

# Degrading Scientific Standards to Get the Defensive Gun Use Estimate Down

By Gary Kleck

*In this article, Florida State University Professor Gary Kleck responds to critics of the National Self-Defense Survey, which found that there are approximately 2.5 million defensive gun uses per year in the United States.*

## 1. Introduction

It has now been confirmed by at least 16 surveys, including the 1993 National Self-Defense Survey (NSDS) of Kleck and Gertz (1995), 12 other national surveys, and 3 state-wide surveys, that defensive use of firearms by crime victims is common in the United States, probably substantially more common than criminal uses of guns by offenders. The estimates of the annual number of defensive uses of guns in the United States range from 760,000 to 3.6 million, with the best estimate, derived from the NSDS, being 2.5 million, compared to about a half a million incidents in which offenders used guns to commit a crime (Kleck 1997, pp. 149-160, 187-189; see also the more recent Centers for Disease Control and Prevention study of Ikeda, Dahlberg, Sacks, Mercy, and Powell 1997, which estimated 1.0 million defensive gun uses linked with burglaries in which the intruder was seen, compared to 0.9 million such incidents derived from the Kleck-Gertz survey, 1995, pp. 184-185, estimates within sampling error of each other).

It has also been consistently and repeatedly confirmed that defensive gun use (DGU) is effective: crime victims who use guns for self-protection are less likely to be injured or lose property than otherwise similar victims in otherwise similar crime situations who either do not resist at all or who use other self-protection strategies (the body of evidence is reviewed in Kleck

1997, pp. 170-175). In recent years, it has become increasingly rare that critics dispute the claim that DGU is effective.

Instead, pro-control critics have focussed their efforts on their claim that, despite the enormous body of evidence indicating otherwise, DGU is actually rare. Thus, they argue, it is of little consequence for gun control policy that DGU is effective, since it is so infrequent. The critics' discussion of the topic of the frequency of DGU is strident, polemical, and extreme. For example, Philip Cook and his colleagues baldly describe large estimates of DGU frequency as a "mythical number" (1997, p. 463). Likewise, an article by David Hemenway (1997a) was brazenly titled "The Myth of Millions of Annual Self-Defense Gun Uses." In another article by Hemenway (1997b), his title implicitly took it as given that DGUs are rare, and that surveys indicating the opposite grossly overstate DGU frequency. For Hemenway, the only scholarly task that remained was to explain *why* surveys did this: "Survey Research and Self-Defense Gun Use: An Explanation of Extreme Overestimation." Finally, McDowall and Wiersema (1994), although well aware of the large number of surveys yielding large DGU estimates, nevertheless flatly concluded, in extremely strong terms, that "armed self-defense is extremely rare" (p. 1884). This conclusion was based entirely on a single survey, the National Crime Victimization Survey (NCVS), which did not even directly ask respondents about defensive gun use.

These critics do not mainly support the low-DGU thesis by affirmatively presenting relevant empirical evidence indicating few DGUs. The only empirical evidence affirmatively cited in support of the low-DGU thesis is the uniquely low estimates derived from the NCVS. The critics appear in no way embarrassed by the fact that the only national estimate they can cite in support of their theory is a survey that does not even ask respondents the key question—whether they have used a gun for self-protection. Instead, the critics get around the large volume of contrary survey evidence by pronouncing all of it invalid and

insisting that all surveys (excepting the NCVS?) grossly overstate the frequency of DGU.

## **2. The Degradation of Scientific Standards through the Use of One-Sided Speculation**

This negative strategy depends almost entirely on one-sided speculation about errors in surveys that supposedly cause overestimation of DGUs. This strategy represents an irresponsible degradation of scientific standards. The central guiding value of scientific inquiry is the primacy of empirical evidence. Advocates of the low-DGU thesis invert this principle by treating speculation as if it can trump evidence. To the extent that one-sided speculative criticism of evidence comes to be accepted as a respectable tool for assessing evidence bearing on public policy issues, the practice will reinforce the already altogether too common practice of simply ignoring or discounting evidence inconsistent with one's political prejudices, and make it virtually impossible to dislodge people from well-entrenched but erroneous positions.

In direct contradiction of scientific principles, the plausibility of speculation commonly relies on the *absence* of relevant evidence, since this is what makes it impossible to decisively rebut the speculation. With respect to both good research and bad, there is no upper limit on the amount of speculative criticism that can be directed at the work. Indeed, precisely because it is speculative, this sort of critique is just as easily applied to good research as to bad.

The only thing worse than criticizing on the basis of speculation is to do it in a persistently one-sided fashion, since this sort of critique is useless for separating the wheat from the chaff or providing scholars with a basis for knowing which are the findings to which they should give greatest weight in drawing conclusions. Indeed, in Hemenway's case, his style of critique perverts the truth-seeking process by selectively attacking the best available research, in hopes of undercutting its credibility,

without applying the same standards to more flawed research yielding contrary findings.

For example, it is a useful exercise to contrast Hemenway's assessment of the NSDS results with his uncritical citation (Hemenway 1997b, p. 1442) of findings from a bizarre study (Kellermann et al. 1995) in which the authors assessed the frequency of DGUs linked with home invasion crimes entirely on the basis of the number of times victims volunteered information about such DGUs to Atlanta police. According to the Atlanta Police Department, the offense report forms that their officers fill out do not include a box or other place calling for information about victim weapon use, nor are officers trained or required to ask crime victims about such things. Thus, information about victim weapon use, no matter how common it might in fact be, would almost never appear in police offense reports (a fact reported in the journal that published the Kellermann article—see Fotis 1996; confirmed by Kooi 1997). Nevertheless, solely on the basis of Atlanta Police Department offense reports, Kellermann and his colleagues concluded that DGUs almost never occurred in connection with home invasion crimes, because they were almost never mentioned in the offense reports!

Having made no effort to uncover any DGUs in a way likely to locate any, Kellermann et al. saw nothing wrong with concluding that they almost never occur. Hemenway likewise treated the results of this study as if they indicate something about how often DGUs actually occur in connection with this sort of crime (“in only 3 cases [1.5%] was a victim able to use a firearm in self-defense”—p. 1442). He evidently either could not see any flaws in Kellermann's reasoning, or did not feel obliged to point them out to readers, if uncritically citing these obviously *non sequitur* conclusions could be used to advance his arguments. Apparently no study could be too transparently and fatally flawed, if it supported the rare DGU thesis.

While this kind of scholarship is to be deplored, it might be less destructive if there were equally numerous and influential

advocates on both sides of the debate. At least then, all relevant evidence would eventually get a fair hearing somewhere, and the truth would have some chance of emerging from this adversary process. The reality, however, is that academic gun control believers greatly outnumber skeptics. Consider, for example, the members of the Criminology Advisory Board of the *Journal of Criminal Law and Criminology*, which published Hemenway's attack on the NSDS. The Board includes such pro-control luminaries as Richard Block, Alfred Blumstein, Roland Chilton, Philip Cook, Jeffrey Fagan, Rosemary Gartner, John Hagan, Richard McCleary, Steven Messner, Daniel S. Nagin, Lawrence Sherman, Wesley Skogan, and Marvin Wolfgang, but does not include even one scholar who has publicly expressed skepticism about gun control (see p. vii of the Summer 1997 issue).

If scholars are allowed to indulge in one-sided speculation that inevitably leads to conclusions preordained by their biases, impressions about the evidence will be determined largely by the numbers of advocates publishing articles, rather than the strength of the evidence. And if compatibility with prevailing ideological positions is allowed to determine the outcome of the debate, it will become impossible to overturn false established ideas and difficult in general to change scholars' minds about anything. This article presents an analysis of this method of assessing evidence, and a rebuttal of the criticisms of large estimates of DGU frequency.

### **3. How the Scholarly Community Has Handled the DGU Frequency Issue**

There has probably been more outright dishonesty in addressing the issue of the frequency of DGU than any other issue in the gun control debate. Faced with a huge body of evidence contradicting their rare-DGU position, hard-core gun control supporters have had little choice but to simply promote the unsuitable NCVS estimate and to ignore, attack, or discount everything else. Authors writing in medical and public health journals are typically the most crudely dishonest—they simply

withhold from their readers the very existence of a huge volume of contradictory evidence. For example, Kellermann and his colleagues discussed the issue of DGU in a recent paper, but omitted any mention of any of the surveys indicating large numbers of DGUs. Instead they cited only the NCVS estimate (1995, p. 1761). Even if Kellermann and his colleagues did not know of all 15 of the other surveys that had been conducted by the time their article was written, they clearly knew of the existence of at least six contradictory surveys, since these early surveys were reviewed in a source that Kellermann et al. cited and presumably had read (see their note 24, citing Kleck 1988). Thus it is fair to say that Kellermann and his colleagues knowingly withheld from their readers information from at least six surveys contradicting their low-DGU claims.

Since the readers, referees and editors of medical journals ordinarily know little about violence outside of the misleading bits of information they obtain from other medical/public health outlets, authors writing for these journals can ordinarily freely suppress contrary information in this way without fear of exposure or censure. Further, editors have insured near-total censorship of contrary information through their own publication decisions (see Kates, Schaffer, Lattimer, Murray, and Cassem 1995 for a review of how medical and public health journals suppress information hostile to a pro-control position). And although these journals sometimes provide for expression of contrary views in letters to the editor, editors of the journals have refused to publish even brief letters challenging the rare-DGU thesis.<sup>1</sup>

Pro-control writers publishing in criminological and social science outlets are marginally more sophisticated, “fuzzing over” the extent of contrary evidence through the vagueness of their references to the magnitude of the evidence, and through one-sided and selective critiques of the sources of the contradictory evidence. For example, Reiss and Roth (1993) concealed the extent of the contradictory evidence by vaguely referring to “a

number of surveys” that implied larger estimates (p. 265) and then dropping the matter, with no detailed further discussion of any of these surveys. Then, later in their essay, they uncritically accepted the unreliable NCVS estimates at face value (p. 266), effectively ignoring all the contrary sources. At the time they wrote, there were a least *eight* other surveys yielding DGU estimates, all radically higher than the NCVS estimate, surveys that they knew about because they had been reviewed in sources they cited.

Likewise Cook (1991) blandly referred to “a number of surveys” yielding large DGU estimates, but without mentioning how numerous these surveys were, and giving detailed attention to only one of them. McDowall and Wiersema (1994) censored even more severely; they gave their readers the false impression that conclusions in an earlier article (Kleck 1988) were based on results of a single survey. It is clear that McDowall and Wiersema were aware of at least seven of these other surveys, since they were reviewed in one of the sources they cited (Kleck 1991, p. 146, cited in their note 11).

Once large estimates of DGU frequency became too numerous and widespread to simply ignore, adherents of the rare-DGU thesis shifted to another tactic, which will be discussed at length herein. On those rare occasions when they briefly and very partially address some of the contrary evidence, they counter evidence with one-sided speculation rather than better empirical information. Cook (1991, pp. 54-55) set the pattern, speculating that surveys yield high DGU estimates because respondents telescope incidents into the recall period. “Telescoping” refers to respondents reporting events as having happened during the recall period (e.g. in the year prior to the interview), though they actually occurred earlier. This error contributes to overestimates of the number of times the experience occurred during the recall period.

While some respondents undoubtedly do telescope DGUs into the recall period, this error would not lead to an overestimate of DGU incidence unless the effects of telescoping exceeded the

effects of recall failure, i.e. respondents forgetting or intentionally failing to report genuine DGUs. Cook offered no evidence that any DGU surveys or indeed any crime-related surveys, are afflicted by more telescoping than recall failure.

The relevant technical literature indicates that the relative size of recall failure effects (mostly forgetting) compared to telescoping effects grows with increasingly long recall periods, moving estimates in the direction of a net *undercount* (Sudman and Bradburn 1973; Woltman, Bushery and Carstensen 1984). Since recall failure and telescoping effects appear to be about equal in surveys of crime victimization with a one year recall period (Dodge 1970), this means that for recall periods of five years (used in the Hart, Mauser, and Kleck-Gertz surveys discussed in Kleck 1997), there should be a net undercount of crime-related events such as DGUs, not the overcount Cook hinted at.

Cook labeled the alleged shortcomings of a survey by the Peter Hart organization as “severe” (p. 55) without offering any evidence whatsoever concerning how much effect any alleged flaw would have on DGU estimates. He did not explain how technical problems can be rated as “severe” if one does not even know if they are even minimally consequential.

Reiss and Roth (1993) later picked up on Cook’s theme, essentially repeating his unsupported and one-sided speculations about telescoping, adding in another equally unsupported and one-sided speculation that significant numbers of respondents might have erroneously characterized incidents as DGUs that did not involve any actual use of a gun. Reiss and Roth speculated that many respondents so radically misunderstood the question pertaining to defensive *uses* of guns that they reported incidents in which they merely “brought the gun nearby in anticipation of an encounter that never occurred” (p. 265). Similarly, McDowall speculated that respondents might have thought that merely carrying a gun for protection constituted actually using it for self-defense (1995, p. 137). Kleck and Gertz (1995) tested these

speculations and found little support for them—respondents claiming a DGU nearly all directly confronted their adversaries and, at minimum, pointed their guns at them or referred to the guns verbally in a threatening manner. No more than 13 of 222 cases (6%) initially reported as DGUs were “no-encounter” cases of the sort imagined by Reiss and Roth or by McDowall.

