# Adjusting the National Crime Victimization Survey's Estimates of Rape and Domestic Violence for "Gag" Factors

Ann L. Coker, PhD[1] and Elizabeth A. Stasny, PhD[2]

[1] Department of Epidemiology and Biostatistics, School of Public Health, University of South Carolina

[2] Department of Statistics, Ohio State University

# Introduction

## Purpose and Objectives

The **purpose** of this NIJ Data Resource Program funded project was to use statistical modeling techniques to estimate rape and domestic violence rates adjusting for interviewing conditions under which the National Crime Victimization Survey (NCVS) was administered. We found that the type of interview (personal or telephone) and the presence of another (particularly a spouse) influences or "gagged" the reporting of rape and domestic violence in the NCVS.

Our **specific objectives** were (1) to identify factors associated with reporting rape or domestic violence, (2) to identify factors associated with conditions of the interview (i.e. the type of interview (telephone or in-person) and who was present during the personal interview), and (3) to use statistical modeling techniques (expected maximization E-M algorithms) to estimate the annual incidence rates of rape and domestic violence adjusting for the conditions of the interview and other factors associated both with reporting the incident and the conditions of the interview.

## Why use the National Crime Victimization Survey?

We proposed adjusting the NCVS estimates of rape and domestic violence because this survey can uniquely provide data estimating crime rates including those reported to police and those not reported. As the majority of both rape and domestic violence are not reported to police, the NCVS provides more complete data than does the FBI Uniform Crime Report (UCR). The NCVS was designed to enumerate crime victimization in the USA. This survey, conducted by the US Census Bureau for the US Justice Department's Bureau of Justice Statistics, is large (>60,000 households) and was designed to be nationally representative (BJS, 1983). Although there are limitations of using the NCVS to estimate rates of both rape and domestic violence, this data set is the only on-going large and nationally representative survey to directly ask individuals whether they have or not have been victims of specific crimes.

## Why do NCVS rates need adjusting?

Reported rates of both rape and domestic violence vary widely. Annual **rape incidence rates** range from 0.7 per 1,000 women, reported from the Uniform Crime Report (UCR; 1986), to 9.0 per 1,000 in a San Francisco area study (Russell 1983). Rape incidence rates were 1.2 per 1,000 from the NCVS (1985). Crime Victims Research and Treatment Center (1992) reported a rate of 7 rapes per 1,000 women. Annual **domestic violence incidence rates** ranged from 2.2 per 1,000 women (NCVS data - BJS, 1986) to 160 incidents per 1,000 reported using the National Family Violence Survey (1986). The variance in reported rates may be due to differing definitions of rape and of domestic violence, differing purposes for which the data were collected, differing interviewing methods, and differing detail regarding the incident (victim-offender relationship, level of violence between intimates, completed rape - meaning forced sexual intercourse).

NCVS rape and domestic violence rates are under-estimates because: 1) whether a rape (sexual assault) occurred was not directly asked of NCVS respondents prior to the 1989 redesign (BJS 1989), 2) the definition of "criminal" rape or domestic violence was left to the respondent who may define rape differently relative to the legal definition or to other surveys, and 3) personal interviews may not be conducted confidentially (alone) and telephone interviews may not be sufficiently private. Both the telephone and personal interviews are conducted such that the respondent must tell the interviewer about the nature of the victimization in her own words; there are no yes/no questions characterizing the incident. Yes/no questions regarding a crime incident, but particularly rape or domestic violence, would allow victims to report these crimes in telephone interviews even with spouses present in the home during the interview. The NCVS does address privately issues by instructing interviewers to conduct interviews individually; the majority (>45%) are interviewed individually. Further, the issue of the privacy during the NCVS interview is amenable to statistical modeling techniques to estimate the rates of rape and domestic violence using existing data on both the mode of the interview and the persons present during the personal interview. This statistical adjustment then is the focus of our work.

### Who benefits from these findings?

Two distinct group will gain from the results of this study: the criminal justice community and the public health community. The **criminal justice community** needs to know the actual rates of rape and domestic violence, not only those reported through UCR, to **evaluate** how well the criminal justice community meets the needs of victims and how well criminal justice interventions may reduce the rates of rape or domestic violence. The NCVS is currently used as the baseline and only population-based data source to estimate rates of rape for the **Centers for Disease Control and Prevention** *Healthy People 2000* report (US Public Health Service, 1990). Interventions to reduce rape (or domestic violence) can and will use the NCVS data to **evaluate** their efficacy.

## Findings

### Unadjusted violence rates (Table 1 and 2)

We used the NCVS unweighted data for years 1986-1990. Only women 16 years of age or older were included in our analyses. We excluded proxy interviews. In total, we identified 428 rapes, 1942 incidents of domestic violence, 13,429 other assaults, 88,950 incidents of breaking and entry, and 28,780 other crimes reported by in the NCVS. A total of 542,072 women reported no crime incident during this period. The crude crime rates were as follows: rape, 0.66 per 1,000 women; domestic violence, 3.00 per 1,000; other assaults, 20.81 per 1000; and breaking and entry, 137.51 per 1,000. These frequencies and rates are presented in Tables 1 and 2.

