TECHNICAL ISSUES: CODING PROTOCOLS, FORMATTING, DATES AND
COMPATIBILITY OF FILES

The records were entered by the police both chronologically and alphabetically. At the transcription phase sequential record numbers were assigned according to their order in which cases appeared in the police Homicide Books.[11] Therefore, the case entries in the police books are neither wholly chronological nor wholly alphabetical. The record numbers are roughly chronological by letter of the alphabet. The earlier numbers refer to earlier cases, and to cases early in the alphabet in that chronological segment. Yet there are clumps of cases, such as those involving "unknown" victims, which all appear under the letter "U". The entry of a sequential unique record number for each case was critical to track cases. The record number entered at the transcription stage has been the case identifier for all subsequent rounds of research.

The records were kept according to the following system: In a large book all of the victims whose surname began with A would be entered chronologically by day, month and year of death under the letter A. For example, Armstrong, James, Dec. 21, 1870, [followed by additional information on the homicide] and the next entry might be Abbott, Mary, April 30, 1873, followed by Abbott, Henry, [no relation to Abbott, Mary] June 13, 1875, the next chronological entry under the letter A.

This system created a double index: the date of death chronologically in one column and the name of the victim alphabetically in a parallel column. It is a victim-based, chronological system. The name of the victim and even the approximate date of death are sufficient to locate the case with ease. In the coded file it is now possible to identify cases by name of defendant, as well as by other variables. Within each letter classification the entries are not alphabetical but chronological, by date of death of the victim, to the end of the alphabet for that particular book. There are three separate homicide books, with the number of entries increasing sharply after 1918. Each new book started over at the beginning of the alphabet for that period. Each initial entry was followed by a space, allowing for the subsequent recording of information on the disposition of the case and the sentence. Time presumably elapsed between entries of the initial date and

tification and analysis. The frequencies and other aggregate statistics reported in this Introduction are from the combined file with duplicates removed, as of September 2002.

[11] There were a few instances where names or dates were not in chronological order. Record numbers correspond to the order in which cases were entered in the books. Book I covered the period 1870-1910 and contains cases 1-2847; Book II covered 1911-1924 and contains cases 2848-5624; Book 3A included cases 5625-7480; Book 3B cases 7481-9391; and Book 3C cases 9392-11453.

subsequent developments.

Here are some typical examples, preceded by the added record number:

[310] 1910, Aug. 18, Burks, James. 34 yrs. old. Shot dead during a quarrel at 325 N. Wood St. by Will Church who escaped. 28 prect. Kenneth Taylor, alias Frank Smith, colored brought back from Holland, Mich., and held on $1,000 bond [EO1]. Aug. 18, by Judge. William, alias "Ham" Church, colored, arrested and held by coroner's jury Aug 22, 1910. Joliet Pen. Jan 28, 1911, for manslaughter. Judge Kavanagh.

[322] 1878, Aug. 12 Connors, James, young hoodlum, died, home, 537 S. Union St., as result of bullet wound received Aug. 10, while attempting to rescue two prisoners under arrest by Off. John McTigue of West Twelfth St. Station.

[399] 1900, May 29, Cameron, Frank, notorious swindler, fatally shot, Illinois Central Depot, Harvey, Ill., by Thomas O'Neill, who was arrested. Cameron died in Chicago Hospital May 31. O'Neill was held in bonds of $20,000, June 11, by Judge Dunne and was acquitted by Jury in C.C. (Judge Smith.) Nov. 28, 1900.

The system instituted by the police for recording cases was orderly and simple, allowing for cases to be retrieved easily by name of victim and/or date of offense.[12] These entries illustrate the richness of the data set and some typical ambiguities. All three cases include date of offense, name, gender, age of victim, weapon, and place of homicide. Two cases include the name of defendant. Two of the three cases include details and dates for arrest and disposition and the name of the judge. The name of the judge was presumably another tracking variable for the police keeping the log, telling them where they needed to show up to testify.

There are also typical ambiguities in these cases, the kinds of issues which were resolved in the weekly coders' meetings. For example, in case No. 322, is Officer McTigue the shooter? There is no indication of any arrest or disposition for this killing, or if it was by a police officer. Nor is

---

[12] The system of recording developed by the police allowed for a large amount of information to be compressed into a few lines: date of offense, age, gender and race of victim and defendant, names of victim and defendant, circumstances of the offense (e.g. "during a quarrel"), sometimes the name of the arresting officer and the name of the sentencing judge, and in many cases several dates related to legal decision making, in addition to the date of crime; for example, the date of arrest, the date of the coroner's verdict, occasionally the amount of and/or date of the grant of bond; the date of the grand jury decision, and the date of sentence and the receiving institution.

there any reference to another person, or an escapee being the shooter. That Officer McTigue was the shooter seems to be a reasonable inference. A contemporaneous newspaper account might resolve this ambiguity and add additional facts.

Similarly, in cases No. 322 and No. 399, a code was created for a category "known to the police," in order to capture organized crime cases and the fact that many victims were known to the police.[13] Both cases would have been coded positively on that variable, although from the records it appears that these were different kinds of persons whose activities were "known to the police." A later researcher may identify and then recode or reinterpret these cases and impose a different structure upon that information.

