ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

ICPSR 4351

Uniform Crime Reports [United States]: Supplementary Homicide Reports, 1976-2003

James Alan Fox
Northeastern University. College of Criminal Justice

Codebook

Inter-university Consortium for Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

Table of Contents

HOMICIDE TRENDS IN THE UNITED STATES

WEIGHTING AND IMPUTATION PROCEDURES FOR THE 1976-2002 CUMULATIVE DATA FILE

Bureau of Justice Statistics

James Alan Fox
BJS Visiting Fellow, and
The Lipman Family Professor of Criminal Justice
Northeastern University

and

Marianne W. Zawitz
BJS Statistician

April 9, 2004

Most of the data used in *Homicide Trends in the United States* are from the FBI's Supplementary Homicide Reports (SHR) which provides detailed, incident-level data on nearly all murders and non-negligent manslaughters in the United States. These reports include information on the month and year of an offense, on the reporting agency and its residential population, county and Metropolitan Statistical Area (MSA) codes, geographic division, and population group, on the age, race, and sex of victims and offenders, and on the victim-offender relationship, weapon use and circumstance of the crime. Except for some slight modification in 1980, the record layout and variable definitions in the SHR data have remained unchanged since 1976 when the reporting format underwent a major revision.

This document describes adjustments used in this analysis for handling missing data in the SHR that result from agency failure to file reports and incomplete records that are missing certain information about the incident, victim or offender.

Correcting for Missing Records

Law enforcement agencies voluntarily report both Uniform Crime Reports (UCR) summary data and SHR incident data to the FBI on a monthly basis. The offense data in the UCR includes counts of murder and nonnegligent manslaughter as well as seven other Index crimes. The number of murders and nonnegligent manslaughters is based on the number of victims. For the most part, each month agencies report a total number of offenses in each Index crime category. In some instances agencies may not report each month or may not report at all which results in missing data when aggregating to the state or national level. After imputing for missing data, the FBI publishes estimates for the nation as a whole as well as for individual states for all Index offenses, including murder and nonnegligent manslaughter. (For additional information on the imputation methods used by the FBI, see "Bridging Gaps in Police Crime Data" by Michael Maltz, 10/99, NCJ 176365, http://www.ojp.usdoj.gov/bjs/abstract/bgpcd.htm) These annual UCR state and national estimates of homicide volumes are used as benchmarks for assessing the completeness of the SHR data file and to adjust SHR victim or offender counts as needed.

Not all of the murders and nonnegligent manslaughters reported in the UCR are included in the SHR. The SHR file appears to be just over 90% complete, although, as shown in Table 1, the level of completeness of the SHR has generally diminished in recent years. To correct for missing SHR records for the national and regional numbers used in this analysis, the UCR murder and nonnegligent manslaughter estimates have been used to weight SHR records (that is, state and national weights have been assigned so that tabulations total to the respective estimated counts). This weighting process assumes that the missing records are not systematically different from those available in the file. While the systematic exclusions for certain years of a state like Florida or a city like Washington, D.C. may cause some concern, the fact that missingness occurs in both large and small jurisdictions lends support for applying these benchmark weights.

Correcting for Incomplete Records

Even for the 90% of SHR records that are available for analysis, certain variables have non-trivial rates of missingness. At one extreme, characteristics of an agency (e.g., region, population group) are always complete. Victim data, though sometimes missing, are absent at such a low rate that standard approaches for handling missing data (specifically, listwise deletion) hardly biases analyses of patterns and trends in victimization. Specifically, as shown in the top panel of Table 2, victim age is missing in 1.70% of homicides, race in 1.00% and sex in 0.13%. Overall, 2.45% are missing on at least one of these measures.

At the other extreme, a significant problem in using SHR data to analyze offender characteristics is the sizable and growing number of unsolved homicides contained in the data file. Overall, 26 percent of the SHR offender records describe the perpetrator as unknown (based on situation codes), and this percentage has grown from just under 20 percent in 1976 to nearly 30 percent by the mid-1990s. Even when the offender is known to the police, moreover, the characteristics about the offender may still be unavailable. Table 2 shows specifically that as many as 26.39% of offender records contain no information about the perpetrators, and 31.29% of records are not complete in terms of the assailant's age, race and sex.

Ignoring unsolved homicides and missing offender data seriously understates calculated rates of offending overall and by particular sub-groups of the population, distorts trends over time among these same sub-groups, and biases observed patterns of offending to the extent that the likelihood of missingness of offender data is associated with offender characteristics.

While it is not possible to determine directly whether case solution and thus missingness in offender data are associated with offender characteristics themselves, some indication about the pattern of missingness can be derived from the examining the extent to which the likelihood of case solution is related to victim and incident variables. As shown in Table 3, case solution rates are lowest for homicides against young adult victims as well as for elderly victims. Solution rates are also lower for incidents involving black or male victims. As expected, solution rates decrease with increasing population size and urbanness. In part as a consequence of urbanness differences, solution rates in the South are much higher than other regions. Finally, whether or not a gun is used to commit the homicide does not appear to impact upon rates of case solution.

A weighting strategy based on available information about the victims (age, race and sex) murdered in both solved and unsolved homicides is used to adjust for missing offender data in *Homicide Trends in the United States*. Through this imputation algorithm, the demographic characteristics of unidentified offenders are inferred on the basis of similar homicide cases--similar in terms of the victim and incident profile--that had been solved. In other words, offender profiles for unsolved crimes are estimated based on the offender profiles in solved cases matched on victim age, sex and race, region, urbanness, weapon and circumstances.

The weighting procedure is accomplished by establishing adjustment groups within a large, multi-directional cross-tabulation. For each cell, we tally the number of offenders (N_c) and the number having complete offender data (n_c), and use as a weight the inverse proportion of complete cases:

$$w_c = \frac{N_c}{n_c}$$

Next these adjustment cell weights are applied to the offender records based on their cell membership (i.e., based on victim age, race, sex, location type, weapon and circumstances) and whether or not the offender information is complete. That is,

$$w = \begin{cases} W_c & \text{if offender data are complete} \\ 0 & \text{otherwise} \end{cases}$$

Finally, the weights are adjusted or “raked” post-stratification so that marginal weighted counts by year and state match fixed values.

This weighting approach is applied to the entire offender record so that cases with missing offender age, race or sex are excluded by virtue of their zero case weights. As a consequence, partial offender information is discarded, causing some slight inefficiency in the approach. While it would be possible to retain partial offender information, this would require separate weights for each offender characteristic. Finally, all non-zero weights are further increased slightly to account for the small percentage of cases unassigned to any adjustment cell because of their being missing on one or more victim characteristic.

In any analysis of weighted data, offenders with incomplete age, race or sex information are dropped due to assigned zero weights. Offenders with complete age, race and sex information all have weights at or above 1.0, and become proxies for excluded cases, matched on victim characteristics, state and year. Thus, for example, an offender with an imputation weight of 2.0 would count in any analysis as if he/she were two offenders. The entire distribution of the weighting variable is shown in Table 4.

Table 5 demonstrates the impact of applying the imputation weights by comparing the distribution of offender characteristics (age, race and sex) using the adjustment weights with those using listwise deletion of missing data. The approach boosts the abundance of youthful offenders; offenders under the age of 25 represent 47.3% of the imputed distribution, compared to 44.9% for the distribution using listwise deletion. The percentage of black offenders grows to 52.1%, compared to 51.1% without weighting. The sex distribution, which by any measure overwhelmingly favors male perpetrators, shifts slightly from 88.0% to 8865% (clearly indicating a ceiling effect governing this percentage).

Table 1: SHR Records Pertaining to Incident, Victim and Offender Counts

Year	All Homicides		Murder & Non-Negligent Manslaughter					UCR Benchmark Weight
	SHR Data File		SHR Data File			UCR Estimated		
	Records	Victims	Incidents	Victims	Offenders	Victims	Offenders	
1976	16,744	17,406	15,951	16,605	17,995	18,780	20,352	1.1310
1977	17,825	18,586	17,277	18,032	19,228	19,120	20,388	1.0603
1978	18,546	19,308	17,957	18,714	20,075	19,560	20,983	1.0452
1979	20,576	21,417	19,756	20,591	22,247	21,460	23,186	1.0422
1980	21,911	22,786	21,002	21,860	24,859	23,040	26,201	1.0540
1981	20,152	20,931	19,284	20,053	21,794	22,520	24,475	1.1230
1982	19,413	20,288	18,622	19,485	21,219	21,010	22,880	1.0783
1983	18,690	19,426	17,954	18,673	20,315	19,310	21,008	1.0341
1984	17,168	17,858	16,574	17,260	18,781	18,690	20,337	1.0829
1985	17,348	18,131	16,763	17,545	18,831	18,980	20,371	1.0818
1986	19,089	19,849	18,510	19,257	20,722	20,610	22,178	1.0703
1987	17,747	18,509	17,205	17,963	19,396	20,100	21,703	1.1190
1988	17,846	18,546	17,277	17,971	19,841	20,680	22,832	1.1507
1989	18,812	19,588	18,184	18,952	20,898	21,500	23,708	1.1344
1990	20,404	21,246	19,451	20,273	22,889	23,440	26,465	1.1562
1991	21,817	22,656	20,863	21,676	24,807	24,700	28,268	1.1395
1992	22,753	23,793	21,701	22,716	25,382	23,760	26,549	1.0460
1993	23,320	24,336	22,175	23,180	26,116	24,530	27,637	1.0582
1994	22,231	23,246	21,093	22,084	25,075	23,330	26,490	1.0564
1995	20,099	21,193	19,154	20,232	22,667	21,610	24,211	1.0681
1996	17,053	17,829	16,203	16,967	19,350	19,650	22,410	1.1581
1997	15,929	16,726	15,052	15,836	17,918	18,210	20,604	1.1499
1998	14,313	14,975	13,556	14,209	16,177	16,970	19,320	1.1943
1999	12,792	13,511	12,299	13,011	14,587	15,522	17,402	1.1930
2000	13,220	13,856	12,597	13,230	15,059	15,517	17,662	1.1729
2001	14,239	17,695	13,369	14,080	15,949	15,980	18,101	1.1348
2002	12,309	12,940	11,598	12,217	16,204	13,543	17,963	1.3263
1976-2002	492,346	516,631	471,427	492,672	544,909	545,720	603,826	

Table 2: Patterns of Missingness in Victim and Offender Data

SHR File	Characteristic Age	Race	Sex	Cases	Percent
Victim File					
	●	●	●	480,616	97.55%
	●	●	○	43	.01%
	●	○	●	3,555	.72%
	●	○	○	60	.01%
	○	●	●	7,052	1.43%
	○	●	○	12	0.00%
	○	○	●	809	0.16%
	○	○	○	525	0.11%
			Total	492,672	100.00%
Offender File					
	●	●	●	374,990	68.71%
	●	●	○	33	.01%
	●	○	●	2,637	.48%
	●	○	○	161	.03%
	○	●	●	21,740	3.98%
	○	●	○	120	.02%
	○	○	●	2,050	.38%
	○	○	○	143,989	26.39%
			Total	545,720	100.00%
With complete victim data					
	●	●	●	370,350	67.86%
	●	●	○	29	.01%
	●	○	●	1,352	.25%
	●	○	○	152	.03%
	○	●	●	20,885	3.83%
	○	●	○	114	.02%
	○	○	●	1,882	.34%
	○	○	○	137,865	25.26%
With incomplete victim data					
	●	●	●	4,640	0.85%
	●	●	○	4	0.00%
	●	○	●	1,285	0.24%
	●	○	○	9	0.00%
	○	●	●	855	0.16%
	○	●	○	6	0.00%
	○	○	●	168	0.03%
	○	○	○	6,124	1.12%
			Total	545,720	100.00%

Key: ● Data available; ○ Data missing

Table 3: Case Solution Rates by Victim, Location and Incident Characteristics

Variable	Category	Solved Cases	Percent	Unsolved Cases	Percent
Victim Age	0-13	21,751	86.8%	3,314	13.2%
	14-17	19,203	71.1%	7,794	28.9%
	18-24	86,071	68.3%	40,030	31.7%
	25-34	107,268	69.5%	46,964	30.5%
	35-49	87,693	72.6%	33,032	27.4%
	50-64	35,386	71.3%	14,238	28.7%
	65+	19,694	68.3%	9,122	31.7%
Victim Race	White	196,819	72.6%	74,227	27.4%
	Black	172,241	69.1%	77,133	30.9%
	Other	8,007	71.9%	3,134	28.1%
Victim Sex	Male	282,702	69.6%	123,459	30.4%
	Female	94,366	75.3%	31,035	24.7%
Location	Large city	117,284	62.9%	69,093	37.1%
	Medium City	82,682	69.3%	36,668	30.7%
	Small city	47,544	78.1%	13,318	21.9%
	Suburban	81,955	74.4%	28,174	25.6%
	Rural	47,603	86.8%	7,241	13.2%
Region	Northeast	57,342	63.2%	33,440	36.8%
	Midwest	71,408	69.6%	31,148	30.4%
	South	166,179	76.4%	51,357	23.6%
	West	82,139	68.1%	38,548	31.9%
Weapon	Non-gun	136,360	71.6%	54,080	28.4%
	Gun	240,708	70.6%	100,413	29.4%

Table 4: Distribution of Imputation Weights

Weight	Weights	
	N	Pct
Exactly 0	175,370	32.1%
1.00 - 1.99	298,864	54.8%
2.00 - 2.99	55,485	10.2%
3.00 - 3.99	11,843	2.2%
4.00 - 4.99	3,296	0.6%
5.00 - 5.99	826	0.2%
6.00 - 6.99	33	0.0%
7.00 - 7.99	2	0.0%
8.00 +	1	0.0%
Total	545,720	100.0%
Mean	1.11	
Standard deviation	0.92	
Median	1.21	
75th %tile	1.60	
90th %tile	2.17	
95th %tile	2.65	
99th %tile	3.79	
Maximum	9.22	