Although there is little empirical basis for these critics’ speculations about the gun use surveys, even if there had been, this would not constitute a sound basis for concluding that the far lower NCVS estimates of DGU frequency are either approximately valid or that they are closer to the correct number than estimates derived from the many other surveys yielding high figures. The speculations about the latter surveys simply do not concern flaws that are serious or common enough to account for such an enormous difference as exists between the NCVS estimates and all other estimates.

For example, Kleck and Gertz (1995) cited direct evidence from Census Bureau research on the NCVS that surveys of crime victimization experiences result in about a 21% telescoping rate—is, estimates will be about 21% too high due to people remembering events as having occurred in the recall period that actually occurred earlier (pp. 171-172). It is absurd to suggest that this rate of telescoping could account for more than a negligible share of, for example, the 30-to-1 difference between the NSDS and NCVS estimates. On the other hand, it is a simple matter to attribute the enormous discrepancy to radical underreporting in the NCVS, since there is already ample evidence of similarly radical underreporting of other violence-related events in this survey, including domestic violence, rapes, and gunshot woundings linked with criminal assaults (Cook 1986; Loftin and MacKenzie 1990).

Survey expert Tom Smith rejected the 21% estimate of telescoping, claiming that the telescoping rate “is more likely to be around 50%” (1997, p. 1468), and even computed adjusted estimates of DGU frequency based on this fanciful rate of error. As support for his 50% figure, he cited three sources of research

on telescoping (see his footnote 42). Two of these sources did not even concern surveys of crime victimization experiences, or indeed anything related, or even similar, to crime. One study pertained to health surveys (Anderson et al. 1979), and another concerned surveys about consumer expenditures on household repairs (Neter and Waksberg 1964). The degree of telescoping obviously is heavily dependent on the subject matter being asked about, so estimates of telescoping linked with one topic can reveal nothing about the frequency of telescoping in connection with another topic, unless the topics are very similar.

Smith did not offer any explanation for why he thought research on surveys on health matters and consumer household repair expenditures was more relevant than the Census Bureau research directly bearing on surveys of crime experiences that had already been cited by Kleck and Gertz.

Smith's third source (Cantor 1989) did briefly address telescoping in surveys of crime victimization experiences, specifically the NCVS, but did not support a claim of a 50% telescoping rate. Smith apparently simply misread this source, since its author directly stated that it was not possible to separately estimate telescoping from the data he examined, since telescoping was but one component in a set of survey errors. In sum, there was no foundation whatsoever for Smith's claims that there is likely to be a 50% rate of telescoping concerning survey reports of DGUs, and no reason to believe that telescoping is any higher than the 21% rate cited by Kleck and Gertz.

In any case, even if some of the critics' speculations about flaws in DGU surveys had been correct and consequential, it is not helpful or honest to speculate only in one direction, such as speculating only about flaws that *might* artificially push DGU estimates up. If one is not willing to seriously consider errors in both directions, one is simply engaging in "adversary scholarship" or "sagecraft" (Tonso 1983), an enterprise aimed not at discovering the truth, but rather at buttressing predetermined positions.

Speculation about the flaws in surveys indicating large numbers of DGUs resemble UFO buffs' beliefs that the federal government captured aliens from other worlds at Roswell, N.M., in 1947. The reason most people do not share these beliefs about UFOs is not that the beliefs can be proven false; they cannot, since it is impossible to prove a negative. Rather, most people reject them because there is no credible evidence that they are true. It is the same with speculations about DGU surveys' supposed flaws. Since it is impossible to prove a negative, one cannot prove that massive misreporting of nonexistent DGU incidents does *not* occur in surveys. There is, however, no evidence whatsoever that such massive misreporting *does* occur. There is an unlimited number of things that humans are capable of imagining existing, but almost all of these things do not in fact exist. It is the main business of science to separate what really exists in the world from that which is merely a logical possibility.

Faced with overwhelming survey support for the idea that DGUs are common, some pro-control scholars belatedly adopted the view that surveys simply cannot yield any useful information about how often DGUs occur. A cynic might conclude that, faced with defeat on the field of empirical evidence, they suddenly developed a radical skepticism toward all survey estimates. For example, prior to 1995, Philip Cook uncritically cited the very low NCVS survey estimates of DGUs (Cook 1991, p. 56; Cook and Moore 1994, p. 272) as solid evidence that DGUs were in fact rare. As late as 1994 he stated, based *solely* on survey research, that "self-defense with a gun is a rare event in crimes like burglary and robbery" (Cook and Moore 1994, p. 275). Then, preliminary frequencies on the DGU questions in the 1994 Police Foundation survey (Cook and Ludwig 1997) became available in early 1995 and the results of the Kleck-Gertz survey were published in December of 1995. Thus, in 1995 it became evident that good quality national surveys, including the 1994 Police Foundation survey that Cook helped design and analyze (eventually published as Cook and Ludwig 1997), were likely to continue indicating the DGUs occurred quite often.

By no later than May of 1996 Cook had radically altered his position to the view that “surveys are a decidedly flawed method for learning about the frequency with which innocent victims of crime use a gun to defend themselves” (Cook and Ludwig 1996). Not only did Cook thereby dismiss all previous survey evidence, but also any evidence that might be generated by surveys in the future. Further, he went beyond stating this position on the accuracy of the scientific evidence—he also forestalled policy use of *any* future evidence on the prevalence of DGU by asserting that “even if we could develop a reliable estimate of [DGU] frequency, it would only be of marginal relevance to the ongoing debate over” gun control (Cook and Ludwig 1996).

Since surveys are the only way we have of measuring the frequency of DGUs, Cook had thereby transformed the claim that DGUs are rare into a nonfalsifiable proposition, i.e. an assertion that, even if it were false, could not, under Cook’s standards, be shown to be false. Note, however, that this radical turnabout in views came about only after the National Self-Defense Survey (NSDS) (Kleck and Gertz 1995) and his own Police Foundation survey (Cook and Ludwig 1996; 1997) had both yielded estimates of annual DGUs, based on large-scale, high-quality national surveys specifically designed to estimate DGU frequency, in the millions .

The Police Foundation survey, while based on a sample only half that of the NSDS, was modeled after, and otherwise comparable to, the NSDS, and included even more questions getting at details of alleged DGUs. It strongly confirmed the results of the Kleck-Gertz NSDS, yielding estimates, where comparable, of annual DGU frequency that were within sampling error of those obtained by Kleck and Gertz (Cook and Ludwig 1997, esp. pp. 62-63). Faced with estimates that he himself had helped develop, but which radically contradicted his earlier acceptance of the very low NCVS estimates, Cook flatly refused to accept the verdict of the evidence. Instead, he and his coauthor indulged in numerous evidence-free pages of one-sided

speculation about how suspected flaws in their and other surveys *might have* led to errors in DGU estimates. They noted a few inconsistencies in responses of their respondents but failed to establish how or why these would lead to a net overestimate of DGU frequency. Equally important, by almost exclusively focussing (by their own admission—see Cook and Ludwig 1996, p. 118) on possible sources of false positives, they failed to make any case for why false positives should outnumber false negatives, such as respondents concealing or forgetting DGUs.

Cook and Ludwig claimed to have established inconsistencies between their results and other statistics, concluding that their large DGU results were therefore implausible. In all cases, their reasoning was fallacious. For example, they cited data on the number of people treated in emergency rooms for nonfatal gunshot wounds and asserted that their own survey's estimates of criminals wounded during DGUs were implausibly high in comparison. In fact, the two sets of numbers are perfectly consistent once one acknowledges that criminals wounded by victims are unlikely to seek medical treatment, since medical personnel are required to report gunshot wounds to police, and most such wounds are survivable without professional medical treatment (Kleck 1997, Chapter 1). Cook and Ludwig dealt with the possibility that most criminals wounded by gun-wielding victims do not receive emergency room treatment by simply announcing that "we find that possibility rather unlikely" (1996). They did not even bother to provide their readers with a rationale for this arbitrary pronouncement, never mind any supporting evidence.

Their assessment might have been based on either of two unsupported premises: (1) a typical GSW is so serious that people suffering such a wound could not substitute self-treatment for professional treatment without placing their lives in peril, or (2) criminals are ignorant of, or indifferent to, the fact that medical personnel treating their wounds would report GSW patients to the police. Unless one accepts these dubious premises, it hard to see how one could reasonably assume that all, nearly all, or even

most criminals wounded during DGUs would seek treatment at an emergency room.

Cook and Ludwig likewise claimed that the estimated number of DGUs connected with particular types of crimes were inconsistent with NCVS estimates of the total number of crimes of a given type, with or without DGUs. For example, they claimed to have shown that the estimated number of DGUs linked with rapes exceeded the total number of rapes, as estimated by the NCVS. One fatal flaw in their reasoning had already been anticipated in a passage in the original article reporting the NSDS estimates (Kleck and Gertz 1995, pp. 167-168), a passage that Cook and Ludwig evidently chose to ignore. That passage noted that the reasoning later applied by Cook and Ludwig relied on the assumption that the universe of events covered by the NSDS (and thus Cook and Ludwig's survey) was a subset of the universe of events covered by the NCVS. This assumption is implausible. As noted in that passage, "a large share of the incidents covered by our survey are probably outside the scope of incidents that realistically are likely to be reported to either the NCVS or police" (p. 167).

It is likely that only a minority of all crime incidents get reported to the NCVS. Therefore, no matter how large the estimated number of DGUs is in a gun survey, the number could still be a plausibly small share of *all* crime incidents, including both those effectively covered by the NCVS and those not covered. Consequently, comparing DGU estimates with NCVS crime estimates can tell us nothing about whether the former are plausible. Ignoring Cook and Ludwig's one-sided speculations and fallacious reasoning, and paying close attention to their empirical results, leads to the conclusion that their survey strongly supported the assertion that DGUs are very common.

Among pro-gun control scholars, the most active in pushing the rare-DGU thesis has been public health scholar David Hemenway, who has presented a critique of DGU survey estimates in a series of overlapping articles (Cook, Ludwig and

Hemenway 1997; Hemenway 1997a; Hemenway 1997b). The most extensive of these papers (Hemenway 1997b) encompassed all of the significant criticisms made of DGU survey estimates, both by Hemenway and by Cook, McDowall, Reiss and Roth, and others. Therefore, the rest of this paper is devoted to a point-by-point refutation of Hemenway's criticisms of the DGU estimates generated by the 1993 National Self-Defense Survey (Kleck and Gertz 1995), as presented in Hemenway's article in the Summer 1997 issue of the *Journal of Criminal Law and Criminology*.

#### **4. The Hemenway Critique of the National Self-Defense Survey**

Hemenway's paper was not an attempt to produce a balanced, intellectually serious assessment of estimates of defensive gun use. Instead, his critique served the narrow political purpose of "getting the estimate down," for the sake of assisting the gun control cause. An honest, scientifically based critique would have given balanced consideration to both flaws that would tend to make the estimate too low (e.g., people concealing DGUs because they involved unlawful behavior, and the failure to count any DGUs by adolescents), and to those that contribute to making them too high. Equally important, it would have given greatest weight to relevant empirical evidence, and little or no weight to idle speculation about possible flaws. Hemenway's approach was precisely the opposite—one-sided and almost entirely speculative. Readers who have any doubts about the degree to which Hemenway's paper was imbalanced could carry out a simple exercise to assess this claim: count the number of lines Hemenway devoted to flaws tending to make the estimate too high and the number devoted to flaws making the estimate too low.

Hemenway's one-sided determination to fixate only on possible sources of overestimation was so strong that he failed to recognize even the most conspicuous sources of underestimation. He claimed that Kleck and Gertz obtained an estimate of gun

ownership prevalence in their sample that was “outside the range of all other national surveys” (p. 1434), to the low side, yet was oblivious to the implication of this for DGU estimates—since DGUs are obviously more common among gun owners, any underrepresentation of gun owners in the survey sample would contribute to an *underestimate* of DGUs.<sup>2</sup>

He likewise noted the underrepresentation of blacks in the NSDS sample (p. 1434), a problem nearly universal in national surveys, yet did not note the implication that underrepresentation of highly victimized subsets of the population would necessarily imply an underrepresentation of persons who had occasion to engage in acts of self-defense, including use of a gun for self-protection. Similarly, Hemenway asserted that the NSDS gives too much weight to persons who are the only adult in their household (p. 1434), yet apparently was not aware that persons who live alone or in smaller households are *less* likely than others to be victims of crimes like burglaries (U.S. Bureau of Justice Statistics 1996, p. 28), and that he was therefore noting a problem likely to contribute to an *underestimation* of DGUs.

Likewise, Hemenway made no mention of the even more obvious fact that surveys confined to adults (as all of the DGU surveys were) by definition exclude all self-reports of DGU experiences by adolescents. Since rates of gun carrying are as high among adolescents as among adults (Kleck and Gertz 1998, pp. 200-201), and persons age 12-17 claim about 24% of all violent victimizations (U.S. Bureau of Justice Statistics 1997, pp. 6, 8), this problem alone could cause surveys to miss as much as a quarter of all DGUs. Nor did Hemenway acknowledge other obvious sources of *underestimation* that Kleck and Gertz had explicitly noted, such as the omission of persons without telephones, who are poorer and thus more likely to be crime victims than others (Kleck and Gertz 1995, p. 170).

The political function of this sort of advocacy scholarship is clear. While high estimates of DGU frequency do not constitute an obstacle to moderate controls over guns such as laws

requiring background checks, they constitute a very serious obstacle to advocacy of gun prohibition. Disarming the mass of noncriminal prospective crime victims would, if high DGU estimates are even approximately correct, result in large numbers of foregone opportunities for defensive uses of guns that could prevent deaths, injuries, and property loss. To acknowledge high DGU frequency would be to concede the most significant cost of gun prohibition. Hemenway's paper was an attempt to neutralize concerns about such costs and to provide intellectual respectability for positions identified with Handgun Control Incorporated (HCI), the nation's leading gun control advocacy group.