The homicide books seemed to have had as one purpose the creation of a record of accurate and precise information for police officers who testified at coroner's hearings, grand jury hearings or trial. This may be why those dates are recorded. The details as to dates and the names of arresting officers and judges suggests that the records served as cribs and authentication for testifying officers, who perhaps had no other records, or allowed an officer who wasn't present at the arrest or homicide scene to testify accurately as to events. The detail and accuracy of these records was undoubtedly encouraged and reinforced by the presence of civic organizations such as the City Crime Commission and by the periodic movements towards reform of city and county government during the period.[14]

Neither the date of the offense, nor the date of death, nor the alphabetical name of the victim alone was an organizing principle for the entire data base.[15] Record number became the link between all versions of the

---

[13] This dataset will be valuable for researchers studying the history of the police force as an institution, and the social and cultural practices surrounding law enforcement during the period. For example, the fact that so many defendants and victims were "known to the police" implies that the police had a relationship with that community, and were not anonymous outsiders.

[14] Researchers studying police procedures will find the arrest data a rich source of information about the practices of the police, especially if these data can be supplemented by newspaper reports of arrests or crimes. Surprisingly, the number of entries for date of arrest is only 3,603 in the Defendant file, although there are over 9,000 cases with a name of defendant. The practice generally seems to have been to have the arrest occur at the precinct, even if the arrest was at the scene.

[15] Typically the police file includes the address where the homicide occurred, or where the victim was found, weapon or manner of death, and often the surrounding circumstances of the offense; for example, fight in a saloon, accidental shooting, husband killed wife and children. In many cases there is information on the motive or stated reason for the killing; for example, shot dead while resisting arrest; "footpad" (a highway man who robs on foot); from assault with wooden leg by a fellow prisoner; struck and instantly killed by an automo-

data base: the text file of the case summary; the Crime File, the Victim File, the Defendant File, and the SDCI, all four of the quantitative files. Record number was entered separately as a tracking number in each data file. Each separate case file also had its own identifier, a date or a name, in addition to the record number.[16] The supplementary data collection instrument (SDCI) was created to capture aspects of the crime for specialized interest and to build in verifying codes for important variables.[17] For example, relationship between victim and defendant was coded in four different ways in four different variables.[18]

The inclusion of identifying dates as dates, when they appeared in the log, is important for linkages to other data sets, such as the coroner's files, court records, or files in the prison system. Transferring the dates from and

---

bile at cor.[sic]; shot and killed his wife and son and then committed suicide, and so forth. These brief descriptions are recorded in the string, or text files, preserving the original expressive language.

For those cases where there was no date of death or date of homicide, the fifteenth of the month was entered as the default date. There were few such cases, less than twenty-five. Those cases were usually the 'unknown" victims. If there is no name of the victim, the record indicates unknown victim; however, gender, and race and often estimated age of victim, were included in that case record. The proportion of unknown victims is small. A total of 193 victims were listed as "unknown" in the victim file, and an additional fifty-two records are blank, and presumably unknown, for name of victim. That totals 245 victim files with no names out of a total of 11,341 entries in the victim file, with 11,076 named victims in the entire victim file. Information from the coroner's records might increase that number of unknown victims, or the contemporaneous coroner's records might supply an identification of some of those victims.

[16] At the weekly coding meetings additional numerical codes could be added to the instruments consistently, and interpretations would be finalized. The establishment of a protocol in which the entries were recorded exactly as they appeared in the handwritten records, without any alteration or interpretation, insured that later investigations and interpretation could be imposed uniformly, according to the research design or interest of the researcher. Data was initially in Access and then transferred into Excel, accommodating both text files and numerical files. In addition to creating a new quantitative file, we preserved identifying and descriptive language and, very importantly, identifying dates.

[17] Examples would be offenses related to prohibition, a more detailed breakdown for accidents and manslaughters, details on areas of special interest, strikes, riots, organized crime, and a recode and check on other variables of interest. Within these three separate instruments were overlapping data points for important variables such as, relationship between defendant and victim, circumstances of the offense, location of the offense, weapon, ethnic identification of victim or defendant, or variables relating to the social status of the victim or defendant. The purpose of creating overlapping data points was to create internal checks on coding and data entry.

[18] Relationship of victim and defendant was quantified in four separate precoded variables, including the current definition of relationship of victim and defendant which is used in the FBI Uniform Crime Reports. Quantitative variables were created for relationship of victim and defendant, then defendant and victim, and we created a number of subcategories for stranger relationship and relationships outside of the family.

into Access for dates prior to 1900 turned out to be an especially thorny technical problem, requiring the assistance of highly skilled programmers.[19] Year of homicide is a critical variable in the analysis of incidence, in testing hypotheses about fluctuations in the rate of homicide, and in linking clusters of homicides[20] to neighborhoods and specific historical events, such as strikes and riots and the imposition of Prohibition.[21] The analysis of the incidence of homicide has traditionally focused upon the month, and the day of month, the day of the week, and even the time of day, which was rarely included here. For example, an analysis of alcohol related cases could focus on day of the week or day of the month.[22] The dates, especially the date

---

[19] Access could not read the dates prior to 1900 as dates. And if the dates were to be useful as analytical variables, they had to be read accurately and sequentially as dates. The Project especially wishes to thank Bill Hayward, Director, Analytical Studies and Michael Jones, IT Support Specialist, both of Northwestern University Information Technology, and Mortezema A. Rahimi, Vice President for Information Technology at Northwestern University, for their expert assistance in transforming the dates and in combining all four of the files into a single file for analysis. When the right formula was discovered, James Anderson reentered the dates for more than 1,100 cases prior to 1900. The dates were always able to be read in Excel, but they could not be analyzed in SPSS.

