Journal of Criminal Law and Criminology

Volume 87
Issue 4 *Summer*

Article 6

Summer 1997

Illegitimacy of One-Sided Speculation: Getting the Defensive Gun Use Estimate Down

Gary Kleck

Marc Gertz

0091-4169/97/8704-1446
THE JOURNAL OF CRIMINAL LAW & CRIMINOLOGY
Vol. 87, No. 4
Printed in U.S.A.

THE ILLEGITIMACY OF ONE-SIDED SPECULATION: GETTING THE DEFENSIVE GUN USE ESTIMATE DOWN

GARY KLECK & MARC GERTZ*

INTRODUCTION

It is obvious to us that David Hemenway (H) had no intention of producing a balanced, intellectually serious assessment of our estimates of defensive gun use (DGU). Instead, his critique serves the narrow political purpose of "getting the estimate down," for the sake of advancing the gun control cause. An honest, scientifically based critique would have given balanced consideration to flaws that tend to make the estimate too low (e.g., people concealing DGUs because they involved unlawful behavior, and our failure to count any DGUs by adolescents), as well as 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. H's approach is precisely the opposite—one-sided and almost entirely speculative. Readers who have any doubts about the degree to which H's paper is imbalanced might carry out a simple exercise to assess our claim—count the number of lines H devotes to flaws tending to make the estimate too high and the number devoted to flaws making the estimate too low. We submit that the ratio is over 100-to-1, i.e., almost entirely devoted to speculations about why the estimate is too high.

The political function of this advocacy scholarship is clear. While high estimates of DGU frequency do not constitute an obstacle to moderate controls over guns, they constitute the most serious obstacle to advocacy of gun prohibition. Disarming the mass of noncriminal prospective crime victims would, if high DGU estimates are even ap-

* Professors of Criminology and Criminal Justice, School of Criminology and Criminal Justice, Florida State Univesity, Tallahassee, Florida. The authors wish to acknowledge the skillful editorial contributions of Sally Gertz. The authors also note that severe space limits prevented them from presenting their entire response to Hemenway. A longer version rebutting all of his claims is available from the senior author.

proximately correct, result in large numbers of foregone opportunities for 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. H's paper is 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.

H has close ties to HCI through two key staff members of HCI's "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,[1] while H has co-edited a strongly pro-control propaganda tract with Dennis A. Henigan, legal counsel to HCI and CPHV.[2]

H's political intentions and strong feelings are also evident in his overstatements and in the grandiose conclusions he draws from weak or irrelevant evidence and fallacious reasoning. He does not get past his title before making his first overstatement, claiming that he had established, without benefit of any new empirical evidence, that our estimates are too high and that they are "extreme overestimates."[3] He states in his first paragraph that "it is clear that [the Kleck and Gertz] results cannot be accepted as valid."[4] He incorrectly claims that "all checks for external validity of the Kleck-Gertz finding confirm that their estimate is highly exaggerated,"[5] when in fact these checks have repeatedly confirmed our estimates.

DGUs usually involve unlawful possession of a gun by the gun-wielding victim, and sometimes other illegalities as well,[6] a point H does not dispute. Yet, in making the extraordinary and counterintuitive claim that there is a social desirability bias to people reporting their own illegal behavior,[7] H insists that such a desirability bias is not

---

[1] *See, e.g.*, David Hemenway & Douglas S. Weil, *Phasers on Stun*, 9 J. POL'Y ANALYSIS & MGMT. 94 (1990); David Hemenway & Douglas S. Weil, *Less Lethal Weapons*, WASH. POST, May 14, 1990 (Op-Ed); Douglas S. Weil & David Hemenway, *Loaded Guns in the Home*, 267 JAMA 3033 (1992); Douglas S. Weil & David Hemenway, *I Am the NRA*, 8 VIOLENCE & VICTIMS 353 (1993); Douglas S. Weil & David Hemenway, *A Reply to Kleck*, 8 VIOLENCE & VICTIMS 377 (1993); Douglas S. Weil & David Hemenway, *Violence in America: Guns*, 268 JAMA 3072 (1992).

[2] DENNIS A. HENIGAN ET AL., GUNS AND THE CONSTITUTION: THE MYTH OF THE SECOND AMENDMENT (1995).

[3] David Hemenway, *Survey Research and Self-Defense Gun Use*, 87 J. CRIM. L. & CRIMINOLOGY 1430 (1997).

