researchers have commonly relied on the SHR to study victimization at the individual, situational, and macro-levels of analysis, these data have several well-documented weaknesses (Maltz, 1999), which include the inability to capture all information when multiple offenders and victims are involved (Loftin, Kindley, Norris, & Wiersema, 1987), missing homicide incidents from precincts that do not report (Fox, 2004; Langford, Isaac, & Kabat, 1998; Pridemore, 2005), and missing or incorrect values from reported cases either through error or lack of information (Flewelling, 2004; Fox, 2004; Maxfield, 1989; Williams & Flewelling, 1987). Research seeking to validate SHR data based on other similar data sources, such as the National Vital Statistics System (Loftin et al., 2008), or the internal records of local law enforcement agencies (Braga, Piehl, & Kennedy, 1999), has found important discrepancies across data sources. Some have suggested methods of addressing systematic biases in SHR data, including the replacement of missing values through statistical imputation techniques (Flewelling, 2004; Fox, 2004). Importantly, though, the SHR remains the most comprehensive and easily accessible source of official data available to researchers studying homicide (Maltz, 1999).

For this study, the population of homicide victims from Seattle, Washington, were identified from the SHR data available from the Seattle Police Department's website and local media sources. Other research has discussed the use of open-source information to create a sampling frame (Chermak, Freilich, Parkin, & Lynch, 2012; Freilich et al., 2014; LaFree & Dugan, 2007), so appropriate methods for this will not be discussed here. Twenty-two homicide victims were reported in the SHR for Seattle in 2011, all were coded as victims of murder and non-negligent manslaughter. After further investigation, two were discarded for this analysis. One incident was removed because it was a duplicate. The second homicide was removed, as no corroborating information to its occurrence could be found in any open-sources, official or otherwise.³ One victim killed in the final days of 2010 was included as she appears in the SHR for the month her body was discovered in 2011. Finally, there was only one homicide incident where two individuals were murdered.

Open-Source Data

Using the names of each victim, multiple sources were searched to collect as much open-source information as possible on each homicide victim. To

demonstrate the variation in responses across open-sources, they are disaggregated into eight distinct sources. The *Seattle Times* and *SeattlePI* represent the primary newspapers in the region and therefore their results were reported separately. While the *Seattle Times* is in print and online with a print circulation of approximately 250,000 in 2011 (Lamm, 2011), the *SeattlePI* was exclusively an online newspaper in 2011. Although there are multiple news aggregators, LexisNexis Academic was utilized as an open-source resource. At the time of the search, LexisNexis's news search accessed content from thousands of sources. Neither the *Seattle Times* nor the *SeattlePI* were in the LexisNexis search. Twenty local news blogs that focused on neighborhood events made up the Community Blogs category, 22 additional news publications that represented weekly and monthly newspapers and magazines made up the Other Print News category, and six local television news stations were included to create the Television News category.⁴

A public data aggregator, which collects information on individuals that is available through electronic means, such as residence history and criminal history, and an ancestry database that collects records related to births, deaths, marriages, divorces, obituaries, military service, and educational enrollment was used. This information is typically available through websites for a nominal monthly fee, while the information they provide varies based on many different factors. Disaggregating the data provided by these sources allows for an understanding of how much information a researcher might be likely to find from each type of source. All sources were searched over the period of a single week in late 2013. The goal was to ensure that there was a sufficient lag in time (approximately 2 years) between the time the homicide occurred and when it was open-source searched.

The information from each of these sources was published and accessible online, either through their websites or secondary databases like LexisNexis. As the purpose of the research is to present information on the potential of open-source materials as sources of victimization data, the original layout of the data in print sources, or the order of presentation of television news, was not relevant. Admittedly, valuable information is lost through the access of electronic data, as photographs and videos that could have been present in the original format are often not replicated due to copyright or archival limitations. Also, there is no expectation that these sources are mutually exclusive from each other, as blogs often insert information from larger news outlets into their posts, and television news stations are increasingly partnering with newspapers and other local media to link to or repost stories on their websites.

Coding

Information about the victims, homicide offenders, and the incidents were coded using the data found from each open-source type. As there was only one researcher collecting and coding data, inter-coder reliability was not measured. The variables chosen were based on their availability in the official data for comparative purposes or their potential usefulness in the study of victimization based on prior literature or current theory. For variables such as victim race or age, the information was either known or unknown. For other variables, such as marital status, the interpretation of the open-source is not as simple as being known or unknown, and the coding procedure could vary based on the research question. For example, if a researcher was curious as to whether a victim was married, they could code that there either is, or is not, evidence that the victim was married at the time of the incident based on the open-source materials. In this scenario, there is no missing data, as a person is either coded as having evidence of being married or not. However, if the research question entails knowing the person's exact relationship status (e.g., married, divorced, dating, single), coding and determining the appropriate way to deal with unknown and missing responses becomes more complicated. There is little reason for a newspaper to report that a teenager has a boyfriend/girlfriend unless they were also involved in the victimization event. An obituary would most likely not mention this, nor would it be available in ancestry records. However, a news outlet would be more likely to report that an elderly victim was married at the time they were killed, not to mention there would be an increased chance of identifying this relationship in an obituary, ancestry database, or in public records. With these circumstances, it would be difficult to code a person as definitively single as this type of information would not be directly stated in a source, although it could be inferred through the lack of mention. For this research, both methods are utilized so that the differences can be examined and the implications for theory testing and data validity discussed.

After collecting the documents from all of the open-source materials, attributes of each homicide victim were coded for each source. A targeted follow-up search was conducted to fill in missing values not found in the original open-sources. For the targeted searches, no restrictions were placed on the type of open-source materials that could be used, allowing for an understanding of when data are missing and what is the likelihood that new information will be found outside of news materials. The responses were also compared across the official data source and the data from the open-sources and only sources with the exact response for a variable were counted as a match. For example, if the victim was reported in the official data as being 33 years of

age at the time of death, but the open-source reported that they were 31 years of age, then this would not be considered a match. Data were then aggregated into five tables, allowing for the comparison between data originating from open-sources and the official data, as well as a comparison within types of open-source data.