### Differing violence rates by setting of the NCVS interview (Table 3)

We found that crime reporting rates did differ by the setting of the interview and the type of crime reported (See Table 3). Rape was reported 2.4 times less frequently in telephone compared with in-person interviews independent of who was present during the interview. Similarly domestic violence was reported 3.1 times less frequently in telephone versus in-person interviews. Other assaults were reported 50% less frequently and breaking and entry were reported 33% less frequently in telephone

versus in-person interviews. We noted that the magnitude of the "gagging" effect was smaller for both other assaults and breaking and entry compared with that for rape and domestic violence. We further explored the "gagging" effect of who was present during the interview for in-person interviews only because it is only for these interviews that this information was available. Compared to the ideal setting, women interviewed alone, rape was reported 2.8 times less frequently when a spouse was present but 1.35 times more frequently when others (excluding the spouse) were present. Similarly, domestic violence was reported 5.6 times less frequently if a spouse was present for the interview but 1.8 times more frequently if others were present. These results support the specificity of these "gag" factors for rape and domestic violence as these factors are not associated with great differences in reporting other assaults or breaking and entry. Compared with those interviewed alone, those interviewed with a spouse present were 1.2 times **more** likely to report other assaults and were no more likely to report breaking and entry. Compared with those interview alone, those interviewed with others present (excluding the spouse) were 1.9 times more likely to report other assaults and 1.4 times more likely to report breaking and entry.

| Table 1. Frequency Distribution of Crimes Reported by Settings of the Interviews: NCVS 1986-1990 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type of Interview | Who was present during interview | | Type of Crime (Frequencies) | | | | | |
| | Person | Frequency | Rape | Domestic Violence | Other Assault | Breaking and Entry | All other crime | No crime reported |
| Phone | Unknown | 469567 | 224 | 893 | 8777 | 60870 | 16972 | 398803 |
| In-Person | Alone | 81452 | 96 | 452 | 1637 | 11718 | 5655 | 67549 |
| | Spouse only | 23520 | 6 | 17 | 431 | 2807 | 627 | 20259 |
| | Spouse and others | 14522 | 10 | 21 | 491 | 2352 | 630 | 11648 |
| | Others (not spouse) | 57790 | 92 | 559 | 2123 | 11203 | 4896 | 43813 |
| All In-Person | | 177284 | 204 | 1049 | 4682 | 28080 | 11808 | 143269 |
| All Interviews | | 646851 | 428 | 1942 | 13459 | 88950 | 28780 | 542072 |

| Table 2. Crime Reporting Rates by Settings of the NCVS Interviews (1986-1990) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type of Interview | Who was present during interview | | Type of Crime (Rates per 1,000) | | | | | |
| | Person | Frequency | Rape | Domestic Violence | Other Assault | Breaking and Entry | All other crime | No crime reported |
| Phone | Unknown | 469567 | 0.48 | 1.90 | 18.69 | 129.63 | 36.14 | 849.30 |
| In-Person | Alone | 81452 | 1.18 | 5.55 | 20.10 | 143.86 | 69.43 | 829.31 |
| | Spouse only | 23520 | 0.26 | 0.72 | 18.32 | 119.35 | 26.66 | 861.35 |
| | Spouse and others | 14522 | 0.69 | 1.45 | 33.81 | 161.96 | 43.38 | 802.09 |
| | Others (not spouse) | 57790 | 1.59 | 9.67 | 36.74 | 193.86 | 84.72 | 758.14 |
| All In-Person | | 177284 | 1.15 | 5.92 | 26.41 | 158.39 | 66.60 | 808.13 |
| All Interviews | | 646851 | 0.66 | 3.00 | 20.81 | 137.51 | 44.49 | 838.02 |

## "Risk factors" for reporting violence crimes (Table 4)

From the crude comparisons (see Table 3), it is impossible to determine whether differences in reporting are solely a function of the interview setting or of other risk factors for experiencing a crime which are also associated with the setting of the interview (confounders). We explored the possibility that demographic confounders may explain these differences in crime reporting rates by "gag" factors. To do this we first characterized correlates of reporting specific crimes, then correlates of the actual interview setting, and finally conducted logistic regression analyses characterizing the association between the "gag" factors interview setting and the specific crime categories adjusting for these potential confounders (see Table 3). The following describes these findings.