[4] *Id.*

[5] *Id.* at 1431.

[6] *See* Gary Kleck & Marc Gertz, *Armed Resistance to Crime*, 86 J. CRIM. L. & CRIMINOLOGY 150, 156, 174 (1995).

[7] Hemenway, *supra* note 3, at 1430-31.

only plausible, but that it is likely.[8] By the end, without having provided a scintilla of credible supporting evidence, H concludes that our research 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."[9] It is 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.

I. The Illegitimacy of One-sided Speculation: An Ounce of Evidence Outweighs a Ton of Speculation

H's critical technique is simple: one-sided, and often implausible, speculation about flaws that might have afflicted our research, and that might have been consequential enough to significantly affect our conclusions. H devotes his attention almost exclusively to suspected flaws that might have contributed to the overestimation of defensive gun use (DGU) frequency. He either ignores well established sources of underreporting, or briefly and superficially discusses them only for the sake of dismissing them.[10] When H speculates about sources of response error that are plausible, he offers no rationale for why the problems should lead to more false positives than false negatives. Instead he simply conjures up reasons they might lead to false positives. As support for his one-sided speculations H even cites other people guilty of the same dubious practice.[11]

All research is flawed. Known flaws should be identified and their likely impact 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. Unfortunately, in both good research and bad, there is no upper limit on the amount of speculative criticism that can be directed at the work, and thus this sort of critique is just as easily applied to good research as to bad.

---

8 *Id.* at 1438 (asserting that "the likelihood of social desirability response bias (self-presentation bias) is clear").

9 *Id.* at 1444.

10 *See, e.g., id.* at 1439.

11 *See id.* at 1432 nn.11-12 (citing McDowall et al. and Reiss and Roth).

II. Red Herrings and the Issue Not Addressed

Much of H's paper is a red herring in that it implicitly misstates the central technical question about our estimates. Much of it is devoted to elaborate speculations about why people might falsely claim to have used a gun defensively, as if it were somehow in dispute that there are some false positives.[12] He inaccurately hints that we unreasonably ignored the possibility that some of our respondents (Rs) provided false positives.[13]

We assume as a matter of course that our survey is like all other surveys in that some Rs give inaccurate responses to questions, and that these errors include both false positives and false negatives. The central question is not whether there are false positives, nor even how many false positives there are, but rather what the relative balance is between false positives and false negatives. Because H makes no effort to assess the frequency of false negatives,[14] it is logically impossible for him to draw meaningful conclusions about whether our estimates were too high or low.

III. The Nature of False Positives

It is hard to discern exactly what kinds of false positives H thinks most often show up in all these gun use surveys. He waffles 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 aggressive use of a gun); or (3) unconsciously distorting real events. He seems to have doubts himself about possibility (1) occurring very often, hastening to assure readers that false responders do not necessarily have to lie,[15] but is otherwise unwilling to commit himself to the relative frequency of these types of misreports.

It is worth emphasizing how difficult it was for our Rs to falsely report a completely nonexistent event as a DGU. Unlike the UFO example that H insists is somehow parallel to reports of DGUs,[16] a respondent who wanted to falsely report a nonexistent DGU could not qualify as having had such an experience merely by saying "Yes." Rather, respondents had to provide as many as nineteen internally consistent responses covering the details of the alleged incident. In

---

12 *See, e.g., id.* at 1430, 1438-40.

13 *E.g., id.* at 1439 ("according to K-G, . . . none of the 88,800 individuals who have not had a gun use are reporting having had one").

14 This judgment specifically includes H's section V. *See id.* at 1435-37.

15 *Id.* at 1435.

16 *Id.* at 1437.

short, to sustain a false DGU claim, Rs had to do a good deal of 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 R to speak a single inaccurate syllable: "No." The point is not that false positives were impossible, but rather that it was far harder to provide a false positive than a false negative.

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

Regarding possibility (2), we noted that most of the DGUs were linked with the types of crimes—burglaries, robberies, and sexual assaults—where there is little opportunity for participants to be honestly confused about who was the victim and who was the offender.[17] While a few Rs 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 Rs could report an account of a real burglary, robbery, or sexual assault in which they were aggressors and somehow honestly distort it into a DGU incident.