Table 5: Offender Characteristics without and with Imputation Weights

Offender Characteristic	Without Imputation			With Imputation Weights	
	N	Pct	Adj. Pct	N	Pct
Offender Age					
Under 14	2,099	0.3%	0.5%	2,935	0.5%
14-17	41,241	6.8%	9.9%	63,881	10.6%
18-24	144,193	23.9%	34.5%	218,608	36.2%
25-34	121,156	20.1%	29.0%	173,039	28.7%
35-49	76,815	12.7%	18.4%	103,818	17.2%
50-64	23,982	4.0%	5.7%	31,088	5.1%
65+	8,141	1.3%	1.9%	10,458	1.7%
Missing	186,198	30.8%			
Total	603,826	100.0%	100.0%	603,826	100.0%
Offender Race					
White	205,868	34.1%	46.9%	277,458	46.0%
Black	224,425	37.2%	51.1%	314,318	52.1%
Other	8,658	1.4%	2.0%	12,050	2.0%
Missing	438,951	27.3%			
Total	164,875	100.0%	100.0%	603,826	100.0%
Offender Sex					
Male	390,771	64.7%	88.0%	535,214	88.6%
Female	53,220	8.8%	12.0%	68,612	11.4%
Missing	159,835	26.5%			
Total	603,826	100.0%	100.0%	603,826	100.0%

CODEBOOK FOR THE

SUPPLEMENTARY HOMICIDE REPORTS FILE, 1976-2003

James Alan Fox
College of Criminal Justice
Northeastern University
Boston, MA 02115
e-mail: j.fox@neu.edu

October 1, 2005

SUPPLEMENTARY HOMICIDE REPORTS
CUMULATIVE VICTIM AND OFFENDER FILES, 1876-2003
Variable Descriptions and Codes

Important note: The two cumulative SHR files (one structured by victim record and the other by offender record) contain cases for murder and non-negligent manslaughter, manslaughter by negligence, justifiable homicide, and killings related to the September 11, 2001 terrorist attack. In order to conform to published homicide statistics from the FBI and those available at the Bureau of Justice Statistics website (http://www.ojp.usdoj.gov/bjs/homicide/homtrnd.htm), only incidents of murder and non-negligent manslaughter (HOMTYPE = 1) should be utilized.

Variable *Codes*

STCODE Alphanumeric state code

These are equivalent to postal codes, except for Nebraska, (coded NE) to avoid confusion with the abbreviation commonly used for New England.

STNUMBER Numeric state code

ORI Agency ORI code

These are (unique within state) identification codes given each reporting agency. For most states, the first three digits represent the county number (ordered alphabetically) and the last two digits represent the reporting agency. The latter two generally are SP for State Police, 00 for Sheriff, 01 for county police, and on for each local police agency.

POPGROUP Population group and subgroup

Groups 1-7 are cities; 8-9 are counties. The second digit (applicable for groups 1, 8 and 9) is for subgroup.

1 Cities 250,000 and over
11 Cities 1,000,000 +
12 Cities 500,000-999,999
13 Cities 250,000-499,999

2 Cities between 100,000 and 249,999
3 Cities between 50,000 and 99,999
4 Cities between 25,000 and 49,999
5 Cities between 10,000 and 49,999
6 Cities between 2,500 and 9,999
7 Cities under 2,500
8 Non-SMSA counties
  81 Non-SMSA counties 100,000 +
  82 Non-SMSA counties 25,000-99,999
  83 Non-SMSA counties 10,000-24,999
  84 Non-SMSA counties under 10,000
  85 State Police
9 SMSA counties
  91 SMSA counties 100,000 +
  92 SMSA counties 25,000-99,999
  93 SMSA counties 10,000-24,999
  94 SMSA counties under 10,000
  95 State Police
0 Possessions (Puerto Rico, Guam, etc.)

DIVISION Geographic region

1 New England (CT, ME, MA, NH, RI, VT)
2 Middle Atlantic (NJ, NY, PA)
3 East North Central (IL, IN, MI, ON, WI)
4 West North Central (IA, KS, MN, MO, NE, ND, SD)
5 South Atlantic (DE, DC, FL, GA, MD, NC, SC, VA, WV)
6 East South Central (AL, KY, MS, TN)
7 West South Central (AR, LA, OK, TX)
8 Mountain (AZ, CO, ID, MT, NV, NM, UT, WY)
9 Pacific (AK, CA, HI, OR, WA)
0 Possessions

MONTH Month of offense

YEAR Year of offense (1976, 1977, 1978 etc.)

POP Population of place

COUNTY County code

SMSA SMSA/MSA code

SMSAIND		SMSA/MSA indicator
	0	Non-suburban
	1	Suburban

A suburban agency is defined as all group 9's and groups 4 thru 7 that have SMSA/MSA numbers and are not core cities

AGNAME		Agency name
STNAME		Abbreviated state name
HOMTYPE		Type of Homicide
	1	Murder & Non-negligent manslaughter
	2	Negligent Manslaughter
	3	Justifiable Homicide
	4	Related to 9/11/01 Terrorist Attack
SITUAT		Situation of offense
	1	Single victim/Single offender
	2	Single Victim/Unknown offender(s)
	3	Single victim/Multiple offenders
	4	Multiple victims/Single offender
	5	Multiple victims/Multiple offenders
	6	Multiple victims/Unknown offender(s)
INCNUM		Incident number

These are unique numbers per jurisdiction per month. That is, each month a jurisdiction numbers its first homicide 001, its second 002, and so on, until the end of that month. Therefore, every homicide can be uniquely identified by State, ORI, year, month, and this incident number.

VCTCNT		Number of victims in offense
OFFCNT		Number of offenders in offense
VICAGE, OFFAGE		Age of victim, Age of Offender
	00	Under one year old (formerly NB and BB)
	01-98	Age in years (99 and older coded as 98)
	99	Age unknown

VICSEX, OFFSEX		Sex of victim, Sex of offender
	1	Male
	2	Female
	9	Unknown
VICRACE, OFFRACE		Race of victim, Race of offender
	1	White
	2	Black
	3	American Indian
	4	Asian and Pacific Islander
	9	Unknown
WEAPON		Weapon used
	11	Firearm, type not stated
	12	Handgun--pistol, revolver, etc.
	13	Rifle
	14	Shotgun
	15	Other gun
	20	Knife or cutting instrument--ax, icepick, etc.)
	30	Blunt object--hammer, club, etc.
	40	Personal weapon--hands, feet, teeth, etc.
	50	Poison
	55	Pushed or thrown out window
	60	Explosives
	65	Fire
	70	Narcotics and drugs
	75	Drowning
	80	Strangulation--choking, hanging, drowning, etc.
	85	Asphyxiation
	99	Unknown
RELATION		Relationship of victim to offender

A man killing his spouse, for example, is properly coded as "wife" (02). Note: This gives the relationship of the first victim to the offender(s). Consequently, in multiple-victim incidents, e.g., a family massacre, certain misleading values may result.

Within family:

01	Husband
02	Wife
03	Common-law husband
04	Common-law wife
05	Mother
06	Father
07	Son
08	Daughter
09	Brother
10	Sister
11	In-law
12	Stepfather
13	Stepmother
14	Stepson
15	Stepdaughter
16	Other family

Outside family but known to victim:

21	Neighbor
22	Acquaintance
23	Boyfriend
24	Girlfriend
25	Ex-husband
26	Ex-wife
27	Employee
28	Employer
29	Friend
30	Homosexual relationship
31	Other known

Offender not known to victim:

40	Stranger

Unknown Relationship:

99	Relationship cannot be determined

CIRCUM Circumstances

Felony-type (e.g., felony murder):

02	Rape
03	Robbery
05	Burglary
06	Larceny
07	Motor vehicle theft
09	Arson

10	Prostitution and commercialized vice
17	Other sex offense
18	Narcotics and drug laws
19	Gambling
26	Other felony-type not specified

Other than felony-type:

32	Abortion
40	Lover's triangle
41	Child killed by babysitter
42	Brawl due to influence of alcohol
43	Brawl due to influence of narcotics
44	Argument over money or property
45	Other arguments
46	Gangland killing
47	Juvenile gang killing
48	Institutional killing
49	Sniper attack
60	Other

Suspected felony-type:

70	Suspected felony-type

Justifiable homicides:

80	Felon killed by citizen
81	Felon killed by police

Unable to determine circumstances:

99	Unknown circumstances

SUBCIRC Sub-circumstances

Only if applicable (justifiable homicides):

1	Felon attacked office
2	Felon attacked fellow officer
3	Felon attacked citizen
4	Felon fleeing
5	Felon killed during crime
6	Felon resisting arrest
7	Undetermined

WTUS All cases weighted to reflect UCR estimated US homicide counts

WTST All cases weighted to reflect UCR estimated state homicide counts

WTIMPUS All cases weighted to reflect UCR estimated US homicide counts and imputed offender data.

Missing Data Problems in the SHR

Imputing Offender and Relationship Characteristics

JAMES ALAN FOX
Northeastern University

Although the so-called "dark figure" crime measurement problem has never been a major concern for homicide researchers, the Supplementary Homicide Reports (SHR) as well as other local data series on murder still are plagued by other kinds of missing data issues. Most prominent is missingness in data pertaining to offender characteristics as well as to victim-offender relationship that results from uncleared cases. Ignoring unsolved homicides would, of course, seriously understate calculated rates of offending by particular subgroups of the population, would distort trends over time among these same subgroups, and would bias observed patterns of offending to the extent that the likelihood of missing offender data is associated with offender characteristics. This article presents several approaches for overcoming missing data problems in the 1976-2001 cumulative SHR data file. First, a weighting procedure is described that uses characteristics of known offenders to serve as proxies for those of unidentified perpetrators. The weighting procedure included in the SHR file archived at ICPSR as well as an enhanced version are both presented and compared. Next, a "hot-deck" imputation strategy is applied to fill in missing offender attributes based on similar cases for which the offender is known. Finally, the matter of imputing victim-offender relationship data is discussed. Because this form of missingness cannot be assumed to occur at random, an ad-hoc procedure for estimating the number of intimate homicides among the pool of unsolved slayings is presented.

Keywords: imputation; missing data; Supplementary Homicide Reports

Although not nearly as well-known or widely cited as the "Crimes Known to the Police" data drawn from the Uniform Crime Reports' (UCR) Return A form, the Supplementary Homicide Reports (SHR) provide detailed, incident-level data on nearly all

AUTHOR'S NOTE: This article has been prepared with support from the Bureau of Justice Statistics (grant 2001-BJ-CX-X029). The opinions expressed in the article are those of the author and do not necessarily reflect the position of the United States Department of Justice. The author wishes to thank Michael D. Maltz, Alex Piquero, Donald B. Rubin, Maria Tcherni, Marianne Zawitz, and several anonymous reviewers for their helpful advice.

HOMICIDE STUDIES, Vol. 8 No. 3, August 2004 214-254
DOI: 10.1177/1088767904265592

murders and nonnegligent manslaughters in the United States. These reports include information on the month and year of an offense, on the reporting agency and its residential population, county and MSA codes, geographic division, and population group, on the age, race, and sex of victims and offenders, and on the victim-offender relationship, weapon use, and circumstance of the crime. Except for some slight modification in 1980, the record layout and variable definitions in the SHR data have remained unchanged since 1976 when the reporting format underwent a major revision.[1]

The SHR data do not suffer from the same kind of pervasive underreporting (so-called "dark figure") that plagues Return A data on nonfatal crimes. They are not, however, without their limitations.

The SHR form, unfortunately, does not solicit certain key pieces of information (e.g., the specific type of location, such as home, store, or street) and contains too little specificity in circumstance codes (and too many cases are, as a result, classified as "other"). Also, a certain level of missing or inaccurate information exists in the data. The relationship codes occasionally have the roles of victim and offender interchanged (i.e., the relationship of the offender to the victim is apparently sometimes coded rather than that of the victim to the offender). For example, there are a few cases of male "wives" and female "husbands" likely resulting from confusion with regard to the direction in which the spousal relationship should be coded.

In addition, by nature of the inherent data structure, the victim-offender relationships and, to a lesser extent, weapon and circumstance codes in multiple-victim crimes can be misleading. Specifically, because weapon, victim-offender relationship, and circumstance data are associated with offender strings in the data records, this information cannot physically vary across multiple victims. That is, if the same offender murders a number of victims, the structure of the data collection forms prescribes that the relationship of the offender to the first victim (often chosen arbitrarily) be coded for this offender. Thus, for example, in 1977, a Redondo Beach, California, woman killed her husband and three step-children by burning down the family home. Appropriately in this case, the weapon was coded as "fire" for all four victims, but the relationship of victim to offender was coded as "stepdaughter" for all victims—two 8-year-old White females, a 7-year-old White male, and a 40-year-old White male. Fortunately, this problem is not pervasive throughout the data set. Relatively few incidents involve multiple victims, and many of these entail relationship codes that could or should be invariant across victims (e.g., "son," "friend," "stranger," or "unknown").

CORRECTING FOR MISSING HOMICIDE RECORDS

Like the UCR generally, the SHR program is voluntary on the part of law enforcement agencies across the country. Therefore, agencies may not submit supplementary homicide data, and there is no easy way to track and capture these missing reports. Also, on occasions, an agency might fail to report homicide data for one or more months in the year. As an extreme example, in 1976, New York City supplied reports for only the first 6 months of the year.

