

Background Memo

Subject: **Guide to Homicide Databases and Publically Available Research Using Original Homicide Records and Documents**

Note: This memo was prepared in June of 2012 and does not include the most recent data for 2012.

This memo was prepared by Michael Lombardo and Jarret Baughman, Northwestern University Interns, Center of Wrongful Convictions, Bluhm Legal Clinic, Northwestern University School of Law, Northwestern University, Chicago, Illinois.

The purpose of this Guide is to refer users of the Northwestern University Law School Capital Crimes Database, the Illinois Murder indictments web site, to other homicide datasets and research, available online from public sources of information and to introduce different research approaches.

Many homicide specific datasets have been created using varied types of source information. We have looked at several prominent homicide-specific datasets to analyze their differences and similarities. These datasets incorporate various locations during different periods of time, and our dataset of Illinois murder indictments can be compared and contrasted with the data presented by these projects. In the interest of future research, persons interested in our dataset should be made aware of these similar datasets and the possible strengths and weaknesses of each. Each researcher used different methods of data collection as well as different ways of organizing their datasets.

Wesley G. Skogan: Crime Decline in Chicago <http://skogan.org/>

Wesley G. Skogan has been a faculty member at Northwestern University since 1971, and holds joint appointments in the Political Science Department and the University's Institute for Policy Research. His research focuses on the interface between the public and the legal system, in crime prevention, victim services, and community-oriented policing.

In his report, "Reflections on Declining Crime in Chicago," Wesley Skogan examines recent crime trends in Chicago. The report challenges popular conclusions about why this could be happening. Some researchers have hypothesized that the decline was due to the increase in prison populations, increase in police, or the economic conditions. The data that Skogan used was drawn from reports issued by the Chicago Police Department from 1991-2005. The definition of aggravated assaults is UCR-compatible. This is not the same process that is used for internal purposes. Murder and homicide are interchanged throughout.

Skogan found that there was a widespread decline in crime in Chicago from 1991-2005. 264 out of 269 police beats experienced decreases in crime. He found that homicide dropped by 52%. Skogan says that the crime drops were not due to one single factor. He says that over the fourteen years many different factors, possibly at different times and places, were responsible for the decline. Skogan posits that all of the social and economic factors described may have been working to this end result of a decline in crime. In the report he described demography, economic conditions, incarceration, policing, drugs, guns, and gangs. He argues it was most likely “a mix of law enforcement and community factors” that caused the decline. He diminishes the value of many of these factors in order to show that it was not one specifically. After using data to show how certain factors may not have contributed as much as believed to the decline, he uses his judgment to decide which combinations were most likely responsible. At the end of the report, Skogan calls for further research, claiming that more information needs to be gathered in the community domain.

National Archive of Criminal Justice Data (NACJD)

<http://www.icpsr.umich.edu/icpsrweb/NACJD/>

National Archive of Criminal Justice Data (NACJD) was established in 1978 under the Inter-university Consortium for Political and Social Research (ICPSR). The ICPSR is part of the Institute of Social Research at the University of Michigan. NACJD is a special topic archive at the ICPSR. Special topic archives are funded by agencies outside of the ICPSR. Sponsoring agencies for the NACJD are the Bureau of Justice Statistics, National Institute of Justice, Office of Juvenile Justice and Delinquency Prevention, and the Bureau of Justice Assistance. Most of the data collections of the special topic archives are available to the public.

Researchers deposit their data on the website. The NACJD then reviews the deposit form and electronic files. NACJD researchers check for consistency and confidentiality. Changes will not be made to data values except to ensure confidentiality. The original author must authorize any changes to the data. The ICPSR staff then develops a processing plan. The plan outlines confidentiality issues, how it is released, level of quality review necessary, and the structure of final documentation files.

