

# Measuring Civilian Defensive Firearm Use: A Methodological Experiment

David McDowall,<sup>1</sup> Colin Loftin,<sup>1</sup> and Stanley Presser<sup>2</sup>

---

Estimates of the incidence of victim gun use from the National Crime Victimization Survey (NCVS) are consistently lower than are those from other studies. To examine the divergence, we conducted a survey that gauged the impact of methodological differences between the NCVS and the other studies. For half of the sample, we asked questions from the NCVS, followed by questions from the other surveys. For the other half of the sample, we presented the questions in the reverse order. We examined two hypotheses: (1) survey methods account for the divergent results, and (2) the questions cover unrelated activities. The results provided some support for the first hypothesis, but respondents also reported many more defenses to the questions from the other surveys than to the NCVS questions. Consistent with the second hypothesis, this suggests that the NCVS and the other surveys measure responses to largely different provocations.

---

**KEY WORDS:** defensive gun use; firearms; survey research; experimental designs; crossover designs.

## 1. INTRODUCTION

The National Crime Victimization Survey (NCVS) provides much lower estimates of the frequency of civilian defensive gun use than do other studies. Recent NCVS data yield an annual estimate of 116,000 cases of armed defense against criminals. The other surveys find as many as 4.7 million cases (Cook and Ludwig, 1998; Kleck and Gertz, 1995). This difference has attracted a large amount of scholarly attention (e.g., Cook *et al.*, 1997; Hemenway, 1997a, b; Kleck and Gertz, 1997; Smith, 1997), and it has both scientific and policy implications.

For policy, an accurate count of gun defenses would be useful in comparing the costs and benefits of widespread firearm ownership (Cook and Ludwig, 1998; Kleck, 1988). Firearm defenses may deter potential crimes

<sup>1</sup>Violence Research Group, School of Criminal Justice, University at Albany, State University of New York, Albany, New York 12222.

<sup>2</sup>Department of Sociology, University of Maryland, College Park, Maryland 20742.

and reduce injuries to victims. These social benefits would then help balance the costs inflicted by armed offenders. Conclusions about the relative sizes of the positive and negative effects depend on estimates of the frequency of victim gun use.

For science, finding the source of the difference between the surveys would be useful for efforts to measure crime. Besides armed defense, the NCVS also diverges from competing surveys in its estimates of other quantities. This is especially the case for the incidence of rape and domestic assault (Bachman and Taylor, 1995; Koss, 1992; Lynch, 1996). Understanding why the surveys disagree about armed defense may provide insights into the other differences as well.

Most attempts to reconcile the NCVS with the other surveys focus on reasons why one data source might be inaccurate. Critics of the NCVS note features of its design that could discourage reporting and deflate its estimates (e.g., Kleck, 1988; Kleck and Gertz, 1995). Critics of the other surveys stress aspects of their designs that could lead to false reports (e.g., Cook and Ludwig, 1998; Cook *et al.*, 1997; Hemenway, 1997a, b).

Unfortunately, no third data source exists for comparison. In cases like these, accuracy explanations cannot usually resolve the issues (Lynch, 1996). Accuracy explanations also neglect other mechanisms that might generate the divergence, especially the possibility that the NCVS and competing surveys do not measure the same phenomena.

Another strategy for studying the differences is to compare survey designs experimentally. This approach allows one to vary some features of the studies while holding other features constant. Experiments cannot show that the results from either of a pair of surveys are correct. Because they alter only selected aspects of the research designs, however, they should help in locating the source of divergent findings.

In this paper we use an experiment to study the disparity between the NCVS and the other surveys. In the following sections, we first discuss the two sets of estimates and consider possible explanations for their differences. We then describe our strategy for comparing the studies, and present our methods, hypotheses, and results. We find evidence that the other surveys cover actions that are largely—although not completely—different from the defenses measured by the NCVS. The final sections of the paper consider the implications of these findings for estimating firearm defense and for comparing the NCVS with other studies.

## **2. ESTIMATES OF THE FREQUENCY OF ARMED SELF-DEFENSE**

Kleck and Gertz (1995) found 14 surveys that included items on firearm use against criminals. While the studies varied in their details, most used

**Table I.** Study Designs, Kleck and Gertz Survey and National Crime Victimization Survey

	Kleck and Gertz Survey	National Crime Victimization Survey
Target population	U.S. population, age 18 and over	U.S. population, age 12 and over
Sample design	Stratified sample of telephone numbers, random digit dialing, one-time survey	Stratified area sample of housing units, both telephone and personal contacts, panel survey
Sample size	4977 completed interviews	43,000 housing units (about 80,000 individuals) in 1997
Response rate	61% of eligible telephone numbers answered by human being	95% of eligible housing units (89% of eligible individuals) in 1997
Sponsorship	Research Network (private organization)	U.S. Census Bureau for U.S. Bureau of Justice Statistics
Estimate of defensive gun use	2,549,862 annual incidents, 1992	64,615 annual incidents using 1987–1990 data from original survey; 116,398 annual incidents using 1993–1994 data from redesigned survey
Possible concerns about accuracy	Susceptibility to false-positive responses; memory telescoping; estimates disagree with external evidence	Respondent doubts about confidentiality; no direct questions about defensive firearm use; exclusion of minor offenses

some form of the following question (Kleck and Gertz, 1995, p. 161):

Within the past five years have you yourself or another member of your household used a gun, even if it was not fired, for self-protection or for the protection of property at home, work, or elsewhere? Please do not include military service, police work, or work as a security guard.