Hemenway has close ties to HCI through two key staff members of its "educational" branch, the Center to Prevent Handgun Violence (CPHV). His closest and most frequent collaborator on gun-related research is Douglas Weil, currently Research Director of CPHV, with whom Hemenway has co-written at least five articles on gun topics (Hemenway and Weil 1990a; 1990b; Weil and Hemenway 1992; 1993a; 1993b). (Interestingly, Hemenway did not include Weil, his erstwhile closest collaborator, among those he thanked in his acknowledgements, presumably for their comments on earlier drafts of his paper [Hemenway 1997b, p. 1430], as if to distance himself from an HCI employee). Hemenway also has contributed to, and co-edited, a strongly pro-control 96-page propaganda tract with Dennis A. Henigan, legal counsel to HCI and CPHV (Henigan, Nicholson, and Hemenway 1995). This obscure tract presented a note-for-note rendition of the HCI/CPHV view of the Second Amendment, a view sharply at variance with virtually all scholarly research on the topic (see Reynolds 1995 for a review of the Second Amendment literature).

In one of his articles coauthored with Weil, Hemenway claimed that their survey data showed that the National Rifle Association (NRA) misrepresents the gun control views of its own members. Kleck pointed out in a published critique that many of those respondents that Weil and Hemenway treated as

NRA members probably were not, since their figures overstated known NRA membership by a factor of three. This accurate claim is oddly parallel to the inaccurate one Hemenway has since directed at Kleck's work, the main difference being that NRA membership is exactly known, and so it was indisputable that Weil and Hemenway's data grossly overstated NRA membership.

Hemenway's political intentions and strong feelings were evident in his wild overstatements and the grandiose and unwarranted conclusions he drew from weak or irrelevant evidence and fallacious reasoning. He did not get past his title before making his first overstatement, claiming that he had established, without benefit of any new empirical evidence, not only that the NSDS estimates were too high but that they were "extreme overestimates" (Hemenway 1997b, p. 1430). He then announced in his first paragraph that "it is clear that [the Kleck and Gertz] results cannot be accepted as valid" (p. 1430). He went on to falsely claim that "all checks for external validity of the Kleck-Gertz finding confirm that their estimate is highly exaggerated" (p. 1431), when in fact these checks have repeatedly confirmed the conclusion that DGUs are common.

DGUs usually involve unlawful possession of a gun by the gun-wielding victim, and sometimes other illegalities as well (Kleck and Gertz 1995, pp. 150, 156, 174), a point Hemenway did not dispute. Yet, he made the extraordinary and counterintuitive claim that there is a social desirability bias to people reporting their own illegal behavior (Hemenway 1997b, p. 1431)—that is, people will falsely report DGU experiences because they believed this would present them in a more positive, socially desirable light. Hemenway insisted that such a desirability bias is not only plausible, but that it is likely: "the likelihood of social desirability response bias (self-presentation bias) is clear" (p. 1438). By the end of his paper, without having provided any credible supporting evidence, Hemenway concluded that the NSDS was afflicted by an "enormous problem of false positives"

(persons claiming a DGU who did not have one) and “massive overestimation,” flatly stating that “the Kleck and Gertz survey results do not provide reasonable estimates about the total amount of self-defense gun use in the United States” (p. 1444). It was an impressive achievement to be able to arrive at such high-powered conclusions without the inconvenience of gathering or even citing *any* new empirical evidence.

### **5. The Illegitimacy of One-sided Speculation: An Ounce of Evidence Outweighs a Ton of Speculation**

Hemenway’s critical technique, like that of Cook, McDowall, Reiss and Roth, and other proponents of the rare-DGU thesis, was simple: one-sided, and often implausible, speculation about flaws that *might* have afflicted DGU surveys, and that might have been consequential enough to significantly affect their estimates. As a typical example of this technique, he speculated that people claiming DGU experiences might have been mentally ill, hinting that such states of mind would cause people to invent nonexistent DGUs, due to their “different perception of reality” (p. 1435). He did not provide any evidence that even one of the DGU-reporting respondents in the NSDS or any of the other DGU survey was in fact mentally ill, or reported false information about DGUs because of such illness. Indeed, he did not even report any evidence indicating that large numbers of respondents in *any* survey are mentally ill. Hemenway’s idea of supportive evidence was merely to cite estimates of the share of the general population that is thought to suffer from mental illness. It was sufficient for Hemenway that large numbers of DGU reporters *could have* been mentally ill. The mere hypothetical possibility was treated as seriously as actual empirical evidence. The fact that he had no basis for believing that even one DGU reporter in the NSDS or any other DGU survey was mentally ill, or invented a nonexistent event, was effectively treated by Hemenway as a relatively unimportant detail.

Nor did he explain why the “different perception of reality” of mentally ill people would cause them to develop long, detailed,

and internally consistent accounts of nonexistent DGUs. One would think that many forms of mental illness would make it harder for people to provide such consistent-but-false accounts, while disorders such as paranoia would be at least as likely to cause people to withhold information about real events from strangers who called them up on the phone as to motivate them to fabricate nonexistent events. If someone were suffering from a variety of schizophrenia, such as paranoia, why would they invent or falsely recall events featuring their own illegal behavior? Would it not be more common that such persons would be suspicious of the intentions of interviewers and be especially likely to withhold accounts of DGUs that really occurred? And if both kinds of false responses were given, as we assume is the case, why should the former kind be more common than the latter? If it is not, then Hemenway's citation of data on the prevalence of mental illness cannot support his argument that DGUs are overestimated.

Hemenway even speculated that respondents reporting DGUs were deliberately lying for the explicit purpose of boosting DGU estimates, in order to advance their political beliefs opposing gun control (p. 1439). Our point here is not that it is impossible for this sort of thing to happen; certainly one cannot logically rule it out. Rather, the point is that Hemenway's critique was filled with similar speculations about a long string of hypothetical, logically possible sources of false positives, but devoid of any empirical evidence that even one respondent in an actual survey had actually provided a false DGU account due to any of the hypothetically possible causes of such accounts, never mind evidence that enough such errors occurred to substantially distort DGU estimates. It bears repeating that a virtually unlimited number of things are possible in the world, and can be imagined by the human mind, but almost none of the hypothetical possibilities are in fact a part of the world.

The reliance on musings about logically possible errors in the absence of supporting evidence would not be quite so bad had

Hemenway made even a minimal effort at balance in considering the full range of errors possible in surveys. Unfortunately, he devoted his imaginative powers exclusively to thinking up flaws that might have contributed to the overestimation of defensive gun use (DGU) frequency, while either ignoring well established sources of underreporting, or briefly discussing them only for the sake of superficially dismissing them (e.g., p. 1439). Even when Hemenway speculated about sources of response error that are plausible, he offered no rationale for why the problems should lead to more false positives than false negatives. Instead he simply conjured up reasons why they might lead to false positives. As support for his one-sided speculations, Hemenway even cited *other* people guilty of the same dubious practice (p. 1433, notes 11 and 12, citing McDowall et al. 1992 and Reiss and Roth 1993).

All research is flawed. Known flaws should be identified and their likely consequences carefully assessed. Speculation about flaws can play a role in the pursuit of truth by motivating researchers to gather better empirical evidence less afflicted by the flaws. Speculation by itself, however, should not be given any weight in assessing evidence. An ounce of evidence, even though flawed, outweighs a ton of speculation.

## **6. Deceptive Claims and Insinuations in the Hemenway Critique**

Unable to develop any empirical evidence of false positives in the DGU surveys, Hemenway resorted to simply inventing false details about the surveys and the conclusions drawn from them by their authors. Unable to develop valid criticisms of the research actually conducted, he created imaginary straw man versions of it that he could criticize.

For example, Hemenway misrepresented the implications of Kleck and Gertz' findings concerning how many people thought they had saved lives through DGU. He claimed that "the K-G results imply that many hundreds of thousands of murders should have been occurring when a private gun was not available for

protection” (p. 1443). Hemenway in fact knew that the Kleck-Gertz results did *not* imply such a thing, since the authors had explicitly stated (Kleck and Gertz 1995, p. 176) that they had only asked people about their *perceptions* of the likelihood that their DGU had saved a life, and that the results did *not* imply how many murders did not occur as a result of a gun being available for protection: “how many of these were truly life-saving gun uses is impossible to know” (p. 177).

Kleck and Gertz explained why it is not surprising that DGU is so common relative to criminal gun use, noting there are far more gun-owning victims than gun-owning criminals (1995, p. 180). Hemenway characterized this explanation as “nonsensical” because “criminals are more rather than less likely than victims to possess guns” (1997b, p. 1443). He offered no supporting evidence for this “fact,” apparently because he made it up.

Kleck and Gertz were referring to the huge potential for victim gun use in crime incidents, based on the much higher number of prospective victims who *own* guns than criminals, rather than the number who possessed guns during crime incidents, something we do not know from any source. It is possible Hemenway did not understand this, and that his claim referred instead to the distinct issue of gun possession during crime incidents. The NCVS not only does not directly ask victims whether they actually use guns for self-protection, but does not in any way ask whether victims possessed guns during the incident. Nor does any other national survey establish relative gun possession levels during crime incidents among victims and offenders.

Concerning ownership of firearms, the only survey to ask a representative national sample of criminals about gun ownership found, in 1991, that only 24% of state prison inmates personally owned a gun in the month before they were arrested for the offense that got them sent to prison (U.S. Bureau of Justice Statistics 1993, p. 19), while the 1989 General Social Survey indicated that 31% of the general U.S. adult population personally

owns a gun (Kleck 1991, p. 52). While one might selectively speculate that incarcerated criminals underreport gun ownership more than noncriminals, the best available evidence nevertheless indicated, at the time Hemenway wrote, that criminals are *less* likely to own guns than noncriminals, exactly the opposite of what Hemenway flatly stated as fact. Unless he was consciously lying, Hemenway apparently simply did not bother to check whether what he was claiming was correct or supported in any body of empirical evidence.

It would also be wrong to assume that few potential victims carry guns away from home, and conclude therefore that guns are too rarely available in public places to be used very often by victims during crime incidents. The NSDS indicated that each year over 7 million U.S. adults carry guns on their person for self-protection for an average of 138 days per year, implying nearly one billion person-days of such carrying (Kleck and Gertz 1998), compared to 0.7-1.6 million DGUs in public places (Kleck and Gertz 1995). Thus, there are about 1,000 times as many instances of defensive carrying as would be needed to account for all of the DGUs that the NSDS estimated occur in public places each year. The NSDS estimates of carry prevalence are not unique: a 1993 survey by the strongly pro-control Gallup firm found an even higher prevalence of defensive gun carrying on the person (Kleck 1997, Ch. 6; Kleck and Gertz 1998). Consequently, there is good reason to expect huge numbers of victims would not only own guns but would possess them at the time they were victimized.

Hemenway also misled readers by quoting Kleck and Gertz out of context in a way that suggested that they somehow felt that the NCVS was a good survey for estimating DGU frequency (p. 1441), when their position was actually the reverse. On pp. 156-157 of their article, Kleck and Gertz had written that (1) years of careful refinement and evaluation had made the NCVS an excellent vehicle for getting respondents to report illegal things that other people had done to them, but that (2) it was singularly ill-suited to getting people to admit possibly illegal

things (such as DGU) that they themselves had done. Hemenway quoted only the first part of this statement (see text attached to his note 46), a bit of creative editing that served to invert the sense of the passage.

In some instances, Hemenway's speculations about alleged problems were unconscionable since he knew that Kleck and Gertz had already directly addressed them and had presented evidence contradicting the speculation, and Hemenway had offered no rebuttal of the evidence, or argumentation as to why it was invalid or irrelevant. For example, he speculated (p. 1438) that respondents might have reported incidents "in which they were afraid, they retrieved a gun, and nothing bad happened." Kleck and Gertz had explicitly addressed this issue in the article (1995, pp. 162-163) and stated that they had insured the respondents claiming a DGU had (1) actually confronted an adversary, (2) had actually done something with their gun (e.g. pointed it at an adversary), and (3) could state a specific crime (i.e. "something bad") that they thought was being committed against them. In short, Hemenway falsely hinted that Kleck and Gertz did nothing to rule out this sort of report as a DGU.

Hemenway claimed the Kleck and Gertz did little to reduce what Hemenway imagined to be a huge overestimation bias. Since there was no reason to believe such a thing existed when the NSDS was designed, and even less reason to believe it now, this is comparable to saying that Kleck and Gertz did nothing to prevent demons from possessing their interviewers. With a convenient vagueness, Hemenway did not say precisely what he thought Kleck and Gertz should have done to reduce this supposed bias, and therefore does not specify anything they failed to do.

In any case, the claim is false. On p. 161 of their article Kleck and Gertz explained that "all interviews in which an alleged DGU was reported by the respondent were validated by supervisors with call-backs" and, on p. 163, that Kleck "went through interview sheets on every one of the interviews in which

a DGU was reported, looking for any indication that the incident might not be genuine.” They also reported on p. 172 that they debriefed their interviewers after the calling was finished, asking them about possible false reports and found that “only one interviewer spoke with a person he thought was inventing a nonexistent event.” It would be more accurate to say that they did virtually *everything* that could ethically be done to guard against false reports.