Researchers can download data after logging in using a Facebook account. Only users connected to ICPSR authorized institutions have access to the data. The NACJD uses a Survey Documentation Analysis (SDA) program developed and maintained by University Cal-Berkeley. NACJD also provides webinars and tutorials for using SDA. SDA allows users to browse the codebook, perform statistical procedures, manipulate variables, recode variables, compute new variables, list new variables, create subsets of variables and cases, and download the entire collection.

The NACJD encourages researchers to deposit computer-readable data with NACJD for long-term preservation and for use in analysis by other social science researchers. Depositing data here would allow the NACJD to prepare a study description as well as format the data for analysis using their SDA. The following studies are listed on the NACJD homicide resource page.

Homicides in Chicago 1965-1995 (NACJD)

<http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/6399?archive=NACJD&q=homicide&x=0&y=0>

Rebecca Block, Illinois Criminal Justice Information Authority, and Richard Block, Loyola University Chicago, compiled the "Homicides in Chicago Database". This database contains two datasets. These datasets contain information from every homicide file in the Murder Analysis Report (MAR) from the Chicago Police Department. The homicides contained in the datasets took place from 1965-1995. Data up to 2000 has been added to update the dataset. The files in each dataset underwent a review process where files were altered to limit disclosure. Specific information such as date of death and date of injury were deleted and the ages of the victims and offenders were placed into age categories. Both datasets include information regarding the census tract, community area, police district, and police area. Each homicide is also coded according to motive, circumstance, and situation. These datasets can be analyzed using the NACJD's Survey Documentation Analysis.

Part 1 of the database is a dataset containing victim-level files. The victim-level contains 23,817 records. Data provides information on the relationship from victim to offender. This dataset provides variables such as time of the crime, place of the homicide, type of weapon, motivation for the incident, drugs, alcohol, gangs, child abuse or a domestic relationship, offender identification, information on the death of the victim. There is only one record for each victim. However, there is information for up to five offenders included with each victim record. If there is a repeat offender, it is included with each victim record.

Part 2 of the database contains Offender-Level Data. Using the victim-level files would not allow for risk analysis or the creation of offender rates because of the repetition of offenders. There are 26,030 offender records. Offenders were reorganized to create a single record for each offender. Many of the same variables used in the victim-level files were also used with offenders. The offender records contained demographic information about each offender as well as the offender's relationship with the victim. Information on their location, drug use, and the weapon are the same as in the victim-level file.

The homicides included are defined by the police investigation stage without regard to later justice decisions. Some offenders were not convicted because of lack of evidence but are still included.

The homicides included in this dataset are identified at the police investigation stage. Each homicide has been coded for use in regression analysis. The codebook for each dataset includes the frequencies of each coded variable.

State of Washington Study, HITS (NACJD)

<http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/06134/version/1>

Robert Keppel, Washington State Attorney General's Office and Joseph Weis, University of Washington, Center for Law and Justice completed a study that was also featured on the Homicide Resources page on the NACJD's website. The Improving the Investigation of

Homicide and the Apprehension Rate of Murderers in Washington State contains information on all solved murders that occurred between 1981 and 1986. The study was funded by National Institute of Justice.

This data is a subset of the Homicide Tracking System (HITS), which maintains a computerized database by the State of Washington. Police, sheriffs, medical examiners, coroner's office, prosecuting attorneys, and other departments provide the data for HITS voluntarily. According to the Attorney General's website HITS began in 1987 and continues today. The Keppel and Weis study contains data from 1981-86. HITS was originally funded by the National Institute of Justice. In 1991 the Washington State Legislature mandated that HITS track all violent crimes. HITS was designed to provide law enforcement officials with information regarding incidents of murder, provide analysis of these incidents such as links between victims, offenders, and cases, and provide investigators with resources, including names of experts and technical assistance in investigating such cases. Data on HITS is entered with standard forms to facilitate the uniform collection of information. Unlike the study by Keppel and Weis, HITS contains information about both solved and unsolved murders. Data for HITS is submitted voluntarily from police in 273 jurisdictions and medical examiners, coroners, and prosecuting attorneys in 39 counties.

