Reliability and validity of cross-national homicide data: a comparison of UN and WHO data

Catrin Andersson and Lila Kazemian

Department of Law and Criminology, Sheffield Hallam University, Sheffield, UK; Department of Sociology, John Jay College of Criminal Justice, NY, USA

ABSTRACT

Data reliability and validity are methodological concerns in cross-national analyses of crime, but there is little agreement on which source of data provides the most reliable estimates. Moreover, few studies have examined the potential threat to validity posed by unclassified deaths. The current study aims to (1) assess the reliability of cross-national homicide data from the United Nations (UN) and the World Health Organization (WHO); and (2) investigate the impact of unclassified deaths on the validity of WHO data. Findings indicate that UN and WHO homicide rates differ in magnitude, but produce similar outcomes. The UN data produce more robust results and statistical models with less error. The WHO data are more stable and reliable over time, and better suited for longitudinal analyses. Analyses drawing on WHO data should not disregard unclassified deaths because their inclusion provides a more accurate estimate of the true number of homicides.

ARTICLE HISTORY

Received 31 March 2017
Accepted 2 August 2017

KEYWORDS

Cross-national; homicide; unclassified deaths; reliability; validity; World Health Organization; United Nations

Introduction

In 2012, the United Nations (UN) documented 437,000 homicides worldwide (UNODC: Global Study on Homicide, 2013). According to the most recent Global Study on Homicide, homicide rates, and their fluctuations across time, vary by region. For instance, decreases in homicide rates were observed in Europe while increases in violence have been noted in the Americas, though these appear to be stabilising. Approximately, half of the homicides reported in 2012 occurred in non-Western nations, despite the fact that these countries only make up 11% of the overall world population (UNODC: Global Study on Homicide, 2013).

Various studies have examined variations in cross-national homicide rates. These studies have been confronted with a number of methodological challenges, such as small sample size and missing or inconsistent data. In addition, researchers have raised concerns about the reliability and validity of cross-national data (Bennett & Lynch, 1990; Rubin, Culp, Mameli, & Walker, 2008; Schaible, 2012). Testing the reliability of cross-national data has become increasingly challenging, as the availability of sources of homicide data has declined since the seminal work of Bennett and Lynch in 1990. In the past, the International Criminal Police Organization (INTERPOL) data were the most frequently used source of crime data (Bennett & Lynch, 1990), but these data are no longer available. The only remaining sources of cross-national homicide data are compiled by the UN and the World Health Organization (WHO). Scholars do not agree on which source provides the most reliable data.
The validity of homicide data is equally important in cross-national research. Homicides, unlike other crimes, leave a body behind, making this offense more visible and detectable by the authorities (Ouimet & Montmagny-Grenier, 2014). As a result, homicide data are believed to be characterised by greater external validity when compared with other types of crimes (Lauritsen, Rezey, & Heimer, 2016). Notwithstanding the fact that this distinctive feature of homicide lends to more valid cross-national comparisons, homicide data are not immune to validity concerns, particularly at the cross-national level (Bennett & Lynch, 1990). Unclassified deaths (i.e., deaths of an unknown cause) pose a significant threat to the validity of homicide data. Researchers have investigated the influence of unclassified deaths on the distribution of suicides (Björkenstam et al., 2014), but this issue has seldom been examined in homicide research. Unclassified deaths are likely to result in an underestimation of homicide events, potentially undermining the validity of findings. The detailed nature and consistent data collection procedures (i.e., completed death certificates) of the WHO data offer an opportunity to examine the potential influence of unclassified deaths in analyses of homicide rates. There have been few tests of reliability and validity of the UN data, but these data continue to be used frequently in homicide research (e.g., Chu & Tusalem, 2013; Wolf, Gray, & Fazel, 2014).

The current study aims to address knowledge gaps in cross-national research in two ways. First, drawing on recent data (1995–2012), it assesses the reliability of cross-national homicide data from the UN and the WHO (n=56). Using robust indicators identified in the cross-national homicide literature, we assess whether multivariate analyses of homicide rates produce similar results when using UN versus WHO data. Second, through a multivariate analysis of the predictors of homicide rates, the study investigates the validity of the WHO data by comparing results with and without the inclusion of unclassified deaths. With the exception of one Russian study (Andreev, Shkolnikov, Pridemore, & Nikitina, 2015), no homicide study has, to our knowledge, examined the potential threat posed by unclassified deaths to the validity of cross-national homicide data.

Methodological issues in homicide research

Studies on the reliability and validity of crime data

Most rigorous methodological studies on national crime data have emerged from the United States (Berg & Lauritsen, 2016; Lauritsen et al., 2015; Loftin & McDowall, 2010). These studies have investigated the degree of convergence in longitudinal crime trends (Lauritsen et al., 2015; Lynch & Addington, 2006; McDowall & Loftin, 2007, as well as the reliability and validity, of crime data across different sources).