In these data we investigated correlates of reporting rape, domestic violence, other assaults and breaking and entry. The correlates available for investigation were primarily demographic in nature. Our results are presented in Table 4. Compared with women not reporting rape, those reporting were younger (mean age 26.3; standard deviation 8.8), more likely to be non-white, to have annual incomes of less than $15,000, to be unemployed, to rent rather than own their home, to live in households with more than 5 persons, not to be currently married, to live in urban areas, within SMSAs, to move more than 5 times in the last three years, and not have been in sample by NCVS last interview. Resulting correlates of reporting domestic violence were similar to those for those rape. Compared with women not reporting domestic violence, those reporting were younger (mean age 29.5; standard deviation 9.1), more likely to have a high school or less, to have annual incomes of less than $15,000, to be unemployed, to rent rather than own their home, **not** to be currently married, to live within SMSAs, to move more than 5 times in the last three years, and not have been in sample by NCVS last interview. Those reporting other assaults were also younger (mean age 31.5; standard deviation 13.3), to be non-white, to have annual incomes of less than $15,000, to be unemployed, to rent rather than own their home, **not** to be currently married, to live within SMSAs, to move more than 5 times in the last three years, and not have been in sample by NCVS last interview. Correlates of reporting breaking and entry differ slightly from those of reporting the various assaults; having a higher income (>$15,000) was

positively associated with breaking and entry. However, neither higher education, home ownership, nor being employed were similarly correlated with reporting breaking and entry.

## Adjusted rate ratios for association between interview setting and reporting crime (Table 3)

After identifying correlates of reporting specific crimes, we calculated adjusted rate ratios for reporting crimes by interview setting (Table 3). These rate ratios were adjusted for the interviewee's age, her race, family income, home ownership, number of persons living in the household, whether she lived within an SMSA, and whether her current housing unit was in NCVS sample last interview. From these adjusted rate ratios, we again observed that relative to in-person interviews, all crime categories were reported less frequently during telephone interviews, however, the magnitude of the adjusted rate ratios were smaller than the crude. The adjusted rate ratios for phone interviews and reporting rape was no longer statistically significant (95% CI 0.13-1.57). Interview by telephone do appear to be a significant "gag" factor in reporting domestic violence; domestic violence was reported 60% less frequently in telephone compared with in-person interviews after adjusting for demographic characteristics of who completes a phone versus an in-person interview. Although both other assaults and breaking and entry were less frequently reported in telephone interviews, after adjusting for demographic characteristics, the magnitude of the relative "gag" factor was less than 20%. When looking at who was present during the in-person interview, with the referent group being interviewed alone, the importance of the spouse's presence as a "gag" factor on reporting domestic violence and, to a less extent, on reporting rape, becomes clear. If a spouse was present during an interview, the woman was 5.3 times less likely to report domestic violence than if she has been interviewed alone. Although not statistically significant, women interviewed with a spouse present were 2.2 times **less** likely to report rape than those interviewed alone (95% CI 0.6-7.7). Reporting other assaults or breaking and entry were not similarly associated with a consistent nor marked "gagging" of reports based on the presence of a spouse.

**Table 3: Rate Ratios for Relative Rates of Crime Reporting by Setting of the NCVS Interviews (1986-1990)**

| Type of Interview | Who was present during interview | Rate Ratios and 95% Confidence Intervals | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Rape | | Domestic Violence | | Other Assaults | | Breaking and Entry | |
| | | Crude | Adjusted* | Crude | Adjusted* | Crude | Adjusted* | Crude | Adjusted* |
| Phone | Unknown | 0.41<br>0.29-0.60 | 0.76<br>0.47-1.23 | 0.32<br>0.27-0.38 | 0.61<br>0.49-0.75 | 0.67<br>0.62-0.72 | 0.81<br>0.74-0.88 | 0.75<br>0.73-0.78 | 0.86<br>0.83-0.89 |
| All In-Person | Varies | 1.00<br>REF | 1.00<br>REF | 1.00<br>REF | 1.00<br>REF | 1.00<br>REF | 1.00<br>REF | 1.00<br>REF | 1.00<br>REF |
| In-Person | Alone | 1.00<br>REF | 1.00<br>REF | 1.00<br>REF | 1.00<br>REF | 1.00<br>REF | 1.00<br>REF | 1.00<br>REF | 1.00<br>REF |
| | Spouse | 0.36<br>0.13-0.98 | 0.45<br>0.13-1.57 | 0.18<br>0.09-0.34 | 0.19<br>0.09-0.40 | 1.21<br>1.04-1.42 | 1.23<br>1.03-1.47 | 0.93<br>0.87-1.00 | 0.91<br>0.84-0.98 |
| | Others (not spouse) | 1.35<br>0.78-2.34 | 0.97<br>0.47-2.00 | 1.76<br>1.39-2.23 | 1.21<br>0.90-1.62 | 1.86<br>1.64-2.11 | 1.42<br>1.23-1.65 | 1.43<br>1.36-1.51 | 1.14<br>1.07-1.21 |

* Adjusted for interviewee's age, her race, family income, home ownership, number of persons living in the household, whether she lived within an SMSA, and whether her current housing unit was in NCVS sample last interview.