This kind of misunderstanding of real events in a way that falsely qualifies them as DGUs is more plausible in connection with male-against-male assault incidents, such as when people prefer to characterize their partly aggressive, partly defensive behavior in "mutual combat" incidents as purely defensive in character. We addressed this latter possibility in our article and showed that it could not account for more than a small fraction (probably less than a tenth) of the incidents we counted as DGUs.[18] H does not rebut that evidence.

IV. Raising the Dead: Resuscitating the NCVS Estimates of DGU

H contrasts National Crime Victimization Survey (NCVS) estimates of DGU with our estimates,[19] but is evasive as to why he does this. He never explicitly says that he considers the NCVS estimates accurate, perhaps because he knows this position is indefensible.[20]

---

17 Kleck & Gertz, *supra* note 6, at 174.

18 *Id.*

19 Hemenway, *supra* note 3, at 1432.

20 Kleck & Gertz, *supra* note 6, at 153-57.

But if the NCVS estimates are *not* accurate, what is the point of citing them in the context of a challenge to our very different estimates, and asserting that the NCVS is the "gold standard" for estimating criminal victimization?[21]

On the other hand, if H really does believe the NCVS estimates are even 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[22] and David McDowall[23] have ceased portraying the NCVS estimates as valid. Instead, they have shifted to the agnostic views that no survey, including the NCVS, can yield meaningful estimates[24] or that "the frequency of firearm self-defense is an issue that is far from settled."[25] Either view is 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 H'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."[26]

We 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.[27] Still, H uses the NCVS estimates as a standard against which he judges the DGU estimates of other surveys.[28] He falsely claims that the NCVS asks "about self-defense gun use,"[29] when in fact, as we pointed out, the NCVS never directly asks about DGU.[30] Instead it merely provides Rs with an opportunity to volunteer information about a DGU in response to a general question about self-protection actions. Nor does H acknowledge that the NCVS is the only survey that ever has yielded annual DGU estimates under 700,000, and that its estimates, centering

---

21 Hemenway, *supra* note 3, at 1441.

22 *See* PHILIP COOK & MARK C. MOORE, GUN CONTROL IN CRIME, 267-94, 566-71 (James Q. Wilson & Joan Petersilia eds., 1994); Phillip J. Cook, *The Technology of Personal Violence, in* 14 CRIME AND JUSTICE 1 (Michael Tonry ed., 1991).

23 *See* David McDowall & Brian Weirsema, *The Incidence of Defensive Firearm Use of U.S. Crime Victims, 1987-1990*, 84 AMER. J. OF PUB. HEALTH 1982 (1994).

24 *See* Phillip Cook & Jens Ludwig, *You Got Me: How Many Defensive Gun Uses Per Year?* (Paper Presented at the Annual Meetings of the Am. Soc'ty of Criminology, Chicago, Ill., Nov. 20, 1996).

25 David McDowall, *Firearms and Self-Defense*, 539 ANNALS 130, 138 (1995).

26 Don B. Kates et al., *Guns and Public Health: Epidemic of Violence, or Pandemic of Propaganda?*, 62 TENN. L. REV. 513, 537 (1995) (quoting Cook).

27 Kleck & Gertz, *supra* note 6, at 153-57.

28 Hemenway, *supra* note 3, at 1432.

29 *Id.*

30 Kleck & Gertz, *supra* note 6, at 155.

around 80,000, are far below those generated by at least *fifteen* other surveys.[31] Instead, he inverts reality by falsely hinting that it is the Kleck-Gertz estimate which is the deviant result.

V. Fallacious Reasoning: Hemenway's "Checks on External Validity"

In our article, we cautioned against two kinds of fallacious reasoning. Instead of avoiding these errors, H knowingly embraces them. The fallacious arguments involve a misapplication of *reductio ad absurdum* argumentation, based on the misperception that estimates from our survey are inconsistent with known crime counts and the erroneous assumption that the NCVS provides correct estimates of the absolute frequency of crime.

H argues that our estimates are implausible because our survey implies a number of DGUs occurring in connection with burglaries that exceeds the total number of burglaries of occupied residences estimated by the NCVS.[32] This argument rests on an unacknowledged assumption that the universe of DGU events sampled by our survey is a subset of the universe of crime events covered by the NCVS. However, we had explicitly warned in our 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 either to the NCVS or police."[33] 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.