Results

Comparing Between Official and Open-Source Data

Tables 1 and 2 present the comparison of victim characteristic response rates between the SHR and open-source data. The Victims column reports what percentage of the 20 homicide victims was found in each open-source. For the open-sources, the two major newspapers had at least one article that mentioned every victim, while the category with the lowest percentage of coverage was the Other Print News, which only had articles that mentioned 60% of the victims. The three characteristics of interest reported in Table 1—race, sex, and age—are reported as the percent of victims whose characteristic were found in the source (response), the percent of that characteristic discovered in both the official data and the open-source (overlap), and of those victimization characteristics that overlap both sources, the percent where the coded characteristic is the same in both the open-source and official data (agree).

Importantly, open-sources identified as many characteristics as official data, as reported in Table 1, although agreement between sources was not always 100%. To be more specific, the official data reported race of the victim for 100% of cases and the highest response rate for race found for a singular type of source was from Television News and the *SeattlePI* categories at 50% each, with a coding agreement with the official data of 90%. The open-source with the lowest response rate was LexisNexis, with only 10% response and overlap, and a 50% agreement. Combining all open-source categories, the race of 90% of the victims was identified and targeted follow-up searches for the missing values ended with a 100% response rate, 100% overlap, and a 95% agreement with the SHR for how race was coded across the 20 victims. Sex was identified 100% of the time in the SHR, as well as the *Seattle Times* and the *SeattlePI*, with 100% agreement in coding choices. The lowest response rate for sex came from the Ancestry Database and Public Data Aggregator sources, in which sex was only identified for 40% of the victims. No targeted searches were necessary for sex and age as the open-sources searched identified age for 100% of the victims. For age, the *Seattle Times* once again reported on 100% of the characteristics, with 95% agreement with the

Table 1. Comparison of Victim Characteristics Between Official Data And Open-Sources ($N = 20$).

	Victims			Race			Sex			Age		
	Response ^a (%)	Response (%)	Overlap (%)	Agree (%)	Response (%)	Overlap (%)	Agree (%)	Response (%)	Overlap (%)	Agree (%)	Response (%)	Overlap (%)
Official data	100	100	—	—	100	—	—	100	—	—	100	—
All open-sources ^b	100	90	90	89	100	100	100	100	100	100	100	100
Seattle Times	100	15	15	67	100	100	100	100	100	100	100	100
SeattlePI	100	50	10	90	100	100	100	100	85	85	85	85
LexisNexis	85	10	10	50	80	80	100	70	70	70	70	70
Community Blogs	65	40	40	88	60	60	100	55	55	100	55	100
Other Print News	60	30	30	83	55	55	100	50	50	100	50	100
Television News	85	50	50	90	85	85	100	75	75	100	75	100
Public Data Aggregator	85	35	35	71	40	40	100	85	85	94	85	85
Ancestry Database	65	25	20	100	40	40	100	60	60	92	60	60
Targeted search ^c	—	100	100	95	—	—	—	—	—	—	—	—

^a"Response" is the percent of victims whose characteristic were found in the source, "Overlap" is the percent of that characteristic discovered in both the official data and the open-source, and "Agree" measures for those victimization characteristics that overlap both sources, the percent where the coded characteristic is the same in both the open-source and the official data.

^bThe All open-sources category represents the aggregate of responses for each open-source into one category.

^cA targeted, follow-up search was conducted when the pre-selected open-sources could not identify the characteristic under study for each victim.

Table 2. Comparison of Incident Characteristics Between Official Data and Open-Sources (N = 20).

	Multiple Offenders			Relationship			Weapon			Circumstance			Inter-Racial		
	Response Overlap			Agree			Response Overlap			Agree			Response Overlap		
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
Official data	60	—	—	55	—	—	85	—	—	80	—	—	60	—	—
All open-sources	80	60	100	80	55	100	85	85	75	80	75	—	45	30	83
Seattle Times	80	60	100	70	45	100	85	85	70	75	70	—	15	15	67
SeattlePI	80	60	100	70	50	100	85	85	80	80	75	—	10	10	100
LexisNexis	50	40	100	30	20	100	65	65	55	55	55	—	10	10	50
Community Blogs	40	30	100	30	25	100	40	40	40	40	40	—	20	20	75
Other Print News	45	35	100	45	30	100	45	45	35	40	35	—	20	15	50
Television News	65	45	100	70	50	100	70	70	60	65	60	—	30	20	75
Public Data Aggregator	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Ancestry Database	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Targeted search	—	—	—	—	—	—	—	—	—	—	—	—	75	55	91

SHR (measured as being within 1 year of each other), the same percentages for all sources combined.

For Table 2, which compares incident-level characteristics, the open-sources identified as much, or more, of the characteristics being studied as the official data did. Part of this can be explained by the fact that incident-level characteristics have a greater chance of having missing values from the official data based on whether a suspect is ever identified, at what point during the investigation the police department fills out the SHR report, and whether they update the original report based on the final outcome of the investigation. The first variable, whether there were multiple offenders, had a response rate of 60% for whether this was known and all open-sources were able to identify 80% of the time whether multiple offenders were involved in an incident with 100% agreement for the 60% of overlapping cases. It is important to note that in the 20% of cases where the response is not known, this is not an artifact of the open-source reporting, but reflects the reality of police work. In all these cases, no arrests were made and although open-source information may have identified a person of interest, the police never officially acknowledged these individuals as suspects. This same issue lowers the response rate for the victim–offender relationship and whether the victimization was inter- or intra-racial in nature.

The relationship between the victim and primary offender was only reported as known 55% of the time in the official data, whereas open-sources identified the relationship 80% of the time. This is similar to the overall pattern of unknown or missing victim–offender relationships found in the SHR, which was approximately 41% between 1980 and 2012 (Puzzanchera, Chamberlin, & Kang, 2014). The *Seattle Times*, *SeattlePI*, and Television News had the highest response rates, while LexisNexis and the Community Blogs had the lowest. All open-sources had a 100% agreement with the official data for the cases that overlapped. Interestingly, for the five relationships that the official data did not have that were found in open-sources, there was 100% agreement between the open-sources. The weapon used to kill the victim had an 85% response rate in the official data, *Seattle Times*, and *SeattlePI*. Agreement levels for the overlapping cases ranged from 75% to 88%. The circumstances of the victimization were reported 80% of the time in the SHR and discovered 80% of the time across all open-sources. No agreement is given for the circumstance variable, as the responses given by the SHR are not mutually exclusive and can be vague (e.g., “other arguments”).