Weighting each survey to the national level, Kleck and Gertz obtained estimates of between 770,000 and 3,610,000 cases of armed defense per year.

Although these surveys were often of low quality, more trustworthy work produces similar results. Kleck and Gertz themselves conducted the most thorough study of the issue. They interviewed a national probability sample of 5000 households, specified explicit reference periods for gun use, and collected descriptions of the incidents. They counted only cases in which civilians said that they had used guns against criminal offenders. Table I summarizes the basic structure of the Kleck and Gertz survey.

Kleck and Gertz's estimate was among the largest of the group: 2,550,000 annual cases of defense. Cook and Ludwig (1996, 1998) replicated Kleck and Gertz's methods and obtained an estimate of 4.7 million cases. A confidence interval around Cook and Ludwig's estimate included the Kleck and Gertz value, however, so the difference in findings could be due to chance. In a more specialized analysis, Ikeda and colleagues (1997) obtained results that also agreed with the other studies.

The NCVS provides a much smaller estimate of the frequency of gun defense. The NCVS uses a probability sample of about 43,000 housing units in the United States (see Table I). Units remain in the sample for 3 years, and interviewers contact residents at 6-month intervals. When respondents report an attempted or completed crime, they answer additional questions about the incident.

If victims say that they saw an offender, NCVS interviewers ask, "Did you do anything with the idea of protecting yourself or your property while the incident was going on?" Interviewers code the responses into one or more categories, including "attacked offender with gun; fired gun" and "threatened offender with gun."

Using NCVS data for 1987 through 1990, McDowall and Wiersema (1994) found an average of 65,000 cases of firearm defense per year. With different weights and data through 1992, Rand (1994) estimated an annual average of 82,000 cases.

Following a major questionnaire revision, NCVS crime estimates increased in 1993 and later years (Rand *et al.*, 1997). Applying the methods of McDowall and Wiersema (1994) to 1993 and 1994 NCVS data, McDowall *et al.* (1998) obtained a yearly estimate of 116,000 incidents of defense.

Obviously, the NCVS results differ from the other surveys. The largest estimate from the NCVS is less than one-sixth the value of the smallest estimate from the others. Perhaps the best comparison is between the 1993–1994 NCVS and the Kleck and Gertz study. Here the results differ by a factor of 22.

### **3. EXPLANATIONS FOR THE DIFFERING ESTIMATES**

#### **3.1. Accuracy Explanations**

Most explanations for the differences between the NCVS and the other surveys center on possible inaccuracies in the data. Kleck and Gertz (1995, 1997) argue that NCVS respondents often conceal their gun use, leading the survey to underestimate the frequency of firearm defense. The federal government sponsors the NCVS, and interviewers record the names of sample members. Respondents might doubt the legality of their defensive actions, and so fear prosecution if they report them.

All of the other surveys had private sponsors. Most used random digit dialing to contact sample members. These factors may have led respondents to believe that their answers were anonymous.

Smith (1997) notes additional features of the NCVS that might reduce its estimates. Most important, the survey does not directly ask about victim gun use. Although NCVS interviewers encourage respondents to describe their actions, they do not prompt for specific forms of defense. The other surveys included questions about self-protection with a gun, increasing the chances that respondents would recall such incidents. The NCVS also excludes trespassing and other minor offenses. Some respondents to the other surveys reported armed defense against these crimes.

While these features of the NCVS may lead it to underestimate gun defense, features of the other surveys may lead them to overestimate it. Cook and Ludwig (1998, 1996; Cook *et al.*, 1997) stress the effects of memory telescoping on the other survey estimates. The NCVS uses its panel design to bound the reference period for reports to the last 6 months. The other surveys asked about incidents in a fixed interval, but some persons may have mistakenly mentioned earlier defenses.

In addition, Hemenway (1997a, b) shows that rare event estimates are highly vulnerable to false-positive responses. He points out that only 66 of the 4977 sample members in the Kleck and Gertz survey reported protective gun use in the previous year. Each incident then carries a heavy weight, and a few false-positive reports could wildly inflate the results. Hemenway argues that some persons in the other surveys probably viewed gun use as a socially desirable action, leading them to make false claims of resistance. The NCVS asks about defense only to the small group of respondents who report attempted or completed victimizations. It is therefore less susceptible to this bias.