The sample that Keppel and Weis used contains only the solved murders that occurred during the five year time period. Data includes crime evidence, victim and offender characteristics, geographic locations, weapons, and vehicles. There are three parts to this particular data collection. Dataset 1 contains raw data for victims, Dataset 2 contains raw data for offenders, and Dataset 3 contains raw data for incidents.

Homicides in New York City, 1797-1999 (NACJD)

<http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/03226/version/1>

Eric Monkkonen formerly of the University of California-Los Angeles department of History and Policy Studies compiled a dataset titled: "Homicides in New York City, 1797-1999 [And Various Historical Comparison Sites]." The goal was to create a site-specific, individual-based data series that could be used to examine major social shifts related to homicide, such as mass immigration, urban growth, war, demographic changes, and changes in laws.

Monkkonen's dataset can be found on the NACJD website and is broken up into three sub-datasets. The first dataset's variables include counts of New York City homicides, arrests, and convictions, as well as the homicide rate, race or ethnicity and gender of victims, type of weapon used, and source of data. The second dataset's variables include the date of the murder, the age, sex, and race of the offender and victim, and whether the case led to an arrest, trial, conviction, execution, or pardon. The third dataset contains annual homicide counts and rates for various comparison sites including Liverpool, London, Kent, Canada, Baltimore, Los Angeles, Seattle, and San Francisco. Homicide data from these comparison sites come from a variety of research projects done by others and is often supplemented by public records.

The basic approach to the data collection was to obtain the best possible estimate of annual counts and the most complete information on individual homicides. For the individual data, information from the period of 1798-1862 was drawn from the coroner's indictments, which are

held by the New York City Municipal Archives, and daily newspapers -- to 1874 -- searched for incidents. For some years, the inquests were included in New York Superior (or General Sessions) Court Minutes. Many years have been summarized by genealogist Kenneth Scott. Coroner inquests vary in completeness. Monkkonen has used information on 1,773 homicides prior to 1875.

For the years 1853-1858, Monkkonen used two sources, the coroners' inquests and newspapers, to compare name by name which victims are in both sources and which in only one. Using these three numbers which are only in the coroners' inquests, only in the newspaper, and those in both, Monkkonen used an estimation technique known as "capture-recapture," to estimate the homicides not in either source.

For 1968-1976, individual data are from Marc Riedel and Margaret Zahn's *Trends in American Homicide, 1968-1978: Victim-Level Supplementary Homicide Reports*. For 1976-1992, Monkkonen used James Alan Fox's *Uniform Crime Reports [United States]: Supplementary Homicide Reports, 1976-1992*. After 1992, data have been regularly updated by the researchers at the ICPSR.

Annual counts — whether total homicides per year or more specific kinds of homicides per year — come from multiple sources. For post 1930 data, the FBI Uniform Crime Reports or the Supplementary Homicide Reports are the starting point. These have been supplemented with other official counts, ranging from the New York City Police Department to the City Inspector in the early nineteenth century. When there are discrepancies, Monkkonen opted for the source giving the higher count on the assumption that missing information is almost always going to bias towards an undercount. For most years prior to 1976, multiple sources -- police reports to vital statistics reports -- have had to be used in order to generate the more specific kinds of annual rates, say by gender of victim or by kind of weapon.

No regression analysis or other analysis has been done to this dataset by Monkkonen. This database allows one to access a large database of homicide cases in one jurisdiction over a lengthy period of time. This dataset does use coroner's indictments to determine cases of homicide which bears some relation to your research. Likewise, this dataset does contain relatively modern homicide information which is similar to our current project.