In the United States, homicide data are gathered through two main sources: the Uniform Crime Report’s Supplemental Homicide Reports (SHR), and the Fatal Injury Reports included in the Center for Disease Control and Prevention’s National Vital Statistics System (NVSS) (U.S. Department of Justice, Bureau of Justice Statistics, 2014). The SHR are drawn from criminal justice agencies, while the NVSS is public health data. Both sources capture homicide incidents, but they each pose distinctive reporting issues. Homicides are more likely to be under-reported in the SHR data, while these events may be misclassified in the NVSS (Wiersema, Loftin, & McDowall, 2000). The issues characterizing U.S. data have direct relevance for cross-national comparisons, given the similar underlying structure of different sources of international homicide data (i.e., criminal justice vs. public health data).

NVSS consistently reports higher rates of homicide when compared with SHR data. This difference may be due to disparities in definitions (Wiersema et al., 2000) or the fact that unlike the NVSS, criminal justice agencies report data voluntarily for the SHR (U.S. Department of Justice, Bureau of Justice Statistics, 2014). Research has found significant disparities between the
SHR and NVSS. Data from the 1980s suggest that only 22% of US counties had perfect agreement on estimates from the two sources (Wiersema et al., 2000).

Analyses investigating the reliability and validity of international homicide data are even more scarce than those conducted at the national level. Bennett and Lynch (1990) evaluated the reliability of cross-national crime data using a sample of 31 nations between the years 1975 and 1980, drawing on data from INTERPOL, the UN, the WHO, and the Comparative Crime Data File (Archer & Gartner, 1984). Their analyses showed that multi-year averages of homicide rates produced similar results across sources. However, the WHO homicide rates were generally substantially lower than UN homicide rates, leading the authors to conclude that the WHO data were more reliable in cross-sectional analyses. Bennett and Lynch (1990) did not examine homicide trends over time. The authors concluded that the selection of data source for cross-national analyses should be based on the completeness of data.

Data availability is only one of the obstacles faced by cross-national scholars. Cross-national analyses can be influenced by data irregularities, which can impact both reliability and validity. In 2001, Huang re-evaluated Bennett and Lynch’s (1990) data using two triangulated methods, and found that the UN homicide data displayed the most discrepancies between mean rates, standard deviations, and adjusted values when compared with the WHO data. Similarly, in a study conducted by Rubin et al. (2008), the UN data were reported to have up to 24% of missing items, as well as a greater frequency of “anomalies” (i.e., large fluctuation in homicide rates) when compared with the INTERPOL data. Rubin et al. (2008) called for more efforts to assess the validity of the UN data.

The question of whether homicide data originating from public health (WHO) or criminal justice (UN) sources are more reliable remains unresolved. A handful of studies have examined the reliability and validity of homicide data, dating back to the seminal work of Bennett and Lynch (1990) nearly three decades ago. While many scholars favour WHO data because of its professed reliability (Levchak, 2015; Messner, Raffalovich, & Sutton, 2010; Nivette & Eisner, 2013), the use of UN data is also prevalent in cross-national research (Chu & Tusalem, 2013; Pridemore & Chamlin, 2006; Wolf et al., 2014). Although the UN data are characterised by increased irregularities when compared with the WHO data, it remains unclear whether Bennett and Lynch’s (1990) key finding (i.e., that the UN and the WHO homicide data produce similar results) continues to hold with more recent homicide data.

The potential influence of unclassified deaths

The WHO’s mortality data are collected rigorously but nonetheless include a significant proportion of unclassified deaths (otherwise referred to as Events of Undetermined Intent, or EUI). These events are unclassified due to insufficient information surrounding the circumstances of the death. In these cases, at the time that the expert completed the death certificates, it was unclear whether the incident was a homicide, a suicide, or an accident. It has been argued that events may be misclassified into this category intentionally, and that “…there are reasons to believe that in some nations at some times this category may be employed to purposely misclassify homicide and suicide deaths” (Andreev et al., 2015, p. 13). The WHO homicide data may be misrepresented or underestimated due to this often neglected mortality category. These unclassified deaths were purported to make up at least one-third of unclassified homicides in a Russian sample (Andreev et al., 2015). Unclassified deaths reveal valuable information about the reporting practices of different countries. Some European countries report relatively high rates of unclassified deaths (e.g., 12% in the United Kingdom, 10% in Poland, and 8% in Sweden; Andreev et al., 2015). These figures are higher in the developing world. More than 20% of injury deaths in many Asian, African, and Middle Eastern countries are unclassified (Bhalla et al., 2010).

The exclusion of unclassified deaths may lead to an underestimation of mortality rates (Andreev et al., 2015). While countries vary in their rates of reported unclassified deaths, the exclusion of these events from cross-national analyses may impact the validity of the findings.
Some studies have provided detailed discussions of issues relevant to unclassified deaths (Bhalla, Harrison, Shahraz, & Fingerhut, 2010), namely classification accuracy (Prinsloo, Matzopoulos, Laubscher, Myers, & Bradshaw, 2016) and implications for data quality (Hu & Mamady, 2014; Prinsloo et al., 2016). Andreev and colleagues (2015) re-classified Russian mortality data in an attempt to establish the true cause of death in unclassified cases. Drawing on detailed Russian death records ($n > 3$ million) and using a multinomial logistic regression model, the authors produced an estimated classification probability (ECP). ECP measures the probability that an unclassified incident was either a non-transport accident, a homicide, or a suicide. This analysis drew on specific information about each case (e.g., sex, age, and event characteristics) in order to predict the correct type of death. The unclassified deaths were distributed as follows: non-transport accidents: 48%; homicides: 33%; and suicides: 16%. While these results from Russia do not necessarily generalise to all countries, the findings suggested that a sizeable proportion of unclassified deaths may be homicides.

