For rare events, overestimation is likely even if the misclassification is not random.

The Myth of Millions of Annual Self-Defense Gun Uses: A Case Study of Survey Overestimates of Rare Events

David Hemenway

A criminologist has been using national self-report surveys to estimate the incidence of self-defense gun use in the United States (Kleck 1991). His most recent estimate is that civilians use guns in self-defense against offenders more than 2.5 million times each year (Kleck and Gertz 1995). This figure has been widely cited, not only by the National Rifle Association but in the media and in Congress. A criminologist in Canada is touting high self-defense estimates for that country, derived from similar self-report surveys (Mauser 1995).

All attempts at external validation of the 2.5 million figure show that it is an enormous overestimate (Hemenway, in press). For example, in 34% of the times a gun was used for self-defense, the offender was allegedly committing a burglary. In other words, guns were reportedly used by defenders for self-defense in approximately 845,000 burglaries. From sophisticated victimization surveys, however, we know that there were fewer than 6 million burglaries in the year of the survey and in only 22% of those cases was someone certainly at home (1.3 million burglaries). Since only 42% of U.S. households own firearms, and since the victims in two thirds of the occupied dwellings were asleep, the 2.5 million figure requires us to believe that burglary victims use their guns in self-defense more than 100% of the time.

A more reasonable estimate of self-defense gun use during burglary comes from an analysis of Atlanta police department reports. Examining home invasion crimes during a four-month period, researchers identified 198 cases of unwanted entry into a single-family dwelling while someone was at home (Kellermann, Westphal, Fisher, and Harvard 1995). In 6 of these cases, an offender obtained the victim's gun. In only 3 cases (1.5%) was a victim able to use a firearm in self-defense.

The estimate of 2.5 million self-defense gun uses per year leads to many other absurd conclusions. For example, the number of respondents who claim to have used a gun against rape and robbery attempts suggests that victims of these attempted crimes are more likely to use a gun against the offender than the attackers are to use a gun against the victim—even though the criminal chooses the time and place for the attack, most citizens do not own guns, and very few carry guns.

How could the survey estimate be so far off? The 2.5 million figure comes from a national random-digit-dial telephone survey of 5,000 dwelling units (Kleck and Gertz 1995), in which slightly over 1% of the individuals surveyed reported that they themselves had used a gun in self-defense during the past year. Using that percentage to extrapolate to the entire population of 200 million adults gives approximately 2.5 million uses.

Other telephone surveys (of 600 to 2,000 respondents) also yield high estimates, often of millions of annual self-defense gun uses (Kleck 1991). All of the surveys have very serious methodological deficiencies, but the most important problem, never sufficiently considered, is that the researchers are attempting to estimate a rare event. That fact alone leads to the likelihood of extreme overestimation.

Misclassification

Surveys have many potential sources of bias (e.g., nonresponse bias). One source of bias is the possibility of misclassification. Incorrect classification

comes from a wide variety of causes including miscoding, misunderstanding, misremembering, misinterpretation of events, mischief, or downright mendacity.

Virtually every survey has problems with accuracy. The problem can be particularly acute for self-report surveys, in which virtually all questions have some incorrect responses. For example, respondents substantially overreport seat-belt use and often inaccurately report about whether they voted and for whom they voted in past elections. Some people do not report truthfully about such mundane details as their age, height, or weight. A literature review of the validity of self-report responses characterizes 83% to 98% accuracy rates to questions about possession of an automobile, a home, a driver's license, or a library card as "quite high" (Wentland and Smith 1993, p.19).

Table 1—Results of a Screening Test

	Truth		
Screen	Actual positive	Actual negative	Total
Positive	a (true positive)	b (false positive)	$a + b$
Negative	c (false negative)	d (true negative)	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

False Positives

In a survey seeking to determine the incidence of a rare event, the potential for misclassification is asymmetric; random misclassification will lead to an overestimation bias. For even minor amounts of misclassification, the overestimation can be substantial.

Courtesy of Richard Wentworth

Personal Presentation Bias

For example, assume that the actual incidence in the population is .2%. In a random survey, on average, for every 1,000 respondents, 998 will have a chance to be misclassified as a false positive. On average, however, only two respondents could be misclassified as a false negative.