For a variety of reasons (including data accuracy, completeness, and compatibility), the SHR also fails to provide full coverage of states. As shown in Table 1, several states are missing for whole or part-years, requiring some type of adjustment before using these data for purposes of trend analysis. Some states, such as Florida and Kansas, are frequently absent from the SHR database due to compatibility problems with state-run programs and the federal initiative. A few sparsely populated states, such as North and South Dakota, Vermont, and Montana, rarely submit SHR data for all 12 months, but this can easily be the result of zero homicide counts for the missing months. The far right column of Table 1 shows the calculated probability of a zero count month, based on an assumption that the monthly counts follow a Poisson distribution.[2]

Whereas Table 1 focuses on missingness at the entire state level, in virtually all states there are agencies—some large but most small—that report only sporadically. Table 2 displays similar monthly counts for cities with typically high homicide counts. Specifically, these 37 agencies have SHR homicide counts large enough that the probability of having a homicide-free month is

(text continues on p. 220)

TABLE 1
Monthly SHR Filings for States, 1976-2001

	Year																										*Aggregate for 1976-2001*			
State	*76*	*77*	*78*	*79*	*80*	*81*	*82*	*83*	*84*	*85*	*86*	*87*	*88*	*89*	*90*	*91*	*92*	*93*	*94*	*95*	*96*	*97*	*98*	*99*	*00*	*01*	*Total*	*UCR Homicides*	*SHR Victims*	*%*	p(*m* = 0)
MT	10	10	11	12	11	8	1	11	11	11	2	0	9	12	1	11	9	0	0	10	0	8	7	5	9	10	189	802	411	51.2	.0765
KS	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	6	0	0	0	0	0	0	12	12	234	3,531	2,134	60.4	< .0001
MS	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	7,943	4,918	61.9	< .0001
FL	12	12	12	12	12	12	12	12	12	12	12	12	0	0	0	0	12	12	12	12	1	0	0	0	0	0	193	30,388	19,279	63.4	< .0001
NB	12	11	12	12	11	12	11	11	12	12	12	10	12	10	12	12	12	11	9	9	9	10	9	11	9	8	281	1,353	952	70.4	.0131
NM	12	12	12	12	12	2	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	302	3,837	2,759	71.9	< .0001
SD	8	7	6	6	4	6	7	7	6	6	9	8	6	7	8	7	5	11	6	7	5	4	4	8	7	7	172	340	260	76.5	.3363
VT	8	3	8	4	4	4	4	4	8	10	7	8	7	8	7	8	8	1	2	7	6	6	4	7	6	4	153	341	264	77.4	.3352
IL	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	28,932	22,685	78.4	< .0001
DC	12	12	12	11	11	12	12	12	12	12	10	12	12	12	12	12	12	12	12	12	0	12	0	0	0	1	249	7,358	5,892	80.1	< .0001
KY	12	12	12	12	12	12	12	12	12	12	12	6	0	12	11	1	1	12	12	12	12	12	12	12	12	12	271	6,996	5,606	80.1	< .0001
IN	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	9,977	8,309	83.3	< .0001
GA	12	12	12	12	12	11	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	311	18,122	15,177	83.7	< .0001
TN	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	12,020	10,347	86.1	< .0001
IA	12	12	12	11	12	12	12	12	12	11	12	10	12	12	12	0	11	12	12	12	12	11	12	12	11	12	293	1,489	1,294	86.9	.0085
NH	10	10	7	8	9	8	10	10	6	9	9	11	9	12	8	11	9	10	9	11	9	0	9	8	9	9	230	574	509	88.7	.1589
LA	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	16,575	14,725	88.8	< .0001
MA	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	5,082	4,545	89.4	< .0001
OH	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	16,604	14,951	90.0	< .0001
ME	8	11	11	11	10	10	11	10	8	9	10	12	11	11	7	0	0	5	12	9	12	10	10	11	8	9	236	683	616	90.2	.1120
DE	11	11	12	12	12	11	11	12	10	10	12	11	11	10	10	12	7	9	8	4	11	8	11	8	10	10	264	805	731	90.8	.0758
AL	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	1	12	12	12	12	12	12	8	9	12	294	11,343	10,355	91.3	< .0001
MN	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	3,044	2,819	92.6	< .0001
MO	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	11,859	10,997	92.7	< .0001
WY	11	7	9	11	11	11	11	12	10	11	10	7	7	12	7	9	10	9	9	9	8	9	7	5	8	7	237	545	507	93.0	.1743
NY	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	47,791	44,492	93.1	< .0001

(continued)

TABLE 1 (continued)

State	Year																											Aggregate for 1976-2001			
	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	Total	UCR Homicides	SHR Victims	%	p(m = 0)
WI	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	0	12	12	12	300	4,502	4,203	93.4	< .0001
AK	11	10	12	12	11	12	12	12	12	12	12	12	11	11	11	11	12	11	11	12	11	12	11	12	12	11	299	1,251	1,171	93.6	.0181
ND	7	3	6	9	4	8	4	7	5	4	5	5	7	2	3	3	6	7	1	5	6	5	6	6	3	6	133	213	200	93.9	.3055
NV	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	3,480	3,302	94.9	< .0001
PA	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	17,888	17,128	95.8	< .0001
WA	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	5,708	5,487	96.1	< .0001
AR	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	5,425	5,250	96.8	< .0001
CT	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	3,823	3,704	96.9	< .0001
CO	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	4,818	4,683	97.2	< .0001
TX	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	50,427	49,068	97.3	< .0001
WV	11	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	11	12	310	2,551	2,494	97.8	.0003
NC	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	15,676	15,345	97.9	<.0001
OK	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	6,439	6,315	98.1	< .0001
AZ	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	7,734	7,596	98.2	< .0001
HI	12	11	12	12	11	12	11	12	12	12	12	12	12	11	12	12	12	11	11	12	12	12	12	12	11	12	305	1,245	1,223	98.2	.0185
MI	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	23,090	22,700	98.3	< .0001
OR	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	3,186	3,134	98.4	< .0001
ID	12	12	12	12	11	11	11	12	12	10	11	10	12	10	11	9	12	9	12	11	11	12	11	11	10	11	288	891	877	98.4	.0575
MD	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	12,048	11,913	98.9	< .0001
CA	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	76,432	75,942	99.4	< .0001
RI	11	10	12	10	12	12	12	10	12	12	11	12	10	11	12	12	12	10	12	11	9	11	11	11	4	5	277	921	917	99.6	.0522
UT	11	10	11	12	11	12	12	12	10	12	12	12	11	11	12	12	11	12	12	12	12	12	12	12	12	12	302	1,405	1,399	99.6	.0111
VA	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	12,057	12,006	99.6	< .0001
NJ	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	10,328	10,308	99.8	< .0001
SC	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312	8,473	8,556	101.0	< .0001

NOTE: SHR = Supplementary Homicide Reports; UCR = Uniform Crime Reports.

TABLE 2

Monthly Supplementary Homicide Reports Filings for Cities With High Homicide Counts, 1976-2001

	Year																									
City	*76*	*77*	*78*	*79*	*80*	*81*	*82*	*83*	*84*	*85*	*86*	*87*	*88*	*89*	*90*	*91*	*92*	*93*	*94*	*95*	*96*	*97*	*98*	*99*	*00*	*01*	*Total*
Atlanta	12	10	12	11	10	11	12	12	12	12	12	8	12	12	12	12	10	12	12	12	12	12	12	12	12	12	300
Baltimore	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	10	12	12	310
Birmingham	12	11	12	11	12	12	12	12	12	12	12	9	12	12	12	12	1	12	12	12	12	12	12	0	0	12	272
Boston	12	10	12	12	12	12	12	12	12	12	9	10	8	12	10	12	12	12	12	12	12	11	11	12	12	12	297
Charlotte	12	12	11	12	12	11	12	11	12	10	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	307
Chicago	12	12	12	12	12	12	12	12	0	0	9	1	12	12	6	12	12	12	12	12	12	12	12	12	12	12	268
Cleveland	11	12	11	12	12	12	12	12	12	12	12	12	12	12	12	11	12	12	12	12	12	12	12	12	12	12	309
Columbus	12	9	12	12	12	11	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	308
Dallas	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312
Denver	12	12	12	12	12	12	12	12	12	12	12	12	12	11	12	12	12	12	12	12	12	12	12	11	12	12	310
Detroit	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312
Fort Worth	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	11	12	311
Gary	11	11	12	11	12	12	12	12	12	12	10	10	12	12	12	12	12	12	12	9	4	11	12	12	12	12	293
Houston	12	12	12	12	10	8	0	12	12	12	12	12	10	12	12	12	12	12	12	12	12	12	12	12	12	12	292
Indianapolis	12	12	12	12	12	12	12	12	12	12	12	10	12	10	12	12	12	12	12	12	12	12	12	12	11	12	307
Jacksonville	10	12	12	12	12	12	12	12	12	12	12	12	0	0	0	0	12	12	12	11	1	0	0	0	0	0	190
Kansas City	12	12	12	12	12	12	12	12	12	12	12	12	11	12	12	12	11	12	12	12	12	12	12	12	12	12	310
Las Vegas	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312
Long Beach	12	11	12	12	12	12	12	12	11	12	12	12	12	12	12	12	12	12	12	12	12	11	12	12	11	11	307
Los Angeles	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312
Memphis	12	9	12	11	12	12	10	1	1	12	12	12	12	12	12	11	12	12	12	12	12	12	12	12	12	12	283
Miami	11	12	12	12	12	12	12	12	12	12	12	12	0	0	0	0	12	12	12	12	1	0	0	0	0	0	192
Milwaukee	12	12	12	12	12	12	12	12	12	12	11	12	12	12	12	12	12	12	12	12	12	12	0	12	12	12	299
Nashville	11	12	12	11	12	12	12	12	12	12	12	11	12	12	12	12	12	12	12	12	12	12	12	12	12	12	209
New Orleans	12	10	12	12	11	12	12	12	12	12	12	8	12	12	9	0	12	12	12	12	12	12	12	12	12	12	290
New York City	6	12	12	12	11	10	11	12	12	11	10	11	11	10	10	11	12	12	12	12	12	12	12	12	12	12	292
Newark	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312
Oakland	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312
Oklahoma City	12	11	12	12	12	12	12	12	12	11	12	12	12	11	12	12	12	12	12	12	12	12	12	12	12	12	309
Philadelphia	11	12	12	12	12	12	12	12	12	12	9	10	10	12	12	12	12	12	12	12	12	12	12	12	12	12	304
Phoenix	11	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	311
Richmond	12	12	11	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	311
San Antonio	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312
San Diego	12	11	12	12	11	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	11	12	309
San Francisco	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	312
St. Louis	12	12	12	12	12	12	12	12	12	12	12	11	12	12	12	12	12	12	12	12	12	12	12	12	12	12	311
Washington, DC	12	12	12	11	11	12	12	12	12	12	10	12	12	12	12	12	12	12	12	12	0	12	0	0	0	0	248

less than 1%.[3] Apparently, few agencies provide data consistently for every month.

For agencies with smaller populations, on the other hand, it is quite possible for 1 or more months to be homicide free and, thus, for there legitimately to be fewer than 12 monthly reports for a calendar year. Although the SHR database does not flag nonreports (that is, whether it represents no homicides or missing data), the Return A of monthly tallies of all index offenses would rarely be zero even for relatively small jurisdictions. The Federal Bureau of Investigation's (FBI) Uniform Crime Reporting Section monitors these monthly filings and uses a split strategy for imputing offense counts for nonreporting or partially reporting law enforcement agencies. Moderate levels of underfiling are handled through weighting, whereas more serious levels of missing data are predicted based on the data from similar and nearby jurisdictions.[4] After imputing Return A offense counts for missing reports, the FBI produces and publishes estimates for the nation as a whole as well as for individual states for all index offenses, including homicide. These annual state and national estimates of homicide volumes can then be used as benchmarks for assessing the completeness of the SHR data file and to adjust SHR victim or offender counts as needed.

As shown in Table 3, over the 26-year period, 1976-2001, the SHR cumulative file contains 480,037 records, of which 459,829 represent incidents of murder and nonnegligent manslaughter. The remaining 20,208 records pertain to negligent manslaughter (e.g., children playing with a gun), justifiable homicides (e.g., police officer killing a fleeing felon), or deaths linked to the September 11 terrorist attack on New York City's World Trade Center and elsewhere.[5] The 459,829 incidents, many of which involve multiple victims and/or offenders, include data for 480,455 victims and 532,177 offenders. By matching the available victim counts with the national estimates produced by the FBI through their Return A imputation strategy, the SHR file appears to be slightly more than 90% complete, with annual rates of completeness ranging from a high of 96.7% in 1983 to a low of 83.7% in 1998. It is also noteworthy and troubling that the level of completeness of the SHR has generally diminished in recent years.