**Table 4: Correlates of Reporting Crimes to NCVS Interviewers by type of crime among women ages 16+ (excluding proxy interviews)**

| Correlate | Odds Ratios and 95% Confidence Intervals[b] | | | |
|---|---|---|---|---|
| | Rape[a] | Domestic Violence[a] | Assault[a] | Breaking and Entry[a] |
| Age (Mean ± STD) | 26.31 ± 8.82 | 29.47 ± 9.12 | 31.52 ± 13.28 | 35.37 ± 14.98 |
| Age (in years) | 0.96 (0.94-0.98) | 0.96 (0.95-0.97) | 0.97 (0.97-0.98) | 0.98 (0.98-0.98) |
| Race: Non-White vs. White | 2.00 (1.30-3.03) | 1.25 (0.99-1.59) | 1.30 (1.18-1.41) | 1.16 (1.12-1.20) |
| Education: >12 years vs 12 years | 1.09 0.75-1.59 | 0.75 0.62-0.90 | 1.02 0.95-1.09 | 1.35 1.32-1.39 |
| Income ≤ $15,000 vs > $15,000 | 2.55 (1.72-3.79) | 2.45 (2.05-2.93) | 1.14 1.05-1.22 | 0.88 (0.85-0.91) |
| Job in last 6 months: No vs Yes | 3.25 1.89-5.58 | 2.51 1.88-3.36 | 2.14 1.90-2.40 | 2.06 1.96-2.17 |
| Home ownership: Rent vs. Own | 3.57 (2.44-5.26) | 4.00 (3.45-4.76) | 1.75 (1.64-1.85) | 1.52 (1.47-1.56) |
| # Persons in Household: > 5 vs ≤ 5 | 2.55 (1.00-6.47) | 0.87 (0.42-1.80) | 2.40 (2.00-2.87) | 2.03 (1.86-2.22) |
| Current Marital Status: Not Married vs Married | 3.85 2.50-5.88 | 4.76 3.85-5.88 | 1.33 1.25-1.43 | 1.18 1.14-1.20 |
| Urban vs Rural Residence | 2.09 1.24-3.52 | 1.17 0.96-1.44 | 1.45 1.33-1.57 | 1.60 1.55-1.66 |
| Location in SMSA: Yes vs. No | 2.22 (1.54-3.20) | 1.28 (1.07-1.53) | 1.47 (1.37-1.57) | 1.48 (1.43-1.52) |
| Number of moves in last 5 years: >3 vs ≤3 | 2.81 1.80-4.41 | 3.25 2.66-3.97 | 1.98 1.79-2.19 | 1.82 1.74-1.91 |
| Same Household last interview: No vs Yes | 3.13 (2.04-4.76) | 4.00 (3.33-5.00) | 1.85 (1.72-2.04) | 1.82 (1.75-1.89) |

[a] Comparison group includes those reporting no crime incident or another crime type.

[b] 95% CI confidence intervals adjusted for the NCVS design effect (1.92).

## Initial Model for "Gag" Factors

We start our modeling considering only who was present during the interview, what crime was reported, and whether the interview was conducted in person or by telephone. We use four classifications for who, in addition to the woman being interviewed, was present during the interview: 1) a spouse/significant other and no one else, 2) spouse /significant other and at least one other person, 3) at least one person but no spouse/significant other, and 4) no one present. The three crime reporting classifications we use are: 1) rape (or domestic violence) and possibly some other crime, 2) other crime, and 3) no crime. The summaries of the data are provided below in Table 5A and 5B. Because interviewer observation is used to obtain the information on who is present during an interview, we do not know who was present during telephone interviews.

**Table 5A**: Observed Data (Frequencies) for Reporting Rapes

| | In Person Interview | | | | Telephone |
|---|---|---|---|---|---|
| Type of Crime | Spouse | Spouse and Other | Other | Alone | Interview |
| Rape | 6 | 10 | 92 | 96 | 224 |
| Other Crime | 3221 | 2834 | 13649 | 13626 | 66986 |
| No Crime | 20293 | 11678 | 44049 | 67730 | 402357 |

**Table 5B**: Observed Data (Frequencies) for Reporting Domestic Violence

| | In Person Interview | | | | Telephone |
|---|---|---|---|---|---|
| Type of Crime | Spouse | Spouse and Other | Other | Alone | Interview |
| Domestic Violence | 17 | 21 | 562 | 452 | 893 |
| Other Crime | 3210 | 2823 | 13179 | 13270 | 66317 |
| No Crime | 20293 | 11678 | 44049 | 67730 | 402357 |

The data in Tables 5A and 5B is not, of course, the "truth". Some women may not report an incidence of rape or domestic violence to a NCVS interviewer under any circumstances. There is no information in the data to help us adjust for such cases. There are two cases which we may be able to adjust for. Namely, the differential reporting of crimes in telephone versus personal interviews, and the differential reporting of crimes depending on who is present during the interview. We now describe an initial model for the observed data that takes into account the potential under reporting of rapes and domestic violence incidence based on telephone interviews and the presence of a spouse during the interview.

We first suppose that, if we could only observe it, the complete data underlying Table 5A would tell us if a crime actually occurred and whether it was reported, was not reported because of the presence of a spouse, or was not reported because the interview was conducted over the telephone. In addition, if the interview was conducted over the telephone, the complete data would include information on who was present for the interview. Thus, the complete (but unobserved) data for reporting rapes would be as shown in Table 6.

