

APPENDIX A: METHODOLOGY AND RESEARCH PROTOCOLS

TRANSFORMING THE HANDWRITTEN POLICE RECORDS INTO A QUANTITATIVE DATA SET: METHODOLOGY AND PROTOCOLS

An article describing the laborious, heroic work at the Illinois State Archives which preserved these original records and put them onto microfilm was published in the *Chicago Tribune* in August of 1998.¹ This was the announcement of a new source of information on homicide with exciting research possibilities.² Seeing the actual homicide books at the Chicago Police Archives weeks later brought the realization that this set of cases constituted a systematic record of all homicides over an uninterrupted sixty-year period and as such presented a special opportunity from homicide researchers, criminologists and historians.³

All of 1999 was required to transcribe the microfilm of handwritten files into a text file, with the important addition of entering a record number for each case. At this initial stage the transcription protocol was simple:

¹ See Charles M. Madigan, *Crime—Chicago Style*, CHI. TRIB., Aug. 19, 1998, at 1:

The Archives got the books from the Chicago Police Department at 11th and State Streets, where it found them sometime in the 1980's when it was making a routine visit . . . They were in terrible shape. The paper crumbled in your hands, Hopkins [Dorothy Hopkins, the conservator from the State Archives] said, and certainly would not have been able to withstand much more handling. As she removed each of the old pages, cleaned them, de-acidified them, sealed them and prepared them for remounting, she found herself reading more and more about bloody murder in Chicago and wondering what it must have been like . . .

Id. This Project is greatly indebted to the Chicago Police for keeping and preserving these original records, and to the Illinois State Archives, for the laborious work of preserving these records and for making these archives generally available.

² An article in the Chicago Tribune on August 19, 1998 included contemporaneous photographs. The photographs published in the August 19, 1998, article were from the Chicago Tribune photographic archives. To date no police photographs from Chicago have been located. An archive of homicide police photos from New York City in the post World War I period had been a rich source of historical and archival information. See LUC SANTE, EVIDENCE (1992). Those photographs of homicide victims in situ, taken by the police for the purpose of collecting information for a murder investigation, were unexpected reservoirs of historical information, richly expressive of the contemporaneous society: "Nothing in the reams of photographic documentation I'd sorted through . . . had prepared me for this. Here was a true record of the texture and grain of a lost New York, laid bare by the circumstances or murder. Lives stopped by a bullet were frozen by a flash of powder. . . ." *Id.*

³ Even in the form of a microfilm record, as it is now available from the Illinois State Archives for a nominal fee, this archive of handwritten records, more than 11,000 entries over sixty years, is a unique resource for teachers, scholars and the general public.

transcribe the records *exactly* as they appear in the handwritten police log, including misspellings, idiosyncratic punctuation, seeming contradictions in the information about cases, for example, the date of arrest entered as prior to the date of the homicide, and create a sequential record number for each separate case entry, including a separate record number for each “unknown” victim. A threshold decision was made to create record numbers by case entry in the police file, and not to create separate files for multiple victims or multiple defendants. Instead independent variables recorded the number of victims and number of defendants. At the conclusion of this phase there was a sequenced text file of more than 11,400 homicide cases,⁴ ordered by record number, date of death, and name of victim.

The entry of a sequential record number for each case at the transcription was the key to the next stage: the transformation of the text file into a quantitative data file. The Project was fortunate in having a single, highly trained person transcribe the entire database under this protocol.⁵ There was close control over the process of transforming the handwritten micro-filmed files into a text file and remarkably few errors have emerged. This laborious process allowed for a preview of the contents of the file and gave a sense of the uniformity and consistency of the cases as the data collection instruments were developed.

The data collection instruments were designed to capture the standard variables of interest to criminologists, historians and legal scholars. Specially crafted, new variables were introduced to address idiosyncrasies specific to the time period and to this dataset.⁶ From the outset, the Project was committed to preserving the contemporaneous integrity of the case summaries, by transposing the exact language, even the informal language, into the

⁴ The initial number of more than 11,400 included some duplicate entries; for example, cases involving police as victims appeared as double entries in the homicide books. When duplicates were removed in August of 2002, the dataset numbered slightly over 11,000. See tbl. 1, *infra* Appendix B.