TABLE 3
SHR Records Pertaining to Incident, Victim, and Offender Counts, 1976-2001

	All Homicides		*Murder & Nonnegligent Manslaughter*					
	SHR Data File		*SHR Data File*			*UCR Estimated*		
Year	*Records*	*Victims*	*Incidents*	*Victims*	*Offenders*	*Victims*	*Offenders*	*UCR Benchmark Weight*
1976	16,744	17,406	15,951	16,605	17,995	18,780	20,352	1.1310
1977	17,825	18,586	17,277	18,032	19,228	19,120	20,388	1.0603
1978	18,546	19,308	17,957	18,714	20,075	19,560	20,983	1.0452
1979	20,576	21,417	19,756	20,591	22,247	21,460	23,186	1.0422
1980	21,911	22,786	21,002	21,860	24,859	23,040	26,201	1.0540
1981	20,152	20,931	19,284	20,053	21,794	22,520	24,475	1.1230
1982	19,413	20,288	18,622	19,485	21,219	21,010	22,880	1.0783
1983	18,690	19,426	17,954	18,673	20,315	19,310	21,008	1.0341
1984	17,168	17,858	16,574	17,260	18,781	18,690	20,337	1.0829
1985	17,348	18,131	16,763	17,545	18,831	18,980	20,371	1.0818
1986	19,089	19,849	18,510	19,257	20,722	20,610	22,178	1.0703
1987	17,747	18,509	17,205	17,963	19,396	20,100	21,703	1.1190
1988	17,846	18,546	17,277	17,971	19,841	20,680	22,832	1.1507
1989	18,812	19,588	18,184	18,952	20,898	21,500	23,708	1.1344
1990	20,404	21,246	19,451	20,273	22,889	23,440	26,465	1.1562
1991	21,817	22,656	20,863	21,676	24,807	24,700	28,268	1.1395
1992	22,753	23,793	21,701	22,716	25,382	23,760	26,549	1.0460
1993	23,320	24,336	22,175	23,180	26,116	24,530	27,637	1.0582
1994	22,231	23,246	21,093	22,084	25,075	23,330	26,490	1.0564
1995	20,099	21,193	19,154	20,232	22,667	21,610	24,211	1.0681
1996	17,053	17,829	16,203	16,967	19,350	19,650	22,410	1.1581
1997	15,929	16,726	15,052	15,836	17,918	18,210	20,604	1.1499
1998	14,313	14,975	13,556	14,209	16,177	16,970	19,320	1.1943
1999	12,792	13,511	12,299	13,011	14,587	15,522	17,402	1.1930
2000	13,220	13,856	12,597	13,230	15,059	15,517	17,662	1.1729
2001	14,239	17,695	13,369	14,080	15,949	15,980	18,101	1.1348
1976-2001	480,037	503,691	459,829	480,455	532,177	528,579	585,720	

NOTE: SHR = Supplementary Homicide Reports; UCR = Uniform Crime Reports.

To correct for missing SHR records, national estimates and state-level estimates of homicide victimization published annually by the FBI have been used to weight SHR data records (that is, state and national weights have been assigned so that tabulations total to the respective estimated counts). This weighting process, included in archived SHR files for 1976-1997 (see Fox, 2000) and for subsequent annual updates, assumes of course that the missing records are not systematically different from those available in the file. Although the systematic exclusions of a state like Florida or a city like Washington, D.C., may cause some concern, the fact that missingness occurs in both large and small jurisdictions gives us some reason for comfortably applying these benchmark weights.

Homicide undercounting, either because an agency fails to supply SHR information or because it supplies information for only part-years, can reasonably be assumed missing completely at random (MCAR). The MCAR assumption holds that missing data are a random subset of the underlying data universe and, thus, that the probability that a data element is missing does not depend on the value itself or that of any other variable (see Allison, 2002; Little & Rubin, 2002). Specifically,

$$P(Y \text{ is missing} \mid X_1, X_2, \ldots, X_k, Y) = P(Y \text{ is missing}).$$

This assumption holds that the likelihood that a record and its associated data are missing (and thus particular characteristics of the victim, offender, and incident are missing) is not correlated with the data values themselves or any other variables used in an analysis. Even though agency characteristics may be related to the completeness of an agency's reporting patterns, we assume that this is not associated with the characteristics of the incident themselves.

The MCAR assumption allows us to weight all cases upward for a particular year (for the U.S. overall and for each state) based on published benchmarks, regardless of victim, offender, or incident characteristics. Table 3 shows year-by-year tallies of SHR records, victims and offenders represented by these records, estimated homicides published by the FBI, and the yearly benchmark weights used for the nation as a whole. The benchmark

$$W_t = \frac{UCR_t}{SHR_t}$$

where UCR_t is the estimated count of murder and nonnegligent manslaughter for the nation (or for state i) in year t, and SHR_t is the number of available victim records for the nation or state i in that same year. The national benchmark weights are calculated separately (rather than aggregating across states) to avoid problems associated with missing states.

WEIGHTING OFFENDER RECORDS

Even for the 90% of SHR records that are available for analysis, certain variables have nontrivial rates of missingness. At one extreme, characteristics of an agency (e.g., region, population group) are always complete. Victim data, although sometimes missing, are absent at such a low rate that standard approaches for handling missing data (specifically, listwise deletion) hardly bias analyses of patterns and trends in victimization. Specifically, as shown in the top panel of Table 4, victim age is missing in 1.70% of homicides, race in 1.00%, and sex in .13%. Overall, 2.43% are missing on at least one of these measures.

At the other extreme, a significant problem in using SHR data to analyze offender characteristics is the sizable and growing number of unsolved homicides contained in the data file. Overall, 26.2% of the SHR offender records describe the perpetrator as unknown (based on situation codes), and this percentage has grown from slightly less than 20% in 1976 to nearly 30% by the mid-1990s. Even when the offender is known to the police, moreover, the characteristics about the offender may still be unavailable. Table 4 shows specifically that as many as 26.29% of offender records contain no information about the perpetrators, and 31.12% of records are not complete in terms of the assailant's age, race, and sex.

Ignoring unsolved homicides and missing offender data would, of course, seriously understate calculated rates of offending overall and by particular subgroups of the population, would distort trends over time among these same subgroups, and would bias observed patterns of offending to the extent that the likelihood of missingness of offender data is associated with offender characteristics.

TABLE 4
Patterns of Missingness in Victim and Offender Data

	Characteristic				
SHR File	*Age*	*Race*	*Sex*	*Cases*	*%*
Victim file					
	•	•	•	468,714	97.56
	•	•	○	41	.01
	•	○	•	3,444	.72
	•	○	○	57	.01
	○	•	•	6,888	1.43
	○	•	○	12	.00
	○	○	•	781	.16
	○	○	○	518	.11
Total				480,455	100.00
Offender file					
	•	•	•	366,563	68.88
	•	•	○	31	.01
	•	○	•	2,553	.48
	•	○	○	153	.03
	○	•	•	20,894	3.93
	○	•	○	111	.02
	○	○	•	1,960	.37
	○	○	○	139,912	26.29
Total				532,177	100.00
With complete victim data					
	•	•	•	362,051	68.03
	•	•	○	27	.01
	•	○	•	1,307	.25
	•	○	○	147	.03
	○	•	•	20,071	3.77
	○	•	○	105	.02
	○	○	•	1,796	.34
	○	○	○	133,928	25.17
With incomplete victim data					
	•	•	•	4,512	.85
	•	•	○	4	.00
	•	○	•	1,246	.23
	•	○	○	6	.00
	○	•	•	823	.15
	○	•	○	6	.00
	○	○	•	164	.03
	○	○	○	5,984	1.12
Total				532,177	100.00

NOTE: • = data available; ○ = data missing; SHR = Supplementary Homicide Reports.

Although we cannot directly determine if case solution and thus missingness in offender data are associated with offender characteristics themselves, we can explore whether this pattern of missingness is at all tied to victim and incident variables. As shown in Table 5, case solution rates are lowest for homicides against young adult victims as well as for elderly victims. Solution rates are also lower for incidents involving Black or male victims. As expected, solution rates decrease with increasing popu-

TABLE 5
Case Solution Rates by Victim, Location, and Incident Characteristics

Variable	*Category*	*Solved Cases*	*%*	*Unsolved Cases*	*%*
Victim age	0-13	21,016	86.7	3,233	13.3
	14-17	18,682	71.2	7,559	28.8
	18-24	83,458	68.5	38,383	31.5
	25-34	104,659	69.8	45,316	30.2
	35-49	85,098	72.8	31,799	27.2
	50-64	34,465	71.4	13,825	28.6
	65+	19,199	68.2	8,961	31.8
Victim race	White	191,250	72.6	72,125	27.4
	Black	167,663	69.4	73,958	30.6
	other	7,664	71.9	2,992	28.1
Victim sex	male	274,962	69.8	118,821	30.2
	female	91,615	75.2	30,254	24.8
Location	large city	113,747	63.0	66,730	37.0
	medium city	80,311	69.5	35,171	30.5
	small city	46,257	78.4	12,766	21.6
	suburban	79,667	74.5	27,317	25.5
	rural	46,597	86.8	7,092	13.2
Region	Northeast	55,588	63.1	32,561	36.9
	Midwest	69,467	69.9	29,945	30.1
	South	162,339	76.6	49,702	23.4
	West	79,183	68.2	36,868	31.8
Weapon	non-gun	132,590	71.5	52,776	28.5
	gun	233,987	70.8	96,299	29.2

lation size and urbanness. In part as a consequence of urbanness differences, solution rates in the South are much higher than other regions. Finally, whether or not a gun is used to commit the homicide does not appear to affect rates of case solution.

To examine further whether missingness occurs completely at random or rather is associated with key variables of the event, I have performed a logistic regression of whether or not a case is solved (and thus offender information is available) on victim attributes (age, race, and sex), place characteristics (location urbanness and region), as well as weapon (gun/non-gun). For the dependent variable, offender records missing on offender age, race, or sex are coded as 1, whereas cases with complete offender information are coded as 0. Other incident characteristics besides those used here as covariates—specifically, circumstances (felony, argument, etc.) and victim-offender relationship (e.g., intimate, family, friend/acquaintance, stranger)—are rarely available when an offender is unknown and thus are not used to model the likelihood of missingness of offender data. In addition, circumstances and victim-offender relationship are often an outcome of case solution rather than a contributor.

Based on the results shown in Table 6, missingness is associated with various victim/incident characteristics. Homicides of teen or young adult victims and Black victims are more likely to go unsolved. Overwhelmingly, urbanness has the strongest association with solvability (Wald $\chi^2 = 9096.50$, 1 *df* for the full model), in part accounting for the effects of race and gender. Region also associates strongly with offender data missingness, with the largest contrast between the Northeast and the South. This regional effect may partly reflect differences in urbanness, even though this variable is explicitly included in the model as well.

WEIGHTING SHR OFFENDER DATA

To adjust for unsolved homicides in the archived SHR files, I fashioned a method for offender imputation using available information about the victims murdered in both solved and unsolved homicides. The age, sex, and race of victims as well as the year and state of the offense can provide valuable, albeit imperfect, information about the characteristics of unidentified assailants. If a 15-year-old Black male is killed by an unidentified person, for example, it is much more plausible that the perpetrator would have also been a young Black male than, say, a middle-aged White female.

Through this imputation algorithm, the demographic characteristics of unidentified offenders are inferred on the basis of similar homicide cases—similar in terms of the victim profile and state and year of the offense—that had been solved. In other words, offender profiles for unsolved crimes are estimated based on the offender profiles in solved cases matched on victim age, sex, and race as well as year and state. (Note that year and state were used to ensure that marginal counts match, but later in this article, I accomplish this without using unnecessarily all high-level interactions with year and state.)

Using victim covariates (and later location and incident covariates) to condition the weighting process makes an assumption about the nature of missingness that is far less rigid than MCAR. The so-called "missing at random" (MAR) assumption holds that the probability of missingness does not depend on the

TABLE 6
Logistic Regression of Offender Data Missingness on Victim, Location, and Incident Characteristics

		Model 1			*Model 2*			*Model 3*		
Variable	*Category*	*b*	*S.E.*	*Odds Ratio*	*b*	*S.E.*	*Odds Ratio*	*b*	*S.E.*	*Odds Ratio*
Constant		–1.832	.019	.160	–1.257	.021	.285	–1.246	.021	.288
Victim age			χ^2 = 3104.85, 6 *df*			χ^2 = 3021.40, 6 *df*			χ^2 = 3070.86, 6 *df*	
	younger than 14[a]									
	14-17	.904	.023	2.469	.847	.024	2.333	.873	.024	2.394
	18-24	1.023	.020	2.782	.997	.020	2.709	1.021	.020	2.777
	25-34	.972	.020	2.643	.966	.020	2.626	.988	.020	2.686
	35-49	.838	.020	2.313	.870	.020	2.386	.889	.020	2.432
	50-64	.919	.021	2.506	.971	.022	2.641	.986	.022	2.681
	65+	1.119	.023	3.061	1.181	.023	3.257	1.188	.023	3.281
Victim race			χ^2 = 419.05, 2 *df*			χ^2 = 42.44, 2 *df*			χ^2 = 48.41, 2 *df*	
	White[a]									
	Black	.129	.006	1.138	.043	.007	1.044	.046	.007	1.047
	other	.053	.022	1.055	–.025	.023	.976	–.026	.023	.974
Victim sex			χ^2 = 807.59, 1 *df*			χ^2 = 391.64, 1 *df*			χ^2 = 427.78, 1 *df*	
	male[a]									
	female	–.217	.008	.805	–.154	.008	.857	–.163	.008	.850
Urbanness						χ^2 = 9093.13, 4 *df*			χ^2 = 9096.50, 4 *df*	
	large city[a]									
	medium city				–.209	.008	.812	–.210	.008	.811
	small city				–.671	.011	.511	–.673	.011	.510
	suburban				–.423	.009	.655	–.424	.009	.654
	rural				–1.173	.014	.310	–1.171	.014	.310
Region						χ^2 = 3036.85, 3 *df*			χ^2 = 2975.44, 3 *df*	
	Northeast[a]									
	Midwest				–.264	.010	.768	–.261	.010	.770
	South				–.453	.009	.636	–.448	.009	.639
	West				–.100	.010	.905	–.097	.010	.907
Weapon									χ^2 = 53.95, 1 *df*	
	nongun[a]									
	gun							–.051	.007	.950
Model			χ^2 = 5467.04, 9 *df*			χ^2 = 22310.92, 16 *df*			χ^2 = 22364.79, 17 *df*	

a. Reference category.

value itself, once a range of covariates ($X_1, X_2, \ldots, X_k$) related to the likelihood of missingness is controlled (see Allison, 2002). That is,

$$P(Y \text{ is missing} \mid X_1, X_2, \ldots, X_k, Y) = P(Y \text{ is missing} \mid X_1, X_2, \ldots, X_k)$$

Thus, for example, it may be true that the probability of missingness for offender race is greater for Black perpetrators. However, once factors such as victim race are controlled, we can reasonably assume that the correlation between offender race and its propensity for missingness largely disappears.