Cook and Ludwig (1998) provide empirical support for the hypothesis of flaws in the other surveys. They show that these studies often suggest outcomes that are contrary to external evidence. Their survey's estimate of the frequency of armed defense against rape, for example, exceeded the total number of completed or attempted rapes reported to the NCVS. Their estimate of the number of criminals shot by armed victims was larger than independent estimates of all serious firearm injuries in the nation. Like Hemenway, Cook and Ludwig suggest that the other surveys encouraged false reports.

### 3.2. Content Explanations

Accuracy explanations call attention to differences in methods that might bias the estimates. If one could eliminate these differences, both the NCVS and the other surveys should produce identical results.

Another explanation for the divergent findings also is possible, however: the NCVS and the other surveys yield different estimates because they measure different outcomes. In particular, the other surveys allowed preemptive actions. Here respondents used firearms to protect against *possible* threats to their safety. The potential victims acted before they suffered harm, basing their actions on rapid judgments about what might happen next. The respondents were unharmed, and they could have no more than suspicions about the possible offender's motives.

In contrast, the NCVS asks about defense only after establishing that respondents were victims of attempted or completed crimes. Any defensive gun use clearly occurred during crimes that were in progress. Criminal victimizations are probably less frequent than suspicious encounters. One would accordingly expect fewer reports of firearm defense to the NCVS than to the other surveys.

This class of possibilities is not entirely at odds with explanations that focus on accuracy. It can account for the differences that Cook and Ludwig found between the other surveys and external evidence. If one drops Hemenway's suggestion that many respondents *deliberately* misrepresented their actions, it is also consistent with the possibility of false positive reports. It is different from accuracy arguments, however, because it does not assume that either set of estimates is incorrect.

## 4. METHODS

### 4.1. General Research Design

One empirical strategy for studying the divergent estimates is to compare results from the NCVS and the other surveys after removing their methodological differences. Using this approach, we conducted a survey in which we held constant the major differences that accuracy explanations stress. We then asked respondents questions from both the NCVS and the other studies.

Our basic research design was a standard crossover experiment (e.g., Jones and Kenward, 1989). We randomly selected half the sample to first receive a set of questions like those in the NCVS. After the NCVS questions, we then used items from the other surveys to find if respondents could recall additional instances of self-defense with a gun.

For the remaining half sample, we reversed the procedures. We first asked the questions from the other surveys, then used NCVS items to discover any additional crimes that resulted in armed defense.

To eliminate major methodological differences between the NCVS and the other surveys, we modified the procedures from each. NCVS interviews

first use a screening instrument that asks about broad classes of crimes. If respondents report victimizations during the screening, they provide verbal narratives and complete lengthy incident forms for each offense. NCVS personnel use the incident forms to classify the crimes by type and to discard reports that are outside the survey's scope.

Our study also presented respondents with the NCVS screening form. If sample members reported one or more crimes, we collected narratives for these offenses and asked a few questions from the incident form.

The NCVS only covers incidents in the last 6 months, but we specified a 1-year interval. While the NCVS uses its panel design to verify that any reports occurred within the reference period, our survey was not similarly bounded. Our survey also did not drop crimes that were outside the scope of the NCVS.

Finally, if victims did not mention firearm defense when they described an incident, we asked them if they used a gun. Smith (1997) suggests that the NCVS underestimates the frequency of firearm defense because it does not explicitly ask about gun use. To the degree that Smith's hypothesis is correct, the direct question about firearm resistance should eliminate the difference in responses between the NCVS and the other surveys.

We also made minor changes to the procedures from the other surveys. The most important of the other set of questions asked:

Within the past 12 months, have you yourself used a gun, even if it was not fired, to protect yourself or someone else, or for the protection of property at home, work, or elsewhere?

This is largely identical to the standard question from the other surveys, but the reference period is 1 year rather than 5 years. The question also refers to the respondent alone, rather than to all household members. These changes are necessary so that the NCVS and other questions include the same actors and times.

Unlike most of the other surveys, we did not instruct respondents to omit military, police, or security work. Instead, we collected narratives for each offense and classified the incidents by type ourselves.

These changes should control for the differences in methods that accuracy explanations have emphasized. We asked each set of questions under identical sponsorship. Each set of questions was equally susceptible to memory telescoping. We ensured that estimates from each question set included the same range of crimes. In contrast to the NCVS, we asked victims about gun defense if they did not mention it themselves.

## 4.2. Survey Design

Firearm defense is an infrequent event, and one would expect few reports even in a large sample. Obtaining population estimates of defense

was not a goal of our study, so we traded a representative sample for a higher reporting rate.

Because gun ownership is a strong correlate of firearm resistance (e.g., Kleck and Gertz, 1996, p. 187), we selected a national sample from commercial lists of likely gun owners. Of the eventual respondents, 83% did report the presence of a gun in their home. Unless our respondents understood the questions differently than other persons, this sample will allow a reasonable comparison of measurement methods.