⁵ The entire database of 11,400 plus entries was transcribed into a Word file by Elizabeth Olds over the course of nine months. The Project benefited enormously from having one highly skilled, careful researcher perform this often depressing and tedious work. Periodic meetings established transcription protocols, and during the transcription stage the data collection was planned.

⁶ The decision to enter data in Access for conversion into Excel was guided by its ease of entry and the availability of string variables for key words, for names, addresses and for other words preserving language expressive of the character of the offense. Later data were transferred into Excel, SPSS and other statistical programs, depending upon the preferences of the individual researcher. For this first round of publications all researchers began with the individual Excel files, prior to the preparation of the combined version in SPSS. Each researcher then crafted a subset according to the topic of interest. The next set of papers will begin with the combined SPSS file.

quantitative file. String files for words, language, and phrases describing the circumstances, persons and manner of killing were incorporated into the quantitative design.

Three independent coding instruments were initially set out.⁷ Weekly meetings with coders and managers addressed contradictions and resolved ambiguities and variations in data interpretation as much as possible.⁸ The Supplementary Data Collection Instrument (SDCI) was created to capture information on prohibition and organized crime cases, and on other subsets of special interest. Surprisingly, only 100 cases have been initially identified as being related to Prohibition.⁹ After many technical challenges, all four files are now combined in a single file, containing all cases from 1871 to 1930, capable of being read by the standard statistical programs.¹⁰

⁷ They divided the data collection as follows: data pertaining primarily to the circumstances of the crime, (CRIME FILE), the characteristics of the victim, (VICTIM FILE), and the characteristics of the defendant and the disposition of the defendant's case (DEFENDANT FILE). The purpose of having three files was to allow for cross checks on accuracy of recording information and coding, and also because different kinds of researchers would be interested in different aspects of the case. Each file began with the record number as an identifier, and overlapping variables were included in each instrument as a check on coding and data entry. The consistency of the record number was a key check on coding and data entry. A Supplementary Data Collection Instrument (SDCI) was created a few months later with some alternative formulations of important variables: for example, relationship between victim and defendant, age of the victim; and to capture variables of special interest, for example, prohibition and mob related cases, a detailed breakdown of accidents and manslaughters by type and character. The initial programs for the pre-coded data collection instruments were written by Denise Duffy, then at Information Services at Northwestern University. The SDCI was designed later by the authors of this Introduction.

⁸ Coding began after pretesting the instruments in the spring and summer of 2000, with law students and college students as coders. Regular meetings were held to institutionalize consistency and to identify other problems. The case summaries were available as a reference, making it easy to systematize interpretations.

⁹ That number may increase with a more focused analysis. The purpose of the SDCI was to identify and disaggregate special subsets of the database, for example, to achieve greater precision in the coding of manslaughters, to reformulate and duplicate some key variables, such as relationship between victim and defendant, in a form more closely aligned with current definitions used by the FBI and other criminologists, and to comb the database for indicators of governmental corruption and improper influence of organized crime on the criminal justice system, especially during prohibition. John H. Lyle, a felony courts judge in the 1920's and member of the Chicago City Council reported that 530 gangsters were slain in the Chicago area, including 190 murders in 1920 and 50 gang slayings in 1924; and 399 murders in 1928. JOHN H. LYLE, *THE DRY AND LAWLESS YEARS* 122 (1960). The Illinois Crime Survey, on the other hand, reports 45 gang murders in the City of Chicago in 1926, and 37 in 1927. Arthur Lashly, *Homicide (in Cook County)*, in *THE ILL. ASS'N FOR CRIMINAL JUSTICE, THE ILLINOIS CRIME SURVEY* 611 (John H. Wigmore ed. 1929).

¹⁰ Because this combined file was not available to the researchers writing papers for this volume, each researcher worked with an individually crafted subset of cases. Each paper describes how that particular set of cases was created, and what methods were used for iden-