The weighting procedure is accomplished by establishing adjustment groups within a large, multidirectional cross-tabulation. Using seven categories for victim age (younger than 14, 14-17, 18-24, 25-34, 35-49, 50-64, and 65+), three victim race groups (White, Black, and other), two groups for victim gender, 26 year categories, and 51 state classifications produces a five-dimensional cross-tabulation with thousands of cells. For each cell ($i = 1, 2, \ldots, 7$ for victim age, $j = 1, 2, 3$ for victim race, $k = 1, 2$ for victim gender, $I = 1, 2, \ldots, 26$ for year, and $m = 1, 2, \ldots, 51$ for state), we tally the total number of offenders (N_{ijklm}) and the number having complete offender data (n_{ijklm}) and use as a weight the inverse proportion of complete cases:

$$W_{ijklm} = \frac{N_{ijklm}}{n_{ijklm}}.$$

Next, these adjustment cell weights are applied to the offender records based on their cell membership (i.e., based on victim age, race, sex, year, and state) and whether the offender information is complete. That is,

$$W = \begin{cases} W_{ijklm} & \text{if offender data are complete} \\ 0 & \text{otherwise.} \end{cases}$$

This weighting approach is applied to the entire offender record so that cases with missing offender age, race, or sex are excluded by virtue of their zero case weights. As a consequence, partial offender information is discarded, causing some slight inefficiency in the approach. Although it would be possible to retain partial offender information, this would require separate weights for each offender characteristic. Finally, all nonzero weights are further increased slightly to account for the small

TABLE 7
Distributions of Imputation Weights

	Initial Weights		*Revised Weights*	
Weight	N	%	N	%
Exactly 0	157,391	30.3	157,381	30.3
1.00-1.99	314,483	60.5	294,208	56.6
2.00-2.99	30,513	5.9	53,104	10.2
3.00-3.99	8,892	1.7	10,963	2.1
4.00-4.99	3,432	.7	3,010	.6
5.00-5.99	1,643	.3	761	.1
6.00-6.99	1,050	.2	2	.0
7.00-7.99	559	.1	2	.0
8.00-8.99	423	.1	1	.0
9.00-9.99	282	.1		
10.00-29.99	642	.1		
30.00-49.99	88	.0		
Total	519,432	100.0	519,432	100.0
Mean	1.11		1.13	
Standard deviation	1.14		.91	
Median	1.22		1.23	
75th percentile	1.42		1.60	
90th percentile	1.93		2.16	
95th percentile	2.54		2.64	
99th percentile	4.82		3.77	
Maximum	46.73		8.08	

percentage of cases unassigned to any adjustment cell because of their being missing on one or more victim characteristics.

In any analysis of weighted data, offenders with incomplete age, race, or sex information are dropped due to assigned zero weights. Offenders with complete age, race, and sex information all have weights at or above 1.0 and become proxies for excluded cases, matched on victim characteristics, state, and year. Thus, for example, an offender with an imputation weight of 2.0 would count in any analysis as if he or she were two offenders.

Table 7 provides some basic descriptive information about the weighting variable itself, with percentages based on cases having complete victim information. The weight variable (the initial weights shown in the first panel) is, of course, 0 for the 30.3% of offender records that are incomplete in terms of offender age, race, or sex. Most of the nonzero weights applied to solved cases are relatively small. Fewer than 10% exceed 2. The distribution of weights is clearly skewed, with a number of cases weighted over 10. These exceedingly large weights are a result of certain adjustment cells having relatively few cases and a large percentage of these with missing offender information.

Large case weights generally do not affect data analyses of national patterns and trends as they appear on very few records. One notable exception occurs with adjustment cells associated with Washington, D.C., an agency that has had a poor record of providing offender detail. In certain years, there were so few cases with known offender information that just a couple of perpetrators came to serve as proxy for a large number of unknown offenders in that city. The resulting weights can be so inflated that even analysis of thousands of cases can be biased. For example, an analysis of changes in the offender age distribution tracked over time contained a noticeable spike in the age curve traced to a couple of cases with weights over 50 (see Fox & Zawitz, 2003).

To correct for this problem, a few excessively large weights (weights over 50) were zeroed out and redistributed across remaining cases with similar victim characteristics, essentially reweighting cases after collapsing the adjustment cross-tabulation for state.

Table 8 shows the distribution of offender age, race, and sex for known cases (nonimputed) using listwise deletion (adjusted percentages in the left panel) and then for the data with initial corrected weights applied (middle panel). After weighting, the percentage of young offenders increases slightly (from 44.3% to 46.3% for offenders in the combined 14-24 age group). The race and gender distributions are altered very slightly, with the share of Black offenders increasing from 51.2% to 51.8% and that of male perpetrators expanding from 87.9% to 88.7%.

The effect of weighting is potentially greater in calculating rates per 100,000 and tracking them over time. Table 9 displays homicide offending rates per 100,000 by age of offender for the years 1976 to 2001. The magnitude of the rates increases, if only by virtue of using benchmark weights to compensate for missing data. But exploring changes over time does reveal some effect that differential adjustment weights appear to have. For the 1985-1993 time period, for example, when youth homicide rates grew precipitously, the change in imputed rates is somewhat greater than that using known cases only. For the 14- to 17-year-old group, the rate of offending using only known cases increased 168%, whereas the (initial) imputed rate jumped 205%.

TABLE 8
Offender Characteristics Without and With Imputation Weights

	Without Imputation			*With Initial Imputation Weights*		*With Revised Imputation Weights*	
Offender Characteristic	N	%	*Adj. %*	N	%	N	%
Offender age							
Younger than 14	2,054	.4	.5	2,753	.5	2,873	.5
14-17	40,347	6.9	9.9	60,334	10.3	62,363	10.6
18-24	139,693	23.8	34.4	210,865	36.0	211,100	36.0
25-34	118,104	20.2	29.1	169,722	29.0	168,228	28.7
35-49	74,664	12.7	18.4	101,338	17.3	100,657	17.2
50-64	23,387	4.0	5.8	30,613	5.2	30,252	5.2
65+	7,981	1.4	2.0	10,095	1.7	10,246	1.7
Missing	179,489	30.6					
Total	585,720	100.0	100.0	585,720	100.0	585,720	100.0
Offender race							
White	199,869	34.1	46.9	271,417	46.3	269,227	46.0
Black	218,347	37.3	51.2	303,328	51.8	304,960	52.1
Other	8,318	1.4	2.0	10,974	1.9	11,533	2.0
Missing	159,186	27.2					
Total	585,720	100.0	100.0	585,720	100.0	585,720	100.0
Offender sex							
Male	379,275	64.8	87.9	519,340	88.7	518,653	88.5
Female	52,082	8.9	12.1	66,380	11.3	67,067	11.5
Missing	154,363	26.4					
Total	585,720	100.0	100.0	585,720	100.0	585,720	100.0

TABLE 9

Trends in Homicide Offending Rates (per 100,000) by Age Without and With Imputation Weights

	14-17			*18-24*			*25+*		
Year	*No Weights*	*Initial Weights*	*Revised Weights*	*No Weights*	*Initial Weights*	*Revised Weights*	*No Weights*	*Initial Weights*	*Revised Weights*
1976	8.2	10.9	11.4	17.3	22.7	22.9	7.9	10.0	9.9
1977	7.4	10.4	10.7	16.5	23.0	22.8	7.5	9.7	9.7
1978	7.6	10.3	10.5	17.4	23.9	24.0	7.7	9.8	9.7
1979	8.5	12.3	12.4	19.1	26.6	26.8	7.9	10.4	10.4
1980	8.6	13.0	13.0	19.9	30.1	30.1	7.9	11.4	11.4
1981	8.5	11.4	12.1	18.8	26.3	26.6	8.0	10.8	10.7
1982	7.4	10.7	11.1	17.4	24.6	25.0	7.2	10.0	9.8
1983	6.8	10.0	10.2	15.9	22.6	23.0	6.6	9.0	8.9
1984	6.2	8.7	9.1	15.2	22.3	22.2	6.3	8.6	8.6
1985	7.2	10.1	10.5	15.5	21.9	22.1	6.3	8.5	8.4
1986	8.6	12.0	12.7	16.9	24.3	24.3	6.6	9.0	8.9
1987	8.7	12.6	12.8	16.7	25.1	25.1	6.1	8.4	8.4
1988	10.7	16.4	16.7	18.5	27.7	27.7	6.0	8.3	8.3
1989	12.0	17.3	18.4	20.7	31.3	31.4	5.8	8.1	8.0
1990	16.3	25.4	25.7	23.8	35.9	36.2	6.0	8.5	8.4
1991	17.6	27.3	28.1	27.0	42.6	42.6	5.7	8.2	8.2
1992	17.4	26.9	27.9	24.9	39.6	39.5	5.2	7.7	7.6
1993	19.3	30.8	31.3	26.5	42.4	42.6	5.0	7.4	7.3
1994	19.1	29.3	31.1	25.3	40.6	40.7	4.7	7.0	6.9
1995	15.4	24.1	25.1	23.7	38.1	37.7	4.4	6.5	6.4
1996	13.1	19.6	20.9	23.4	36.5	36.8	4.1	6.0	5.8
1997	10.8	16.9	17.5	21.1	34.3	34.0	3.7	5.5	5.5
1998	8.6	12.9	13.7	20.2	31.9	31.7	3.7	5.3	5.3
1999	7.1	11.1	11.1	18.0	28.6	28.4	3.3	4.7	4.8
2000	6.0	9.6	9.5	17.5	27.9	27.8	3.3	4.8	4.8
2001	5.8	9.0	9.2	17.9	28.7	28.3	3.3	4.8	4.8

Notwithstanding this particular distinction during the historically unusual period of the late 80s and early 90s, the imputation weights appear to have limited effects on conclusions with regard to the characteristics of offenders. Apparently, the imputation algorithm changes the demographic distributions only to a very minor degree. Actually, this may be a good thing; a sizable difference would rest on the unproven validity of the assumption that solved and unsolved cases are similar after controlling for the variables used to construct the weights.

EXPANDING THE RANGE OF ADJUSTMENT FACTORS

The effect of weighting using adjustment cells may also be understated to the extent that the covariates used to define the adjustment cells fail to capture important differences among categories of homicide that are related to the missingness of offender data and their unobserved distributions. Year and state both were used to ensure marginal equivalence geographically and over time. Yet, these may unnecessarily deplete the available cases within each victim class—unnecessarily inflating the instability of the imputation weights. It is still possible to ensure proper representation of states and individual years by adjusting weights after the fact (as is done to correct for cases excluded for lack of victim data), without having to force the model to incorporate explicitly all higher level interactions involving year and state with other covariates.

As an alternative, the weighting approach can incorporate other information about the locale and crime that is related to characteristics of offenders. When choosing adjustment factors, it is not so much their effect on the probability of offender missingness that is critical but their correlation with offender characteristics themselves. Therefore, even though some factor might correlate strongly with the likelihood of case solution, it may not correlate at all with the variables we wish to impute, that is, offender age, race, and sex.

Table 10 displays not only the extent of offender data missingness by a range of victim and incident characteristics but also their association with offender characteristics when available. Specifi-

TABLE 10
Effects of Victim and Offense Covariates on Offender Characteristics

		Offender Age (%)							Offender Race (%)					Offender Sex (%)			
Covariate	N	*% Missing*	*< 18*	*18-24*	*25-34*	*35-49*	*50+*	*V*	*% Missing*	*White*	*Black*	*Other*	*V*	*% Missing*	*Male*	*Female*	*V*
Victim age								.228					.057				.186
0-13	22,014	14.2	13.6	39.1	33.7	11.9	1.7		12.7	55.1	42.4	2.6		12.2	67.3	32.7	
14-17	28,451	28.1	35.4	45.2	11.8	5.8	1.8		24.4	44.6	52.8	2.5		23.7	95.4	4.6	
18-24	127,438	32.0	12.6	52.1	24.5	8.5	2.3		27.7	42.8	55.2	2.0		27.1	93.7	6.3	
25-34	150,723	31.7	6.6	30.8	39.8	18.0	4.8		27.7	44.5	53.6	1.9		27.1	89.2	10.8	
35-49	115,528	28.8	7.3	22.6	27.3	32.5	10.4		25.4	49.2	49.0	1.8		24.8	85.2	14.8	
50-64	47,663	29.6	7.4	23.6	24.0	22.4	22.6		26.9	51.1	47.1	1.8		26.4	82.9	17.1	
65+	27,615	31.9	10.1	25.3	22.8	17.9	23.9		29.9	54.9	43.6	1.5		29.3	85.7	14.3	
Victim race								.035					.680				.061
White	265,772	28.0	10.4	32.7	28.9	19.6	8.4		25.0	83.1	15.7	1.2		24.3	89.6	10.4	
Black	242,768	32.3	10.3	36.1	29.3	17.1	7.1		28.0	6.4	93.3	.3		27.6	85.7	14.3	
Other	10,892	28.3	13.5	35.3	28.4	16.8	6.0		25.4	26.0	18.2	55.8		24.2	90.6	9.4	
Victim sex								.158					.072				.027
Male	407,075	31.1	11.4	37.0	28.7	16.6	6.3		27.0	45.0	53.2	1.9		26.4	87.4	12.6	
Female	112,357	26.0	7.2	25.2	30.5	24.3	12.7		24.4	53.3	44.5	2.2		23.9	89.5	10.5	
Region								.048					.185				.057
Northeast	88,697	40.8	10.9	36.9	29.1	16.8	6.4		34.3	43.9	55.0	1.2		33.0	90.5	9.5	
Midwest	99,617	29.7	12.0	35.9	28.2	16.8	7.1		27.3	35.5	63.2	1.3		27.0	87.3	12.7	
South	211,909	25.7	9.0	31.8	29.5	20.4	9.3		21.8	42.6	56.4	1.0		21.5	86.1	13.9	
West	119,209	30.0	11.5	36.1	29.1	17.0	6.4		28.1	66.5	28.7	4.8		27.3	89.9	10.1	
Urbanness								.067					.192				.051
Large city	187,589	39.1	12.5	38.3	28.5	15.2	5.5		32.9	35.2	63.2	1.6		32.1	90.1	9.9	