Finally, the inter-racial variable (whether victim and offender were of a different race), could be determined in 60% of the SHR cases, but only 45% of the open-sources, with percent agreement ranging from 50% to 100%. Part of this low response for open-sources is based on the search protocol, which

only used the victims' names to search the open-sources and systematically excluded any open-source documents that listed information about the suspect without mentioning the victim's name. Due to this, the inter-racial variable was the only variable reported in Table 2 that had improved responses from the targeted follow-up searches. Most important when considering the utility of open-source data is that when searches were conducted with no limits as to where and how information was searched, the response rate increased to 75%, 55% overlap, and 91% agreement.

Comparing Within Open-Source Data

The characteristics in Tables 3 and 4 report response rates across victim and incident characteristics not available in official data. These results highlight the ability of open-source data to augment official data through the identification and coding of variables that might be of theoretical interest, but are not systematically reported in official data. The first characteristic in Table 3, victim's residence, that is the town where the victim lived when killed, was identified in 95% of cases. The Other Print News category had the lowest response rate (30%), while the Public Data Aggregator identified the current residence of the victim 80% of the time. Forty-five percent of victims were identified as having a criminal history and once again the Public Data Aggregator had the best response rate. In 65% of the cases, the offender occupation, or lack thereof, was known, with the *Seattle Times* having the highest response rate of 60%.

To offer an example in the variation of the response rates based on how one codes open-source information, the individual's relationship status was coded in two different ways. First, marital status recorded the percentage of victims whom anything about their relationship status was known, whether married, divorced, dating, single, or in a long-term relationship. The second method was only to identify the percent of individuals that was reported as married at the time of their death. With targeted follow-up searches, the marital status was identified for 70% of victims, while evidence of being married was found with only 15% of victims. The "evidence of" coding rule was also used for whether the victims had children, with open-sources identifying 35% of victims having evidence of children, with that rate increasing to 45% with targeted searches. Finally, educational level was found in 30% of open-sources, once again increasing by 10% when considering targeted searches.

Table 4 presents the results of the incident-level characteristics not found in the official data. First, the most important piece of incident-level information related to theory testing, specific addresses, was identified for 100% of the victims in open-sources. The importance of this cannot be overstated. The

Table 3. Comparison of Victim Characteristics Within Open-Sources ($N = 20$).

	Residence (%)	Criminal History (%)	Occupation (%)	Marital Status (%)	Married (%)	Children (%)	Education (%)
All open-sources	95	45	65	65	15	35	30
Seattle Times	70	20	60	55	15	30	15
SeattlePI	65	25	45	35	10	25	15
LexisNexis	50	5	25	15	0	15	15
Community Blogs	40	15	20	25	5	15	5
Other Print News	35	5	25	30	10	15	15
Television News	60	15	20	30	5	20	20
Public Data Aggregator	80	45	—	—	—	—	—
Ancestry Database	—	—	—	30	5	—	10
Targeted searches	100	—	—	70	—	45	40

Table 4. Comparison of Incident Characteristics Within Open-Sources (N = 20).

	Address (%)	Location (%)	Date/Day (%)	Time (%)	Activity (%)	Victim Precipitation (%)
All open-sources	100	100	85	65	95	10
Seattle Times	70	95	85	50	75	10
SeattlePI	95	95	85	60	75	10
LexisNexis	50	70	65	35	50	5
Community Blogs	55	55	45	25	40	5
Other Print News	30	55	45	15	45	5
Television News	65	80	70	40	65	10
Public Data Aggregator	—	—	—	—	—	—
Ancestry Database	—	—	—	—	—	—
Targeted searches	—	—	—	70	—	—

SHR only ties victims to the county where they were murdered and the department that reported the crime. To have block-level data for a population or sample being studied allows for the merging of macro-level Census data, which in turn increases a researcher's ability to test theoretical constructs related to victimization. In addition, the data can be used for analyses that control for spatial clustering. Although individual- and situational-level characteristics vary in their response rates, to have the ability to geocode data is a very powerful research tool. It also provides opportunities for multi-level analyses, which experts have identified as lacking in the field of victimology (Lauritsen, 2010; Lauritsen & Laub, 2007). The *SeattlePI* had the highest response rate for address, reporting this piece of information in 95% of the cases, while the Other Print News had the lowest rate of 30%. In addition, the type of location (e.g., residence, business, park) also had a 100% response rate across all open-sources.

The exact date of the incident, which would also allow for the calculation of the day of the week, was only reported across all open-sources for 85% of the cases. However, with the missing 15%, the date of the incident will most likely never be known, as a several days to several weeks gap occurred between the last date the victim was seen alive and the day their body was found. In 70% of the cases (65% without targeted searches), the time of victimization was known within 15 min. The Other Print News category had the worst response rate (15%), while the *SeattlePI* had the highest (60%). If a researcher coded the victimization time in larger blocks (e.g., 00:00-05:59; 06:00-11:59), they would be able to increase the response rate by an additional 15%. The activity in which the individual was engaged in at the time of the victimization was also coded (home; work; leisure), with a 95% response rate across all sources. Finally, in 10% of the cases, evidence of victim precipitation was found, with very little variation across the open-sources. Victim precipitation was operationalized as the victim engaging in threatening behavior immediately prior to their victimization in which the offender had cause to fear for their safety. For this variable, except in the 20% of victimization events where the circumstances are not known, nor probably will ever be known, there is strong evidence that the 10% of victims who engaged in victim precipitation is representative of an affirmative response for this characteristic in the population.

Finally, Table 5 is a standardized agreement score for each source comparing whether the responses to that variable agree with the responses in other sources. This agreement measure was then standardized so that each agreement score could be compared across variables. It is important, however, to acknowledge that although open-sources might have a high level of agreement between each other, this does not necessarily mean that the response on which they agree is the

correct one. For example, if an Associated Press news release states that a homicide victim is 35 years of age, but they are really 25 years of age, this incorrect information could be picked up from multiple open-sources and repeated. In this scenario, although all open-sources might consistently report that the victim was 35 because each collected information on the homicide from the Associated Press report, they would all be consistently wrong. The first column for each variable includes the official data, while the second column is the agreement score without accounting for the relationship of these variables to the official data. A positive score indicates that the source agreed with the other sources more often than the average score, whereas a negative score indicates that the source agreed with the others less often than the average.