## Table 6: Form of Unobserved Complete Data

**Personal Interviews**

| | | Spouse | Spouse+Other | Other | Alone |
|---|---|---|---|---|---|
| Rape | Reported | $y_{1111}$ | $y_{1112}$ | $y_{1113}$ | $y_{1114}$ |
| | Not Reported-Spouse Present | $y_{1121}$ | $y_{1122}$ | - | - |
| Other Crime | Reported | $y_{1211}$ | $y_{1212}$ | $y_{1213}$ | $y_{1214}$ |
| | Not Reported-Spouse Present | $y_{1221}$ | $y_{1222}$ | - | - |
| No Crime | Reported | $y_{1311}$ | $y_{1312}$ | $y_{1313}$ | $y_{1314}$ |

**Telephone Interviews**

| | | Spouse | Spouse+Other | Other | Alone |
|---|---|---|---|---|---|
| Rape | Reported | $y_{2111}$ | $y_{2112}$ | $y_{2113}$ | $y_{2114}$ |
| | Not Reported-Spouse Present | $y_{2121}$ | $y_{2122}$ | - | - |
| | Not Reported-Phone Interview | $y_{2131}$ | $y_{2132}$ | $y_{2133}$ | $y_{2134}$ |
| Other Crime | Reported | $y_{2211}$ | $y_{2212}$ | $y_{2213}$ | $y_{2214}$ |
| | Not Reported-Spouse Present | $y_{2221}$ | $y_{2222}$ | - | - |
| | Not Reported-Phone Interview | $y_{2231}$ | $y_{2232}$ | $y_{2233}$ | $y_{2234}$ |
| No Crime | Reported | $y_{2311}$ | $y_{2312}$ | $y_{2313}$ | $y_{2314}$ |

Note that some outcomes are impossible, for example, not reporting because the spouse/significant other is present when the woman was alone for the interview. Such impossible outcomes are denoted by

a dash in Table 6. The complete data for reporting domestic violence would be similar to that for rapes and, hence, is not shown here.

We now present a model to describe the probabilistic relationship between the underlying complete data and the data actually observed. We use the following notation:

$\omega_{ij}$ = probability of crime status I and interview status j

where I = 1 if rape, 2 if other crime, 3 if no crime

j = 1 if spouse, 2 if spouse+other, 3 if other, 4 if alone

and $\Sigma_i \Sigma_j \omega_{ij} = 1$

p = probability of a telephone interview

1 - $\rho$ = probability crime not reported because of telephone interview

1 - $\xi$ = probability crime not reported because spouse is present

The probabilities underlying the complete data can then be given as in Table 7.

## **Table 7**: Probabilities Underlying Unobserved Complete Data

Personal Interviews

| | | Spouse | Spouse+Other | Other | Alone |
|---|---|---|---|---|---|
| Rape | Reported | $(1-p)\xi\omega_{11}$ | $(1-p)\xi\omega_{12}$ | $(1-p)\omega_{13}$ | $(1-p)\omega_{14}$ |
| | Not Reported-Spouse Present | $(1-p)(1-\xi)\omega_{11}$ | $(1-p)(1-\xi)\omega_{12}$ | - | - |
| Other Crime | Reported | $(1-p)\xi\omega_{21}$ | $(1-p)\xi\omega_{22}$ | $(1-p)\omega_{23}$ | $(1-p)\omega_{24}$ |
| | Not Reported-Spouse Present | $(1-p)(1-\xi)\omega_{21}$ | $(1-p)(1-\xi)\omega_{22}$ | - | - |
| No Crime | Reported | $(1-p)\omega_{31}$ | $(1-p)\omega_{32}$ | $(1-p)\omega_{33}$ | $(1-p)\omega_{34}$ |

Telephone Interviews

| | | Spouse | Spouse+Other | Other | Alone |
|---|---|---|---|---|---|
| Rape | Reported | $p\xi\rho\omega_{11}$ | $p\xi\rho\omega_{12}$ | $p\xi\rho\omega_{13}$ | $p\xi\rho\omega_{14}$ |
| | Not Reported-Spouse Present | $p(1-\xi)\rho\omega_{11}$ | $p(1-\xi)\rho\omega_{12}$ | - | - |
| | Not Reported-Phone Interview | $p\xi(1-\rho)\omega_{11}$ | $p\xi(1-\rho)\omega_{12}$ | $p(1-\rho)\omega_{13}$ | $p(1-\rho)\omega_{14}$ |
| Other Crime | Reported | $p\xi\rho\omega_{21}$ | $p\xi\rho\omega_{22}$ | $p\xi\rho\omega_{23}$ | $p\xi\rho\omega_{24}$ |
| | Not Reported-Spouse Present | $p(1-\xi)\rho\omega_{21}$ | $p(1-\xi)\rho\omega_{22}$ | - | - |
| | Not Reported-Phone Interview | $p\xi(1-\rho)\omega_{21}$ | $p\xi(1-\rho)\omega_{22}$ | $p(1-\rho)\omega_{23}$ | $p(1-\rho)\omega_{24}$ |
| No Crime | Reported | $p\omega_{31}$ | $p\omega_{32}$ | $p\omega_{33}$ | $p\omega_{34}$ |