On p. 1439, after noting that Kleck and Gertz had concluded that virtually all respondents in the NCVS who in fact had a DGU experience fail to report the experience to NCVS interviewers, Hemenway asserted that “there is certainly no precedent for this extreme pattern of lying.” This is a falsehood in two ways. First, it is a mischaracterization of the Kleck-Gertz conclusion, since they only wrote that “virtually none of the victims who use guns defensively tell interviewers about it in the NCVS” (1995, p. 168). They did not assert that this was due to lying. Quite the contrary—they had explicitly pointed out (p. 155) that since the NCVS never directly asks respondents explicitly about DGU, it is not even necessary for a respondent to lie in order for the DGU to go unreported. The NCVS makes it easy for this to happen, since all a respondent need do to conceal a DGU is to remain silent about their gun use, refraining from volunteering the information in response to an unspecific prompt about the respondent’s possible protective actions. Thus, precedents about levels of “lying” are irrelevant to the arguments they made.

Second, if one generously assumed that Hemenway merely expressed himself badly, and was only claiming that underreporting, due to any causes, of this magnitude was unprecedented, then he knew this claim was false. Specifically in connection with the very survey in question, Kleck and Gertz cited prior research indicating that the NCVS appeared to miss approximately 97% of rapes and sexual assaults, and over 90% of spousal assaults (p. 168, citing the review by Loftin and MacKenzie 1990). Since Hemenway did not rebut (or mention)

this prior research, but had been made aware of it by the Kleck-Gertz article, he knew that there was indeed ample precedent for believing that the NCVS could miss nearly all DGUs.

Hemenway invented still another claim about Kleck and Gertz' conclusions. He asserted that the Kleck and Gertz claimed "that *many* responders who actually did use a gun in self-defense in the past year forgot to report it on their survey" (p. 1440, emphasis added). He did not cite a page where the authors made this claim, because there is no such page. While there undoubtedly are at least a few respondents who did forget a minor DGU, Kleck and Gertz argued that the main reason respondents would fail to report a DGU was because people are reluctant to report experiences in which they engaged in criminal behavior, or behavior others might define as criminal (pp. 156-157, 171).

Hemenway also deceived by omission when discussing telescoping as a source of overestimation in the NSDS (p. 1439), as if it were a flaw in the survey that he had discovered. What he did not say is that Kleck and Gertz had already addressed this issue in their article, used prior research to estimate its likely magnitude, and had shown that it was likely to have only a minor impact on estimates (Kleck and Gertz 1995, pp. 171-172).

Hemenway also misled his readers when he claimed that Kleck and Gertz "do not provide detailed information about their survey methodology" (p. 1433), since he knew that they did in fact provide unusually detailed information about the methods used in the NSDS, including such arcane information as procedures for taking indirect reports from proxies, selection of interviewers, random monitoring of interviews, rates of validation call-backs by supervisors, and details of the sampling procedures (Kleck and Gertz 1995, pp. 160-163). Indeed far more detail was provided than is customary in journal articles reporting survey results.

Noteworthy here is Hemenway's hypocrisy in criticizing (pp. 1433-1434) Kleck and Gertz for not reporting details that he

never reported in his own published survey reports, such as methods for weighting data, or survey organization procedures for handling busy signals or answering machines (contrast Hemenway's criticisms with the sketchy information provided in Hemenway et al. 1995; Hemenway and Richardson 1997, pp. 188-190; Weil and Hemenway 1992; 1993a).

In this case, Hemenway's insinuation was that the absence of details on some extremely specialized technical matters in the Kleck-Gertz report somehow indicated there were in fact problems with how the matters were handled. Yet, the criticism was so devoid of content that Hemenway did not even bother saying why any of these hypothetical problems, even if they had existed, would have caused the DGU estimate to be too high, and thus why his insinuations had any bearing on the topic at hand.

Hemenway also misrepresented the conclusions of other scholars to generate spurious support for his positions. For example, he miscited David Cantor (1989) to support his theory of extraordinarily high rates of telescoping in DGU surveys, contrasting these surveys with the NCVS. Unlike most surveys, including the DGU surveys, the NCVS is a "bounded" survey in that the same respondents are repeatedly interviewed at six month intervals and asked about crime experiences that occurred in the six months since the previous interview. This serves to establish a clear "bound" on the time period respondents are supposed to speak about, eliminating the telescoping that afflicts unbounded surveys. As support for his claim of high telescoping in the DGU surveys (all of them unbounded), Hemenway reported that "Unbounded rates of reported victimization are typically 30% to 40% higher than bounded rates" (p. 1439), citing Cantor.

What Hemenway did not pass on to his readers was Cantor's explicit conclusion that one could *not* attribute all of this difference to telescoping by respondents in the unbounded interviews, and that some of it was due to *underreporting* in the bounded interviews. Cantor stated, in terms clear enough that Hemenway could not have honestly misunderstood, that it was

impossible to tell how much of the 30-40% difference was due to telescoping. Thus, Hemenway's tactic was to raise the issue of telescoping, cite the 30-40% discrepancy figure, and then let readers "draw their own conclusions" that this represented the level of telescoping. If one more honestly recognized that only part of this discrepancy is due to telescoping, and assumed, for example, that only half of it is due to telescoping, one would arrive at a telescoping rate of 15-20%, i.e. almost exactly the same as the 21% figure cited by Kleck and Gertz.

In this same vein, Hemenway mischaracterized the published opinions of pro-control scholars as part of an effort to exploit prestige bias by invoking the name of the well-respected National Research Council (NRC). He alleged that a report by the NRC "finds" that Kleck's earlier estimates "appear exaggerated." This is a mischaracterization, since this was not a "finding" of the NRC, or of any of its panels, but merely a personal opinion expressed by an NRC report's authors, Albert Reiss and Jeffrey Roth (1993). These authors had no relevant evidence of their own, and simply relied on the same technique of one-sided speculation that Hemenway later used, in a none-too-subtle effort to "get the estimate down."

The Reiss-Roth opinions were, in any case, irrelevant to the purposes of Hemenway's paper, which was intended as a critique of the Kleck-Gertz survey, conducted after the Reiss-Roth report was written, rather than an assessment of the many less sophisticated early surveys reviewed in the Kleck papers that Reiss and Roth addressed. This passage appears to serve no purpose other than to provide Hemenway with an excuse to cite someone else's outdated and equally unfounded personal opinions that "Kleck's conclusions rest on limited data and assumptions" (p. 1432).

The citation of the Reiss-Roth critiques of older studies also ignored the fact that estimates have gotten larger as methods have been improved and the problems cited by Reiss and Roth (and Kleck 1991, pp. 108-111) were solved. The expectation of

critics that problems in the surveys were inflating DGU estimates was contradicted by the simple fact that the more technically sound the surveys became, the larger the DGUs estimates got (compare Cook and Ludwig 1997 and Kleck and Gertz 1995 with the pre-1991 surveys critiqued in Reiss and Roth 1993, and summarized in Kleck 1997, pp. 187-189).

### **7. Red Herrings and the Issue Not Addressed**

Much of Hemenway's paper was a red herring in that it implicitly misstated the central technical question about survey estimates of DGU frequency. Much of it was devoted to elaborate speculations about why people might falsely claim to have used a gun defensively, as if it were somehow in dispute that some respondents might have provided false positive responses (pp. 1430, 1438-1440). He inaccurately hinted that Kleck and Gertz unreasonably ignored the possibility that some of their respondents provided false positives (p. 1439), a claim that served to portray them as being as doctrinaire and unreasonably one-sided as Hemenway was.

We assume as a matter of course that the NSDS was like all other surveys in that some respondents gave inaccurate responses to questions, and that these errors included both false positives and false negatives. The central question is not whether there were some false positives, nor even how many false positives there were, but rather what the relative balance was between false positives and false negatives. Survey estimates cannot be too high unless false positives outnumber false negatives, and cannot be "extreme overestimates" unless false positives *greatly* exceed false negatives. Because Hemenway made no effort to assess the frequency of false negatives, it was logically impossible for him to say what this balance was, and therefore impossible to draw meaningful conclusions about whether the NSDS estimates were too high or low.

Another red herring in Hemenway's paper (pp. 1431-1433) was his discussion of eight earlier surveys Kleck had carefully critiqued in *Point Blank* (1991, pp. 104-111). Kleck and Gertz

made great efforts to fix as many of the problems of those surveys as they could when they conducted the NSDS. What, then, was the point of Hemenway citing criticisms of those surveys as “Background” (p. 1431), if not to score a few cheap debating points by hinting that what was flawed in the earlier surveys must also be flawed in the Kleck-Gertz survey? If the NSDS did indeed still share some flaws with those earlier surveys, it was unnecessary to bring up the flaws of the earlier surveys; Hemenway could have simply addressed these flaws in connection with the NSDS and documented that it had a given problem. On the other hand, if some criticisms applicable to those earlier surveys did *not* apply to the NSDS, it was dishonest to cite critiques of the former that mostly addressed flaws that were fixed in the NSDS, in a context where readers would assume that they were relevant to the NSDS.

Similarly, Hemenway tried to get some mileage out of the fact that the NCVS has larger sample sizes than those in the DGU surveys (p. 1432), even though the only effect this has on estimates is that it reduces random sampling error (and thus the width of an interval estimate). It does not affect, on average, the size of the estimate, which is what Hemenway was challenging. Since Hemenway did not dispute this point, he presumably knew that his observations about the huge NCVS sample sizes were irrelevant to the issue at hand, but may have hoped to score some cheap points with readers who did not know this.

Hemenway also could not resist citing (p. 1443) some irrelevant research that purported to show that a gun in the home raises the risk of homicide (contrary to Hemenway’s phrasing, the study did not merely claim that a gun was “associated with” an increased risk). Hemenway believed that this finding was somehow inconsistent with the NSDS findings on the number of people who believed their DGU might have saved a life. In fact, all that this research (Kellermann et al. 1993) accomplished was that it reconfirmed the commonplace finding in criminological research that the same things that increase one’s risk of violent

victimization also increase the probability that one will acquire a gun for self-protection, and that there will therefore sometimes be a positive association between victimization risk and gun ownership, even if the latter has no impact on the former. (For extended critiques of this study, see Kates et al. 1995, pp. 268-276; Kleck and Hogan 1997; Kleck 1997, Ch. 7). There is, in fact, nothing in this study's data that is incompatible with the assertion that the net causal effect of owning a gun is, on average, to reduce the likelihood one will become a victim of homicide.

Note however, that citation of this study would be a red herring even if one believed that keeping a gun in one's home does increase the risk of homicide victimization, since it would not imply anything about whether actual defensive *use* of guns saves lives or how often it might do so, never mind how often people believe they saved a life with a DGU. It is perfectly possible that DGU saves lives with great frequency, but that, with even greater frequency, guns in a person's home somehow contribute to the likelihood of one resident of a home killing another.

## 8. The Nature of False Positives

It is hard to discern exactly what kinds of false positives Hemenway believed show up most often in all these DGU surveys. He waffled on the issue of whether people are: (1) consciously inventing nonexistent events; (2) consciously but honestly misrepresenting accounts of real events that did not really involve DGU (e.g., they involved an aggressive use of a gun that the respondent wrongly regarded as defensive); or (3) unconsciously distorting real events. He seemed to have doubts himself about possibility (1) occurring very often, hastening to assure readers that false responders do not necessarily have to lie (p. 1435), but was otherwise unwilling to commit himself to the relative frequency of these types of misreports.

It is worth emphasizing how much trouble NSDS respondents had to go to in order to falsely report a completely nonexistent event as a DGU. Hemenway cited a survey in which 10% of the

respondents told interviewers that they had seen something they thought was a spacecraft from another planet (p. 1438), insinuating that one could reasonably expect similarly large numbers of people to falsely claim to have used a gun for self-protection. Unlike respondents in the UFO survey, however, respondents in the NSDS who wanted to falsely report a nonexistent DGU could not qualify as having had such an experience merely by saying “Yes.” Rather, they had to provide as many as 19 internally consistent responses covering the details of the alleged incident. In short, to sustain a false DGU claim, respondents had to do a good deal of very agile mental work, and stay on the phone even longer. On the other hand, all it took to yield a false negative was for a DGU-involved respondent to speak a single inaccurate syllable: “No.” The point is not that false positives were impossible but rather that it took far more time and trouble to provide a false positive than a false negative.

Consider also the context in which Hemenway imagined all these false reports to have been provided. Randomly selected people were called unexpectedly, and questioned rapidly by total strangers, for no more than 15 minutes, with one question immediately following another. There was no prolonged opportunity to invent a nonexistent event, rehearse inaccurate details, or to otherwise get a false story straight. Respondents providing a false positive account of a DGU had to be not only dishonest but very quick-witted and creative as well.

Regarding possibility (2), Kleck and Gertz (1995, p. 174) noted that most of the reported DGUs were linked with the types of crimes—burglaries, robberies, and sexual assaults—where there is little possibility of participants being honestly mistaken about who was the victim and who was the offender, or whether their gun use was genuinely defensive. While some respondents may well have consciously misrepresented aggressive actions as defensive, and a very few might have consciously invented entirely fictitious events, it is hard to see how respondents could report an account of a real burglary, robbery, or sexual assault in

which they were aggressors and somehow unconsciously or honestly distort their own criminal, aggressive use of a gun into a “defensive” use.