When examining the results, we see that official data had levels of agreement that were higher than average in all categories except for the multiple offenders category. However, only in the race category, did the official data agree with the other sources more often than one of the open-sources. The official data also tied with the open-source data for having the highest standardized agreement score for the sex variable. When comparing across variables, we see that the *Seattle Times* had lower levels of agreement in comparison with other sources when measuring race, but some of the highest levels of agreement with age, sex, weapons, and multiple offenders. Once again, we can see the impact of content delivery and editorial policies across sources and the impact it has on the ability to collect certain variables in comparison with other data sources.

Discussion

This case study found that for each variable open-sources reliably captured as much (or more) of the information presented in official data. Agreement between responses was not 100% for every variable, which means that either miscoding occurred in the official data or misrepresentation of the facts of the homicide occurred in the open-sources. To this point, the use of open-sources allowed for the entry of information into the data set that may not have been available when the official data were collected, or even months later when law enforcement agencies had the opportunity to update information. Importantly, variables not available in official data that are necessary for theory testing were identified in open-sources. In addition, materials collected from open-sources could be used by researchers to recode variables found in official data, such as the circumstance variable, to have a better understanding of the situational characteristics of the victimization event. Open-sources can also be used to validate the information coded in official data, using multiple sources to triangulate the responses of interest.

As far as the variation within open-sources, the comparisons of certain characteristics demonstrated that the major print and online newspapers in the Seattle area provided the highest response rates for each variable. The Television News category also had relatively high response rates. The Community Blogs, which reported almost exclusively on events within a constrained geographic area and often linked to the larger media sources, and the Other Print News sources, had the lowest response rates. This is understandable, as many of these sources concentrated on reporting news to a specific demographic, whether limited through geography or interest. If a homicide victim was killed in a community without a representative blog, their victimization would most likely not be mentioned in Community Blogs. Similarly, if the victim's personal characteristics did not fit into the geography or interests of the demographics of those reading articles found within the Other Print News category, then the likelihood of their murders being reported also decreased. Editorial guidelines also seem to affect certain variables, such as race and whether the incident was inter-racial. Many newspapers restrict how and when race and ethnicity can be mentioned, especially with suspects and persons of interest in ongoing criminal investigations. In addition, there is little or no reason to mention the race or ethnicity of a victim unless it is related to their victimization. In this circumstance, if a photograph of the victim or suspect is not provided with the open-source material, race can be especially difficult to ascertain. Even when a photograph is presented, there will be error introduced into the measurement as coders would be making a "best guess" as to the individual's racial and ethnic background. More informal open-sources, such as social media sites, suffer similar structural inequities based on who has access to resources to post information on the web and the variation in the size of online social networks. One can imagine a scenario where an 18-year-old homicide victim with a social media site and hundreds of "friends" would be more likely to have relevant (and much more irrelevant) information posted on their social media page than someone with fewer friends, not to mention those who have no social media presence at all. When discussing the temporal distribution of open-source data, even in Seattle, the sources utilized for this case study are not available across all time periods. Community blogs are a relatively recent innovation in news media, while electronic access to television news stories through a channel's website also does not span very far into the past. Archival newspaper data are somewhat more accessible, for example the *Seattle Times* archive is available from 1966 to the present and the *Seattle Post-Intelligencer* or *SeattlePI* are available from 1921 to the present as either print or online archives.

Other characteristics that may be of theoretical interest, such as criminal history, marital status, and victim precipitation had varying success in the

open-sources. Forty-five percent of the victims had evidence of a criminal history discovered mostly through the Public Data Aggregator category. It is difficult to determine how reasonable this number is, as open-sources would rarely, if ever, state that the victim had no criminal history. This absence of information contrary to the characteristic under study is an ongoing issue with open-source data. In most cases, only information affirming the characteristic will ever be found.

Similar issues with the measurement of marital status, which were already briefly discussed, should be reiterated. Homicide victimization risk has been shown to vary based on the marital status of a person, as well as the subtype of homicide, for example, domestic violence versus stranger homicides (Breault & Kposowa, 1997; Kposowa & Breault, 2005; Sampson & Wooldredge, 1987; Stack, 2002). To study the impact of this characteristic on homicide risk, researchers must have a valid and reliable way to measure the construct. Using open-sources to collect information on variables such as marital status does allow for a systematic coding schema that could potentially capture the differences between being single, living with a partner long term, or even living with a same-sex partner in a state where marriage and civil-unions are not available. Whereas an official data source might list a same-sex couple as single based on legal technicalities, a researcher might code them as married, or the equivalent, being more concerned with the impact of a monogamous relationship on a dependent variable, such as victimization risk. Another example where the measurement of a construct might not be available in official data, but is important from a theoretical perspective and possibly available in open-source data, would be the victim-offender relationship of ex-boyfriend or ex-girlfriend. Finally, although victim precipitation had a low response rate, it is reasonable to assume that, except for the victimization events where the circumstances were not known, this 10% captures the totality of victims whose actions may have played a role in the homicide incident.

For the data collected through the open-sources, as is with all data sources, there are weaknesses, although, on average, no worse than official data and somewhat better depending on the variable, missing values are a concern in open-sources. This case study was conducted under ideal circumstances of data availability, purposefully conducted in a jurisdiction with one of the lowest homicide rates for an American city of its size. It is not surprising that media cover most homicide events, as they are rare forms of serious violence. Jurisdictions with higher homicide rates will not report all victimizations in open-sources, nor is there any reason to believe that all victims and their characteristics have an equal chance of being reported. In fact, there is a strong argument to make that missing values in open-source data are not

missing at random. Research on newsworthiness has shown that in certain populations, victimization characteristics increase or decrease the likelihood that the news will report on the event. For example, homicides including minority victims and individuals from communities of a lower socio-economic status are not as likely to be covered by the media (see, for example, Chermak, 1995; Gruenewald, Pizarro, & Chermak, 2009; Schildkraut & Donley, 2012). Future open-source research will need to account for biases in victimization data stemming from such systematically missing information.

A similar point of weakness in open-source data is that it is not uniformly available across all temporal and geographic locations. The number of newspapers, community blogs, and television news channels in this study is unique to Seattle. Studies in other regions will have varying open-sources from which to draw. Access, or lack thereof, to these types of sources is not equally distributed across the U.S. Rural areas with low population densities will have fewer media outlets collecting data and reporting on homicide victimization than more population dense suburban and urban areas. However, they will also have fewer homicides on which to report. Other types of open-source data that come from government sources, such as court records, Department of Corrections reports, or law enforcement statistics are collected and distributed by local and state jurisdictions with no national standards or protocols as to whether they are easily accessible electronically.