In a Swedish study examining the influence of unclassified deaths on suicide rates, Björkenstam and colleagues (2014) argued that it is not good practice to ignore unclassified deaths, or to simply combine these events with suicide data. The authors recommended that analyses be replicated with the inclusion of unclassified deaths. Björkenstam et al. (2014) further argued that homicides were less likely to be misclassified when compared with suicides, though this statement is solely based on a 1989 study that investigated the validity of suicide data (O’Carroll, 1989). Other scholars have argued that there was no reason to believe that unclassified deaths were primarily suicides (Värnik et al., 2009). What is known for sure is that the correct classification of deaths leads to more accurate data, which are crucial for the development of effective violence prevention measures (Hu & Mamady, 2014).

In summary, drawing on samples of varying sizes and from different time periods, studies have suggested discrepancies between UN and WHO homicide data. However, researchers have been largely silent on which source of homicide data is the most reliable. The state of knowledge on the validity of homicide data is equally nebulous. Scholars have called for validity tests of the UN-recorded homicides (Rubin et al., 2008), but the data required to conduct such tests are not available. The current study aims to address some of these knowledge gaps in cross-national research.

The current study

This paper investigates the reliability of cross-national data by examining whether trends and findings about homicide remain consistent when drawing upon different sources, an issue that has been neglected in criminological research (Liem & Pridemore, 2011). This study also assesses the validity of cross-national data by examining the influence of deaths without a known cause, included in the WHO’s mortality data.

Drawing on homicide data from 1995 to 2012 ($n = 56$ countries), the current study has three main objectives: (1) to compare the UN and WHO homicide trends, (2) to assess whether the predictors of cross-national homicide rates vary when using UN versus WHO data, and (3) to estimate the validity of the WHO data by replicating analyses inclusive of unclassified deaths. To our knowledge, no such efforts have been undertaken with recent cross-national homicide data.

Methods

Data

The UN

The UN defines homicide as an “unlawful death purposefully inflicted on a person by another person” (“UNODC: Global Study on Homicide,” 2013). The units of analysis, which vary across countries, are body count and the number of homicide incidents. These two methods of counting
homicides may result in discrepant data because a homicide “event” can implicate more than one body but would nonetheless be regarded as one homicide incident. In comparison, the body count method is based on the number of bodies, irrespective of the number of homicide events. The UN data exclude infanticides, and some nations include attempted homicides (Lynch & Pridemore, 2011).

The UN collects crime data in two ways. First, the crime survey (i.e., United Nations Surveys on Crime Trends and the Operations of Criminal Justice Systems, or CTS) is collected through criminal justice agencies in UN member nations every year. If there are any reported discrepancies in the collected data (i.e., a 30 or more per cent change in the crime statistics from one year to the next), the UN requests an explanation from the reporting agency (Lynch & Pridemore, 2011). Second, a more complete set of data is collected and disseminated by a large division of the UN, the United Nations Office on Crime and Drugs (UNODC). These data are drawn from multiple sources, including the CTS survey, police forces, and the WHO. Since 2011, a growing number of countries have developed national registration systems for homicide statistics, resulting in improved cooperation with the UN’s homicide data collection programme (UNODC: Global Study on Homicide, 2013; for a more comprehensive review of homicide data from the UN, see Lynch & Pridemore, 2011).

The WHO

According to the WHO, homicides are defined as “injuries inflicted by another person with intent to injure or kill, by any means” (World Health Organization, 2014). This definition is based on the premise that homicides are intentional. However, because the WHO cannot enforce compliance with the definition, Smit, De Jong, and Bijleveld (2012) argued that deaths are likely to include both intentional and unintentional homicides; body count is the measurement unit. Certain deaths are excluded from the WHO data: legal intervention deaths (i.e., caused by law enforcement or military personnel), war-related deaths, lives lost as a result of dangerous driving, assistance with suicide, abortion, and homicide attempts. The WHO collects mortality data based on death certificates that are completed by medical doctors or, in certain countries, police officials (Smit et al., 2012).

Comparability issues between the UN and WHO data

There are some tangible differences between the UN and WHO data. The UN data are reported by criminal justice agencies, whereas the WHO data are generally collected by health professionals. The two sources employ different definitions of homicide and draw on different measurement units (i.e., body count in the WHO data vs. body count and/or number of homicide incidents in the UN data; Lynch & Pridemore, 2011). The two data sources also exclude different categories of deaths (infanticides in the UN data vs. various other types of death in the WHO data; see above). It has been argued that disparities in findings across the different data sources may be due to operational differences relating to the types of incidents included in the homicide category (Smit et al., 2012). In the WHO data, the level of expertise of public health officials is variable (Howard, Newman, & Pridemore, 2000), and they are at times unable to determine the cause of death. Similarly, there are no assurances that the UN’s data collection regulations are respected (Lynch & Pridemore, 2011).