Medium city	115,366	31.6	10.8	35.4	29.5	17.4	6.9		28.3	36.5	61.4	2.0		27.8	87.4	12.6	
Small city	58,690	21.4	9.9	34.6	29.2	18.3	8.0		19.9	47.9	50.0	2.1		19.4	86.6	13.4	
Suburban	105,338	25.1	9.3	31.4	29.4	20.7	9.1		23.5	64.2	34.3	1.5		23.1	86.9	13.1	
Rural	52,449	13.3	7.0	27.1	29.4	24.1	12.4		12.3	65.4	31.2	3.4		12.0	85.4	14.6	
Weapon								.090					.091				.105
Non-gun	186,104	28.1	9.2	33.4	33.0	19.0	5.5		26.3	52.1	45.3	2.5		25.7	83.3	16.7	
Gun	333,328	31.1	11.2	34.8	26.8	18.1	9.1		26.5	43.8	54.6	1.6		25.9	90.4	9.6	
Circumstances								.166					.063				.098
Felony	120,745	36.1	15.1	45.2	27.4	10.2	2.1		31.4	40.3	58.2	1.5		30.8	93.4	6.6	
Argument	196,620	8.9	6.8	28.3	31.3	23.4	10.1		6.5	46.3	51.6	2.0		5.9	85.2	14.8	
Other	202,067	46.9	13.1	36.5	26.5	15.9	8.0		42.8	52.2	45.7	2.1		42.3	88.1	11.9	
Total	519,432	30.1	10.4	34.3	29.1	18.4	7.8		26.5	46.8	51.2	1.9		25.9	87.9	12.1	

cally, the percentage distributions for offender age (in five categories, collapsing the high and low ends of the age range), race, and sex are calculated by victim age, race, and sex, region, urbanness, weapon, and circumstance of the crime. Also, Cramér's V is given as a measure of correlation for each pair of variables.[6]

As expected, offender age correlates fairly strongly with victim age (V = .228).[7] Offender race correlates with victim race (V = .680) and to a lesser degree with urbanness (V = .192). Circumstances also vary by offender age (V = .166) in that younger offenders are more often implicated in felony homicides and those unclassifiable by the police. Region and weapon appear to have smaller effects on offender characteristics but still may have some interactive effects along with other covariates.

The initial weighting approach based on victim age, race, and sex plus year and state has been used for a number of years for the SHR data file (see Fox 2000, 2001, 2002). The approach essentially predicts the characteristics of unknown offenders based on these five variables and all of their interactions. Although it is desirable to maintain marginal counts within year and within state, they may be accomplished by adjustment rather than using them as stratification factors. A more sensitive alternative would use urbanness, region, gun use, and even circumstances along with victim age, race, and sex in the creation of adjustment cells.

Even though there are more factors on which to stratify when replacing state and year with region, urbanness, weapon, and circumstances, the total number of adjustment cells is reduced significantly. Rather than $7 \times 3 \times 2 \times 26 \times 51 = 55{,}692$ cells with the initial weights, the revised approach distributes solved and unsolved cases across $7 \times 3 \times 2 \times 4 \times 5 \times 2 \times 3 = 5{,}040$ cells. Moreover, this avoids having to subdivide cases within states with few homicides or to isolate Washington, D.C., with its particularly large volume of missing offender information.

The right panel of Table 7 describes characteristics of the revised weighting variable. The weights are far less skewed, with none exceeding 8.08, owing to the smaller number of adjustment categories used. More important, the revised approach, by incorporating covariates (especially urbanness) that are more closely correlated with missingness and the offender variables to be imputed, may potentially have greater effects on the resulting offender distributions.

The right panel of Table 8 shows the distribution of offender characteristics (age, race, and sex) using the revised adjustment weights. The revised approach further boosts the abundance of youthful offenders; offenders younger than 25 represent 46.6% of the revised imputed distribution, compared with 46.3% for the initial weights and 44.3% for the distribution using listwise deletion. The percentage of Black offenders grows to 52.1%, compared with 51.8% using the initial weights and 51.2% without weighting. The sex distribution, which by any measure overwhelmingly favors male perpetrators, decreases slightly from 88.7% to 88.5% (clearly indicating a ceiling effect governing this percentage).

HOT-DECK IMPUTATION

Rather than using weights to adjust for unsolved cases, we may instead explicitly fill in the values of missing data on offender age, race, and sex using a variety of predictors or covariates. The so-called "hot-deck" method (see Rubin, 1987) creates victim/incident strata (like the adjustment cells used in the weighting approach) and then fills in for each missing value a random selection from the hot-deck pool of potential donor cases, matched on a set of victim/incident covariates. Although a number of regression-type algorithms exist for predicting the missing values, the hot-deck strategy is particularly well-suited for categorical predictors. The hot-deck method has the added advantage over the weighting approach in that partially complete offender records can be retained while only imputing the specific variables that are missing.

The hot-deck approach follows a series of steps for each cell of the k-way contingency table defined by the string of k covariates (e.g., victim age, race, and sex, urbanness and circumstances) that are to be used to impute missing values of a variable Y (e.g., offender age, race, or sex).

For a given cell c of the k-way contingency table having a total of N_c cases matched on k covariates, let n_c represent the number with known data for a particular variable Y (offender age, race, or sex) and m_c represent the number of missing cases on Y. (Should $n_c = 0$ for any cell, then the table is temporarily collapsed into sequentially fewer dimensions until a donor pool becomes available.)

From the pool of n_c cases with available data for Y, a random sample (with replacement) of n_c is drawn. Recreating a donor pool through replacement, which will likely duplicate some cases while ignoring others, is necessary so as not to bias downward the variance of postimputation distribution of Y.

Next, a sample of m_c cases is selected (with replacement) from the donor pool of n_c scores on Y. These are used to fill in the missing Y values. Finally, this process is repeated independently for each variable to be imputed.

Despite its straightforwardness, the process of sorting cases into donor cells, of resorting when case deficiencies arise, of reconstructing donor pools by replacement sampling, and then of filling in imputation results for each variable with missing data causes the hot-deck method to be exceptionally time-consuming, even with a high-speed processor. In fact, the length of the process grows exponentially as the sample size, number of imputations, and complexity of covariates increase.

Guided by the associations contained in Table 10, victim age, race, and sex, plus urbanness and circumstances were selected to construct donor pools for the hot-deck imputation. Also, the same hot-deck approach was performed successively for 3-year intervals of homicide cases, after which the 3-year imputation subfiles were combined. Not only did the grouping of years greatly shorten the overall computing time, but it also avoided the potential for biasing year-dependent patterns by forcing imputation and donor cases to come from similar time periods.[8]

Table 11 shows offender age, race, and sex characteristics for the known cases, imputed cases, and total file.[9] As with the imputation-weighted results presented earlier, the hot-deck approach increases the percentage of young assailants (from 44.7% to 47.1% of offenders younger than 25), of Blacks (from 51.3% to 52.7%), and very slightly that of males (from 87.9% to 89.1%). Of course, the effect of imputation is constrained by the fact that it can only affect the 30% of cases having missing data. That is, an overall change of one percentage point implies that the imputed cases differed by more than 3%.

This can be seen clearly by comparing the distribution of imputed cases to that of the known offenders. For the imputed

TABLE 11
Hot-Deck Imputation of Offender Characteristics

	Known Cases		*Imputed Cases*		*Total*	
Offender Characteristic	N	%	N	%	N	%
Offender age						
Younger than 18	37,912	10.4	19,749	12.7	57,661	11.1
18-24	124,714	34.3	62,131	39.9	186,845	36.0
25-34	105,778	29.1	44,067	28.3	149,845	28.8
35-49	66,879	18.4	22,395	14.4	89,274	17.2
50+	28,249	7.8	7,558	4.8	35,807	6.9
Total	363,532	100.0	155,900	100.0	519,432	100.0
Offender race						
White	178,939	46.8	57,052	41.6	235,991	45.4
Black	195,908	51.3	77,925	56.8	273,833	52.7
Other	7,407	1.9	2,201	1.6	9,608	1.8
Total	382,254	100.0	137,178	100.0	519,432	100.0
Offender sex						
Male	338,461	87.9	124,291	92.6	462,752	89.1
Female	46,764	12.1	9,916	7.4	56,680	10.9
Total	383,412	100.0	133,881	100.0	519,432	100.0

cases, as many as 52.6% are younger than 25 years of age, 56.8% are Black, and 92.6% are male. Each is several percentage points greater than in comparable distributions of known cases, managing to shift the overall distributions to a somewhat lesser extent. The differences between imputation and known distributions in age and race occur, of course, because of the greater extent of unsolved homicides involving young Black victims in urban areas. When adjusted or imputed, the prevalence of young Black perpetrators grows.

Comparing overall distribution with hot-deck imputation from Table 11 with the revised imputation weighted distribution in the right panels of Table 8 shows remarkably similar results. This is to be expected as both methods, although approaching the task in different ways, use similar sets of covariates to produce the final result.

The hot-deck approach was employed both for the five offender age categories in Table 11 as well as for age as an interval-level measure. Figure 1 depicts the complete offender age distribution of known, imputed, and total cases. Consistent with the results for age categories, the peakedness of the distribution associated with young adult perpetrators is increased by the imputation process.

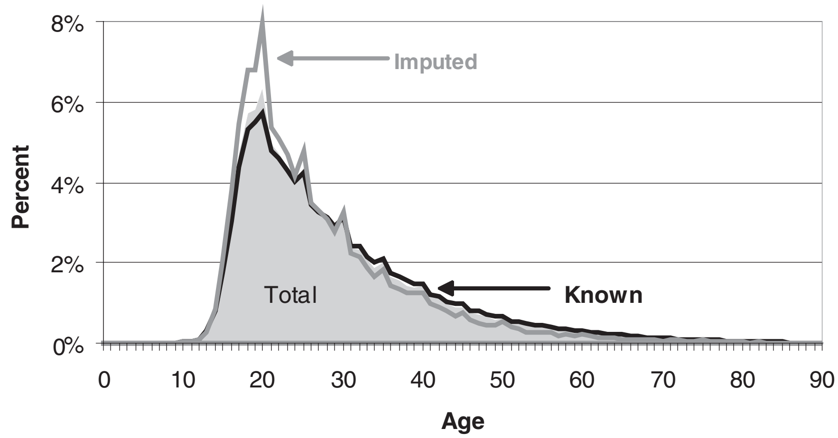

Figure 1: Offender Age Distribution With Hot-Deck Imputation.

IMPUTING INTIMATE HOMICIDES

Concern for missing data in the SHR file extends beyond the age, race, and sex characteristics of perpetrators. Most notably, several researchers and policy analysts have been interested in determining or characterizing the victim-offender relationship information hidden within unsolved homicides.

Unlike the imputation efforts for offender characteristics, it is hardly defensible to assume that victim-offender relationship data are missing at random. To the contrary, we can expect that unsolved homicides are more likely perpetrated by a stranger or an acquaintance than a close friend or intimate, even after controlling for characteristics of the victim or the incident.

The MAR assumption is a rather critical one and has been the subject of some debate. In a recent article imputing victim-offender relationships in Chicago and Los Angeles homicide data sets, Regoeczi and Riedel (2003) argue forcefully that the MAR assumption can safely be invoked for missing victim-offender relationships. They claim not only that there is little counter-indicative evidence but also that certain data even support the validity of the MAR assumption.

While conceding that there are no formal tests of the MAR assumption, Regoeczi and Riedel (2003) point out that "for the data to be MAR, the missingness should be able to be predicted by other variables in the data set" (p. 158). They then proceed to show that missingness is related to variables such as circumstances, location, and so on. Regoeczi and Riedel also contend that "even if stranger homicides are more likely to be missing data on victim/offender relationship than other types of homicides, victim/offender relationship can still be MAR if other variables in the data set can be used to predict this difference" (p. 159). With this as their foundation, Regoeczi and Riedel proceed to show that victim-offender missingness is indeed related to a range of homicide characteristics.

In a strict sense, these assertions are true only if missingness in Y were completely predicted by other variables (X). Otherwise, missingness in variable Y could still be correlated with Y itself, after controlling for X variables. Essentially, if the partial correlation of Y and Prob(Y is missing) holding constant X is nonzero, then MAR does not hold. Of course, this is not an observable partial correlation, yet observing any correlation between P(Y is missing) and X that is less than 1.0 leaves open the possibility of data that are NMAR.

To illustrate this point further, consider the hypothetical data on victim-offender relationship by victim gender shown in Table 12 for a group of 220 male victims and 100 female victims. Of these, the victim-offender relationship is available for 120 of the male victims and 80 of the female victims. Missingness in victim-offender is associated with type of relationship (stranger/nonstranger), having a nonobservable $r = .33$ correlation coefficient.