Another potential obstacle to using open-source data is that the gathering, archiving, and coding of open-source data is time-consuming. A conservative estimation for the data collection and coding process for the 20 homicide victims used for this case study would be a minimum of 5 hr per victim. When researchers are building databases on entire populations or need a large sample for statistical power, 5 hr per case could quickly become overwhelming. However, original data collection, whether quantitative or qualitative in nature, has always taken time and effort and the benefits of one collecting their own data, or augmenting official data, should far outweigh the costs of the time needed to do it properly.

Understanding the validity and reliability of data is also important in social science research. Open-sources present a dilemma on this issue, as systematic studies of the reliability and validity of using multiple, publically available resources have not been conducted. In fact, one can imagine the difficulty in attempting to study something as simple as response agreement across thousands of potential sources. What happens when a newspaper and community blog offer competing information? What happens when the SHR and television news stations present contradictory information on a victim's race? Some efforts have already been made to rank sources by their potential reliability. Terrorism scholar Marc Sageman (2004), for instance, has guided

open-source researchers to grant greater weight to more “trusted” sources, favoring documents stemming from court proceedings subject to cross examination, followed by other court reports, then corroborated information from witnesses closely linked to the case, uncorroborated witness statements, and finally, secondhand information. The open-source data from the Extremist Crime Database used in this study relied on Sageman’s (2004) guiding principles for increasing reliability across sources, considering appellate court proceedings as the most reliable of information and advocacy groups and Internet forums (e.g., blogs, websites) as some of the least reliable of sources (Freilich et al., 2014). When the Extremist Crime Database faced conflicting information for a particular crime variable, those sources considered relatively more reliable were utilized.

It is also important to think about how a researcher can be confident that open-source measures are valid. In a simple attempt to look at Seattle homicide victims in 2011, the official data’s population frame included two homicide victims (9% of the population) that, through open-source searches, were later determined not to be verifiable homicide victims. Due to the relatively low *N*, and assistance from law enforcement, we were able to validate our data. However, it is not difficult to imagine that through the systematic searching of thousands of cases that non-victims would be accidentally included or actual homicide victims would be excluded, especially as the temporal and geographic scope of a study increases. Future research should focus on larger *N* jurisdictions, where official homicide data are available from law enforcement, medical examiner’s offices, or even prosecutors to the assess validity and reliability of the open-source. Through these comparisons, a more rigorous test of the capabilities of open-source data can be conducted. Also, advancements in big data analyses and machine learning offer an interdisciplinary pathway forward for identifying and coding homicide incidents on a national scale. This type of data collection, analysis, and presentation is already occurring in other fields, such as terrorism and public health (National Consortium for the Study of Terrorism and Responses to Terrorism, 2014; Signorini, Segre, & Polgreen, 2011).

Also, open-source data are ever changing. Materials collected at Time 1 will not be the same as those collected at Time 2. When is the appropriate time to access open-source data so that it offers the most complete picture of the topic being researched? How can data be archived and shared so that researchers can support their findings and allow others to conduct follow-up and confirmatory studies using the materials? This study collected data approximately 18 to 30 months after the homicide victimizations occurred and was not able to assess how the validity and reliability of data may have

changed at different periods after the victimization occurred. A research design could be developed, however, that systematically researched a homicide case at specific intervals. For example, one could search for information the day a homicide occurs, a week after it occurs, 1 month after it occurs, and 1 year after it occurs. One could then test the hypothesis about the dynamic nature of open-source data and how changes in data affect its validity and reliability overtime. Although the aforementioned issues need further research and discussion, they are not insurmountable nor do they irrevocably damage open-source research, especially when one considers the strengths of the methodology.

Conclusion

The combination of official data and open-sources offer the possibility of another powerful research tool in studying homicide victimization. Each source could be used to confirm the other and identify potentially missing victims or values, while validating the ones that they do have. This methodology has already been used with the SHR and open-source data for subsets of homicide victimization and errors with both open-sources and official data were identified (see, for example, Huff-Corzine et al., 2014; Overberg, Upton, & Hoyer, 2013). In addition, many of the SHR's challenges, such as non-reporting, non-dynamic data, and missing values could be addressed through data augmentation strategies based on open-sources.

The following steps in the use of open-sources to study homicide victimization, however, should include more methodological studies and analyses of small, specialized populations. It is through comparison with official data, whether from institutions, such as the local police department or the medical examiner's office, that the validity and reliability of open-sources can be quantitatively assessed. A study that uses multiple jurisdictions and time periods would be ideal, presenting variation in open-sources in varying temporal and geographic locations. Once the sample size becomes large enough, researchers can develop theoretically relevant statistical models for data sets created from official data sources or open-sources. If the data sources capture the same information, the results of these models should not change from data set to data set. However, if the models and their coefficients do change significantly between data sets, than this provides evidence that the type of data used can affect the results of these analyses.

In conclusion, as has already been done in other social science subfields, open-sources have the potential to be used to study populations of homicide victims. Researchers interested in homicide victimization could not only use

official data to identify cases but could also conduct open-source searches to draw on the work done by journalists, advocacy groups, and other organizations that post their data online, and through this process further develop their population and code variables of theoretical interest. As access to open-source materials continues to grow at an exponential rate, social scientists will only continue to use this data more frequently to study populations and answer research questions that had historically been limited to exploration through official data. The key to harnessing this wealth of information lies in studying how, and when, it is most effective. In addition, it is important that it is done in a way that furthers an extant body of research, allowing for others to reproduce and confirm such studies. Hopefully, this study is a step forward in utilizing this underutilized resource, adapting it to inform theory testing and development, whether related to homicide victimization, or any other topic in the social sciences.

Appendix

Detailed List of Sources Searched.