## Table 8: Form of Observed Data

| | Personal Interviews | | | |
|---|---|---|---|---|
| | Spouse | Spouse+Other | Other | Alone |
| Rape Reported | $x_{111} = y_{1111}$ | $x_{112} = y_{1112}$ | $x_{113} = y_{1113}$ | $x_{114} = y_{1114}$ |
| Other Crime Reported | $x_{121} = y_{1211}$ | $x_{122} = y_{1212}$ | $x_{123} = y_{1213}$ | $x_{124} = y_{1214}$ |
| No Crime Reported | $x_{131} = y_{1121} + y_{1221} + y_{1311}$ | $x_{132} = y_{1122} + y_{1222} + y_{1312}$ | $x_{133} = y_{1313}$ | $x_{134} = y_{1314}$ |

| | Telephone Interviews |
|---|---|
| Rape Reported | $x_{21} = y_{2111} + y_{2112} + y_{2113} + y_{2114}$ |
| Other Crime Reported | $x_{22} = y_{2211} + y_{2212} + y_{2213} + y_{2214}$ |
| No Crime Reported | $x_{23} = y_{2121} + y_{2122} + y_{2131} + y_{2132} + y_{2133} + y_{2134} + y_{2221} + y_{2222} + y_{2231} + y_{2232} + y_{2233} + y_{2234} + y_{2311} + y_{2312} + y_{2313} + y_{2314}$ |

In the observed data some of the cells from the complete data are mixed up together. Hence we observe only sums of several cells together rather than all 32 possible cells represented in the complete-data table. Table 8 presents the notation for the observed data table and indicates which cell counts from the complete data table are summed together to create the observed data. Again, the observed data for domestic violence would be similar to that for rapes and, hence, is not shown here.

The probabilities underlying the observed data are similarly just the sums of the probabilities underlying the unobserved complete data. They are shown in Table 8.

We may estimate the p, ρ, ξ, and ω parameters under this model using the EM-algorithm (see, for example, Dempster, Laird, and Rubin (1977)). The M-step involves maximizing the complete data likelihood function obtained using the cell probabilities shown in Table 7 and the complete data from Table 6, subject to the constraint that $\Sigma_i\Sigma_j \omega_{ij} = 1$. The likelihood function has a simple, multiplicative form and may be split into four factors, each a function of only of the p, ρ, ξ, and ω parameters.

The likelihood function, written so that the functions of the four types of parameters are obvious, is proportional to the following: (A "•" in a subscript indicates summation over the corresponding index.)

$$(1-p)y_{1\cdots} \times p y_{2\cdots}$$

$$\times \xi y_{1111} + y_{1112} + y_{1211} + y_{1212} + y_{2111} + y_{2112} + y_{2113} + y_{2114} + y_{2131} + y_{2132} + y_{2231} + y_{2232}$$

$$\times (1-\xi)y_{1121} + y_{1122} + y_{1221} + y_{1222} + y_{2121} + y_{2122} + y_{2221} + y_{2222}$$

$$\times \rho y_{2111} + y_{2112} + y_{2113} + y_{2114} + y_{2211} + y_{2212} + y_{2213} + y_{2214}$$

$$\times (1-\rho)y_{2131} + y_{2132} + y_{2133} + y_{2134} + y_{2231} + y_{2232} + y_{2233} + y_{2234}$$

$$\times \omega_{11}y_{1111} + y_{1121} + y_{2111} + y_{2121} + y_{2131} \times \omega_{12}y_{1112} + y_{1122} + y_{2112} + y_{2122} + y_{2132}$$

$$\times \omega_{13}y_{1131} + y_{2113} + y_{2133} \times \omega_{14}y_{1114} + y_{2114} + y_{2134} \times \omega_{21}y_{1211} + y_{1221} + y_{2211} + y_{2221} + y_{2231}$$

$$\times \omega_{22}y_{1212} + y_{1222} + y_{2212} + y_{2222} + y_{2232} \times \omega_{231213} + y_{2213} + y_{2233} \times \omega_{24}y_{1214} + y_{2214} + y_{2234}$$

$$\times \omega_{31}y_{1311} + y_{2311} \times \omega_{32}y_{1312} + y_{2312} \times \omega_{33}y_{1313} + y_{2313} \times \omega_{34}y_{1314} + y_{2314}$$

$$\equiv (1-p)y_{1\cdots} \times p y_{2\cdots} \times \xi a_1 \times (1-\xi)a_2 \times \rho b_1 \times (1-\rho)b_2 \times \{\Pi_{i=1}^{3} \Pi_{j=1}^{4} \omega_{ij} y^{\bullet i \bullet j}\}$$

Because of the form of this likelihood function, maximization may be accomplished separately for the p, ρ, ξ, and ω parameters. The closed-form estimators are shown below.