The Survey Research Center at the University of Maryland conducted the interviews by telephone. Interviewers called each selected number up to 20 times, and a refusal conversion specialist recontacted sample members who initially declined to respond.

The interviewers told respondents that they were calling from the University of Maryland, as part of a study of crime in the United States. They promised the respondents that their answers would remain confidential. Kleck and Gertz (1995) argue that concerns about confidentiality account for the difference between the NCVS and the other survey estimates. All our respondents experienced the same confidentiality conditions. Both sets of questions should therefore produce identical results if Kleck and Gertz are correct.

The original sample consisted of 3714 households. Twelve percent of these ultimately refused to participate, and the interviewers could not contact an additional 6%. In 1% of the households the respondents were ill, did not speak English, or otherwise could not complete the interview. This left 3006 households, an 81% response rate.

The interviewers selected a single respondent from within each household. In a random 75% of the cases, the interviewers asked for the male head of household. In the remaining 25% they asked for the female head. Of the completed interviews, 1522 respondents received the NCVS questions first, and 1484 received the other questions first.

## 5. HYPOTHESES

The results allow us to examine the two general explanations that we discussed earlier. First, the NCVS and the other studies may measure identical phenomena and disagree only because of their procedures. We label this the “methods hypothesis,” because it covers most explanations involving study design and administration.

The procedures are constant for both sets of questions in our study. If the methods hypothesis is true, both sets should produce identical proportions of defensive gun reports. In addition, respondents should mention all incidents of gun use during the first question set.



Our second hypothesis is that the questions measure at least partially different phenomena. Here respondents may be more likely to report some types of incidents to one set of questions than to the other. The NCVS may include only cases where respondents used guns against persons who were trying to victimize them. The other surveys may more heavily cover cases in which respondents used guns to protect against potential crimes. We label this the “domain hypothesis,” because it implies that each set of questions measures a different universe of events.

If the domain hypothesis is correct, some respondents who receive the NCVS questions first will mention additional defensive incidents to the other set. Similarly, some respondents who receive the other questions first will mention additional incidents to the NCVS items. One set of questions may also produce many more defense reports than does the other.

The methods and domain hypotheses do not exhaust all possible relationships between the surveys. In particular, neither fully addresses Hemenway’s (1997a, b) suggestion that some respondents to the other surveys fabricated their reports.

If the fabrication hypothesis is correct, one would expect respondents to describe both fabricated and factual defenses when the other survey questions appear first. This would then leave no incidents to report to the NCVS questions. On the other hand, respondents should report all factual incidents to the NCVS items when they appear first, perhaps followed by additional fabricated reports to the other survey questions. This would create an interaction between question order and question content.

The research design does not allow us fully to explore this possibility, because it is confounded with a differential carryover effect (Jones and Kenward, 1989). One set of questions may more heavily sensitize respondents to the issues than does the other. When the sensitizing set appears first, reports of gun use will be more frequent than when the other set appears first. This will produce the same interaction between content and order that the fabrication hypothesis predicts.

Due to the confounding, we cannot separate a substantively meaningful interaction from differential carryover. We can, however, test the composite hypothesis that neither effect exists.

## **6. RESULTS**

### **6.1. Classification of Cases**

As a shorthand in reporting the results, we call the NCVS items the CRIME questions, because of their stress on victimization. We call the other items the GUN questions, because of their emphasis on firearm use.

In coding the cases, we followed past researchers (e.g., Kleck and Gertz, 1995) by using only a single incident for each respondent. Multiple reports usually involved similar events, and respondents often provided details about only the first defense.

When respondents reported one or more incidents of armed self-defense against a visible offender, we selected the most recent case that did not involve military, police, or security work. If a respondent did not report one of these cases, we selected the most recent of any other type of defensive firearm use.

Table II presents frequency distributions for the two sets of questions by type. The CRIME incidents are simple to classify, because they all involved defense against criminal offenders. The CRIME questions (and the NCVS) include police officers who defended themselves in the line of duty. Since our interest is in civilian gun use, we separated these cases from the others.

The GUN defenses are more diverse. About 29% of the incidents involved law enforcement, security, and military work or hunting and protection against animals. Several respondents mentioned target shooting, saying that they were practicing self-defense. Other persons counted occasions when they carried a firearm, reasoning that they deterred potential criminals. Finally, some respondents described incidents in which they armed themselves against a possible crime but did not confront an offender. In the most common situation, the respondent heard a noise and retrieved a gun, but found no evidence of an offense.