In short, the most fundamental challenge in comparisons of UN and WHO data relates to their divergent inclusion criteria and definitions of homicide. There is no agreement in the cross-national homicide literature as to which source provides the most valid and reliable data. WHO data tend to be favoured because of the rigorous recording method (LaFree, 1999), larger sample sizes, and the exclusion of homicide attempts (Lynch & Pridemore, 2011), while others have argued that the multisource nature of the UN data renders it most appropriately for cross-national comparisons (Ouimet, 2012). These issues are explored in this paper.
**Variables**

**Dependent variables**

The current study employs three dependent variables: UN homicide rates, WHO homicide rates with unclassified deaths, and WHO homicide rates without unclassified deaths; all three rates reflect the number of deaths per population of 100,000 ("UNODC Statistics Online," n.d.; “WHO-CoDQL – Cause of Death Query online,” n.d.; “WHO Mortality Database,” n.d.). Mortality data were obtained for 56 countries between 1995 and 2012. The range of years is limited to this period because the UN did not publish homicide data online prior to 1995. As a result of missing values on the independent variables, a smaller range of years was used for the cross-sectional multivariate analyses, using multiyear averages for the most recent years available (i.e., 2008–2012). Analyses drew on these years because this was the range with the most complete data for the selected countries. In addition, the inclusion of developing nations typically results in a shorter range of years (Lappi-Seppälä & Lehti, 2014). Many researchers have attempted to maximise sample size in cross-national studies (Koeppel, Rhineberger-Dunn, & Mack, 2015). While this may strengthen statistical models and enable the inclusion of more predictors, this practice comes at the cost of exclusion of developing nations. Most cross-national homicide research draws upon smaller samples of mainly Western and modernised nations. Data on unclassified deaths were available between 1998 and 2010 for 42 of the countries included in the overall sample. All dependent variables were transformed from counts into rates using WHO population data (or World Bank data when the WHO data were unavailable).

**Independent and control variables**

Predator and control variables were selected on the basis of indicators that have been consistently found to be associated with variations in cross-national homicide rates (Nivette, 2011; Trent & Pridemore, 2012): poverty (infant mortality rate), Human Development Index (HDI), quality of governance, female labour participation, and male-to-female sex ratio.

The infant mortality rate is defined as the number of infants dying before reaching 1 year of age per 1000 live births; these data were obtained from the World Bank. This indicator commonly serves as a proxy for poverty (Dasandi, 2013; Pridemore, 2011). As a measure of absolute deprivation, poverty is one of the most salient indicators of aggregate homicide rates (Pridemore, 2011; Rogers & Pridemore, 2016).

While many studies have focused on economic measures, such as GDP, these indicators do not capture the social dimension of development (Cao & Zhang, 2017). The Human Development Index (HDI), compiled by the UN, is a composite measure consisting of life expectancy at birth, educational attainment (i.e., adult literacy rate), and adjusted real income (i.e., GDP per capita). Ouimet (2012) argued that the HDI is a better measure of national development because it integrates several indicators and does not focus solely on economic factors. The HDI is also advantageous because data for this indicator are widely available for most countries.²

Recent research has indicated the need to evaluate the effects of political and governmental functioning on levels of violence (Cao & Zhang, 2017). The current study draws on a measure of quality of governance (i.e., the World Governance Indicators), which is available from the World Bank. This variable integrates several indicators, including measures of voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption. Instead of focusing on a single dimension of governance, a composite measure was created through factor analysis, capturing the overall quality of governance (Cao & Zhang, 2017).³

Prior research has also found links between the proportion of women in the work force and violence rates, largely due to the weakened social control in families (Neumayer, 2003). Female labour participation is measured by the percentage of women in the population who are active in the labour force.
Finally, sex-ratio distributions (i.e., higher ratios of men to women) are strongly and positively associated with homicide rates (Chamlin & Cochran, 2006; Chu & Tusalem, 2013), as men are more likely to perpetrate and also to be victims of homicide. The sex ratio refers to the number of males per 100 females in the population (Messner & Rosenfeld, 1997); this information was obtained from the World Bank. All predictor variables were drawn from the year 2007 in order to create a lag with the dependent variables, measured between 2008 and 2012. Descriptive statistics for all variables are presented in Table 1.

**Analytical strategy and procedures**

The current study presents three main analyses. First, trend analyses investigate differences over time between the UN and WHO homicide data patterns. Second, ordinary least squares (OLS) regression and weighted least squares (WLS) regression were carried out on the two dependent variables (i.e., the UN homicide rate and the WHO homicide rate) to compare differences in outcomes between the data sources (see Lauritsen, Rezey, & Heimer, 2016). Third, in order to test for data validity, OLS regressions were conducted, with and without unclassified deaths, to assess whether the inclusion of unclassified deaths produced different results. In two of the models, heteroscedasticity was present, and subsequently, alpha and beta estimates were not BLUE (i.e., best linear unbiased estimation). WLS was used to adjust for this issue, resulting in smaller standard errors than a corrected OLS model (McClendon, 2002).