Category	Source	URL
Community Blogs	<i>Aurora Seattle</i>	http://www.auroraseattle.com/
	<i>Belltown People</i>	http://belltownpeople.com/
	<i>Capitol Hill Seattle</i>	http://www.capitolhillseattle.com/
	<i>Central District News</i>	http://www.centraldistrictnews.com/
	<i>Eastlake Ave</i>	http://eastlakeave.com/
	<i>Fremont Universe</i>	http://www.fremontuniverse.com/
	<i>Magnolia Voice</i>	http://www.magnoliavoices.com/
	<i>Maple Leaf Life</i>	http://www.mapleleaflife.com/
	<i>Montlaker</i>	http://montlaker.com/
	<i>My Ballard</i>	http://www.myballard.com/
	<i>My Wallingford</i>	http://www.mywallingford.com/
	<i>PhinneyWood</i>	http://www.phinneywood.com/
	<i>Queen Anne View</i>	http://www.queenanneview.com/
	<i>Ravenna Blog</i>	http://www.ravennablog.com/
	<i>Southend Seattle</i>	http://southendseattle.com/
	<i>The South Park News</i>	http://thesouthparknews.com/
	<i>U District Daily</i>	http://www.udistrictdaily.com/
	<i>Wedgewood View</i>	http://www.wedgewoodview.com/
	<i>West Seattle Blog</i>	http://westseattleblog.com/
	<i>White Center Now</i>	http://whitecenternow.com/

(continued)

Appendix. (Continued)

Category	Source	URL
Other Print News	<i>Ballard News-Tribune</i>	http://www.ballardnewstribune.com/
	<i>Belltown Messenger</i>	http://www.belltown.org/messenger/archives.html
	<i>City Living Seattle</i>	http://www.southseattlebeacon.com/
	<i>City Living Seattle</i>	http://www.northseattleherald-outlook.com/
	<i>Daily Journal of Commerce</i>	http://www.djc.com/
	<i>El Siete Dias</i>	http://elsietedias.com/
	<i>International Examiner</i>	http://www.iexaminer.org/
	<i>Madison Park Times</i>	http://www.madisonparktimes.com/
	<i>MyNorthwest.com</i>	http://mynorthwest.com/
	<i>Northwest Asian Weekly</i>	http://www.nwasianweekly.com/
	<i>Puget Sound Business Journal</i>	http://www.bizjournals.com/seattle/
	<i>Queen Anne & Magnolia News</i>	http://www.queenannenews.com/
	<i>Queen Anne & Magnolia News</i>	http://www.magnolianews.net/
	<i>Seattle Gay News</i>	http://www.sgn.org/
	<i>Seattle Press</i>	http://www.seattlepress.com/
	<i>Seattle Weekly</i>	http://seattleweekly.com/
	<i>The Capitol Hill Times</i>	http://www.capitolhilltimes.com/
	<i>The North American Post</i>	http://www.napost.com/
	<i>The Seattle Medium</i>	http://seattlemedium.com/
	<i>The Skanner</i>	http://www.theskanner.com/
	<i>The Stranger</i>	http://www.thestranger.com/
	<i>West Seattle Herald</i>	http://www.westseattleherald.com/
Television News	<i>CBS Seattle</i>	http://seattle.cbslocal.com/category/news/
	<i>KCTS 9</i>	http://kcts9.org/
	<i>KING 5</i>	http://www.king5.com/
	<i>KIRO TV</i>	http://www.kirotv.com/
	<i>KOMO News</i>	http://www.komonews.com/
	<i>Q13 FOX</i>	http://q13fox.com/

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Any official data sets or information classified, or protected by local, state, or federal statutes, are not considered open-source data in this study. While some may consider all information collected by government organizations as open-source data, the difficulty in obtaining them in large quantities across multiple jurisdictions is cumbersome, labor-intensive, and thus, prohibitive for most social science researchers. In contrast, we suggest that open-source information that can be located on the web through search engines or specialized websites that are easily collected and archived has the greatest potential for assisting in criminological research.
2. A complete picture of the use of open-sources in the field of criminology is difficult to ascertain as researchers may use certain types of open-source data, such as media accounts or advocacy group reports, to develop or augment their data sets but never label or discuss this as open-source information. A survey of research articles published in *Homicide Studies* between 2008 and 2012 identified nine articles that used in full, or in part, open-sources for data. This accounts for approximately 7% of the published articles.
3. Although the duplication was an error and its removal should not be controversial, the removal of the non-corroborated victim might. Prior research demonstrates that not all homicide victims will be reported in the media and that certain incident- and victim-level characteristics increase or decrease the news-worthiness of the event. Arguably, this missing victim was not reported to the Supplementary Homicide Report (SHR) in error, but instead was never reported in open-sources. After reviewing all open-sources, however, the authors determined that the probability that the additional victim was not reported in the open-sources was low, and that a coding error had most likely occurred in the SHR. This was based on the police department's own statistics presented on their website, which did not match the SHR, and an article in the local newspaper that listed each homicide and corresponding details, comparing their results to official statistics (McNerthney, 2012). In addition, we explored the issue with the senior research analyst in the Seattle Police Department's Research Section who also determined that it was most likely a coding error (L. Atherly, personal communication, December 4, 2014).
4. See the appendix for a more detailed list of sources searched.

References

- Allely, C. S., Minnis, H., Thompson, L., Wilson, P., & Gillberg, C. (2014). Neurodevelopmental and psychosocial risk factors in serial killers and mass murderers. *Aggression and Violent Behavior, 19*, 288-301.