**Table II.** Types of Incidents of Firearm Defense, Gun Questions and CRIME Questions

Type of incident	Number of respondents	Percentage of respondents	Percentage of incidents
GUN questions			
No incident	2851	94.8	—
Civilian against offender, clear	<b>48</b>	1.6	31.0
Civilian against offender, ambiguous	<b>24</b>	0.8	15.5
Law enforcement and security work	30	1.0	19.3
Civilian against possible offender, no contact	20	0.7	12.9
Against animals	13	0.4	8.4
Carries gun for protection only	10	0.3	6.4
Target shooting	8	0.3	5.2
Military duties	2	0.1	1.3
CRIME questions			
No incident	2977	99.0	—
Civilian against offender, clear	<b>24</b>	0.8	82.8
Law enforcement and security work	5	0.2	17.2

Reports of defensive gun use that did not involve encounters with criminals also were common in the other surveys. Kleck and Gertz (1995) and Cook and Ludwig (1997) dropped these incidents from their analyses, and we followed their example. It is notable, however, that more than one-half of the cases in which respondents believed they protected themselves with guns did not qualify under a strict definition of firearm defense.

Removing the incidents that did not involve civilian defense against criminals leaves two categories of GUN cases. In the first category, the respondents had guns when they encountered persons whom they suspected of malicious intent. These make up about one-third of the incidents, and they are the clearest cases of victim gun use.

The other category consists of more ambiguous cases. Here the respondents would not fully describe the incidents, or they gave too few details to allow a firm classification. From their answers to other questions (for example, their assessment of the danger they faced), however, these cases also appear to be defenses against crime.

## 6.2. Hypothesis Tests

As a step toward examining the methods and domain hypotheses, we computed frequency distributions for the GUN and CRIME questions by their order of presentation. The results appear in Table III. Since our major concern is with civilian defense against criminals, the table lists only these cases. The GUN category combines the clear and ambiguous incidents. Overall, respondents reported 72 GUN incidents and 24 CRIME incidents. Of the 96 total incidents, 18 were from persons who reported defenses to both sets of questions.

To test the hypotheses, we estimated a logistic regression model for the Table III frequencies. This equation included two binary independent variables, one for question content (GUN or CRIME question) and one for question order (first or second question set).

According to the methods hypothesis, each set of questions measures identical activities. Both question sets should thus generate the same proportions of defensive incidents, and respondents should report all defenses

**Table III.** Incidents of Firearm Defense, against Offenders, GUN Questions and CRIME Questions by Question Order

	GUN questions first	CRIME questions first	Total
No incident	1434 (96.6%)	1485 (97.6%)	2919
GUN incident only	40 (2.7%)	23 (1.5%)	63
CRIME incident only	6 (0.4%)	9 (0.6%)	15
Both GUN and CRIME incident	4 (0.3%)	5 (0.3%)	9
Total	1484 (100.0%)	1522 (100.0%)	

**Table IV.** Logistic Regression of Firearm Defense on Question Content and Order<sup>a</sup>

Variable	Coefficient	SE	z	Exp (Coeff.)
Interaction model				
Question (Guns = 1)	1.1913	0.3090	3.85	3.29
Order (second = 1)	-0.3137	0.4157	0.75	0.73
Interaction (Question * Order)	-0.1751	0.5242	-0.33	0.84
Constant	-4.6795	0.2685	-17.43	
Main effects model				
Question (Guns = 1)	1.1182	0.2184	5.21	3.06
Order (Second = 1)	-0.4377	0.1899	-2.30	0.64
Constant	-4.6282	0.2148	-21.54	

<sup>a</sup>These are GEE estimates, using an exchangeable correlation structure. The correlation within respondents is  $\rho = 0.22$  for the interaction model and  $\rho = 0.21$  for the main effects model.

to the first set. The coefficient for the question order variable will be large, while the coefficient for the content variable will be zero.

The domain hypothesis states that the questions measure different activities. If this is true, the proportions of defense reports will differ between the two question sets, but they will be identical for each question order. Here the content variable will have a large coefficient, and the coefficient for the order variable will be zero.

Because the same respondents answered both sets of questions, each pair of answers forms a cluster. Conventional maximum-likelihood methods will then produce statistically inconsistent standard error estimates (see, e.g., Cox and Snell, 1989, pp. 106–115). To avoid this problem, we applied the generalized estimating equation (GEE) approach proposed by Liang and Zeger (1986; Diggle *et al.*, 1994). GEE uses quasi-likelihood to model the marginal expectation of the response variable separately from the within-cluster correlation. It provides consistent estimates of the coefficients and their standard errors, with only modest assumptions about the form of the clustering.

Table IV presents the results. To test for a differential carryover effect, we first estimated a model that included an interaction between question content and question order. As we noted earlier, a carryover effect would occur if one question set more heavily sensitized respondents to the issue of armed defense. This would pose problems for the analysis, because differential carryover complicates interpretations of the other coefficients (Jones and Kenward, 1989, pp. 39–51).<sup>3</sup> The estimates for this equation appear in

<sup>3</sup>An equal carryover effect also is possible. This would occur if each question set identically sensitized respondents to the other. Equal carryover is not harmful to inferences about the hypotheses, however, since neither hypothesis depends on the absence of memory prompts for its validity.

the top panel of the table, and they show that the interaction term is statistically insignificant ( $P = 0.74$ ).