An honest misunderstanding of real events in a way that would falsely qualify them as DGUs is more plausible in connection with assault incidents, such as those where people prefer to characterize their partly aggressive, partly defensive behavior in “mutual combat” incidents as purely defensive in character. Kleck and Gertz addressed this latter possibility in their original article and showed that it could not account for more than a small fraction (probably less than a tenth) of the incidents they counted as DGUs (1995, p. 174). Hemenway did not refute that evidence.

Hemenway’s view of the world was that it is full of potential survey respondents who are simultaneously mischievous or delusional, yet also extremely energetic, persistent, mentally agile, and disciplined enough to invent, on short notice, long, complicated, and internally consistent tales for strangers who unexpectedly call them on the telephone. This strange world is not the one familiar to survey researchers. Instead, their world is a more mundane one in which people who incorrectly answer questions about illegal behavior are mostly those who do not want to tell strangers about their own unlawful behaviors and consequently say “no” when the correct answer was “yes.”

## **9. Raising the Dead: Resuscitating the NCVS Estimates of DGU**

Hemenway (1997b, pp. 1431-1432) contrasted NCVS (victim survey) estimates of DGU with the NSDS estimates, but was evasive as to exactly why he did this. He never explicitly stated that he considered the NCVS estimates to be even approximately accurate, perhaps because he knew that this position was indefensible. He made no effort to rebut Kleck and Gertz’ detailed explanation (1995, pp. 153-157) of why the NCVS grossly underestimates DGU frequency, and did not even discuss or mention most of their arguments or evidence. Thus, the

assertion that the NCVS estimate is far too low remains un rebutted. But if the NCVS estimates are *not* accurate, what was point of Hemenway citing them in the context of a challenge to the very different NSDS estimates?

Exploiting the tactic of “maintaining deniability,” Hemenway did not explicitly state that he thought that the NCVS provides an accurate estimate of DGU frequency, but this is bound to be the meaning that was communicated to some readers by his use of the NCVS results. We can see only two possibilities. Either (1) Hemenway recognized that the DGU estimates derived from the NCVS are grossly inaccurate, but dishonestly presented them to readers as if they were reasonably accurate, or (2) he continued to believe they are fairly accurate, despite his inability to rebut Kleck and Gertz’ case for their inaccuracy, but was unwilling to explicitly commit himself to the accurate-NCVS position. In short, he wanted to have it both ways, using the invalid NCVS estimates to cast doubt on large DGU estimates, while preserving the option of later claiming that he was not naive enough to think the NCVS estimates were even approximately correct.

If Hemenway really did believe that the NCVS estimates are approximately accurate, he may well be the last scholar in this field to cling to this belief. After touting the NCVS estimates of DGU for years, even authors as strongly wedded to the rare-DGU position as Philip Cook (Cook 1991; Cook and Moore 1994) and David McDowall (McDowall and Weirsema 1994) have ceased portraying the NCVS estimates as valid. Instead, they have shifted to the agnostic views that (1) no survey, including the NCVS, can yield meaningful estimates (Cook and Ludwig 1996; 1997) or that (2) “the frequency of firearm self-defense is an issue that is far from settled” (McDowall 1995), views incompatible with the position that the NCVS estimates are at least approximately valid and therefore have settled the matter. By December of 1994, Cook had taken a position directly contradicting Hemenway’s seeming acceptance of the NCVS estimates, stating that there are “persuasive reasons for believing

that the [NCVS] yields total incident figures that are much too low” (Kates et al. 1995, p. 537, quoting a December 20, 1994 letter from Cook). Echoing these views, another strongly pro-control scholar, Tom Smith, has written that “it appears that the [DGU] estimates of the NCVSs are too low” (1997, p. 1462).

Kleck and Gertz provided a detailed explanation of why the NCVS grossly underestimates DGU frequency, and noted that its DGU estimates had been repeatedly disconfirmed by other surveys (1995, pp. 153-157). Still, Hemenway gave the impression that he was using the NCVS estimates as a standard against which he judged the DGU estimates of other surveys (Hemenway 1997b, pp. 1431-1432). In this connection, he falsely claimed that the NCVS asks “about self-defense gun use” (p. 1432) when in fact, as Kleck and Gertz pointed out, one of the many problems with the NCVS as a vehicle for estimating DGU frequency is that it never directly asks respondents about DGU (1995, p. 155). Instead it merely provides respondents with an opportunity to volunteer information about a DGU in response to a general question about self-protection actions.

As Tom Smith, Director of the National Opinion Research Center, has noted, specifically in connection with the NCVS: “Indirect questions that rely on a respondent volunteering a specific element as part of a broad and unfocused inquiry uniformly lead to undercounts of the particular of interest” (Smith 1997, pp. 1462-1463).

Nor did Hemenway acknowledge that the NCVS is the only survey that has ever yielded annual DGU estimates under 700,000, and that its estimates, centering around 80,000, are far below those generated by at least *fifteen* other surveys (Kleck and Gertz 1995, pp. 153-159). Instead, he inverted reality by falsely hinting that it was the NSDS estimate that was the deviant result.

It is tempting to think that the NCVS estimates should be given greater credibility than any one survey because the NCVS has been continuously fielded since 1973, and thus could be regarded as, in some sense, a series of surveys rather than just

one survey, which have provided independent confirmation of low DGU estimates. Hemenway himself, however, noted that “consistency of findings is irrelevant when the methodology among...the surveys is similar.” He made this point with respect to the DGU surveys, but it is far more applicable to the NCVS, since great care has been taken to keep the NCVS, despite periodic revisions, consistent over time. In contrast, there was, contrary to Hemenway’s claims, great diversity in methodology among the DGU surveys (Kleck and Gertz 1995, pp. 157-160).

The NCVS is more accurately viewed as a single ongoing survey, with interviews conducted monthly since 1973 by the same government agency, using methods intentionally kept extremely consistent from 1973 right up to a redesign in 1992. Thus, the flaws that afflicted the NCVS for measuring DGUs in 1973 were, for the most part, still with it in 1992 when the McDowall and Wiersema (1994) and Cook (1991) estimates that Hemenway favorably cited (1997b, p. 1432) were generated. We can only heartily agree with Hemenway that reproducing the same result over and over with the same flawed measurement tool does not provide much evidence about anything. Hemenway just got it wrong as to which surveys this observation is best applied to.

### **10. Fallacious Reasoning—Hemenway’s “Checks on External Validity”**

In their original article, Kleck and Gertz cautioned against two kinds of fallacious reasoning. Instead of taking the warnings seriously, Hemenway seems to have treated them as signposts to deceptive arguments that might prove useful for propagandistic purposes. Both fallacious arguments involve a misapplication of *reductio ad absurdum* argumentation, based on the misperception that estimates from the NSDS were inconsistent with known crime counts and the erroneous assumption that the NCVS provides correct estimates of the absolute frequency of crime.

Hemenway argued that the NSDS estimates are implausible because this survey implied a number of DGUs occurring in connection with burglaries that exceeded the total number of burglaries of occupied residences estimated by the NCVS, and thus the DGU estimate was impossible, or at least implausibly high (p. 1441). This argument rested on an unstated assumption that the universe of DGU events sampled by the NSDS is a subset of the universe of crime events covered by the NCVS. However, Kleck and Gertz had explicitly warned in their paper that “a large share of the incidents covered by our survey are probably outside the scope of incidents that realistically are likely to be reported to the NCVS or police” (1995, p. 167). This is true because DGUs typically involve criminal behavior, such as unlawful gun possession, by the gun-using victim, who therefore is often unwilling to report the incident. Once it is recognized that many DGU events are outside the realm of crime incidents effectively covered by the NCVS, it is logically impossible to treat any NCVS estimates as imposing an upper limit on how many DGUs there plausibly could be.

Hemenway’s logic was also fallacious in assuming that one can cast doubt on conclusions based on a large body of data by deriving implausible implications from smaller subsets of the data. The NSDS estimates of total DGUs are likely to be fairly reliable partly because they are based on a very large ( $n=4,977$ ) sample, while any estimates one might derive pertaining to one specific crime type are necessarily less reliable because they rely partly on a far smaller subsample, i.e. the c. 194 sample DGU cases, of which 40 were linked to burglaries.

Hemenway’s *reductio ad absurdum* logic is equivalent to arguing that Gallup presidential election polls cannot accurately estimate the share of the entire electorate voting for the Democratic candidate (something we know they *can* do, usually to within two percentage points—Gallup 1992) because they commonly yield implausible estimates for small subsets of the electorate, such as rural Hispanic Jews. One undoubtedly could obtain implausible estimates of voter preference for the

Democratic candidate, such as 0% or 100%, based on a very small number of sample cases, for many subsets of the population. This would imply nothing, however, about the ability of the survey to estimate voter preferences in the entire population. Thus, even if estimates of DGUs linked to a given specific crime type were implausible, which they are not, this would imply nothing about whether estimates of the total number of DGUs, based on the full sample, are accurate.

Finally, even if one ignored these logical fallacies, Hemenway's argument still would fail, because it depends on an indisputably erroneous assumption. Hemenway stated that "from the NCVS, we know that there were fewer than 6 million burglaries in 1992" (1997b, p. 1441), and made similar statements about rapes (p. 1442). In fact, we do not "know" any such thing. No competent criminologist believes that the NCVS provides complete coverage of all burglaries, or any other crimes, occurring in the U.S. And once one concedes that there may be far more crimes than the NCVS estimates, Hemenway's argument collapses, since it becomes impossible to argue that estimates of the number of DGUs linked to a given type of crime are implausibly high relative to the total number of crimes of that type—we simply do not know the latter number.

In a second variety of this fallacious line of reasoning, Hemenway cited estimates of the number of gunshot wound (GSW) victims treated in emergency rooms and falsely claimed that "K-G report that 207,000 times per year the gun defender thought he wounded or killed the offender" (1997b, p. 1442). In fact, Kleck and Gertz did not compute or report this 207,000 estimate. Quite the contrary—they specifically cautioned against using NSDS data to generate such an estimate because an estimate of defensive woundings would be based (unlike the estimates of DGU frequency in general) on a small sample (the approximately 200 respondents who reported a DGU) and because NSDS interviewers had done no detailed questioning of

respondents regarding why they thought that they had wounded their adversaries.

In any case, there is nothing even mildly inconsistent about this GSW estimate and emergency room data on persons treated for GSWs. Hemenway necessarily made the implicit assumption that DGU-linked woundings are entirely a subset of woundings treated in medical facilities. If one more plausibly assumes that most less serious DGU-linked woundings are not medically treated, the number of medically treated GSWs cannot be used as an upper limit on the number of DGUs that result in a wounding, since DGU-linked woundings would exist largely outside the set of medically treated GSWs. If, for example, the total annual number of GSWs, treated or untreated, was 400,000, there would be nothing implausible about 200,000 of them being DGU-linked, especially in light of the fact that the vast majority of victims of medically treated GSWs linked to alleged “assaults” are known criminals (Kleck 1997, Chapter 1).

It is unlikely that a criminal wounded by a victim during the commission of a crime would seek medical attention for any but the most life-threatening GSWs, since medical personnel are required by law to report treatment of GSWs to the police. Less than a tenth of assault GSWs are life-threatening (Kleck 1997, Chapter 1). Thus, almost all of the DGU-linked woundings of criminals probably lie outside the universe of GSWs treated in emergency rooms and other medical facilities. The number of medically treated GSWs therefore cannot serve as an upper limit on either the total number of GSWs or on the number that occur in connection with a crime victim’s DGU. In sum, since we do not know the total number of crime victimizations such as rapes or burglaries, or the total number of GSWs, we cannot possibly know if any given DGU estimate is implausibly large relative to these unknown (and possibly unknowable) quantities.

It is worth stressing that crucial logical fallacies in Hemenway’s *reductio ad absurdum* arguments were explicitly noted in the original 1995 article by Kleck and Gertz (pp. 167-168, 172-174), before Hemenway presented them. Thus, because

Kleck and Gertz had explicitly warned against making the very arguments that Hemenway would later make, Hemenway was clearly aware of the fatal flaws in his arguments. Since he did not rebut any of the arguments that Kleck and Gertz used to conclude that this line of reasoning was fallacious, it is reasonable to conclude that when Hemenway made his *reductio ad absurdum* arguments, he knew they were fallacious. Thus, his use of these arguments can be reasonably viewed as part of an intentional effort by Hemenway to deceive his readers, and not merely the product of sloppy thinking.

### 11. The UFO Analogy

Perhaps the most bizarre part of Hemenway's paper was the analogy he drew between survey reports of DGUs and reports of contacts with aliens from other planets. Hemenway noted that 10% of respondents in a Gallup survey told interviewers that they had seen an alien spacecraft. Here too Hemenway was dealing in a red herring. No one disputes that some behaviors or experiences can be greatly overestimated in surveys. Rather, the relevant issue is whether DGU happens to be one of those experiences. The extent and kinds of response errors in surveys are heavily dependent on subject matter, so that extent of misestimation with respect to one topic cannot cast any light on the likely degree of error in misestimating another topic unless the topics are very similar.

We assume that most of the 10% of respondents in the UFO survey who responded affirmatively to the spacecraft question were having a little fun with the interviewers, though a few may well have been serious. On the other hand, it is harder to believe that respondents would regard questions about crime victimization and DGUs in so frivolous a light. In addition, Hemenway's analogy ignored the fact that all it took to be counted as an alien spacecraft spotter was the one-syllable response "Yes," while it took as many as 19 logically consistent responses providing details about the incident to be counted as a defensive gun user.