Multiyear averages were created for the years 2008–2012 to control for yearly fluctuations (and to some extent, measurement errors) in homicide rates (Altheimer, 2008; Cao & Zhang, 2017; Chon, 2011; Krahn, Hartnagel, & Gartrell, 1986). Due to the skewed distribution of homicide rates, the natural log was applied (Altheimer, 2008; Elgar & Aitken, 2011; Sun, Chu, & Sung, 2011); the same solution was applied to other independent and control variables that deviated from a normal distribution. These transformations were used in all multivariate analyses. Variance inflation factors were tested for all independent variables, none of which exceeded the standard threshold of 10 for multicollinearity (Neter, Kutner, Nachtsheim, & Wasserman, 1996).

**Results**

**Comparing homicide trends in the UN and WHO data**

Figure 1 shows parallel trends between UN and WHO homicide patterns between 1995 and 2010. We note that despite the definitional differences, the two sources of homicide data seem to share similar patterns across time. However, the magnitude of homicide rates differs between the two sources, with the UN consistently reporting higher rates. Bennett and Lynch (1990) reported similar patterns in their analysis. This may be unsurprising given that the UN data, unlike the WHO, may include homicide attempts. There is no consensus on the effects of these operational differences on the validity of the data. Some have argued that these different definitions of

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (SD)</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN homicide rate</td>
<td>8.52 (13.67)</td>
<td>.38 (Singapore)</td>
<td>59.56 (El Salvador)</td>
</tr>
<tr>
<td>WHO homicide rate</td>
<td>6.98 (11.42)</td>
<td>.30 (Singapore)</td>
<td>52.64 (El Salvador)</td>
</tr>
<tr>
<td>WHO homicide rate with EUI rate</td>
<td>24.09 (17.95)</td>
<td>3.89 (Singapore)</td>
<td>74.77 (South Africa)</td>
</tr>
<tr>
<td>Poverty rate (infant mortality rate)</td>
<td>10.53 (9.71)</td>
<td>2.20 (Singapore)</td>
<td>46.7 (South Africa)</td>
</tr>
<tr>
<td>HDI dummy</td>
<td>.50 (.50)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>WGI – quality of governance</td>
<td>.70 (.94)</td>
<td>−1.23 (Belarus)</td>
<td>2.22 (Denmark)</td>
</tr>
<tr>
<td>Female labour participation (%)</td>
<td>50.53 (6.26)</td>
<td>37.50 (Italy)</td>
<td>61.90 (Denmark)</td>
</tr>
<tr>
<td>Sex ratio (male to female)</td>
<td>95.53 (3.92)</td>
<td>84.61 (Latvia)</td>
<td>103.35 (Costa Rica)</td>
</tr>
</tbody>
</table>

HDI: Human Development Index; WGI: World Governance Indicators.
homicide pose a significant threat to cross-national comparisons (LaFree, 1999; Liem & Pridemore, 2011), while others have found that definitions produce very small discrepancies in statistical effects (Barclay, 2000).

Lappi-Seppälä & Lehti (2014) suggested that longitudinal homicide data can be regarded as reliable if variations do not exceed 10% from one decade to another. Using this standard, the WHO data appear to be more reliable than the UN data; the mean WHO homicide rate ranges between 7 and 8, while the mean UN homicide rate ranges between 9 and 11. The UN reports a drop in homicide rate in the mid-to-late 1990s, which may be a result of the crime decline that has been observed in most nations during this period (Baumer & Wolff, 2014; Eisner, 2008; Lappi-Seppälä & Lehti, 2014). Western and developed nations experienced a shared decrease in violence rates while other countries, such as some Latin American nations and Russia, displayed increases in violence rates from the 1990s onwards (Weiss, Santos, Testa, & Kumar, 2016). These outlier nations are partly responsible for the fluctuations observed in homicide rates over time.

Temporal fluctuations in outlier nations may also explain the more erratic character of UN homicide trends. For example, the UN reported a homicide rate of 142.7 per 100,000 for El Salvador in 1995, while the WHO reported a rate of 45.3 per 100,000. El Salvador experienced civil war during the 1980s, a period characterised by a marked rise in violence. Even after the end of the conflict in the 1990s, the homicide rate remained the highest in the world, possibly due to the normalisation of violence after a period of high instability, brought about with the civil war (Bourgois, 2001).

In short, although the recorded homicide rate is higher in the UN data than in the WHO data, general trends between the two sources are generally quite similar, suggesting that either source could be appropriate for investigating long-term fluctuations in homicide rates.

Trend analysis enables us to visually examine homicide trends. The purpose of this descriptive analysis is to assess whether homicide rates generally follow similar patterns over time across the two different data sources and to identify temporal fluctuations and outliers. The analyses in the following sections investigate issues of reliability and validity across the two sources of homicide data.