$\hat{p} = y_{2\cdots} / (y_{1\cdots} + y_{2\cdots})$ = (# phone interviews) / (# phone + personal interviews)

$\hat{\xi} = a_1 / (a_1 + a_2)$

$\hat{\rho} = b_1 / (b_1 + b_2)$

$\hat{y}_{ij} = y_{\bullet i \bullet j} / y_{\cdots\cdot}$

## **Table 9**: Probabilities Underlying the Observed Data

Personal Interviews

| | Spouse | Spouse+Other | Other | Alone |
|---|---|---|---|---|
| Rape Reported | $(1-p)\xi\omega_{11}$ | $(1-p)\xi\omega_{12}$ | $(1-p)\omega_{13}$ | $(1-p)\omega_{14}$ |
| Other Crime Reported | $(1-p)\xi\omega_{21}$ | $(1-p)\xi\omega_{22}$ | $(1-p)\omega_{23}$ | $(1-p)\omega_{24}$ |
| No Crime Reported | $(1-p)(1-\xi)\omega_{11}$ + $(1-p)(1-\xi)\omega_{21}$ + $(1-p)\omega_{31}$ | $(1-p)(1-\xi)\omega_{12}$ + $(1-p)(1-\xi)\omega_{22}$ + $(1-p)\omega_{32}$ | $(1-p)\omega_{33}$ | $(1-p)\omega_{34}$ |

Telephone Interviews

| | |
|---|---|
| Rape Reported | $p\rho[\xi(\omega_{11} + \omega_{12}) + \omega_{13} + \omega_{14}]$ |
| Other Crime Reported | $p\rho[\xi(\omega_{21} + \omega_{22}) + \omega_{23} + \omega_{24}]$ |
| No Crime Reported | $p(1-\xi)(\omega_{11} + \omega_{12}) + p(1-\rho)[\xi(\omega_{11} + \omega_{12}) + \omega_{13} + \omega_{14}]$ + $p(1-\xi)(\omega_{21} + \omega_{22}) + p(1-\rho)[\xi(\omega_{21} + \omega_{22}) + \omega_{23} + \omega_{24}]$ + $p[\omega_{31} + \omega_{32} + \omega_{33} + \omega_{34}]$ |

The E-step of the EM-algorithm consists of obtaining the expected cell counts for the complete data matrix, Table 6, given the observed data and the current estimates of the p, $\rho$, $\xi$, and $\omega$ parameters.

These expectations are particularly simple in the case of discrete data (see, for example, Little and Rubin (1987)) and amount to proportionally allocating the $x_{ijk}$'s of the observed data as shown in Table 8 to the $x_{ijkl}$ cells of Table 6 according to the current parameter estimates. For example,

$$\hat{y}_{1121} = x_{131}\{ [ (1-\hat{\rho}) (1-\xi) \hat{\omega}_{11}] / [(1-\hat{\rho})(1-\xi) \hat{\omega}_{11} + (1-\hat{\rho})(1-\xi) \hat{\omega}_{21} + (1-\hat{\rho}) \hat{\omega}_{31} ] \}.$$

Other expected cell counts may be found similarly and, hence, are not shown here.

The E- and M-steps of the EM-algorithm are repeated until parameter estimates have converged to the desired degree of accuracy, in our case when all estimated probabilities had relative differences of no more than 0.0001 between two iterations. Convergence occurred in 233 iterations for our application.

Our model has twelve $\omega$ parameters with the single constraint that they sum to 1, and the $\rho$, $\rho$, and $\xi$ parameters. Thus there are 14 free parameters. The observed data of Table 8 has 15 cells with the single constraint that they sum to the total sample size. Hence, we have as many parameters as there are cells of data to estimate those parameters. This means that our model will fit the data exactly (see, for example, Bishop, Fienberg, and Holland (1975)).

The parameter estimates for rapes are shown in Table 10. Notice that the estimated probability of rape, adjusting for the dampening effect of the presence of a spouse and a telephone interview, is 0.00010 + 0.00017 + 0.00036 + 0.00038 = 0.00101. This compares with an estimate of 0.00066 based on the raw data (Table 2).

**TABLE 10**: Estimates for Rapes

$\hat{\omega}_{ij}$

| | Spouse | Spouse + Others | Others | Alone |
|---|---|---|---|---|
| Rape | 0.00010 | 0.00017 | 0.00036 | 0.00038 |
| Other crime | 0.07870 | 0.06906 | 0.07713 | 0.07700 |
| No crime | 0.05388 | 0.01270 | 0.24847 | 0.38204 |

$\hat{p} = 0.725928$ $\hat{\rho} = 0.756694$ $\xi = 0.231564$

The parameter estimates for domestic violence are shown in Table 11. Notice that the estimated probability of domestic violence, adjusting for the dampening effect of the presence of a spouse and a telephone interview, is 0.00025 + 0.00031 + 0.00195 + 0.00157 = 0.00408. This compares with an estimate of 0.00300 based on the raw data (Table 2).