Victim gender also correlates with missingness of victim-offender relationship ($r = .24$), with a higher prevalence of missingness for male victims (54.5%) than female victims (20.0%). Despite the apparently significant role of victim gender as a covariate, it fails to account for missingness in victim-offender relationship entirely (partial r of missingness and relationship type controlling for gender is .29).

Unfortunately, techniques for estimating data that are not missing at random (NMAR) are rather intractable and often include invoking some external or a priori information. Even so, the results are not always superior (see Allison, 2002; Little & Rubin, 2002).

Several analytic efforts have been undertaken to impute victim-offender relationship data in homicide files, however with varying degrees of success (see, e.g., Pampel & Williams, 2000;

TABLE 12
Hypothetical Case of Victim-Offender Relationship Not Missing at Random

	Male Victim		*Female Victim*	
	Missing	*Observed*	*Missing*	*Observed*
Stranger	60	40	10	10
Nonstranger	40	80	10	70

Williams & Flewelling, 1987). A number of regression-type approaches tend to favor or predict modal responses for imputation while ignoring other important, yet less frequent, categories. Because certain categories are never the most likely regardless of values of covariates, the results tend to regress unacceptably toward the mode.

For example, Pampel and Williams's (2000) imputation of victim-offender relationship, based on maximum likelihood, failed to assign any unknown relationship cases to an intimate partner category. That is, there was no combination of covariates for which the intimate partner classification was more predictable than friend/acquaintance or stranger. By contrast, the log-multiplicative association approach proposed by Messner, Deane, and Beaulieu (2002) succeeded in assigning some unsolved cases to intimate homicide by modeling victim-offender relationship as a set of ordered categories ranging from intimate to stranger.

Our focus here is specifically on the intimate/nonintimate distinction rather than a set of relationship codes, that is, the extent to which we can assess the prevalence of intimate homicide (homicides among spouses, ex-spouses, and boyfriends/girlfriends) after accounting for homicides with unknown victim-offender relationship. For the years 1976 to 2001 combined, 37,457 females were murdered by intimates, 52,671 were slain by nonintimates (e.g., siblings, friends, strangers), and another 34,594 were killed by an assailant of an undetermined relationship to the victim. Thus, intimate-partner assailants account for 30.0% of all homicides of female victims, yet 41.6% of the cases in which the victim-offender relationship has been identified. Clearly, decisions about how to handle the missing cases can have a major effect on interpretations concerning the prevalence of homicide involving intimate partners.

Recent analyses of intimate homicide have had to confront the problem of unknown/undetermined relationship. Reports published by the Bureau of Justice Statistics have assumed that the undetermined relationship cases are nonintimate (see Greenfeld et al., 1998; Rennison, 2003). Although it may be defensible to presume that most are nonintimate, treating all unknowns as nonintimate understates the intimate homicide problem. Yet, at the other extreme, using a listwise deletion of undetermined cases in estimating prevalence of intimate partner homicide (that is, distributing unknown cases proportionately between the intimate and nonintimate groups) would most assuredly exaggerate the role of intimate partners in the homicide commission.

As an alternative, we use the following ad-hoc approach, based on assumptions concerning the likely share of intimate homicide among unsolved cases for various subclasses of victim population. For some victims, for example young children, it makes sense to treat all unsolved homicides as nonintimate cases, as it does for certain types of homicide, for example, those classified as felony-related. Furthermore, it is reasonable to make varying assumptions about the mix of intimate and nonintimate homicides within different age and gender groups. Among men, we may assume that the overwhelming majority—in fact, almost all—of the unknown relationships involve nonintimates. We may also assume the likelihood that an unsolved homicide involves an intimate rises and falls with age of victim.

More formally, for a given subclass of the population of homicide victims, let N_1 represent the number of intimate homicides and N_2 the number of nonintimate homicides, neither of which is observable. Of course, N_1 and N_2 sum to N_t, the number of victims within the subclass, which is known. Of the N_1 intimate and N_2 nonintimate cases, M_1 are identified as intimate and M_2 as nonintimate. Furthermore, we assume no misclassification errors (that is, none of N_1 is included in M_2 and none of N_2 is in M_1).

The solve or clearance rates for intimate and nonintimate homicide (more exactly, the probability that intimate and nonintimate homicides are identified as such in the SHR), $P_1 = M_1/N_1$ and $P_2 = M_2/N_2$, are also unobserved.

The approach is to use available information on M_1, M_2, N_t, and the overall victim-offender solve rate, $P_t = (M_1 + M_2)/N_t$, to establish reasonable values of P_1 and P_2, which then can be employed to solve for M_1 and M_2.

It is clearly reasonable that the ability for the police to identify (solve) an intimate homicide is greater than that for nonintimate homicides. That is,

$$P_2 \leq P_1 \leq 1.$$

For certain victim subclasses (e.g., children, felony-murder), we shall assume that all unsolved homicides are nonintimate, that is, that all intimate homicides have been identified ($P_1 = 1$). For other subclasses, we make P_1 a function of P_2, such that

$$P_1 = q\, P_2 + (1 - q), \text{ where } 0 \leq q \leq 1.$$

The factor q governs the mix of intimate and nonintimate homicides among the unsolved cases (unidentified victim-offender relationship) for a given victim subclass. At one extreme ($q = 1$), we have $P_1 = P_2$ (the implausible listwise deletion approach). At the other extreme ($q = 0$), all unsolved homicides are treated as nonintimate ($P_1 = 1$).

Within these extremes, values of q are assigned based on victim gender, age, and circumstance. To provide some guideline for setting q as a function of victim gender and age, we can consider the ratio of intimate to nonintimate homicides among those that are identified (that is, M_1/M_2). As shown in Figure 2, the odds that a homicide will be classified as intimate rather than nonintimate is relatively low for men, yet it does tend to rise somewhat for men older than 25. Among female victims, the odds are modest for teens and seniors, moderate among women ages 18-24 and 50-65, and rather elevated for women ages 25-34 and 35-49. Although there is no exact way to translate the odds into values of q, we shall set the following: $q = 0$ for male victims younger than 25, and $q = .1$ for male victims older than 25; $q = 0$ for females younger than 14, $q = .1$ for women ages 14-17 and 65+, $q = .3$ for female victims ages 18-24 and 50-64, and $q = .5$ for female victims in the 25-34 and 35-49 age ranges. Moreover, for all cases classified as felony-murder, we use $q = 0$, regardless of victim age and gender.

Surely, the choice of q values is somewhat arbitrary, and there is no clear-cut way to calibrate these coefficients from the age/sex-specific odds of an intimate relationship among identified cases. Still, this approach is fairly conservative (with low q values)

Figure 2: Ratio of Intimate to Nonintimate Homicides by Victim Age and Sex.

yet preferred to the alternative of treating all unsolved cases as nonintimate.

The factor q determines how P_1 is established as a weighted average of P_2 (the listwise deletion extreme that would distribute unidentified cases proportionally to intimate and nonintimate categories) and 1.0 (the extreme where all undetermined cases are considered nonintimate). Even the most liberal value of q used here (.5 for nonfelony murder of women in the 25-49 age range) only halves the difference between these two extremes and never comes close to the unlikely listwise deletion assumption. Most victim subclasses (with q = .1 or q = .3) heavily favor the tendency to treat unsolved cases as nonintimate.

Beyond the effect of q on P_1, we also want P_1 to be influenced by the overall value of P. For victim subclasses in which the overall solution rate is high (approaching 1), the share of unsolved cases that can be assigned to intimate homicide diminishes.

Given the assumption and parameter relationship outlined above, we have this pair of equations:

$$P_1 = q\,P_2 + (1 - q)$$

$$M_1/P_1 + M_2/P_2 = N_t.$$

Although they contain just two unknowns (P_1 and P_2), the system is nonlinear and thus cannot be solved algebraically. Alternatively, we use a numerical solution for determining P_1, which then yields an estimate of $M_1 = N_1/P_1$, the number of intimate homicides within a particular victim subclass.

The population of victims was stratified by year (from 1976 to 2001), seven age categories (younger than 14, 14-17, 18-24, 25-34, 35-49, 50-64, and 65+), race (White and non-White), sex (male and female), five levels of urbanness (large city, medium city, small city, suburban, and rural), four regions (Northeast, Midwest, South, and West), and three circumstance categories (felony, argument, and other/undetermined). P_1 and P_2 were then estimated within each cell. Accordingly, P_1 varies by the extent to which police are able to solve homicides of a particular type and within a particular type of locale, that is, P_2 and thus P_1 both are a function of P_t.

Within each cell defined by a combination of the seven covariates, candidate values for P_2 from 0 to 1 in steps of .001 (.1%) were attempted. For each P_2 value, P_1 was calculated based on q, and then a value of N_t was derived predicated on the trial values of P_1 and P_2 and known values of M_1 and M_2. The trial values that produced a calculated N_t closest to the actual count of homicide victims within the cell were then used for that class. With this numerical solution for P_1 and P_2 came cell-specific estimates of N_1 and N_2, which then were aggregated across victim subgroups as needed.

Table 13 shows results of this imputation strategy for estimates of trends from 1976 to 2001. Separately for gender/race combinations, the estimated counts (known and imputed) are displayed along with the two extremes of assigning all missing cases to the nonintimate group and distributing missing cases proportionately to intimate and nonintimate categories. It is not surprising that because of the conservative assumptions about the nature of unsolved killings involving male victims, the estimated counts of intimate homicides of males remain quite close to the number of homicides known to be intimate in relationship. For White and non-White females, however, the estimated numbers are consistently and nontrivially higher than the known cases, as much as 100 per year for Whites and 50 per year for non-Whites (see also Figure 3).

Table 14 compares the known and estimated percentages of all homicides involving intimate partners, separately by age, gender,

TABLE 13
Trends in Intimate Homicide by Victim Sex and Race

	Male						*Female*					
	White			*Non-White*			*White*			*Non-White*		
Year	*Actual* $P_1 = 1$	*Estimate* $P_2 \le P_1 \le 1$	*Listwise Deletion* $P_1 = P_2$	*Actual* $P_1 = 1$	*Estimate* $P_2 \le P_1 \le 1$	*Listwise Deletion* $P_1 = P_2$	*Actual* $P_1 = 1$	*Estimate* $P_2 \le P_1 \le 1$	*Listwise Deletion* $P_1 = P_2$	*Actual* $P_1 = 1$	*Estimate* $P_2 \le P_1 \le 1$	*Listwise Deletion* $P_1 = P_2$
1976	489	495	602	854	861	982	845	898	1,020	727	752	813
1977	476	482	607	804	811	934	828	891	1,009	585	616	695
1978	488	497	728	708	717	916	864	938	1,075	596	639	700
1979	530	541	797	713	723	940	874	957	1,119	603	652	768
1980	491	501	718	720	731	943	905	984	1,161	620	668	748
1981	550	560	739	714	723	865	948	1,044	1,195	614	661	744
1982	506	515	674	625	632	758	942	1,030	1,150	529	578	642
1983	507	514	638	603	609	733	909	982	1,093	547	586	656
1984	440	446	573	537	542	622	932	986	1,089	496	526	592
1985	424	431	548	528	534	629	1,001	1,079	1,207	537	580	644
1986	446	454	599	530	537	660	993	1,072	1,210	582	622	692
1987	422	428	558	505	511	653	963	1,057	1,215	514	552	628
1988	374	380	517	474	481	639	1,006	1,089	1,233	560	614	709
1989	366	373	500	517	525	733	880	948	1,070	513	559	665
1990	391	397	541	459	465	590	948	1,033	1,171	532	594	730
1991	356	363	514	416	422	549	925	1,031	1,205	571	633	750
1992	336	343	485	379	386	557	885	983	1,152	554	615	773
1993	327	333	493	369	376	546	982	1,060	1,223	584	644	807
1994	318	323	498	364	370	526	897	977	1,124	495	549	700
1995	250	255	355	289	295	427	864	948	1,117	430	481	637
1996	259	265	395	254	257	340	851	931	1,045	448	490	574
1997	238	241	302	207	211	297	753	824	925	437	491	580
1998	271	276	360	234	238	299	872	933	1,057	432	474	570
1999	220	223	304	199	204	313	809	882	969	395	449	553
2000	229	233	322	206	211	339	846	917	1,007	378	430	508
2001	204	209	281	192	197	332	792	863	968	390	446	546

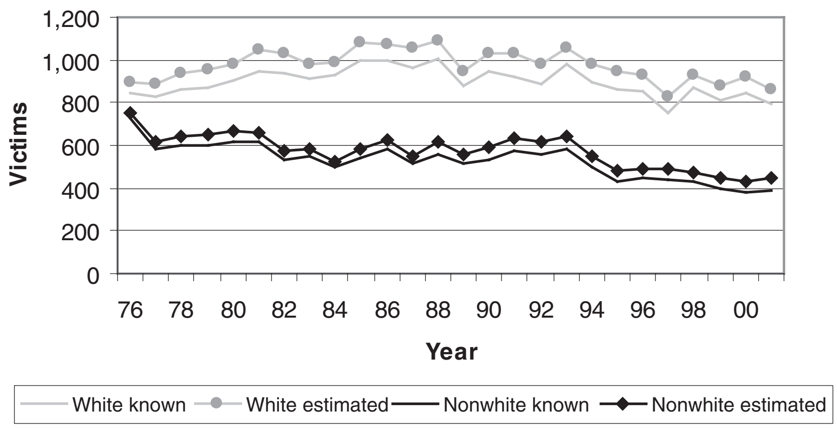

Figure 3: Known and Estimated Female Victims of Intimate Homicide.

and race of victim. For females in the highest risk age groups, the imputation model boosted the calculations by a factor of 10% or more. For example, the percentage of White female victims, ages 35-49, who are murdered by their intimate partners is estimated with imputation to be just short of 50%, rather than 45.0% based only on cases known to involve an intimate assailant. For non-White female victims in the same age group, the percentage of intimate perpetrators increases from 36.7% to 40.4%.