In addition, because the survey is trying to estimate the incidence of a rare event, a small percentage bias can lead to an extreme overestimate. Say that survey findings are a 1% overestimate of the true incidence. If the true incidence of the event were 40%, estimating it at 41% might not be a problem. But if the true incidence were .2%, measuring it as 1.2% would be six times higher than the true rate; and if the true incidence were .1%, measuring it at 1.1% would be a ten fold overestimate.

The overestimation problem can be explained in the context of screening for diseases. In Table 1 the two rows are the response (the screen) and the two columns are the truth (the actual fact). Each respondent can be placed in one of the four categories. In Table 1, a = the number of people who test positive who actually have the disease (true positive), b = the number of people who test positive who do not have the disease (false positive), c = the number of people who test negative who actually have the disease (false negative), and d = the number of people who test negative who do not have the disease (true negative).

Epidemiologists have names for various ratios in this table. Sensitivity is defined as $a/(a + c)$ or the percentage of all true positives accurately detected by the screening test. Specificity is defined as $d/(b + d)$ or the percentage of all true negatives accurately detected by the screening test. Positive predictive value is defined as $a/(a + b)$ or the percentage of the screened positives who are truly positive.

The true incidence for this sample population is $(a + c)/(a + b + c + d)$. The figure derived by a survey, however, would be $(a + b)/(a + b + c + d)$. To determine how close the estimate is to reality, we need to know the divergence between $(a + b)$ and $(a + c)$.

Although the medical profession is generally interested in reducing the size of both b (false positives) and c (false negatives), our interest is whether there is a large absolute difference between the number of false positives and the number of false negatives. A basic epidemiology textbook helps explain why we might expect the size of b and c to differ markedly:

"The predictive value of a screening test is determined not only by factors that determine validity of the test itself (i.e., sensitivity and specificity), but also by the characteristics of the population, in particular the prevalence of the preclinical disease.... For rare diseases, the major determinant of the predictive value positive is the prevalence of the preclinical disease in the screened population. No matter how specific the test, if the population is at low risk for having the disease, results that are positive will mostly be false positives," [assuming less than a 100% specificity rate] (Hennekens and Buring 1987, p. 337).

With a huge number of actual negatives, virtually any screen will pick up a sizable absolute number of false positives. With few actual positives, it is impossible for a screen to pick up many false negatives. It follows that, for events with low incidence, random misclassification will result in the estimated incidence being far greater than the true incidence.

Now let the screen be the response to the Yes/No question, "Have you used a gun in self-defense in the past year." We assume an adult population of 200 million and a sample size of 2,000. In Table 2a we also assume that the truth

is that 200,000 individuals use a gun in self-defense each year. This figure is 2–3 times higher than the estimate from the large-panel National Crime Victimization Surveys, which reduce the false positive problem in two ways. First, they reduce "telescoping" by asking only about events that took place in the six months since the last interview, and second, they only ask about self-defensive actions of those respondents who first answer that someone tried to perpetrate a crime against them.

If there is a random misclassification error of only 1% (a 99% sensitivity rate and a 99% specificity rate), then we expect that 22 people will report a self-defense gun use, which will extrapolate to 2.2 million people, or 11 times the true rate.

The estimate will be extremely sensitive to very small changes in the specificity rate. In Table 2b, we assume that we do not initially know the true rate but that the survey yields 22 respondents who say "Yes" to the self-defense gun question. If 1.05% of respondents are randomly misclassified, the truth would be that just .05% of individuals actually used a gun in self-defense in the previous year, or about 100,000 uses per year for the entire adult population; the 2.2 million figure would be a 21-fold overestimate!

There are many reasons to expect some misclassification in surveys. For example, some people have a different perception of reality than most of us. The best estimates are that in any recent six-month period, one to two million Americans suffered from schizophrenia, one to two million suffered from antisocial personality disorder and another two million suffered from Alzheimer's disease and other cases of dementia. And some small percentage of telephone respondents are undoubtedly drunk at the time of the interview.

There are particular reasons to expect less than a 100% specificity rate on the self-defense gun use question. For example, there is the possibility of some positive social desirability responses, sometimes referred to as personal presentation bias. An individual who purchases a gun for self-defense and then uses it successfully to ward off a criminal is displaying the wisdom of his precautions and his capability in protecting himself, his loved ones, and his property. His action is to be commended and admired.