Reliability of WHO and UN homicide data

Table 2 shows the WLS regression models predicting UN and WHO homicide rates (n = 48). The models appear to produce similar results. Poverty is a slightly stronger predictor in the UN model, while the effects of quality of governance are stronger in the WHO model. It is important to note that indicators are similar in strength and direction, and the small differences observed in coefficients between the two models are relatively marginal. In terms of model fit, there are
more evident dissimilarities. The UN model has a lower root mean square error (RMSE = .58) than the WHO model (RMSE = .73). A lower RMSE indicates a better fit, suggesting that the UN data are more accurate and have less error.

In short, while the trends in Figure 1 showed that UN homicide rates were higher and had a larger standard deviation when compared with WHO rates, the regression analysis suggests that the UN data produce a more accurate model. Bennett and Lynch (1990) found that although there were limited discrepancies between the two sources, the UN data displayed the most irregular data points (Huang, 2001; Rubin et al., 2008). This is slightly inconsistent with the findings of this study. We found that WHO data have more temporal stability and are therefore more suitable for longitudinal investigations. Conversely, the UN data lend to more accurate cross-sectional statistical models.

### Investigating the influence on unclassified deaths in WHO homicide trends

Figure 2 presents WHO homicide trends, with and without the inclusion of unclassified deaths (i.e., EUI). The trends are largely similar, suggesting that a substantial portion of unclassified deaths are likely to be homicides. However, the homicide rate trend is more erratic when including unclassified deaths.

The country with the most variability is South Africa, with a rate fluctuating from 108.5 in 1998 to 13.0 in 2010. South Africa Statistics has noted that external causes of death (accidents, suicides, homicide) have a high degree of misclassification. The agency has urged analysts to interpret the data with caution (Findings from death notification, 2014; Mortality and causes of death in South Africa, 2014) as "many forensic pathologists in South Africa prefer not to indicate

---

**Table 2. Regression models with macro predictors of UN and WHO homicide rates (n = 48).**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>WLS: UN homicide rate Beta (robust standard errors)</th>
<th>WLS: WHO homicide rate Beta (robust standard errors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infant mortality (log)</td>
<td>1.41 (.17)****</td>
<td>.94 (.30)****</td>
</tr>
<tr>
<td>Polity (log)</td>
<td>-1.13 (.40)****</td>
<td>-1.20 (.46)****</td>
</tr>
<tr>
<td>HDI dummy</td>
<td>.76 (.45)</td>
<td>-.60 (.60)</td>
</tr>
<tr>
<td>Fem labour participation</td>
<td>.01 (.02)</td>
<td>-.01 (.03)</td>
</tr>
<tr>
<td>Sex ratio squared</td>
<td>.00 (.00)</td>
<td>.00 (.00)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.81 (2.03)</td>
<td>-1.74 (2.78)</td>
</tr>
<tr>
<td>F</td>
<td>37.72***</td>
<td>25.88***</td>
</tr>
<tr>
<td>R-squared</td>
<td>.88</td>
<td>.81</td>
</tr>
<tr>
<td>RMSE</td>
<td>.58</td>
<td>.73</td>
</tr>
</tbody>
</table>

*p < .10; *p < .05; **p < .01; ***p < .001.

HDI: Human Development Index; RMSE: root mean square error; WLS: weighted least squares; UN: United Nations; WHO: World Health Organization.

---

![Figure 2](https://example.com/figure2.png)
the manner of death for injuries on the DNF [Death Notification Form], resulting in relatively high proportions of undetermined external causes” (Burger et al., 2012, p. 310).

In summary, rates of unclassified deaths do not generally display extreme fluctuations, with the exception of homicide data from South Africa, which appear to be less valid and should perhaps be excluded from cross-national studies.

**Validity of WHO homicide data**

Table 3 presents an OLS regression analysis investigating the predictors of WHO homicide rates, with the inclusion of unclassified deaths. Results are largely similar to the WHO model without unclassified deaths (see second column of Table 2), though some variables have weaker coefficients (e.g., poverty and quality of governance). If it was presumed that the unclassified deaths largely consisted of non-homicide incidents (i.e., accidents and suicides), results would be markedly different from the WHO model without unclassified deaths (Table 2). The weaker strength of some of the coefficients may be a result of the fact that inevitably, some of the deaths consist of accidents and suicides. Nevertheless, we know from prior literature that analyses of homicide and suicide rates tend to produce different results, and that their predictors may be inversely related (e.g., Lee & Pridmore, 2014). The error is lower when including unclassified deaths (RMSE = .48), and thus more accurate, than in the WHO homicide model that excludes unclassified deaths (Table 2).

Of course, it is highly implausible that the combined rate of WHO and unclassified deaths consists solely of homicides (Andreev et al., 2015; Björkenstam et al., 2014; Värnik et al., 2009). Unclassified deaths are central to the discussion of the validity of homicide data from public health sources. These preliminary findings warrant further replication and suggest that studies using WHO homicide, accident, or suicide data should replicate their analyses with the inclusion of a combined rate that integrates unclassified deaths.