## 12. The Positive Social Bias Speculation

Hemenway did not deny or rebut the claim that most of the DGUs reported in the NSDS involved illegal behavior on the part of the respondents (Kleck and Gertz 1995, pp. 155, 171-174). Instead, he simply ignored it, perhaps because he recognized that it would be difficult to persuade readers that survey respondents are biased in favor of *over*reporting their own unlawful behavior. He insisted that the predominant bias surrounding DGU reports is a “social desirability bias,” with respondents making false reports of DGUs to present themselves as “heroic” (Hemenway 1997b, p. 1431).

He ignored the information that Kleck and Gertz provided in their article on the distinctly unheroic character of the DGU accounts provided. What was most striking about the reported events was their banality. If Hemenway’s speculations had merit, false portrayals of heroism should have involved frequent claims of facing down gun-wielding bad guys and exciting shootouts. In fact, respondents reporting DGUs claimed to have faced adversaries with guns in only one in six cases, claimed involvement in a shootout (both parties shooting) in just 3% of the cases, and usually reported opponents with no weapons at all. Likewise, they rarely boasted about their deadly shooting, with only 8% even claiming to have wounded an adversary (Kleck and Gertz 1995, pp. 173, 175).

The more pertinent issue, however, is not how respondents regarded their own actions, but rather how they thought interviewers were likely to regard their actions. Regardless of how respondents may have viewed their alleged DGUs, they would not be likely to falsely report imaginary DGUs or to mischaracterize events as DGUs if they thought that interviewers were inclined to view alleged DGUs in a negative light, and possibly as criminal behavior.

Hemenway offers no reasons why respondents would think interviewers would have favorable views of such actions. All the respondents knew about the interviewers, besides their sex (mostly female), was that they were calling from Florida State

University, and thus were presumably working for college professors, as indeed they were. Thus, respondents who thought about the matter at all were likely to think they were providing information for people generally regarded as liberal intellectuals, hardly the sorts of people likely to provide a sympathetic reception for accounts of DGUs, whether genuine or false. Consequently, there is little logical reason to expect a social desirability bias to operate with many respondents.

In any case, the one-sided focus on social desirability is itself a red herring. The key issue is not whether some respondents might think DGUs are heroic (this is undoubtedly true for at least some people), but rather whether this sentiment is so strong and pervasive that it would, on net, outweigh the seemingly more common and natural tendency to conceal one's illegal behaviors from strangers who call on the phone. By addressing only the social desirability of reporting heroic acts, Hemenway distracted readers from the issue of the relative balance of sources of response errors. He provided no evidence or even argumentation as to why any social desirability effects should outweigh simple concerns about revealing one's unlawful behaviors.

Hemenway did not deny Kleck and Gertz' claim that most DGUs do involve illegal behavior, though he did his best to distract readers' attention from this fact, e.g. by stating that "self-report surveys tend to overestimate rare events which *carry no social stigma*" (Hemenway 1997b, p. 1435). Since when does criminal behavior carry no stigma? If it *does* carry a stigma, and if most DGUs *do* involve criminal behavior, then it something of a puzzle how Hemenway reached the conclusion that not only is there, on net, a positive social desirability bias to reporting DGUs, but that it is clear and obvious that there is such a bias.

### **13. Making Something Out of Nothing—Hemenway's Numerical Exercises**

It would be understandable if some readers of Hemenway's article believed that he did present, in his Section V, evidence on

the relative balance of false positives and false negatives. In fact, this section presented no empirical evidence at all. Instead, Hemenway's numerical examples demonstrated nothing more than that if one arbitrarily *assumes* particular rates of false positives and false negatives, along with extremely low actual DGU rates, one can support the claim that DGU could be greatly overestimated. Hemenway cannot be faulted for his arithmetic. If there were any credibility to the misreporting rates that he assumed out of thin air, they would indeed imply huge overestimates.

Hemenway's argument was fallacious because it was circular—it required that he assume the very conclusions he was trying to support. Specifically, Hemenway assumed as starting points of his exercise that (1) there is a nonnegligible rate of reporting false positives, and (2) DGUs are in fact extremely rare. He stated that “with few actual positives [i.e. few genuine DGUs], it is impossible for a screen to pick up many false negatives,” and that “it follows that, for events with low incidence ... the estimated incidence will tend to be greater than the true incidence” (p. 1436).

All one can validly conclude from this exercise is that there is more *potential* for false positives than false negatives, i.e. that there hypothetically *could* be more false positives than false negatives. Of course, this banal point would apply to estimation of literally any trait that characterized less than half of the population. The problem is that Hemenway did not present any empirical evidence that there were *any* false positives among the cases that Kleck and Gertz treated as DGUs, nor among those so treated in other DGU surveys, never mind the large numbers he assumed.

Whether there actually *are* more false positives than false negatives in surveys of DGU or other crime-related experiences is an issue to which Hemenway never brought any empirical evidence to bear, as distinct from speculations and assumptions. Rather, he jumped from the fact that this potential exists to the non sequitur conclusion that “you *inevitably* [emphasis added]

get a large number of false positives relative to the number of true positives” (p. 1437) and thus an overestimate.

Instead of citing relevant empirical evidence, Hemenway argued indirectly by analogy. Drawing a strained analogy between reporting of diseases in surveys and reporting of illegal behavior like DGUs, he quoted epidemiologists who stated that “if the population is at low risk for having the disease, results that are positive will mostly be false positives” (p. 1436). While that may well be true about reporting of diseases, direct empirical evidence (to be discussed in a later section) indicates that it is clearly not true about the reporting of rare illegal behaviors.

No survey respondent believes that they will be arrested for falsely reporting a disease they do not have, and for most diseases few respondents would expect interviewers to have negative views of the respondent’s health problems. In contrast, much of criminological survey research has been organized around the problem that many respondents do believe they could suffer arrest, or at least embarrassment and other negative consequences, if they reported having committed illegal acts (Hardt and Peterson-Hardt 1977; Hindelang et al. 1981; Kleck 1982). While falsely reporting a disease would typically elicit sympathy, falsely reporting illegal behavior would rarely do so. Observations about the relative frequency of false positives and false negatives in surveys of disease simply have no bearing on the issue at hand.

Note also that, even with respect to diseases, Hemenway was unable to locate *any* examples of overestimating prevalence by a factor of 30, which is what one would have to believe the NSDS did, if one accepts the NCVS estimate of DGU frequency as accurate.

Hemenway’s claim that the NSDS results were “extremely sensitive” to small changes in the specificity rate (the percent of true negatives accurately detected) also relies on assuming the conclusion. The main reason that the example estimates he computed (see his Tables 2A-2C) were so sensitive to the

specificity rate is because Hemenway assumed extremely low actual DGU rates, i.e. he assumed the very conclusion he was trying to support. Thus, instead of using the empirically-based 1.33% estimate Kleck and Gertz obtained, Hemenway assumed imaginary DGU rates of 0.32%, 0.04% and 0.08%, respectively (in his Tables 2(A), 2(B), and 2(C)) (pp. 1444-1445). Because he arbitrarily assumed that there are so few true positives (genuine DGUs), even a handful of false positives could indeed outnumber them and substantially distort the estimates.

For example, in his Table 2(B), the main reason Hemenway's assumed rate of false positives of 1.3% had such a proportionally large distorting effect on the estimate was because he assumed, without any empirical foundation, that the actual DGU prevalence rate was virtually zero, so that just 64 false positives could be 32 times higher than his assumed number of just two (!) true positives, in a sample of 5,000 cases (p. 1445). For what it's worth, the estimates *would* be highly sensitive to the specificity rate, *if* the true DGU rate were as low as Hemenway assumed, but then it is the DGU rate that is at issue.

In our view, a more realistic version of Hemenway's hypothetical scenarios, one more in tune with research on errors in surveys of illegal behavior, might have 48 true positives, 48 false negatives (and thus 96 persons with a genuine DGU), 18 false positives, and 4,886 true negatives in a sample of 5,000 cases, implying 50% test sensitivity (the percent of true positives accurately detected) and 99.6% test specificity. Under this alternative set of hypothetical assumptions, the true DGU prevalence would be 1.92%, while the measured rate would be 1.32%, as was obtained in the NSDS, implying that the true DGU rate was actually 45% higher than the one estimated.

Of course, the question remains, which is the more plausible set of assumptions about the distribution of survey response errors—Hemenway's or ours? Unlike Hemenway, who relied on assumed numbers and strained analogies to the reporting of diseases, we prefer to rely on actual empirical evidence directly

addressing the relative prevalence of different kinds of response error in previous surveys of illegal behavior.<sup>3</sup>

#### **14. Prior Research on the Validity of Survey Estimates of Illegal Behavior**

Hemenway provided a discussion of “misclassification in surveys generally” (pp. 1434-1437) whose most notable feature was its utter silence about surveys concerning illegal behavior and crime-related experiences. While Hemenway cited surveys about height, automobile ownership, diseases, and other topics of negligible similarity to the topic at hand, he said nothing about evidence concerning the validity of responses to questions requiring respondents to report their own illegal behavior. Surely surveys of unlawful and crime-related behaviors are more pertinent to the validity of DGU survey estimates than the surveys Hemenway addressed. We will correct this conspicuous omission.

A large body of empirical evidence indicates that, when asked questions about their own illegal behavior, survey respondents, on net, underreport their involvement, and that false negatives outnumber false positives by a wide margin. The strongest tests of validity on such questions concern illicit drug use. Unlike with other illegal behaviors, there is a strong external criterion that analysts can use to judge the validity of survey self-reports concerning drug use, because consumption of illicit drugs leaves physical traces that can be reliably detected using physiological means such as urine tests and hair assays. Further, illicit drug use may be the only illegal behavior for which validity checks can effectively detect false positives as well as false negatives.

Research using improved chemical tests has repeatedly demonstrated that respondents self-report less drug use in interviews and on questionnaires than is later revealed by hair or urine analysis, even when interviewed under conditions of anonymity and confidentiality (Amsel et al. 1976; Cisin and Parry 1980; Magura et al. 1987; Wish 1987; Baumgartner et al. 1990;

Dembo et al. 1990; Wish and Gropper 1990; Mieczkowski 1990; Mieczkowski et al. 1991, p. 246; Falck et al. 1992; Magura et al. 1992; McNagny and Parker 1992; Feucht et al. 1994; Hindin et al. 1994; Cook et al. 1995; Hoffman et al. 1995; Magura et al. 1995; see Wish et al. 1995 for a general review).

For example, among patients at a walk-in clinic who had positive urine tests for illicit drug use, only 28% had admitted the use in interviews (McNagny and Parker 1992), i.e. actual use was 3.6 times higher ( $100/28=3.6$ ) than reported use. Among a group of juvenile arrestees, while hair analysis indicated 56.8% had used cocaine, only 7.4% self-reported it in interviews (Feucht et al. 1994), implying that actual use levels were 7.7 times higher than self-reports indicated. In a group of youthful jail releasees, while 67% tested positive for cocaine with hair analysis, only 23% self-reported cocaine use in the preceding 90 days, and only 36% reported ever using it (Magura et al. 1995). Among employees of a manufacturing plant, actual drug use prevalence as measured by hair and urine analysis, was 50% higher than the estimate produced by self-reports (Cook et al. 1996).

Some studies separately reported numbers of false positives and false negatives. Among a group of 114 arrestees, 85 of whom later tested positive for cocaine use on hair analysis, 61 falsely denied use in interviews (false negatives), while *none* reported use but tested negative (false positives) (Mieczkowski et al. 1991, p. 246). Likewise, among 86 subjects studied by Baumgartner et al. (1990), there were 16 who falsely denied cocaine use by self-report, but only *one* who reported drug use without a hair assay confirming it, again indicating false negatives are common and false positives close to nonexistent.

These examples could be multiplied, but to no purpose. The evidence is clear that people are far more likely to fail to report illegal behavior in which they have engaged than they are to falsely report illegal behaviors in which they have not engaged, and that self-report surveys therefore *underestimate* illegal behavior. To use Hemenway's epidemiological terms, while "test specificity" probably approaches 100% (i.e. extremely few false

positives), “test sensitivity” is probably less than 50% (i.e. many false negatives).

It is unfortunate there is no way to estimate false positives and false negatives as authoritatively with DGUs as with illicit drug use. We are forced to make do with validity checks on surveys addressing other experiences analogous to DGU. While this is less than ideal, it cannot be seriously argued that surveys of disease, health care, height, weight, and similar topics discussed by Hemenway are as analogous to surveys of DGU as surveys of illegal behavior or crime-related experiences.

### **15. Libeling the NSDS Interviewers**

The interviewers who worked on the NSDS were named individually at the beginning of Kleck and Gertz’ article (1995, p. 150). Without any evidence, Hemenway hinted that these individuals acted unethically, by distorting or inventing responses. In discussing an alleged “limitation” of the NSDS, Hemenway wrote: “the survey was conducted by a small firm run by Professor Gertz. The interviewers knew both the purpose of the survey and the staked-out position of the principal investigator regarding the expected results” (Hemenway 1997b, p. 1433). The unmistakable insinuation was that some of the interviewers faked or altered interviews to create phony accounts of “DGUs” that would please the principle investigator.

To our knowledge, none of the interviewers knew anything about Kleck’s views on DGU or what results he expected, since Kleck did not inform them of those views. Hemenway did not claim to have communicated with even one of the interviewers, to find out what they knew prior to interviewing. Therefore, he had no basis whatsoever for this outrageous charge. It was apparently sufficient for Hemenway that the interviewers *could* have done such a thing in order to publicly hint that they did.