Following standard practice (e.g., Diggle *et al.*, 1994, pp. 153–155), we then estimated the model without the interaction. These results, in the bottom panel of Table IV, provide some support for both the domain and methods hypotheses.

Consistent with the domain hypothesis, respondents were three times more likely to report a GUN incident than a CRIME incident. Even after removing differences in methods, GUN reports thus still greatly exceeded CRIME reports. In addition, controlling for question content, respondents remained about two-thirds as likely to report a defense to the second set of questions as to the first. This value is not far from the equal odds that the domain hypothesis predicts.

Yet the results also partially support the methods hypothesis. Although the difference in reports between the first and the second question sets is substantively small, it is statistically significant. The threefold ratio of GUN reports to CRIME reports also is less than in other comparisons. Variations in methods appear to account for part of the conflict between previous results.

Overall, a conservative conclusion is that differences in methods partially explain the divergence in estimates, but the two sets of questions also measure substantially different actions. While the GUN and CRIME items overlap each other in coverage, each set contains much variation that is unique to itself.

## 7. DISCUSSION

### 7.1. Reconciling the Differences

Our analysis suggests that the NCVS and the other surveys produce different results partly because of differences in their methods and partly because of differences in their content areas. The data do not allow us to find the specific features of the methods and content that contributed most heavily to the divergence. Although our research design is useful in examining broad classes of hypotheses, it restricts us in evaluating particular members of a class.

This limitation is especially noticeable for the methodological differences. Some changes that we made to each survey may have increased measurement accuracy, while other changes may have decreased it.

Directly asking about gun use during the NCVS questions might have prompted respondents to recall incidents they otherwise would not have mentioned. This change would then increase the accuracy of the NCVS responses. On the other hand, eliminating NCVS procedures for controlling

memory telescoping might have resulted in more reports outside the survey's time bounds. This change would reduce accuracy.

Because the changes likely work in both directions, we cannot isolate design features that account for the differences. Still, it is notable that most of the changes we made dropped desirable features of the NCVS. The two sets of surveys may more strongly agree after the changes because these alterations removed NCVS safeguards against false reports.

We also cannot draw final conclusions about the nature of the content differences between the question sets. We are, however, on more solid ground in guessing its source. As we pointed out earlier, the other surveys (and the GUN questions) allowed respondents to report cases in which they used firearms to prevent crimes that had not yet occurred. These surveys relied heavily on respondent judgments about the motives of possible offenders, and motives may be murky if the respondents acted quickly.

In contrast, NCVS interviewers (and the CRIME questions) ask about defense only if respondents have reported attempted or completed victimizations. Here any reports of self-protection are clearly against crimes. The other studies may therefore include a large and heterogeneous class of defenses against suspicious persons, while the NCVS includes a small and tightly focused class of defenses against criminals.

Although our survey changed features of both the NCVS and the other studies, we preserved the basic content of each question set. If the questions yield different results, differences in content are the most likely explanation. This possibility draws support from the respondents' own justifications of their answers. When sample members reported defenses during the second set of questions, the interviewers asked them why they did not mention the cases earlier. We obtained 6 of these reconciliations from persons who received the GUN questions first, and 26 from persons who received the CRIME questions first.

Table V contains the results. Some reconciliations were idiosyncratic and did not fit a pattern. One person said he thought the CRIME questions

**Table V.** Reconciliations, GUN Questions and CRIME Questions

	Number of respondents
Reasons for additional GUN reports	
Don't know/Didn't understand question that way	9
Incident did not involve serious harm	8
Other responses	9
Reasons for additional CRIME reports	
Don't know/Didn't understand question that way	3
Other responses	3

asked only about his home, for example, while another doubted that he could prove he was victimized.<sup>4</sup>

The remaining answers fit easily into a few categories. Respondents most often said either that they did not recall the incidents during the first question set or that they could not explain why they omitted them. These persons implicitly agreed that the questions were similar on the surface, but each set prompted them to remember a different type of event. In addition, several respondents explained their extra GUN reports by saying that they did not involve serious criminal offenses.

The content differences between the two sets of questions highlight a basic ambiguity in the incidents that the other surveys measure. By including defenses against crimes that had not yet occurred, the other surveys rely on respondent perceptions of what would have happened if they had not used a gun. Even if all respondents were completely honest, the accuracy of their perceptions is likely to be variable.

In some cases, the sample members probably were correct in assuming that they were about to become crime victims. If a respondent produced a gun when a stranger approached in a deserted parking lot and asked for money, he or she may well have prevented a robbery. Because episodes like these turned out well for the intended victims, they might not even remember them when answering victimization questions like those in the NCVS.

In other cases, however, the respondents probably misjudged the possible offender's motives. Because the respondents took action quickly, what would have occurred without the gun use is uncertain. The stranger in the parking lot may have been a panhandler, for example, and not a robber. Still, in recounting the incidents, the respondents are likely to emphasize details of the encounters that justify their actions.