**Discussion**

**Summary of findings**

There is ongoing debate about the most reliable source of homicide data. Bennett and Lynch’s (1990) important study found small differences between WHO and UN data. The WHO homicide data have been found to be more reliable than the UN data, but researchers have continued to use the latter in cross-national analyses. This paper set out to examine the reliability and validity of homicide data from the UN and the WHO. First, we investigated homicide trends between the two sources, from 1995 to 2012. The trends were largely parallel but the UN homicide rates were
systematically higher than the WHO rates. In addition, applying the reliability standard suggested by Lappi-Seppälä and Lehti (2014), the WHO data were found to be more reliable than the UN data, with less than 10% variation in the homicide rate across the observation period. The fluctuations in the UN data may be driven by outlier nations, such as El Salvador. Measurement error at all stages of the data collection procedure may also impact reliability across sources; these sources of error include instrumentation error, recorder error, and respondent error (see Bennett & Lynch, 1990, for a more detailed discussion). Our analyses suggested that the WHO data are most appropriate for longitudinal analyses of cross-national homicide data.

Second, we examined the reliability of cross-sectional UN and WHO data by comparing the influence of well-known correlates of cross-national homicide rates across the two sources. The results from these multivariate models were largely similar. The coefficients followed the same direction in both sources, though they were slightly stronger in the UN model. The RMSE, which measures accuracy of the model, was the lowest for the UN model, suggesting that UN data provided the most robust results based on the selected indicators. For cross-sectional analyses with larger and more diverse samples, UN data are preferable because they are more complete and produce more robust statistical outcomes. The small differences in the strength of the coefficients may be due to the fact that the two sources include different types of homicides (e.g., the UN data only include intentional homicides and attempts, while the WHO data also include unintentional homicides). Because the differences were slight, it would be acceptable for researchers to use either source depending on the aims of the study and methodological concerns (e.g., sample size, the inclusion of nations from varying levels of development, etc.).

Finally, the validity of the WHO homicide data was tested in two steps: by comparing trends, and conducting a multivariate analysis to compare findings with and without the inclusion of unclassified deaths in the measure of homicide. The trends were largely similar with and without unclassified deaths, suggesting stability in both rates. The multivariate analysis showed that the inclusion of unclassified deaths produced coefficients that varied in the same direction as analyses excluding these unknown events, but with slightly reduced strength. The model including unclassified deaths was found to have less error. While it is clear that not all unclassified deaths consist of homicides, the similarity in findings suggests that a significant proportion of unclassified deaths are likely to be homicides. It is generally acknowledged that there is a negative association between cross-national suicide and homicide rates and that their covariates tend to vary in opposite directions (Bando & Lester, 2014; Lee & Pridmore, 2014; Stoupel et al., 2005), though recent research has found that this association varies by region (Fountoulakis & Gonda, 2017). Based on this premise, if unclassified deaths mostly consisted of suicides or accidents, we would expect to observe opposite trends in results with and without the inclusion of unclassified deaths, which was not the case.

**Implications of the findings**

Measurement error in homicide data collection efforts mainly stems from definitions of homicide and data reporting mechanisms. Bennett and Lynch (1990) called for more consistent definitions of homicide across different sources, which would facilitate cross-national comparisons. One of the main challenges in adopting more uniform definitions relates to the different nature of the organisations that collect the data. The WHO collects health data used to monitor global health trends, whereas the UN data are collected to inform criminal justice practices. Despite these different mandates, it is not unreasonable to strive for more consistent definitions. After all, the WHO operates within the UN and both organisations share many common goals.

Data reporting and completeness also impact measurement error. Huang and Wellford (1989) noted that because the UN and the WHO did not make data reporting compulsory, data are often incomplete. Bhalla et al. (2010) reported that only 28% of countries in the world had complete national death registration data for the years reviewed in their study. Unsurprisingly, wealthier
nations tend to have more complete data when compared with developing nations. Many countries in turmoil because of war, political upheaval, or high poverty levels are unlikely to have the required infrastructure to prioritise robust data collection systems.

While each nation is responsible for ensuring the quality and completeness of its data, better global awareness about the costs of violence is needed in order to encourage nations to prioritise rigorous data collection efforts. For instance, Homicide Monitor, a project managed by the Igarape Institute, maintains a website granting access to cross-national homicide datasets from the UN. The website also includes visualisation tools. Homicide Monitor aims to inform about the scope and consequences of lethal violence in the world, and to identify ways in which violence can be minimised (“Homicide Monitor,” 2017). In another, smaller scale effort to improve the accessibility and quality homicide data, the European Homicide Monitor, a database of shared homicide data from Sweden, Finland, and the Netherlands, combines comprehensive cross-national homicide data (“The European Homicide Monitor,” n.d.). More efforts of this nature are needed in order to increase cognisance about the importance of collecting valid and reliable homicide data.

Due to cross-national data limitations, it is not possible to reclassify deaths to the appropriate mortality category and assess the influence of unclassified deaths on WHO data, as other scholars have done with national data (i.e., Andreev et al., 2015). Our results have led us to conclude that it may be wise to include unclassified deaths in analyses of cross-national homicide.