Courtesy of Richard Wentworth

Encounters with Aliens

Respondents who misreport are not necessarily deliberately lying; they may be shading the truth, or they may simply perceive and present themselves in a slightly more favorable light than would a purely objective observer. For example, their reported self-defense gun use might well be perceived by the other parties involved as an unprovoked escalation of a simple altercation.

There are also some reasons to expect less than a 100% sensitivity rate. Some respondents may have forgotten, or they may have obtained or used the gun illegally and not wanted to report this fact to an interviewer. The key point, however, is that whenever the true incidence is low, the ratio of the reported incidence to the true incidence depends almost entirely on a test's specificity and very little on its sensitivity.

In Table 2c we assume that 22 individuals report a self-defense gun use, that 1.11% of the true positives are misclassified (98.89% specificity rate), and 100% of the true positives are misclassified (0% sensitivity rate). The actual number of self-defense gun uses per year would still be only 200,000. The reported figure would be a 10-fold overestimate.

Table 2a—Results of a Screening Test

Screen: Response to self-defense gun question	Truth Positive: A self-defense gun use	Truth Actual negative self-defense gun use	Total
Positive	2	20	22
Negative	0	1,978	1,978
Total	2	1,998	2,000

Assumptions:	
Adult population:	200 million
Truth:	200,000 used a gun in self-defense
Sample size:	2,000
Expected "Yes"	2
Sensitivity rate	99%
Specificity rate	99%
(A random 1% of respondents are misclassified.)	
Then:	
Survey incidence:	22/2,000 = 1.1
True incidence:	2/2,000 = .1
Survey overestimation	11 times too high
Estimated total defensive gun use:	2.2 million
True total defensive gun use:.	.2 million (or 200,000)

Table 2b—Results of a Screening Test

Screen: Response to self-defense gun question	Truth Positive: A self-defense gun use	Truth Negative: No self-defense gun use	Total
Positive	1	21	22
Negative	0	1,978	1,978
Total	1	1,999	2,000

Assumptions:

Adult population:	200 million
Sample size:	2,000
Sample incidence:	22
Test sensitivity:	98.95%
Test specificity:	98.95%
(A random 1.2% of respondents are misclassified.)	
Survey incidence:	22/2,000 = 1.1
Actual incidence:	1/2,000 = .05
Survey overestimation	22 times too high
Estimated total defensive gun use:	2.2 million
True total defensive gun use:	.1 million (or 100,000)

Overestimates of Rare Events

Using surveys to estimate rare events typically leads to overestimates. For example, the National Rifle Association reports 3 million dues-paying members, or about 1.5% of American adults. In national random telephone surveys, however, 4–10% of respondents claim that they themselves are dues-paying NRA members. Similarly, although *Sports Illustrated* reports that fewer than 3% of American households purchase the magazine, in national surveys 15% of respondents claim that they are current subscribers.

Consider the most extreme case, when the true incidence is 0%. Then a survey can overestimate but not underestimate the true incidence. In May 1994, ABC News and *The Washington Post* conducted a national random-digit-dial telephone survey of over 1,500 adults. One question asked: "Have you yourself ever seen anything that you believe was a spacecraft from another planet?" Ten percent of respondents answered in the affirmative. These 150 individuals were then asked, "Have you personally ever been in contact with aliens from another planet or not?" and 6% answered "Yes."

Extrapolating to the U.S. population as a whole, we might conclude that 20 million Americans have seen alien spacecraft, and 1.2 million have been in actual contact with beings from other planets.

Continuous Variables and the Tail of the Distribution

Like discrete variables, continuous variables can have serious false-positive problems if they are used to measure the incidence of rare phenomena. For example, survey data is sometimes used to estimate the extent of chronic malnutrition within a population. The incidence of stunted growth—defined as a certain distance below the age- and sex-specific norm—is often used as an objective proxy for the extent of cumulative malnutrition among children.

The National Health and Nutritional Examination Survey III (NHANES III) conducted between 1988 and 1991 asked the main respondent (typically the mother) to report the height of the sample child. For a subsample of over 1,600 children between the ages of 2 and 12, the same child's height was also measured by an anthropometrist. Comparisons were then made between the mother's estimates and the gold-standard anthropometrist measure (Strauss and Thomas 1996).