An interviewer obviously could not accidentally or innocently record an entire false account of a DGU, with as many as 19 logically consistent responses; a single errant mark on an answer sheet would not generate a false positive. Furthermore, as Kleck

and Gertz stated in their article, every single interview in which a DGU was alleged was validated by a call-back by a supervisor (Kleck and Gertz 1995, p. 161). An interviewer-faked incident therefore could not have survived the quality control procedures unless a supervisor colluded. Such a thing could only be accomplished intentionally. How, then, could readers have interpreted Hemenway's remarks except to the effect that he was suggesting that the interviewers were intentionally recording nonexistent interviews, inventing DGUs, or otherwise knowingly distorting responses?

It was reprehensible that Hemenway recklessly impugned the integrity and honesty of these individuals without any facts to support his allegations. His insinuations were irresponsible and offensive. Hemenway owes the NSDS interviewers and supervisors a public apology. It is no defense that he recklessly smeared a set of 14 interviewers as a group, rather than one particular individual. This passage was not only offensive, but diagnostic of the attitude underlying Hemenway's entire critique, i.e. a willingness to write almost anything that might advance his political agenda.

It is worth mentioning in this connection that a colleague of Hemenway's, Deborah Azrael (e.g., see Hemenway, Solnick and Azrael 1995), separately contacted both Kleck and Gertz while Hemenway was preparing his critique, without, however, telling either of them that she was doing it at Hemenway's behest. She contacted Kleck under the guise of setting up his participation in a planned "conference" on guns and violence to be hosted by the Harvard School of Public Health. No such conference was held. In the course of several hours of conversation with Azrael, however, Kleck interpreted the general thrust of her questions to be a "probing" for weaknesses in the NSDS. A major theme of her conversation with Gertz was the search for something ethically dubious in the funding of the research. In short, it seemed to both Kleck and Gertz that Hemenway's colleague was "digging for dirt" at Hemenway's behest.

## 16. The Survey Hemenway Chose Not to Mention

The NSDS estimates were subsequently strongly confirmed by yet another large-sample national survey, sponsored by the National Institute of Justice (NIJ), and conducted under the auspices of the Police Foundation. We can be certain that Hemenway knew about this survey because he served on the NIJ Advisory Committee for the project and was thanked for his comments on a draft of the grant report describing the survey's findings, including its DGU estimates (Cook and Ludwig 1997, p. x). Kleck was the principle consultant on the Police Foundation survey, wrote most of the associated grant proposal and most of the questionnaire, and participated in numerous meetings with Hemenway and Cook.

Hemenway did not mention the results of this survey in his critique, perhaps for an understandable reason: it almost exactly confirmed the NSDS results. The NSDS yielded an estimate of 2.55 million DGUs, using a person-based one-year estimate (Kleck and Gertz 1995, p. 184). The most comparable estimate generated by the Police Foundation survey was 2.45 million, well within sampling error of the NSDS estimate. Many variants of this estimate were even higher (Cook and Ludwig 1997, p. 62).

Hemenway himself had ample opportunity, as a member of the Advisory Committee, to suggest solutions to problems he saw in this survey, or to suggest other steps "to reduce the bias or to validate the findings by external measures," and to show that DGUs are really far less common than so many surveys have indicated. When the Police Foundation survey almost exactly confirmed the NSDS results, Hemenway's response was to suddenly decide that surveys inevitably overstate DGU frequency.

This appears to be a very recent revelation to Hemenway. In repeated and prolonged meetings of the Advisory Committee in 1994, during which the members discussed at length the long series of questions asking about DGUs, Hemenway did not once share his remarkable theory that all that effort was for naught,

and that surveys could not generate even approximately accurate estimates of DGU frequency.

Philip Cook, who also served on the same committee, likewise underwent the same sudden conversion, after the Police Foundation survey yielded DGU estimates every bit as large as those of the NSDS and earlier surveys. Since no new evidence bearing on the ability of surveys to estimate this parameter had come to light since 1994, one can only wonder how and why these revelations came so belatedly to Cook and Hemenway. Cynics might suspect that, metaphorically speaking, once they found they could not win the game, they decided to take their ball and go home.

It is instructive to consider the conspicuously one-sided implications that Hemenway and Cook have derived from their novel theory that surveys are likely to overestimate rare phenomena. Neither of them has acknowledged that one obvious implication is that the National Crime Victimization Survey is likely to overestimate the frequency with which gun crimes are committed, and thus overstate the harm done with firearms.

Most of the Hemenway-Cook arguments for DGU overestimation in surveys (excepting the minor argument concerning telescoping) apply with at least equal force to surveys estimating the frequency of serious crimes, including gun crimes, since such events are also, in absolute terms, quite rare, regardless of whether one accepts evidence indicating that gun crimes are more rare than DGUs.

It is a mildly amusing pastime to go through articles by Hemenway and Cook that push this theory (e.g. Cook, Ludwig and Hemenway 1997, pp. 465-467; Hemenway 1997a; Hemenway 1997b, pp. 1435-1437) and simply substitute “gun crime” for DGU to see how neatly the same theory could be used to argue for survey overestimation of gun crime.

Hemenway and Cook seem to have either missed this implication, or have not chosen to share it with their readers. If Hemenway honestly believed that surveys are likely to overestimate rare phenomena, he would be chastising his friends

at HCI and CPHV for citing NCVS estimates that overstate the frequency of gun crime.

More likely, Hemenway will soon be developing a specialized ad hoc explanation of why his theory applies only to estimates of beneficial uses of guns but not to estimates of harmful uses. It should be stressed that we are not arguing that surveys overestimate gun crime. Rather, surveys almost certainly underestimate both defensive and criminal uses of guns (Kleck and Gertz 1995, pp. 170-171).

In light of Hemenway's claim that "all checks for external validity of the Kleck-Gertz finding confirm that their estimate is highly exaggerated" (Hemenway 1997b, p. 1431), it is hard to see how one could justify Hemenway's calculated decision to withhold from his readers the results of the Police Foundation survey, when it almost exactly confirmed the NSDS estimates, and thus constituted about as strong an external validity check as one could ask for.

It is doubtful whether any evidence or reasoning will ever dissuade Hemenway from his remarkable theory that all surveys are likely to overestimate rare events, so he presumably would justify his decision to not mention the Police Foundation survey by asserting that all surveys are now irrelevant to the issue. But even if one accepted this radical view, the results of the Police Foundation project at minimum established that all Hemenway's speculations about supposed flaws specifically afflicting Kleck and Gertz' NSDS (Hemenway 1997b, pp. 1433-1444) cannot account for their large DGU estimates, since the Police Foundation survey yielded estimates almost identical to those of the NSDS.

This raises the question: what was the point of all of Hemenway's unsupported speculations about flaws supposedly afflicting the NSDS in particular, if he knew that they could not account for the NSDS estimates being as high as they were? Perhaps they were presented in the hope that less rigorous readers would assume that, methodologically speaking, where

there's smoke, there must be fire. Pile on enough criticisms, and readers will assume that at least a few of them must be valid.

Perhaps the only thing more appalling than Hemenway's dishonest ideological diatribe was that fact that a respectable professional journal, the *Journal of Criminal Law and Criminology*, decided to publish it. Its Criminology Editor, John Hagan, attributed his decision to publish the paper to the fact that two or three outside reviewers recommended publication. This was an evasion of editorial responsibility, since all that it takes for an editor to get such recommendations is to select reviewers with strong published views consistent with the author's thesis who are willing to overlook its dishonest tactics, one-sidedness, speculative character, and complete lack of supporting evidence.

In this case, the obvious candidates would be any of the large number of strongly pro-control members of the journal's Criminology Advisory Board (there are at least eleven of them, listed in Section 2 of this article), or others who have also indulged in one-sided speculation on this issue, such as Philip Cook, David McDowall, Albert Reiss, Jeffrey Roth, Steven Messner, Franklin Zimring, and so on.

After Kleck and Gertz supplied Hagan with a long series of documented instances of deceptive claims, red herrings, and inaccuracies in the Hemenway paper, Hagan did not dispute their claims. Instead, he claimed that publishing Hemenway's paper would somehow "contribute" to the gun control debate. To suggest that publishing a long series of falsehoods, inaccuracies, red herrings, irrelevancies, libelous insinuations, and personal ideology disguised as scholarly criticism somehow "contributes" to the scholarly debate over gun use is both bizarre and offensive to the community of scholars who play by the rules and who do not indulge in one-sided speculation as a substitute for even-handed, intelligent assessment of existing evidence and for doing the hard work of getting *better* empirical evidence. Intellectually debased argumentation only muddies the waters and makes the already difficult task of assessing the evidence even more difficult.

## 17. Conclusions—The Political Functions of the DGU Critiques

Hemenway and like-minded critics have failed to cast even mild doubt on the accuracy of the NSDS estimates and other high estimates of DGU frequency. Leaving aside problems with the DGU surveys already noted in the Kleck-Gertz article, the critics' claims have been effectively rebutted. The conclusion that there are large numbers of defensive uses of guns each year in the United States has been repeatedly confirmed, and remains one of the most consistently supported assertions in the gun-violence research area.

Given the political purposes of the critics, however, it is inconsequential that all of their claims have been rebutted. Although it is easy enough to rebut each of Hemenway's claims, the political functions of a piece like this one were served the instant it was published. Even if a "critique" is completely devoid of serious intellectual content, and each of its points are thoroughly refuted in the pages of the publishing journal, once the piece appears in print in a respectable journal, propagandists can cite the publication, either in propaganda tracts or in interviews with reporters, as evidence that "surveys indicating large DGUs have been discredited."

Indeed, this is precisely how the Hemenway piece has already been cited, before it was even published. In a letter to the *Journal of the American Medical Association*, three public health gun control advocates stated that "the reasons that this survey [the NSDS] is incapable of yielding an accurate estimate of defensive gun use are described at length in the Hemenway article" (Vernick, Teret, and Webster 1997, p. 703). Apparently a series of unsupported and one-sided speculations was a sound enough basis for these individuals to reject the findings of at least 15 large-scale, professionally conducted surveys.

We can be confident that ideologues and fanatics will in future cite these one-sided speculations as authoritative proof that large DGU estimates have been "discredited," while pro-control

academics who fancy themselves moderates will conclude that while Hemenway and others like him may have been wrong on some points, they had nevertheless somehow “cast doubt” on the estimates or “raised serious questions” about them.

The critiques can be cited by gun control advocates, pro-control scholars, and reporters alike in good conscience, as part of a “balanced” presentation of the issue. Hemenway’s outrageous and unsupported speculations will be cited in scholarly sources alongside the NSDS estimates, implicitly giving equal weight to careful empirically based estimates and the one-sided speculations of a pro-control extremist. The fact that the balance is completely spurious, and that only one side of the debate can present credible supportive empirical evidence, is politically irrelevant. Since it is highly unlikely that either reporters or the rest of the audience for propaganda will bother to read a rebuttal, the complete lack of any intellectual merit to the DGU critiques will not be evident, and thus will not in any way reduce its political utility.<sup>4</sup>

Thus, critiques of the DGU surveys effectively serve a political, propagandistic function regardless of how one-sided, illogical, intellectually hollow and devoid of empirical support they may be. The critiques can be cited by those who are unwilling to accept the verdict of empirical evidence, providing a fig leaf of respectability to what is basically a political position, that DGUs cannot, and *must* not, be frequent. Left unmentioned will be one simple fact. In all of the critiques, critics did not once cite the only thing that really could legitimately cast doubt on the large DGU estimates: better empirical evidence.

## NOTES

1. For example, when I wrote a brief Letter to the Editor to the *American Journal of Public Health* to point out the journal had published a seriously inaccurate estimate (to the low side) of DGU frequency (McDowall and Wiersema 1994), the editors refused to publish the letter.

2. The claim that the NSDS estimate of household gun prevalence was “outside the range of all other national surveys” (Hemenway 1997b, p. 1434) was, however, false. The NSDS 38% figure was one of three U.S. household gun prevalence figures in the 37-38% range, and one of *eight* in the 37-42% range during 1993-1996, i.e. within sampling error of each other (Kleck 1997, Ch. 3). This falsehood crudely served to present the NSDS results as erratic or deviant, and the survey methods as eccentric.
3. Oddly enough, in his rendition of extreme estimates in surveys covering a wide variety of phenomena, it did not occur to Hemenway to mention his survey with Weil (Weil and Hemenway 1993a) in which he overestimated NRA membership by a factor of three (see Kleck 1993).
4. A Washington Post reporter, Bob Thompson, brought up the critiques of the DGU surveys in interviews with me, and when I offered to send him my written rebuttal of the critiques, he explicitly told me that he was not interested in reading it.

## REFERENCES

- Amsel, Z., D. Mandell, and C. Matthias. 1976. “Reliability and validity of self-reported illegal activities and drug use collected from narcotics addicts.” *International Journal of the Addictions* 11:325-336.
- Anderson, Ronald, Judith Kasper, Martin R. Frankel, and associates. 1979. *Total Survey Error: Applications to Improve Health Surveys*. San Francisco: Jossey-Bass.
- Baumgartner, W.A., J.D. Baer, V.A. Hill, and W.H. Blahd. 1990. “Hair analysis for the detection of drug use in pretrial/probation/parole populations, in *Summary Report to the National Institute of Justice*, pp. 1-18. Summarized in Hindin et al. 1994, p. 775.
- Cantor, David. 1989. “Substantive implications of longitudinal design features.” In *Panel Studies*, edited by Daniel Kasprzk et al. N.Y.: Wiley.
- Cisin, I.H., and H. L. Parry et al. 1971. “Sensitivity of survey techniques in measuring illicit drug use.” Pp. 3-46 in *Developmental Papers: Attempts to Improve the Measurement of Heroin in the National Survey*, edited by J.D. Rittenhouse. Washington, D.C.: U.S. Government Printing Office.
- Cook 1986. “The relationship between victim resistance and injury in noncommercial robbery.” *Journal of Legal Studies* 15:405-416.