This study shares similar limitations to other cross-national empirical studies. The analyses were restricted by the availability of data as well as missing data, which limited the analysis to an 18-year observation period. Like all cross-national studies of homicide, our analysis was limited to countries that have available and complete homicide data, resulting in a relatively small number of nations. Although the analyses carried out in this study included some of the strongest predictors of homicide rates identified in prior research, a smaller sample size inevitably limits the number of covariates that can be included in statistical analyses. As such, the choice of variables to be included in the analyses had to be careful and selective. This is a challenge in all cross-national analyses of crime.

In order to gain better knowledge about the reliability and validity of cross-national homicide data, it is important for future research to conduct ongoing investigations of homicide trends, drawing on emerging and relevant variables, a diverse sample of nations reflecting varying levels of development, and up-to-date data. It is particularly important to monitor developments in homicide data collection efforts in regions that have been traditionally excluded from cross-national comparisons, such as many African nations. Finally, nearly three decades have passed since the publication of Bennett and Lynch’s (1990) important study. The reliability and validity of cross-national homicide data should be investigated at more regular intervals.

To conclude, our findings have suggested that WHO data are more suitable for longitudinal analyses because these data are more stable and reliable over time. However, despite consistently higher homicide rates, the UN data produce statistical models with less error. Our findings partly support those of Bennett and Lynch (1990) and suggest that the UN data are preferable for cross-sectional analyses. When using the WHO data for longitudinal investigations, analyses should be replicated with unclassified deaths. As argued in suicide research, it is good practice to replicate analyses with unclassified deaths for more robust results (Björkenstam et al., 2014). In fact, “… scholars interested in the structural covariates of homicide and suicide rates seem largely unaware of this category and do not account for it in their analyses, which may threaten the validity of these studies (Andreev et al., 2015, p. 13).” Given that a non-negligible number of unclassified deaths are likely to be homicides, our findings suggest that it may be worthwhile to include these events in the homicide category rather than to exclude them.

The drive for reliable and valid homicide data is fuelled by the need for rigorous research that will enable scholars and policymakers to understand the processes and dynamics of violence in a wide array of societies. The implications are crucial, particularly in the context of growing globalisation. Studies of this nature enable researchers to evaluate the effectiveness of different
criminal justice systems and to influence policies and practices that inform efforts to prevent lethal violence (Bennett, 2004). We welcome more research to develop the knowledge base on the reliability and validity of homicide data, particularly with the inclusion of nations that are traditionally excluded from cross-national analyses.

Notes

1. Countries in UN and WHO samples (n = 56): Argentina, Armenia, Australia, Austria, Belarus, Bermuda, Bulgaria, Canada, Colombia, Costa Rica, Croatia, Czech Republic, Denmark, Dominican Republic, El Salvador, Estonia, Finland, France, Georgia, Germany, Greece, Guatemala, Guyana, Hungary, Ireland, Israel, Italy, Japan, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Mauritius, Mexico, Netherlands, New Zealand, Nicaragua, Norway, Panama, Poland, Portugal, Puerto Rico, Republic of Moldova, Romania, Saint Kitts and Nevis, Saint Vincent and Grenadines, Singapore, Slovakia, Slovenia, South Africa, Spain, Sweden, Switzerland, Ukraine, United Kingdom, and Venezuela (Bolivarian Republic of).

2. HDI is scored between 0 and 1. Because of issues of multicollinearity with another key variable (i.e., poverty; VIF > 10), the HDI was recoded into a dummy variable, using cut-offs determined by the UN (Human Development Report, 2014). A value of 0 corresponded to low/medium-level HDI (<.799), and a value of 1 denoted high-level HDI (> .800).

3. An explanatory factor analysis revealed that all the WGI items (i.e., voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption) loaded onto a single factor and reflected the same dimension. A reliability test on the quality of governance factor provided an alpha of .961, indicating high reliability.

4. The Breush–Pagan/Cook–Weisburg test revealed that two of the three models (UN and WHO) failed to meet the OLS assumption.

Notes on contributors

Catrin Andersson graduated with a Ph.D. in Criminal Justice from the Graduate Center/John Jay College of Criminal Justice in 2015. She is currently a lecturer in Criminology at Sheffield Hallam University in the UK. Her research interests include violence at the macro and micro levels, substance use and recovery, comparative research and research methods.

Lila Kazemian is an Associate Professor at John Jay College of Criminal Justice, City University of New York. She earned her Ph.D. from the Institute of Criminology, University of Cambridge, in 2006. Her research work has focused on changes in criminal career patterns across various periods of the life course, the process of desistance from crime, prisoner reentry, and comparative criminology. Her work has been published in the Journal of Quantitative Criminology, the Journal of Research in Crime and Delinquency, Punishment & Society, and the European Journal of Criminology.

ORCID

Catrin Andersson http://orcid.org/0000-0003-4336-4771
Lila Kazemian http://orcid.org/0000-0002-2007-6024

References