These considerations also suggest that some of the defenses respondents reported to the other surveys were themselves criminal acts. Kleck and Gertz (1995) note that many gun defenses involve illegal carrying and possession of firearms by the defender. More generally, the gun use may follow mistaken perceptions of innocuous actions by the supposed criminal. These cases of armed resistance would then legally amount to aggravated assaults.<sup>5</sup>

<sup>4</sup>Two respondents mentioned concerns about confidentiality. Both were explaining why they did not report incidents to the GUN questions.

<sup>5</sup>The other surveys in fact provide evidence that many respondents are incorrect in assessing the risks that they face. Kleck and Gertz (1995), for example, obtained an estimate of 392,000 cases in which respondents reported that someone almost certainly would have been killed without the firearm defense. Yet as Hemenway (1997) notes, fewer than 30,000 homicides occur each year in the United States. It is then extremely unlikely that all of the respondents in the Kleck and Gertz study were correct about the gravity of their situation. Hemenway (1997a, b) and Cook and Ludwig (1998; Cook *et al.*, 1997) also show that the Kleck and Gertz study suggests armed defense against implausibly large fractions of other violent crimes.

Protecting oneself against possible harm is logically distinct from self-defense during a crime. On close inspection, the GUN and CRIME questions are not as alike as they at first appear. They can lead respondents to recall different classes of activities.

## 7.2. Implications of the Differences

A major implication of our results is that the other surveys provide misleading conclusions about the *relative* frequency of armed defense against crime. In an abstract sense, the annual number of defenses is large, whether it is 116,000 cases (1993–1994 NCVS) or 2.5 million cases (Kleck and Gertz). The figure acquires perspective only in comparison to another standard.

McDowall and Wiersema (1994) compared NCVS estimates of firearm defense to NCVS estimates of firearm crime. They found that gun offenses exceeded protective incidents by a factor of 10. This comparison assumes that both estimates are from the same universe of events, so the gun defenses occurred during encounters that the NCVS would record as crimes. This is a reasonable assumption, since the NCVS asks about protection only if respondents report victimizations.

Kleck and Gertz (1995) also compared their estimate of armed defense to NCVS estimates of crime. They concluded that protective incidents exceeded criminal gun use by a factor of three. This comparison again assumes that both estimates are from the same universe of events, so that the defenses their respondents reported occurred during the equivalent of NCVS victimizations. That is, concerns about the legality of the firearm defense aside, Kleck and Gertz assume that respondents would have reported incidents like these to the NCVS.

Our results imply that Kleck and Gertz's assumption is false. Many respondents reported incidents to the second set of questions after they reported no incidents to the first. Some respondents reported incidents to both sets. The NCVS and the other surveys cover partially different classes of events, and the "defenses" in the other surveys were not necessarily against criminals.

A better comparison for the other surveys might then use another standard than NCVS crimes. For example, one might compare the frequency of armed defense to the frequency of encountering frightening or suspicious persons. If these encounters occur more often than do criminal acts, the relative number of firearm defenses would decrease.

Selecting a useful comparison base for the other surveys requires a clear understanding of the provocations for the defensive gun reports that they contain. From our findings, it appears that these range well beyond NCVS victimizations.



## 8. CONCLUSIONS

Our results suggest that while the NCVS and other surveys of firearm defense differ partly because of their methods, much of the difference is due to their questions. Even under largely identical conditions, we obtained many more reports of defense to the questions from the other surveys than to the NCVS questions.

Our study did not directly address the issue of how the methods and questions led respondents in different directions. Research focusing on more limited changes would be helpful in isolating features of the surveys that produced the divergent results. Using cognitive interviews to probe respondent understandings of the questions might shed more light on the incidents. Varying the content and form of the questions could allow closer study of how wording influences reporting.

The research design also did not allow us fully to evaluate Hemenway's (1997a, b) hypothesis that some respondents to the other surveys intentionally invented false reports of gun defense. Consistent with Hemenway's hypothesis, our results support the idea that many reports to the other surveys were false positives (that is, they were not defenses against criminal acts). Yet, as we noted, the possibility that respondents *fabricated* incidents requires an interaction between question content and question order.

Although we did not find such an interaction, the data prevented us from separating the impact of deliberate falsification from a differential carryover effect. We could not examine the falsification hypothesis by itself, and the interaction test has low statistical power (Jones and Kenward, 1989). Additional study of the hypothesis would be useful in clarifying the nature of the defensive gun use reports in the other surveys.

Subsequent research in this area might also use a probability sample from the general population, instead of a sample of likely gun owners. The gun owner sample increased the chances of finding incidents of firearm defense, and it was helpful for the methodological comparisons that were the focus of our analysis. Yet the sample did not allow us to compare directly the magnitude of our estimates to those from other surveys.