- Barker, A. D. (2011). Improvised explosive devices in southern Afghanistan and western Pakistan, 2002-2009. *Studies in Conflict & Terrorism*, 34, 600-620.
- Beard, V., Hunter, S., Kern, L., & Kiley, B. (2014). Death-related crime: Applying Bryant's conceptual paradigm of thanatological crime to serial homicide. *Deviant Behavior*, 35, 959-972.
- Braga, A. A., Piehl, A. M., & Kennedy, D. M. (1999). Youth homicide in Boston: An assessment of Supplementary Homicide Report data. *Homicide Studies*, 3, 277-299.
- Breault, K. D., & Kposowa, A. J. (1997). The effects of marital status on adult female homicides in the United States. *Journal of Quantitative Criminology*, 13, 217-230.
- Chermak, S. M. (1995). *Victims in the news: Crime in American news media*. Boulder, CO: Westview Press.
- Chermak, S. M., Freilich, J. D., Parkin, W. S., & Lynch, J. P. (2012). American terrorism and extremist crime data sources and selectivity bias: An investigation focusing on homicide events committed by far-rights extremists. *Journal of Quantitative Criminology*, 28, 191-218.
- Chermak, S. M., Freilich, J. D., & Suttmoeller, M. (2013). The organizational dynamics of far-right hate groups in the United States: Comparing violent to nonviolent organizations. *Studies in Conflict & Terrorism*, 36, 193-218.
- Chermak, S. M., & Gruenewald, J. (2015). Laying the foundation for the criminological examination of right-wing, left-wing, and Al Qaeda-inspired extremism in the United States. *Terrorism and Political Violence*, 27, 133-159.
- Chermak, S. M., & Gruenewald, J. A. (2006). The media's coverage of domestic terrorism. *Justice Quarterly*, 23, 428-461.
- Coleman, C., & Moynihan, J. (1996). *Understanding crime data: Haunted by the dark figure*. Buckingham, UK: Open University Press.
- Cothren, J., Smith, B. L., Roberts, P., & Damphousse, K. R. (2008). Geospatial and temporal patterns of preparatory conduct among American terrorists. *International Journal of Comparative and Applied Criminal Justice*, 32, 23-41.
- Dugan, L., LaFree, G., & Piquero, A. R. (2005). Testing a rational choice model of airline hijackings. *Criminology*, 43, 1031-1065.
- Elliot, D., & Ageton, S. (1980). Reconciling race and class differences in self-reported and official estimates of delinquency. *American Sociological Review*, 45, 95-110.
- Flewelling, R. L. (2004). A nonparametric imputation approach for dealing with missing variables in SHR data. *Homicide Studies*, 8, 255-266.
- Fox, J. A. (2004). Missing data problems in the SHR: Imputing offender and relationship characteristics. *Homicide Studies*, 8, 214-254.
- Fox, J. A., & Levin, J. (2005). *Extreme killing: Understand serial and mass murder*. Thousand Oaks, CA: Sage.
- Freilich, J. D., Chermak, S. M., Belli, R., Gruenewald, J. A., & Parkin, W. S. (2014). Introducing the United States Extremist Crime Database (ECDB). *Terrorism and Political Violence*, 26, 372-384.
- Graber, D. (1980). *Crime news and the public*. Chicago, IL: University of Chicago Press.

- Gruenewald, J. (2011). A comparative examination of far-right extremist homicide events. *Homicide Studies, 15*, 177-203.
- Gruenewald, J. (2012). Are anti-LGBT homicides in the United States unique? *Journal of Interpersonal Violence, 27*, 3601-3623.
- Gruenewald, J. (2013). Using open-source data to study bias homicide against homeless persons. *International Journal of Criminology and Sociology, 2*, 538-549.
- Gruenewald, J., Chermak, S. M., & Freilich, J. D. (2012). Distinguishing "loner" attacks from other domestic extremist violence. *Criminology & Public Policy, 12*, 65-91.
- Gruenewald, J., & Kelley, K. (2014). Typifying anti-LGBT homicide events by mode of victim selection. *Criminal Justice and Behavior, 41*, 1130-1142.
- Gruenewald, J., Pizarro, J., & Chermak, S. M. (2009). Race, gender, and the newsworthiness of homicide incidents. *Journal of Criminal Justice, 37*, 262-272.
- Gruenewald, J., & Pridemore, W. A. (2012). A comparison of ideologically motivated homicides from the new Extremist Crime Database and homicides from the Supplementary Homicide Reports using multiple imputation by chained equations to handle missing values. *Journal of Quantitative Criminology, 28*, 141-162.
- Hagan, J. (1989). Why is there so little criminal justice theory? Neglected macro- and micro-level links between organization and power. *Journal of Research in Crime & Delinquency, 26*, 116-135.
- Heinonen, J. A., & Wilson, J. M. (2012). Product counterfeiting at the state level: An empirical examination of Michigan-related incidents. *International Journal of Comparative and Applied Criminal Justice, 36*, 273-290.
- Hindelang, M., Hirschi, T., & Weis, J. (1979). Correlates of delinquency: The illusion of discrepancy between self-report and official measures. *American Sociological Review, 44*, 995-1014.
- Horgan, J., & Morrison, J. F. (2011). Here to stay: The rising threat of violent dissident Republicanism in Northern Ireland. *Terrorism and Political Violence, 23*, 642-669.
- Huff-Corzine, L., McCutcheon, J. C., Corzine, J., Jarvis, J. P., Tetzlaff-Bemiller, M. J., Weller, M., & Landon, M. (2014). Shooting for accuracy: Comparing data sources on mass murder. *Homicide Studies, 18*, 105-124.
- Kangaspunta, K. (2003). Mapping the inhuman trade: Preliminary findings of the database on trafficking in human beings. *Forum on Crime and Society, 3*, 81-103.
- Kankaanranta, T., & Muttilainen, V. (2010). Economic crimes in the construction industry: Case of Finland. *Economic Crimes, 17*, 417-429.
- Kelley, K., & Gruenewald, J. (2015). Accomplishing masculinity through anti-LGBT homicide. *Men and Masculinities, 18*, 3-29. doi:10.1177/1097184X14551204
- Kirk, D. (2006). Examining the divergence across self-report and official data sources on inferences about the adolescent life-course of crime. *Journal of Quantitative Criminology, 22*, 107-129.
- Kposowa, A., & Breault, K. (2005). Race, ethnicity and the risk of homicide victimization in the United States: Results from the NHIS-MCD. *Journal of Ethnicity in Criminal Justice, 3*, 21-46.