- Cook, Philip. 1991. "The technology of personal violence." Pp. 1-71 in *Crime and Justice*, volume 14, edited by Michael Tonry. Chicago: University of Chicago Press.
- Cook, Philip, and Jens Ludwig. 1996. "You got me: how many defensive gun uses per year?" Paper presented at the annual meeting of the American Society of Criminology in Chicago, Illinois.
- Cook, Philip, and Jens Ludwig. 1997. *Guns in America*. Report to the Police Foundation on the National Survey of the Private Ownership of Firearms. Washington, D.C.: Police Foundation.
- Cook, Philip, Jens Ludwig, and David Hemenway. 1997. "The gun debate's new mythical number: how many defensive uses per year?" *Journal of Policy Analysis and Management* 16:463-469.
- Cook, Philip, and Mark C. Moore. 1994. "Gun control." Pp. 267-294, 566-571 in *Crime*, edited by James Q. Wilson and Joan Petersilia. San Francisco: Institute for Contemporary Studies.
- Cook, Royer F., Alan D. Bernstein, Thadeus L. Arrington, Christine M. Andrews, and Gordon A. Marshall. 1995. "Methods for assessing drug use prevalence in the workplace: a comparison of self-report, urinalysis and hair analysis." *International Journal of the Addictions* 30:403-426.
- Dembo, R., L. Williams, Eric D. Wish, and J. Schmeidler. 1990. "Urine testing of detained juveniles to identify high-risk youth." National Institute of Justice Research in Brief. Washington, D.C.: National Institute of Justice.
- Dodge, Richard. 1970. "The Washington, D.C. recall study." Reprinted on pp. 12-15 in *The National Crime Survey: Working Papers, Volume I: Current and Historical Perspectives*, edited by Robert G. Lehnen and Wesley G. Skogan. U.S. Department of Justice, Bureau of Justice Statistics. Washington, D.C.: U.S. Government Printing Office.
- Falck, R., H. A. Siegel, M.A. Forney, J. Wang, and R.G. Carlson. 1992. "The validity of injection drug users' self-reported use of opiates and cocaine." *Journal of Drug Issues* 22:823-832.
- Feucht, Thomas E., Richard C. Stephens, Michael L. Walker. 1994. "Drug use among juvenile arrestees: a comparison of self-report, urinalysis and hair assay." *Journal of Drug Issues* 24:99-116.
- Fotis, 1996. Letter to the Editor. *Journal of the American Medical Association* 275:281.

Gallup. 1992. "Gallup poll accuracy record." *The Gallup Poll Monthly* 326:33.

Hardt, Robert H., and Sandra Peterson-Hardt. 1977. "On determining the quality of the delinquency self-report method." *Journal of Research in Crime and Delinquency* 14:247-261.

Hemenway, David. 1997a. "The myth of millions of annual self-defense gun uses: a case study of survey overestimates of rare events." *Chance* 10:6-10.

Hemenway, David. 1997b. "Survey research and self-defense gun use: an explanation of extreme overestimates." *Journal of Criminal Law and Criminology* 87:1430-1445.

Hemenway, David, and Elizabeth Richardson. 1997. "Characteristics of automatic or semiautomatic firearm ownership in the United States." *American Journal of Public Health* 87:286-288.

Hemenway, David, Sara J. Solnick, and Deborah R. Azrael. 1995. "Firearm training and storage." *Journal of the American Medical Association* 273:48-50.

Hemenway, David, and Douglas Weil. 1990a. "Phasers on stun: the case for less lethal weapons." *Journal of Policy Analysis and Management* 9:94-98.

Hemenway, David, and Douglas S. Weil. 1990b. "Less Lethal Weapons" (Op-Ed). *Washington Post* May 14, 1990.

Henigan, Dennis A., David Hemenway, and E. Bruce Nicholson. 1995. *Guns and the Constitution: The Myth of the Second Amendment*. Altheia Press.

Hindelang, Michael J., Travis Hirschi, and Joseph G. Weis. 1981. *Measuring Delinquency*. Beverly Hills: Sage.

Hindin, Rita, Jane McCusker, Maureen Vickers-Lahti, Carol Bigelow, Frances Garfield, and Benjamin Lewis. 1994. "Radioimmunoassay of hair for determination of cocaine, heroin, and marijuana exposure: comparison with self-report." *International Journal of the Addictions* 29:771-789.

Hoffman, J.A., E.D. Wish, J.J. Koman III, S.J. Schneider, P.M. Flynn, and J.W. Luckey. 1995. "Self-reported drug use compared with hair analysis and urinalysis." *Problems of Drug Dependence, 1994*, Volume II: Abstracts.

National Institute of Drug Abuse Research Monograph 153. Rockville, Md.: NIDA.

Ikeda, Rbin M., Linda L. Dahlberg, Jeffrey J. Sacks, James A. Mercy, Kenneth E. Powell. 1997. "Estimating intruder-related firearm retrievals in U.S. households, 1994." *Violence and Victims* 12:363-372.

Kalish, Carol B. 1974. "The Dayton-San Jose methods test." Reprinted on pp. 28-29 in *The National Crime Survey: Working Papers, Volume I: Current and Historical Perspectives*, edited by Robert G. Lehnen and Wesley G. Skogan. U.S. Department of Justice, Bureau of Justice Statistics. Washington, D.C.: U.S. Government Printing Office.

Kates, Don B., Jr., Henry E. Schaffer, John K. Lattimer, George B. Murray, and Edwin W. Cassem. 1995. "Guns and public health: epidemic of violence, or pandemic of propaganda?" *Tennessee Law Review* 62:513-596.

Kellermann, Arthur L., Frederick P. Rivara, Norman B. Rushforth, Joyce C. Banton, Donald T. Reay, Jerry T. Franciso, Ana B. Locci, Janice Prodzinski, Bela B. Hackman, and Grant Somes. 1993. "Gun ownership as a risk factor for homicide in the home." *New England Journal of Medicine* 329:1084-1091.

Kellermann, Arthur L., Lori Westphal, Laurie Fischer, and Beverly Harvard. 1995. "Weapon involvement in home invasion crimes." *Journal of the American Medical Association* 273:1759-1762.

Kleck, Gary. 1982. "On the use of self-report data to determine the class distribution of criminal and delinquent behavior." *American Sociological Review* 47:427-433.

Kleck, Gary. 1988. "Crime control through the private use of armed force." *Social Problems* 35:1-21.

Kleck, Gary. 1991. *Point Blank: Guns and Violence in America*. N.Y.: Aldine.

Kleck, Gary. 1993. "Bad data and the 'Evil Empire': interpreting poll data on gun control." *Violence and Victims* 8:367-376.

Kleck, Gary. 1997. *Targeting Guns: Firearms and their Control*. N.Y.: Aldine.

Kleck, Gary, and Marc Gertz. 1995. "Armed resistance to crime: the prevalence and nature of self-defense with a gun." *Journal of Criminal Law and Criminology* 86:150-187.

- \_\_\_\_\_. 1998. "Carrying guns for protection: results from the National Self-Defense Survey." *Journal of Research in Crime and Delinquency* 35:193-224.
- Kleck, Gary, and Michael Hogan. 1997. "A national case-control study of homicide offending and gun ownership." Revised version of a paper presented at the annual meetings of the American Society of Criminology, Chicago, November 21, 1996.
- Kooi, Roger. 1997. Telephone conversation with Officer Roger Kooi of the Atlanta Police Department, December 18, 1996, and Atlanta Police Department Offense Report forms.
- Loftin, Colin, and Ellen J. MacKenzie. 1990. "Building national estimates of violent victimization." Paper read at the National Research Council Symposium on the Understanding and Control of Violent Behavior, Destin, Florida, April 1-6, 1990.
- Magura, Stephen, Robert C. Freeman, Qudsia Siddiqi, and Douglas Lipton. 1992. "The validity of hair analysis for detecting cocaine and heroin use among addicts." *International Journal of the Addictions* 27:51-69.
- Magura, Stephen, Douglas Goldsmith, Cathy Casriel, Paul J. Goldstein, Douglas S. Lipton. 1987. "The validity of methadone clients' self-reported drug use." *International Journal of the Addictions* 22:727-749.
- Magura, Stephen, Sung-Yeon Kang, and Janet L. Shapiro. 1995. "Measuring cocaine use by hair analysis among criminally-involved youth." *Journal of Drug Issues* 25:683-701.
- McDowall, David. 1995. "Firearms and self-defense." *Annals* 539:130-140.
- McDowall, David, and Brian Wiersema. 1994. "The incidence of defensive firearm use by U.S. crime victims, 1987 through 1990." *American Journal of Public Health* 84:1982-1984.
- McDowall, David, Brian Wiersema, and Colin Loftin. 1992. "The incidence of civilian defensive firearms use." University of Maryland Violence Group Discussion Paper, November 10, 1992.
- McNagny, Sally E., and Ruth M. Parker. 1992. "High prevalence of recent cocaine use and the unreliability of patient self-report in an inner-city walk-in clinic." *Journal of the American Medical Association* 267:1106-1108.

Mieczkowski, Tom. 1990. "The accuracy of self-reported drug use: an evaluation and analysis of new data." Pp. 275-302 in *Drugs, Crime and the Criminal Justice System*, edited by R. Weisheit. Cincinnati: Anderson.

Mieczkowski, Tom, David Barzelay, Bernard Gropper, and Eric Wish. 1991. "Concordance of three measures of cocaine use in an arrestee population: hair, urine, and self-report." *Journal of Psychoactive Drugs* 23:241-249.

Murphy, Linda R., and Richard W. Dodge. 1970. "The Baltimore recall study." Reprinted on pp. 16-21 of *The National Crime Survey: Working Papers, Volume I: Current and Historical Perspectives*, edited by Robert G. Lehnen and Wesley G. Skogan. U.S. Department of Justice, Bureau of Justice Statistics. Washington, D.C.: U.S. Government Printing Office.

Neter, J., and J. Waksberg. 1964. "A study of response errors in expenditures data from household interviews." *Journal of the American Statistical Association* 59:18-55.

Reiss, Albert, and Jeffrey A. Roth (eds.). 1993. *Understanding and Preventing Violence*. Washington, D.C.: National Academy Press.

Reynolds, Glenn Harlan. 1995. "A critical guide to the Second Amendment." *Tennessee Law Review* 62:461-512.

Smith, Tom W. 1997. "A call for a truce in the DGU war." *Journal of Criminal Law and Criminology* 87:1462-1469.

Sudman, Seymour, and Norman Bradburn. 1973. "Effects of time and memory factors on response in surveys." *Journal of the American Statistical Association* 68:805-815.

Tonso, William R. 1983. "Social problems and sagescraft in the debate over gun control." *Law & Policy Quarterly* 5:324-344.

Turner, Anthony G. 1972. "The San Jose recall study." Reprinted on pp. 22-27 of *The National Crime Survey: Working Papers, Volume I: Current and Historical Perspectives*, edited by Robert G. Lehnen and Wesley G. Skogan. U.S. Department of Justice, Bureau of Justice Statistics. Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of Justice Statistics. 1993. *Survey of State Prison Inmates, 1991*. Washington, D.C.: Bureau of Justice Statistics.

\_\_\_\_\_. 1997. *Criminal Victimization in the United States 1994*. Washington, D.C.: Bureau of Justice Statistics.

Vernick, Jon S., Stephen P. Teret, and Daniel W. Webster. 1997. Letter to the Editor. *Journal of the American Medical Association* 278:702-703.

Weil, Douglas, and David Hemenway. 1992. "Loaded guns in the home: analysis of a national random survey of gun owners. *Journal of the American Medical Association* 267:3033-3037.

\_\_\_1993a. "I am the NRA: an analysis of a national random sample of gun owners." *Violence and Victims* 8:353-365.

\_\_\_1993b. "A Reply to Kleck." *Violence and Victims* 8:377.

\_\_\_1992. "Violence in America: Guns." *Journal of the American Medical Association* 268:3072.

Wish, Eric D. 1987. "Drug use forecasting: New York 1984 to 1986." National Institute of Justice, Research in Action. Washington, D.C.: NIJ.

Wish, Eric D., and Bernard Gropper. 1990. "Drug testing by the criminal justice system: methods, research, and applications." Pp. 321-391 in *Drugs and Crime*, edited by James Q. Wilson and Michael Tonry. Chicago: University of Chicago Press.

Wish, Eric. D., Jeffrey A. Hoffman, and Susanna Nemes. 1995. "The validity of self-reports of drug use at treatment admission and at follow-up: comparison with urinalysis and hair assays." Unpublished paper. Center for Substance Abuse Research, University of Maryland, College Park, Md.

Woltman, Henry, John Bushery, and Larry Carstensen. 1984. "Recall bias and telescoping in the National Crime Survey." Pp. 90-93 in *The National Crime Survey: Working Papers, Volume II: Methodological Studies*, edited by Robert G. Lehnen and Wesley G. Skogan. U.S. Department of Justice, Bureau of Justice Statistics. Washington, D.C.: U.S. Government Printing Office.