H's logic is 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. Our estimates of total DGUs are likely to be fairly reliable partly because they are based on a very large sample (n=4977), 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 194 reported DGU incidents, of which about 40 were linked to burglaries.[34] H's *reductio ad absurdum* logic is equivalent to arguing that Gallup presidential

[31] *Id.* at 153-59.
[32] Hemenway, *supra* note 3, at 1441.
[33] Kleck & Gertz, *supra* note 6, at 167.
[34] *Id.* at 184-85.

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)[35] because they sometimes yield implausible estimates for small subsets of the electorate, such as rural Hispanic Jews. 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 all DGUs, based on the full sample, are accurate.

Finally, even if one ignored these logical fallacies, H's argument still fails, because it depends on an erroneous factual assumption. H states that "from the NCVS, we know that there were fewer than six million burglaries in 1992,"[36] and makes similar statements about rapes.[37] 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, H's argument collapses, since it becomes impossible to argue that the number of DGUs linked to a given 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, H cites estimates of the number of gunshot wound (GSW) victims treated in emergency rooms and falsely claims that "K-G report that 207,000 times per year the gun defender thought he wounded or killed the offender."[38] In fact, we did not compute or report this 207,000 estimate, and we specifically cautioned against using our data on GSWs because they were based (unlike our estimates of DGU frequency in general) on a small sample. Moreover, we cautioned because we had done no detailed questioning of Rs regarding why they thought that they had wounded their adversaries.[39]

In any case, there is nothing even mildly inconsistent between this GSW estimate and emergency room data on persons treated for GSWs. H again makes the implicit assumption that DGU-linked woundings are entirely a subset of woundings treated in medical facilities. If one more plausibly assumes that substantial numbers of less serious GSWs are not treated in such facilities, the number of medically treated GSWs cannot be used as an upper limit on the number of

---

35 326 THE GALLUP POLL MONTHLY 33 (Nov. 1992) (summarizing Gallup Poll presidential election accuracy record).

36 Hemenway, *supra* note 3, at 1441.

37 *Id.* at 1442.

38 *Id.*

39 Kleck & Gertz, *supra* note 6, at 173.

DGUs that result in a GSW. If, for example, the total annual number of GSWs, treated or untreated, were 400,000, there would obviously be nothing even mildly implausible about 200,000 of them being DGU-linked, especially in light of the fact that the vast majority of victims of known assault GSWs are criminals.[40]

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.[41] Less than a tenth of assault GSWs are life-threatening.[42] 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 or GSWs, we cannot possibly know if a DGU estimate is implausibly large relative to these unknown (and possibly unknowable) quantities.

VI. THE UFO ANALOGY

Perhaps the most bizarre part of H's paper is the analogy he draws between survey reports of DGUs and reports of contacts with alien spacecraft. H is once again dealing in a red herring. No one disputes that some behaviors or experiences can be greatly overestimated. 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 the extent of misestimation with respect to one topic casts little light on the likely degree of error in misestimating another topic unless the topics are very similar.

We assume that most Rs who respond affirmatively to UFO questions are having a little fun with the interviewers, though a few undoubtedly are serious. On the other hand, we find it harder to believe that Rs would regard questions about crime victimization and DGUs in so frivolous a light. In addition, this analogy ignores the fact that all it took to be counted as a UFO 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

40 GARY KLECK, TARGETING GUNS: FIREARMS AND THEIR CONTROL, ch. 1 (forthcoming 1997).

41 Roberta K. Lee et al., *Incidence Rates of Firearm Injuries in Galveston, Texas, 1979-1981*, 134 AM. J. EPIDEMIOLOGY 511, 519 (1991).

42 KLECK, *supra* note 40, at ch. 1.

user. The analogy H sees escapes us.

VII. The Positive Social Bias Speculation

H does not deny or rebut our observation that most of the reported DGUs involved illegal behavior on the part of the Rs.[43] He simply ignores it, perhaps because he recognizes that it would be difficult to persuade readers that survey Rs are biased in favor of overreporting unlawful behavior. He insists that the predominant bias surrounding DGU reports is a "social desirability response bias"[44] with Rs making false reports of DGUs to present themselves as "heroic."[45] H ignores the information we provided in our article on the distinctly unheroic character of the DGU accounts provided.[46] What was most striking about the reported events was their banality. If H'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, Rs reporting DGUs claimed to have faced adversaries with guns in only one-in-six cases,[47] claimed involvement in a shootout (both parties shooting) in just 3% of the cases,[48] and usually reported opponents with no weapons at all.[49] Likewise, they rarely boasted about their deadly shooting, with only 8% even claiming to have wounded an adversary.[50]

In any case, H is again focussing on a red herring. The issue is not whether some Rs might think DGUs are heroic (this is undoubtedly true for at least a few 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, H distracts readers once again from the issue of the relative balance of response errors. He provides no evidence or even argumentation as to why any social desirability effects should outweigh simple concerns about revealing one's unlawful behaviors.