**TABLE 11**: Estimates for Domestic Violence

$\hat{\omega}_{ij}$

| | Spouse | Spouse + Others | Others | Alone |
|---|---|---|---|---|
| Domestic Violence | 0.00025 | 0.00031 | 0.00195 | 0.00157 |
| Other crime | 0.07873 | 0.06905 | 0.07526 | 0.07587 |
| No crime | 0.05387 | 0.01270 | 0.24847 | 0.38204 |

$\hat{p} = 0.725928$ $\hat{\rho} = 0.756694$ $\xi = 0.233136$

## Model for "Gag" Factors Allowing for Other Factors

The saturated model presented above fits the data exactly but does not allow us to take into account other factors that might be related to being a victim of rape or domestic violence or to a victim's reporting such a crime in an NCVS interview. In this section, we report on the results of extending the model presented above to allow for one of the following other variables: age (less than 30 versus 30 or more), years of education (12 or fewer versus more than 12), income (less than $15,000 versus $15,000 or more), land use (rural versus urban), married at time of interview, number of persons living in household (5 or fewer versus 6 or more), and whether or not the same household was interviewed at the previous interview. If this additional variable was missing for a woman, her record was not used in the analysis. Note that we only considered a single additional variable at a time because the observed data matrix becomes quite sparse, particularly for the rape or domestic violence cells, when more variables are used.

The model is a simple extension of that described above and, hence, will not be presented in detail. The p, $\rho$, and $\xi$ parameters are as defined previously. There is now a third dimension to the $\omega$ parameters, reflecting the additional classification variable. The model must again be fit using an EM-algorithm or some other iterative procedure. There are now 3 degrees of freedom associated with the model, so that a $X^2$ statistic may be used to assess the fit of the model. Table 12 presents the results of the model fitting.

The results in Table 12 are rather disappointing; the model provides a very poor fit in all cases. This is our conclusion even considering that the $X^2$ statistic naturally increases as the total sample size increases, and our sample size of over 600,000 is quite large.

## **Table 12**: Fits of Models When an Addition Variable is Used

(Note: There are 3 degrees of freedom associated with each model.)

| Variable | $X^2$ for Rape Data | $X^2$ for Domestic Violence Data |
|---|---|---|
| Age | 5457 | 5756 |
| Education | 1715 | 2042 |
| Income | 17417 | 17576 |
| Land use | 449 | 802 |
| Married | 2689 | 2984 |
| Persons in household | 184 | 528 |
| Same household last interview | 41982 | 42145 |

## Future Work

Although the model presented above does not fit the data well, it does allow for possible extensions that may provide better fits to the data. Because there are 3 degrees of freedom associated with the above model, we could refine the model to make it more general. For example, there could be two $\xi$ parameters: one for spouse/significant other alone, and one for spouse/significant other and at least one other person present.

Additional modeling could be done with the $\omega$ parameters. Currently we use a saturated model for the $\omega$'s. Some preliminary analysis with modeling the three-way tables including crime reported, who was present for the interview, and one of the seven variables shown in Table 12 suggests that a log-linear model including all two-way interactions fits some of the tables. The tables for which this unsaturated model fits the data are those in which the third classification variable is land use, education,

or married. Thus, additional modeling using one of these variables and the log-linear model having all two-way interactions might be worthwhile.

Because so many of the interviews are telephone interviews, there is a considerable amount of missing data concerning who is present for the interview. A high rate of missing data can lead to estimates with high variances. This suggests that a very important area for future work would include obtaining variance estimates for the point estimates given in Tables 10 and 11. Variance estimates may be obtained using the observed information matrix (see, for example, Baker (1992)). It may also be useful to fit versions of the models proposed above to the data from the personal interviews only, to see if the parameter estimates are similar to those obtained using all the data.

# References


Baker SG. A Simple Method for Computing the Observed Information Matrix When Using the EM Algorithm with Categorical Data. *Journal of Computational and Graphical Statistics*, **1992;1**(1), 63-76.

Bishop YMM, Fienberg SE, Holland PW. *Discrete Multivariate Analysis: Theory and Practice*, 1975, Cambridge, MA: MIT Press.

Dempster AP, Laird NM, Rubin DB. Maximum Likelihood From Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 1977, Ser. B, **39**, 1-38.

Little RJ, Rubin DB. *Statistical Analysis with Missing Data*, 1987, New York: Wiley.

Bishop YMM, Fienberg SE and Holland PW. *Discrete multivariate analysis: Theory and Practice*. The MIT Press, Cambridge, 1975.

Bureau of Justice Statistics. *National crime survey: criminal victimization in the US*. Washington, DC: US Department of Justice, 1983; NCJ-90541.

Bureau of Justice Statistics. *The crime of rape*. Washington, DC: US Department of Justice, 1985 (Mar).

Bureau of Justice Statistics. *Preventing domestic violence against women*. Washington, DC: US Depart ment of Justice, 1986 (August)

Bureau of Justice Statistics. *New Directions for the National Crime Survey*. Washington, DC: US Department of Justice, 1989 (Mar).