The Ohio State University: Historical Violence Database

<http://cjrc.osu.edu/researchprojects/hvd/>

Ohio State's Criminal Justice Research Center website contains a historical violence database that is co-directed by Randolph Roth (Department of History- Ohio State) and Douglas L. Eckberg (Department of Sociology- Winthrop University). Compiling the data made available through the research of many individuals, this database provides data on homicides from different parts of the world (Asia, Australia and New Zealand, Europe, and the United States) over various time periods ranging from as early as 1630 to as late as 1920. Although the primary data available focuses on homicides, there is disaggregated data available on other violent deaths (suicides, accidents, casualties of war) and serious assaults (sexual assaults, attempted murder, arsons).

Roth's individual research focuses on homicides from New England, New Netherlands, Ohio, Georgia, Florida, and Rockbridge County, Virginia over various periods of time. Roth obtained his homicidal data from both legal records (inquests, case files, docket books, minute books, and prison records etc.), and non-legal sources (newspapers, diaries, oral tradition recorded in early town histories, etc.). The list of homicides obtained from the legal records was then matched with the list of homicides found in the non-legal sources in order to determine an estimated homicide rate for the different locations being studied during each specific period of time. To determine the estimated population, Roth obtained data from published volumes of the U.S. Census and supplemented this with various other data sources. In addition, Roth provides estimates of the ethnic composition of each area during the particular period being researched.

Along with Roth's work, this database incorporates the data from several other research projects in order to provide a more comprehensive history of violent behavior. Through this type of meta-analysis, this database provides statistical tables, figures, and analyses that allow one to compare homicide rates to a variety of different factors such as race composition of the population, gun ownership, urbanization, and various other trends. Since data is being incorporated from various research projects, the sources used to acquire homicide data depends on the methodology chosen by each different researcher.

This database is disaggregated based on the different parts of the world the data involves (Asia, Australia and New Zealand, Europe, and the United States). Once a particular region of the world is selected, the database provides a user access to specific studies that the database incorporates. Along with providing either a website link or a citation of where the original research could be accessed, the database provides worksheet files, spreadsheet files, image files and/or text files of original documents depending on the data available.

The worksheet files contain the researchers' coded notes of each case including information such as the details of the crime (type of weapons used, if there was an indictment, court proceedings etc.), the suspect and victim (ethnicity, gender, if they're literate etc.) and all of the sources each researcher used to gather the information such as legal records, newspaper articles, census, and genealogy.

The spreadsheet files consist of an excel spreadsheet in which the state, county, source of the information, category (household, non-domestic, third party etc.), victim (child, infant, adult etc.), victim and assailant race and gender, and whether there was an indictment are all coded. There is a sample codebook to explain what all of the abbreviations in the spreadsheet stand for. This type of spreadsheet allows researchers to identify the specific homicidal trends they are interested in and helps them to organize the data.

The image files gives individuals access to unaltered, photographic images of original documents on violence: coroner's reports, homicide detective reports, newspaper articles, indictments, etc. Also, the database contains text files of original documents which are printed original documents on violence maintained in Microsoft Word. These image and text files are available for some but not all of the data sets.

Although in some cases the research done by other researchers was simply converted into worksheet and spreadsheet, some of the research was expanded upon. For instance, Roth and his students worked through the Chicago Tribune to annotate cases in the Chicago Homicide Project

database and to find information on additional homicides. To date, they have discovered 30% more homicides, 1879-1885, than appear in the database, and they have revised the descriptions of individual cases where new information came to light after detectives filed their initial homicide reports.

This dataset focuses on violent deaths further in the past rather than more modern homicide cases. In addition, although this dataset contains some indictments, many of the cases of violent deaths did not result in indictments. Roth's database claims to be a "living dataset" incorporating data obtained from many different researchers and says that it will add new data from new jurisdictions as it becomes available and encourages researchers to contribute to the website and data analysis.

Statistical tables, figures, and analysis exist on the data they have obtained. Some of the research does include image and text files of original documents when available which allows researchers the raw data necessary to conduct their own analysis. This database does offer analysis comparing homicide rates to other factors.