[20] The work of ethnographers and historians will be critical to test hypotheses about these relationships. *See, e.g.*, GLEN E. HOLT & DOMMINIC A. PACYGA, CHICAGO: A HISTORICAL GUIDE TO THE NEIGHBORHOODS—THE LOOP AND SOUTH SIDE (1979). This is an opportunity to test hypotheses about how much these controversial single violent episodes, such as the Haymarket riots, contributed to the number and character of homicides during the period.

[21] This dataset offers a new perspective from which to analyze the social and political impact of Prohibition, and perhaps to gauge police corruption related to Prohibition during the period. The dataset often identifies "saloon" as the location of homicide, and that identifier with the address of the homicide allows for the possibility of building on existing social and historical research and for cross referencing address with census data. *See, e.g.*, PERRY DUIS, THE SALOON —PUBLIC DRINKING IN CHICAGO AND BOSTON (1983)..

[22] The program has now added day of the week to all records of date of death, and there may be an association between day of the week and homicides in saloons. Consider the following:

> The size of the vice business shocked reform groups. In 1881, the Citizens' League of Chicago stationed observers in five tough saloons in the South Side Levee and saw, by their count, 11,608 men and 1,007 women enter and leave in less than five hours. Another survey conducted a year later concluded that within twenty square blocks were 500 saloons, six 'variety theaters,' more than 1,000 'concert hall,' fifteen gambling houses, between fifty and sixty poolrooms, and fully 500 houses of ill fame harboring over 3,000 women. The police patrolled only the borders, carrying out raids when public opinion demanded it or when a murder or some other extraordinary event makes the need for a visible corrective possible.

Id. at 237.

The streets near to the Stockyards contained many saloons whose practice was to offer free food to workers who purchased liquor. These places became social and political centers, and also places where homicides occurred. Attempts to untangle the relationship between Prohibition and the increase in homicide during the 1920's depends upon having ac-

of the homicidal incident, had to be able to be read sequentially without interruption. An especially valuable feature of this data set is its extraordinary length. To make the data set continuous, the dates prior to 1900 had to be able to be read as sequential dates, as dates which identified the day of the week, and the day and the month.

Dates remained critical identifiers, and as such needed to be able to be read sequentially in Access, to be retrievable as dates in Excel, to be able to be used for time series analysis in SPSS, and to be retrievable as specific dates. Dates had to be transferable into other statistical packages. And the dates had to be capable of being linked to historical events and dates in other data sources and time series.

Reform governments took office at particular times, with the goal of reducing corruption in government. Hopefully there are a sufficient number of dates in the data set to support a research design comparing case progression through stages of legal decision tracking when different political regimes were in power: for example, time from incident or arrest to coroner's decision, to grand jury, to trial and sentence, or execution, in 1885-1900, in comparison to 1910-1925.[23]

CODING PROTOCOLS FOR RACE OF VICTIM AND DEFENDANT

Classifications according to race are at best problematic. The rules for racial classification changed during the time period and within the data set itself. There was no consistency among contemporaneous commentators. Many adopted the term Negro, which appeared more frequently later in the data set.[24] In the earlier years "colored" ("col.") is the most frequently used terms for Blacks.[25] The United States Census changed its method of classification of race and ethnic origin several times during the period.[26] The

---

curate dates and data on location of homicide. Similarly, family homicides or incidents of domestic violence can now be analyzed in terms of day of the week, or month. Were people more likely to get drunk and kill on a Saturday, or at the end of the month, or on payday?

[23] *See* Derral Cheatwood, *Capital Punishment for the Crime of Homicide in Chicago: 1870–1930*, 92 J. CRIM. L. & CRIMINOLOGY 843 (2002) (forthcoming in this volume).

[24] Towards the end of the period "colored" was replaced by "negro," although not consistently. Occasionally there will be a reference to "half breed." The classifications used by the police did not follow the U.S. Census categories for race and ethnicity during the period. Nor did they follow the terminology of the Chicago Crime Commission or other contemporaneous studies of crime which addressed crime and race, or crime and ethnicity. In this dataset "colored" and "negro" were both coded as African American under ethnicity and as black under race.

[25] In this publication, different researchers use different terms for these racial categories. In this Introduction, the term blacks is used to refer to those victims and defendants designated "colored" or "negro" in the dataset.

[26] The term "negro" actually appears for the first time in the 1930 census. In the 1890,