- LaFree, G., & Dugan, L. (2007). Introducing the Global Terrorism Database. *Terrorism and Political Violence*, 19, 181-204.
- LaFree, G., Dugan, L., & Korte, R. (2009). The impact of British counterterrorist strategies on political violence in Northern Ireland: Comparing deterrence and backlash models. *Criminology*, 47, 17-45.
- LaFree, G., Dugan, L., Xie, M., & Singh, P. (2012). Spatial and temporal patterns of terrorist attacks by ETA 1970 to 2007. *Journal of Quantitative Criminology*, 28, 7-29.
- Lamm, G. (2011, May). *Seattle Times* loses print readers, gains online audience. *Puget Sound Business Journal*. Retrieved from <http://www.bizjournals.com/seattle/blog/techflash/2011/05/seattle-times-loses-print-readers.html>
- Langford, L., Isaac, N., & Kabat, S. (1998). Homicides related to intimate partner violence in Massachusetts: Examining case ascertainment and validity of the SHR. *Homicide Studies*, 2, 353-377.
- Lauritsen, J. L. (2010). Advances and challenges in empirical studies of victimization. *Journal of Quantitative Criminology*, 26, 501-508.
- Lauritsen, J. L., & Laub, J. H. (2007). Understanding the link between victimization and offending: New reflections on an old idea. In M. Hough & M. Maxfield (Eds.), *Surveying crime in the 21st century* (pp. 55-75). Monsey, NY: Criminal Justice Press.
- Loflin, C., Kindley, K., Norris, S. L., & Wiersema, B. (1987). An attribute approach to relationships between offenders and victims in homicide. *The Journal of Criminal Law and Criminology*, 78, 259-271.
- Loflin, C., McDowall, D., & Fetzer, M. D. (2008). A comparison of SHR and vital statistics homicides estimates for U.S. cities. *Journal of Contemporary Criminal Justice*, 24, 4-17.
- Maltz, M. D. (1999). *Bridging gaps in police crime data: A discussion paper from the BJS fellows program*. Washington, DC: Bureau of Justice Statistics. Retrieved from <http://www.bjs.gov/content/pub/pdf/bgpdc.pdf>
- Maxfield, M. G. (1989). Circumstances in supplementary homicide reports: Variety and validity. *Criminology*, 27, 671-695.
- Maxfield, M. G., Weiler, B., & Widom, C. (2000). Comparing self-reports and official records of arrests. *Journal of Quantitative Criminology*, 16, 87-110.
- McNerthney, C. (2012, January). List of 2011 homicides in Seattle. *SeattlePI.com*. Retrieved from <http://www.seattlepi.com/local/article/List-of-2011-homicides-in-Seattle-2433596.php>
- Miller, E. (2012). Patterns of onset and decline among terrorist organizations. *Journal of Quantitative Criminology*, 28, 77-101.
- National Consortium for the Study of Terrorism and Responses to Terrorism. (2014). *Global Terrorism Database—Codebook: Inclusion criteria and variables*. Retrieved from <http://www.start.umd.edu/gtd/downloads/Codebook.pdf>
- Overberg, P., Upton, J., & Hoyer, M. (2013, December). USA TODAY research reveals flaws in mass-killing data. *USA Today*. Retrieved from <http://www.usatoday.com/story/news/nation/2013/12/03/fbi-mass-killing-data-inaccuracy/3666953/>

- Parkin, W. S., & Freilich, J. D. (2015). Routine activities and right-wing extremists: An empirical test of ideologically motivated and non-ideologically motivated homicide victimizations. *Terrorism and Political Violence*, 27, 182-203.
- Parkin, W. S., Freilich, J. D., & Chermak, S. M. (2014). Ideological victimization: Homicides perpetrated by far-right extremists. *Homicide Studies*. Retrieved from <http://hsx.sagepub.com/content/early/2014/04/08/1088767914529952>
- Porter, L. E., & Kebbell, M. R. (2011). Radicalization in Australia: Examining Australia's convicted terrorists. *Psychiatry, Psychology and Law*, 18, 212-231.
- Pridemore, W. A. (2005). A cautionary note on using county-level crime and homicide data. *Homicide Studies*, 9, 256-268.
- Puzzanchera, C., Chamberlin, G., & Kang, W. (2014). *Easy access to the FBI's Supplementary Homicide Reports: 1980-2012*. Retrieved from <http://www.ojjdp.gov/ojstatbb/ezashr/>
- Quinet, K. (2011). Prostitutes as victims of serial homicide: Trends and case characteristics, 1970-2009. *Homicide Studies*, 15, 74-100.
- Sageman, M. (2004). *Understand terror networks*. Philadelphia: University of Pennsylvania Press.
- Sampson, R. J., & Wooldredge, J. D. (1987). Linking the micro- and macro-level dimensions of lifestyle-routine activity and opportunity models of predatory victimization. *Journal of Quantitative Criminology*, 3, 371-393.
- Schildkraut, J., & Donley, A. M. (2012). Murder in black: A media distortion analysis of homicides in Baltimore in 2010. *Homicide Studies*, 16, 175-196.
- Signorini, A., Segre, A. M., & Polgreen, P. M. (2011). The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PLoS ONE* 6(5), e19467.
- Smith, B. L., & Damphousse, K. R. (1996). Punishing political offenders: The effect of political motive on federal sentencing decisions. *Criminology*, 34, 289-322.
- Smith, B. L., & Damphousse, K. R. (1998). Terrorism, politics, and punishment: A test of structural-contextual theory and the "liberation hypothesis." *Criminology*, 36, 67-92.
- Smith, B. L., & Damphousse, K. R. (2007). *American Terrorism Study, 1980-2002* (ICPSR04639-v1). Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. doi:10.3886/ICPSR04639.v1
- Sproat, P. A. (2010). Counter-terrorist finance in the UK: A quantitative and qualitative commentary based on open-source materials. *Journal of Money Laundering Control*, 13, 315-335.
- Stack, S. (2002). Southernness, migration and homicide risk: An analysis of individual-level data. *American Journal of Criminal Justice*, 26, 165-180.
- Steffensmeier, D. J., Schwartz, J., & Roche, M. (2013). Gender and twenty-first-century corporate crime: Female involvement and the gender gap in Enron-era corporate frauds. *American Sociological Review*, 78, 448-476.
- U.S. Census Bureau. (2013). *American FactFinder*. Retrieved from <http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml>

- Wiersema, B., Loftin, C., & McDowall, D. (2000). A comparison of Supplementary Homicide Reports and National Vital Statistics System homicide estimates for U.S. counties. *Homicide Studies, 4*, 317-340.
- Williams, K. R., & Flewelling, R. L. (1987). Family, acquaintance, and stranger homicide: Alternative procedures for rate calculations. *Criminology, 25*, 543-560.
- Wilson, J. M., & Dalton, E. (2008). Human trafficking in the heartland: Variation in law enforcement awareness and response. *Journal of Contemporary Criminal Justice, 24*, 296-313.

Author Biographies

William S. Parkin is an assistant professor in the Criminal Justice Department at Seattle University. His research interests include domestic extremism and terrorism, homicide victimization, and the media's social construction of criminal justice issues.

Jeff Gruenewald is an assistant professor in the School of Public and Environmental Affairs at Indiana University-Purdue University Indianapolis. His research interests include homicide, bias crime, terrorism/homeland security, and media representations of violence.