The mean height from the mothers' reports was reasonably close to the mean height as measured by the experts. The variance of the heights was dramatically different, however. The estimate for the standard deviation of reported heights was more than double that for the scientifically measured heights, suggesting large "measurement error" in the mothers' reports.

Cartoons courtesy of Richard Wentworth

Table 2c—Results of a Screening Test

Screen: Response to self-defense gun question	Truth Positive: A self-defense gun use	Truth Negative: No self-defense gun use	Total
Positive	0	22	22
Negative	2	1,976	1,978
Total	2	1,998	2,000

Assumptions:	
Adult population:	200 million
Sample size:	2,000
Sample incidence	22
(No individual who had a true self-defense gun use will admit it.)	
Test sensitivity	0%
Test specificity	98.89%
Survey incidence:	22/2,000 = 1.1
Actual incidence:	2/2,000 = .1
Survey overestimation	11 times too high
Estimated total defensive gun use:	2.2 million
True total defensive gun use:	.2 million (or 200,000)

The greater standard deviation had an enormous impact on estimates of the extent of chronic malnutrition. Based on the actual heights, as measured by the antropometrists, only 1% of the children were stunted. By contrast, because the height distribution as reported by the mother was much fatter, fully 25% of the children were classified as stunted!

Put another way, for every 100 children, one in the survey was actually stunted. For every 100 children classified by the report of the mother, 99 could be a false positive, but only 1 could be a false negative. Random misreporting leads to enormous overestimation of this rare problem.

Conclusion

The argument of this article is not that it is impossible for surveys to underestimate rare events but that random misclassification will lead to overestimates, often extreme ones. Random misclassifications do not "even out." A small change in the specificity rate has a large effect on the estimates, but a large change in the sensitivity rate has a small effect on the estimates.

For rare events, overestimation is likely even if the misclassification is not random. Although there may be many important reasons to expect a high percentage of people to underreport, one small reason to expect even a tiny percentage of responders to overreport may be enough to lead to a substantial overestimate.

Sample estimates are usually presented with confidence intervals that report the likelihood that the true proportion falls within these limits. Such confidence intervals can be extremely misleading, for they assume, among other things, 100% reporting accuracy. Given that some percentage of respondents in virtually all surveys are misclassified, a more informative confidence interval would include an estimate of incorrect classification. For example, if we accept a 5% possibility that as few as 1.4% of respondents were randomly misclassified, the 95% confidence interval for accuracy of the 2.5 million self-defense survey estimate would be 0 to 2.5 million actual uses.

References and Further Reading

Hemenway, D. (1995), "Guns, Public Health and Public Safety," in *Guns and the Constitution*, eds. Henigan, D. A., Nicholson, E. B., and Hemenway, D., Northampton, MA: Aletheia Press.

—— (in press) "Survey Research and Self-Defense Gun Use: An Explanation of Extreme Overestimates," *Journal of Criminal Law and Criminology*, vol. 87 no. 2.

Hennekens, C. H., and Buring, J. E. (1987), *Epidemiology in Medicine*, Boston: Little, Brown.

Kellermann, A. L., Westphal, L., Fischer, L., and Harvard, B. (1995), "Weapon Involvement in Home Invasion Crimes," *Journal of the American Medical Association*, 273, 1759–1762.

Kleck, G. (1991), *Point Blank: Guns and Violence in America, Hawthorne*, NY: Aldine de Gruyter.

Kleck, G., and Gertz, M. (1995), "Armed Resistance to Crime: The Prevalence and Nature of Self-Defense With a Gun," *Journal of Criminal Law and Criminology*, vol. 86, no. 1.

Mauser, G. A., (1995), "Do Canadians Use Firearms in Self-Protection?" *Canadian Journal of Criminology*, 37 (4), 556–562.

McDowall, D., and Wiersema, B., (1994), "The Incidence of Defensive Firearm Use by U.S. Crime Victims, 1987 through 1990," *American Journal of Public Health*, 84, 1982–1984.

Strauss, J., and Thomas, D. (1996), "Measurement and Mismeasurement of Social Indicators," *American Economic Review*, 86, 30–34.

Wentland, E. J., and Smith, K. W. (1993), *Survey Responses: An Evaluation of Their Validity*, Boston: Academic Press.