The validity of this ad-hoc method of imputation turns on the reasonableness of the assumptions as well as the utility of the covariates. To attempt some test of soundness, it may be useful to examine the extent to which the covariates adequately remove bias in measuring the percentage of homicides involving intimate partners caused by the extent of missing data on victim-offender relationship.

I divided police reporting agencies into groups based on the extent of completeness of victim-offender relationship data in their SHR submissions. Specifically, agencies were stratified by the percentage of available victim-offender information: less than 50%, from 50% to 60%, from 60% to 70%, from 70% to 80%, from 80% to 90%, and 90% and more. Within these strata, the percentage of homicides classified as intimate (based only on those cases for which victim-offender relationship data are available) increases monotonically as the clearance rate (more exactly, the victim-offender data availability rate) increases: 14.2%, 14.6%, 15.6%, 16.6%, 20.5%, and 21.6% for the six clearance rate strata,

TABLE 14
Percentage Intimate Homicides by Victim Sex, Race, and Age

		Male Victims					*Female Victims*				
		Known Intimate Victims		*Estimated Intimate Victims*			*Known Intimate Victims*		*Estimated Intimate Victims*		
Victim Race	*Victim Age*	N	%	N	%	*Total Victims*	N	%	N	%	*Total Victims*
White	younger than 14	132	1.8	132	1.8	7,357	151	2.4	151	2.4	6,363
	14-17	49	.6	49	.6	8,786	489	14.7	497	15.0	3,320
	18-24	718	1.7	718	1.7	42,318	3,704	30.2	3,963	32.3	12,252
	25-34	2,587	4.8	2,633	4.9	54,039	7,068	42.4	7,842	47.0	16,675
	35-49	3,967	8.5	4,041	8.7	46,713	7,448	45.0	8,211	49.6	16,548
	50-64	1,915	8.8	1,952	9.0	21,739	2,545	32.2	2,705	34.2	7,906
	65+	540	5.2	553	5.3	10,351	1,911	21.2	1,969	21.9	9,005
	all ages	9,907	5.2	10,079	1.8	191,305	23,316	32.4	25,337	35.2	72,070
Non-White	younger than 14	54	.9	54	.9	5,773	49	1.0	49	1.0	4,755
	14-17	62	.5	62	.5	11,756	338	14.2	344	14.4	2,378
	18-24	1,446	2.5	1,446	2.5	57,077	2,826	27.7	3,014	29.6	10,193
	25-34	4,263	6.7	4,336	6.8	64,001	5,240	34.3	5,822	38.2	15,260
	35-49	4,418	10.4	4,486	10.5	42,621	4,045	36.7	4,451	40.4	11,014
	50-64	1,756	11.5	1,779	11.7	15,211	878	25.6	921	26.8	3,434
	65+	402	6.7	408	6.8	6,039	295	10.7	301	10.9	2,765
	all ages	12,400	6.1	12,570	6.2	202,479	13,671	27.5	14,902	29.9	49,799
Total	younger than 14	186	1.4	186	1.4	13,130	199	1.8	199	1.8	11,118
	14-17	110	.5	110	.5	20,542	827	14.5	841	14.8	5,699
	18-24	2,164	2.2	2,164	2.2	99,395	6,530	29.1	6,977	31.1	22,445
	25-34	6,850	5.8	6,969	5.9	118,040	12,308	38.5	13,664	42.8	31,935
	35-49	8,385	9.4	8,527	9.5	89,334	11,494	41.7	12,661	45.9	27,562
	50-64	3,671	9.9	3,732	10.1	36,951	3,423	30.2	3,626	32.0	11,340
	65+	942	5.7	961	5.9	16,391	2,206	18.7	2,270	19.3	11,770
	all ages	22,308	5.7	22,650	5.8	393,784	36,988	30.4	40,239	33.0	121,869

TABLE 15
Percentage Intimate Homicides by Agency Clearance Rate and Location

Location	*< 50%*	*50-59%*	*60-69%*	*70-79%*	*80-89%*	*90%+*	*Total*
Large city	13.5	11.8	13.1	11.3	12.7	—	12.7
Medium city	12.1	15.8	14.9	16.6	17.4	16.4	15.4
Small city	20.7	14.4	16.0	19.5	19.7	21.1	18.8
Suburban	24.7	18.7	19.8	20.9	22.9	22.2	21.2
Rural	28.4	25.0	22.0	24.0	23.6	21.9	23.2
Total	14.2	14.6	15.6	16.6	20.5	21.6	16.5

respectively. This correlation between the percentage of intimate homicides and percentage of available victim-offender relationships could reflect potential bias in measuring intimate homicide rates resulting from missing data problems or from some specific covariate responsible for this association.

Table 15 displays the effect of clearance rates (i.e., victim-offender relationship data availability) on the percentage of known offenders classified as intimate partners, controlling for location urbanness. The row margin shows the overall effect of clearance on the intimate homicide percentage noted above. The column margin indicates that the percentage of known relationships classified as intimate increases monotonically with declining urbanness. Within urbanness levels, however, percentage intimate homicide levels are generally uncorrelated with data completion rates (see also Figure 4). Thus, using urbanness as a covariate in the imputation model (along with other factors associated with victim characteristics) would appear to minimize the extent of measurement bias caused by nonrandom missing data.

SUMMARY

The SHR data series, despite its limitations, continues to be a rich data source for researchers and policy makers. With attempts to impute missing information, its value can be greatly enhanced, particularly in terms of estimating patterns and trends by offender characteristics and victim-offender relationship.

In this article, we have explored several approaches for compensating for missing data in the SHR. After explicating the

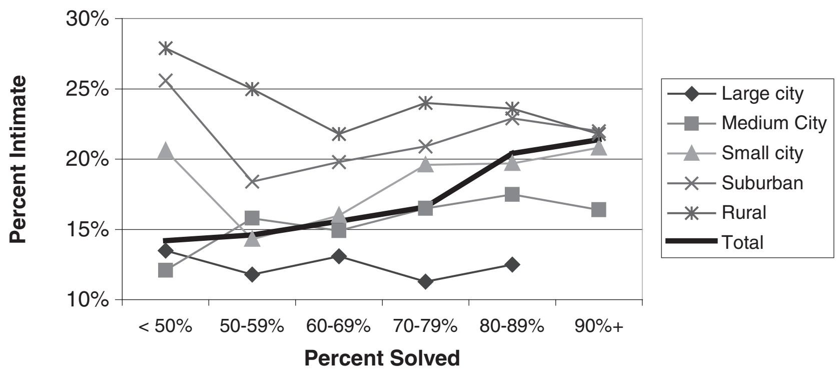

Figure 4: Percentage Intimate Homicides by Agency Clearance Rate and Location.

weighting strategy currently used in the SHR data files held at the National Criminal Justice Data Archive (NCJDA), we revised the weighting scheme based on an improved set of covariates. Not only did eliminating year and state as classification factors avoid the problem of low frequency cells and thus outlier weight values, but the inclusion of additional covariates such as urbanness, region, and circumstances better accounted for deficiencies in known-offender data.

The weighting approach was then compared with a hot-deck strategy that, while similar in logic, has a few methodological advantages in handling partially missing offender data. Overall, the results confirmed those of the revised weighting strategy. In both approaches, the percentage of young, Black, and male offenders increased by a small but important degree once imputation was performed. Although the hot-deck alternative has certain desirable features, it would not be recommended for archiving public use data files because of the inclusion of fictitious data values that could easily be mistaken for real.

After considering imputation of offender data, the article turned to adjustments for missing victim-offender relationship data, specifically the distinction between intimate and nonintimate homicide. Missingness on victim-offender relationship presented some special problems due to the potential bias from nonrandomness in missing reports. Specifically, I was unable to assume that the likelihood of missingness of victim-offender relationship was not related to the actual value of this characteristic, even after controlling for a host of covariates. As a result, an ad-hoc model was proposed based on specific assumptions about the relative solvability of intimate and nonintimate murders. Application of this strategy boosted the number of intimate homicides of at-risk females by about 10%.

In the future, and through NIBRS, efforts to classify homicide types better should ultimately help the imputation process. That is, the SHR limits the range of covariates to three basic victim characteristics (age, race, and sex), characteristics of the reporting agency (urbanness and region), as well as limited and often questionable information about circumstances (see Maxfield, 1989). Greater specificity in homicide data and additional efforts to understand the nature of missingness in the homicide reports should significantly enhance our ability to characterize and track patterns in homicide victimization and offending.

NOTES

1. In 1980, the FBI added ethnicity of victim and offender to the SHR data collection form in an attempt to distinguish Hispanic/non-Hispanic origin from race. The attempt was so unsuccessful in terms of data completeness that by the mid-decade, the FBI backed away from analyzing these data, although some agencies have continued to report these data fields.

2. On occasion, a state will report more homicide cases once all records are submitted by the final close-out date than the number estimated provisionally by the FBI for publication in the *Crime in the United States* annual volume.

3. Using a Poisson distribution, a 26-year count of 1,437 murders (an average of approximately 4.6 per month) yields a probability of a zero monthly count of less than 1%.

4. If an agency has between 3 and 11 monthly Return A (crimes known to the police) submissions for a calendar year, the reported offense counts are weighted upward by the ratio 12/N, where N is the number of complete submissions. For agencies submitting returns for fewer than 3 months, the data are ignored and annual counts for Part I crimes are imputed based on similar-sized agencies with complete returns in the same state (or region if the state offers no matching agencies). For details and discussion of this procedure, see Maltz (1999).

5. Treating these cases as a distinct class of event is consistent with the determination made by the FBI for terrorist-related homicides from September 11 for New York City, the Pentagon area, and Somerset County, PA (see FBI, 2002, pp. 302-307).

6. Some variables used in Table 10 are clearly ordinal, which generally would prescribe using an association measure, like Gamma, that exploits the ordering of categories. However, because the imputation procedures used here disregard category ordering, an association measure for nominal variables is used instead.

7. Overall for 1976-2001, 42.3% of homicides involve a victim and offender who are within 5 years of age of each other.

8. The program SOLAS was used to perform the hot-deck results (see Statistical Solutions, 2001). Even when restricting the number of covariates to the program's limit of five and subdividing the analysis into nine smaller tasks (one for 1976-1977 and then for 3-year intervals from 1978-1980 through 1999-2001), the analysis required in excess of 50 hours of processing time using a Pentium 4 machine, 3 GHz processor, and 1GB RAM.

9. The frequency and totals presented here were not adjusted for missing records or for cases dropped because of missing victim data (as were the weighted distributions previously given in Table 8). Such an adjustment, of course, would not affect the percentages.

REFERENCES

Allison, P. D. (2002). *Missing data* (Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-136). Thousand Oaks, CA: Sage.

Federal Bureau of Investigation. (2002). *Crime in the United States 2001: Uniform crime reports*. Washington, DC: Government Printing Office.

Fox, J. A. (2000). *Uniform crime reports [United States]: Supplementary homicide reports, 1976-1997* (ICPSR Version) [Computer file]. Boston: Northeastern University, College of Criminal Justice [Producer]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [Distributor].

Fox, J. A. (2001). *Uniform crime reports [United States]: Supplementary homicide reports, 1976-1998* (ICPSR Version) [Computer file]. Boston: Northeastern University, College of Criminal Justice [Producer]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [Distributor].

Fox, J. A. (2002). *Uniform crime reports [United States]: Supplementary homicide reports, 1976-1999* (ICPSR Version) [Computer file]. Boston: Northeastern University, College of Criminal Justice [Producer]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [Distributor].

Fox, J. A., & Zawitz, M. (2003). *Homicide trends in the United States*. Retrieved from http://www.ojp.usdoj.gov/bjs/homicide/homtrnd.htm

Greenfeld, L. A., Rand, M. A., Cravan, D., Klaus, P. A., Perkins, C. A., Ringel, C., et al. (1998). *Violence by intimates: Analysis of data on crimes by current or former spouses, boyfriends, and girlfriends* (Document No. NCJ-167237). Washington, DC: Bureau of Justice Statistics.

Little, R.J.A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.

Maltz, M. D. (1999). *Bridging gaps in police crime data* (Bureau of Justice Statistics Fellows Program Discussion Paper, Document No. NCJ-176365). Washington, DC: Bureau of Justice Statistics.

Maxfield, M. G. (1989). Circumstances in supplementary homicide reports: Variety and validity. *Criminology*, *27*, 671-694.

Messner, S. F., Deane, G., & Beaulieu, M. (2002). A log-multiplicative association model for allocating homicides with unknown victim-offender relationships. *Criminology*, *40*, 457-479.

Pampel, F. C., & Williams, K. R. (2000). Intimacy and homicide: Compensating for missing data in the SHR. *Criminology*, *38*, 661-680.

Regoeczi, W. C., & Riedel, M. (2003). The application of missing data estimation methods to the problem of unknown victim/offender relationships in homicide cases. *Journal of Quantitative Criminology*, *19*, 155-183.

Rennison, C. M. (2003). *Intimate partner violence, 1993-2001* (Document No. NCJ-197838). Washington, DC: Bureau of Justice Statistics.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
Statistical Solutions. (2001). *SOLAS: Statistical solutions system for data analysis*. Crosse's Green Cork, Ireland: Statistical Solutions, Ltd.
Williams, K. R., & Flewelling, R. L. (1987). Family, acquaintance and stranger homicide: Alternate procedures for rate calculations. *Criminology*, *25*, 543-560.

James Alan Fox *is the Lipman Family Professor of Criminal Justice at Northeastern University. He is a visiting fellow with the U.S. Department of Justice, Bureau of Justice Statistics, focusing of homicide measurement and analysis.*