This, in turn, limits the conclusions that we can draw from our results. We can conclude that the difference between the NCVS and the other surveys is due partly to their methods and partly to their questions. We cannot, however, estimate the exact proportion of the difference that results from each of these factors. A useful approach here might be to field a supplement to the NCVS that included questions about defensive gun use in noncrime situations as well as during criminal victimizations.

Despite these issues, our findings imply that one should avoid comparing other survey estimates of defense to NCVS estimates of crime. As we

pointed out earlier, a possible use of defensive firearm estimates is in evaluating the costs and benefits of gun ownership. Most discussions of the other surveys assume that each armed defense produces social benefits that negate the costs of a serious crime (e.g., Heston, 1999). But our results suggest that the benefits of these defenses are much less obvious. Comparing dissimilar events cannot provide sound guidance for policy. The other surveys are meaningless without a better understanding of the incidents that they measure.

Given the uncertainty about measures of firearm defense, a reasonable strategy might be to avoid comparing defenses and crimes. Alternatively, one might compare NCVS defense estimates with NCVS crime estimates. In this case, at least, the comparison would be between like quantities.

Beyond the issue of defensive gun use, our analysis has illustrated an approach to studying conflicting survey estimates of criminal events. We suspect that other differences between crime survey findings also are due in part to the incidents that they include. Methodological studies should in the end be more useful in improving measurement than are arguments about the accuracy of a given set of results.

## ACKNOWLEDGMENTS

We would like to thank Michael Maltz and two anonymous referees for their helpful comments and suggestions.

## REFERENCES

- Bachman, R., and Taylor, B. (1994). The measurement of family violence and rape by the redesigned National Crime Victimization Survey. *Just. Q.* 11: 499–512.
- Cook, P. J., and Ludwig, J. (1996). *Guns In America: Results of a Comprehensive National Survey on Firearms Ownership and Use*, Police Foundation, Washington, DC.
- Cook, P. J., and Ludwig, J. (1998). Defensive gun uses: New evidence from a national survey. *J. Quant. Criminol.* 14: 111–131.
- Cook, P. J., Ludwig, J., and Hemenway, D. (1997). The gun debate's new mythical number: How many defensive uses per year? *J. Policy Anal. Manage.* 16: 463–469.
- Cox, D. R., and Snell, E. J. (1989). *Analysis of Binary Data*, 2nd ed., Chapman and Hall, New York.
- Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data*, Oxford University Press, New York.
- Hemenway, D. (1997a). The myth of millions of annual self-defense gun uses: A case study of survey overestimates of rare events. *Chance* 10(3): 6–10.
- Hemenway, D. (1997b). Survey research and self-defense gun use: An explanation of extreme overestimates. *J. Crim. Law Criminol.* 87: 1430–1445.
- Heston, C. (1999). The president's column. *Am. Rifle.* 147(6): 12.
- Ikeda, R. M., Dahlberg, L. L., Sacks, J. J., Mercey, J. A., and Powell, K. E. (1997). Estimating intruder-related firearm retrievals in U.S. households, 1994. *Violence Vict.* 12: 363–372.

- Jones, B., and Kenward, M. G. (1989). *Design and Analysis of Cross-Over Trials*, Chapman and Hall, New York.
- Kleck, G. (1988). Crime control through the private use of armed force. *Soc. Problems* 35: 1–21.
- Kleck, G., and Gertz, M. (1995). Armed resistance to crime: The prevalence and nature of self-defense with a gun. *J. Crim. Law Criminol.* 86: 150–186.
- Kleck, G., and Gertz, M. (1997). The illegitimacy of one-sided speculation: Getting the defensive gun use estimate down. *J. Crim. Law Criminol.* 87: 1446–1461.
- Koss, M. P. (1992). The underdetection of rape: Methodological choices influence incidence estimates. *J. Soc. Issues* 48: 61–75.
- Liang, K. Y., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73: 13–22.
- Lynch, J. P. (1996). Clarifying divergent estimates of rape from two national surveys. *Public Opin. Q.* 60: 410–430.
- McDowall, D., and Wiersema, B. (1994). The incidence of defensive firearm use by U.S. crime victims, 1987 through 1990. *Am. J. Public Health* 84: 1982–1984.
- McDowall, D., Loftin, C., and Wiersema, B. (1998). Estimates of the frequency of firearm self-defense from the National Crime Victimization Survey. Violence Research Group Discussion Paper 20, School of Criminal Justice, State University of New York—Albany.
- Rand, M. R. (1994). *Guns and Crime* (NCJ-147003), U.S. Bureau of Justice Statistics, Washington, DC.
- Rand, M. R., Lynch, J. P., and Cantor, D. (1997). *Criminal Victimization, 1973–95* (NCJ-163069), U.S. Bureau of Justice Statistics, Washington, DC.
- Smith, T. W. (1997). A call for a truce in the DGU war. *J. Crim. Law Criminol.* 87: 1462–1469.