---

43 Kleck & Gertz, *supra* note 6, at 155, 171-74.

44 Hemenway, *supra* note 3, at 1438.

45 *Id.*

46 Kleck & Gertz, *supra* note 6, at 179-80.

47 *Id.* at 185.

48 *Id.*

49 *Id.*

50 *Id.*

VIII. MAKING SOMETHING OUT OF NOTHING: HEMENWAY'S NUMERICAL EXERCISES

It would be understandable if some readers thought that H did present, in his Section V, empirical evidence on the relative balance of false positives and false negatives.[51] In fact, this section presents no empirical evidence at all. Instead, H's numerical examples demonstrate 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 come up with enormous overestimates. We cannot fault H for his arithmetic. If there were any credibility to the misreporting rates he assumes out of thin air, they would indeed imply huge overestimates.

H's argument is fallacious because it assumes the very conclusion he wants to draw—that is, H assumes that there is a nonneglible rate of reporting false positives. The problem is that H does not present any empirical evidence that there were *any* false positives among the cases we treated as DGUs, nor among those so treated in other gun use surveys, never mind the large numbers he assumes.

H states that "with few actual positives, 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."[52] The operative phrase is "tend to be." All one can validly conclude from Section V 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. Whether there actually *are* more false positives that false negatives in surveys of DGU or other crime-related experiences is an issue to which H never brings any empirical evidence (as distinct from speculations and assumptions) to bear. Rather, he jumps from the fact that this potential exists to the non sequitur conclusion that "you *inevitably* get a large number of false positives relative to the number of true positives" and thus an overestimate.[53]

H's claim that our results are "extremely sensitive" to small changes in the specificity rate is another example of argumentation that relies on assuming the conclusion.[54] The main reason that his example estimates in Table 2 are so sensitive to the specificity rate is because H assumes extremely low actual DGU rates, i.e., he assumes the very conclusion he is pushing. Thus, instead of the empirically-

51 *See* Hemenway, *supra* note 3, at 1435-37.

52 *Id.* at 1436.

53 *Id.* at 1437 (emphasis added).

54 *Id.* at 1436.

based 1.33% estimate we obtained,[55] H assumes an actual DGU rate of 0.32% in Table 2(A), 0.04% in 2(B) and 0.08% in 2(C).[56] Because he arbitrarily assumes there are so few true positives, even a handful of false positives can outnumber them and substantially distort the estimates. For example, in Table 2(B), the main reason H's assumed rate of false positives of 1.3% has such a proportionally large distorting effect on the estimate is because he assumes, without any empirical foundation, that the actual DGU prevalence rate is virtually zero, so that just sixty-four false positives can be thirty-three times higher than the assumed number of just two (!) true positives.[57] For what it's worth, the estimates *would* be highly sensitive to the specificity rate, *if* the true DGU rate were as low as H assumes, but then it is the DGU rate that is at issue.

In our view, a more realistic version of H's Table 2—one more in tune with research on errors in surveys of illegal behavior[58]—might have forty-eight true positives, forty-eight false negatives (and thus ninety-six persons with a genuine DGU), ten false positives, and 4,894 true negatives, implying 50% test sensitivity and 99.7% 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 our survey, implying that the true DGU rate was actually 45% higher the one we obtained. Of course, the question remains, which is the more plausible set of assumptions about the distribution of survey response errors—H's or ours? Unlike H, who relies on assumed numbers[59] and strained analogies to the reporting of diseases,[60] 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.

IX. Prior Research on the Validity of Survey Estimates of Illegal Behavior

H provides a discussion of "misclassification in surveys generally"[61] whose most notable feature is that it is utterly silent about surveys concerning illegal behavior and other crime-related experiences. While H discusses surveys about height, automobile ownership, diseases, and other topics of negligible relevance to the topic at

55 Kleck & Gertz, *supra* note 6, at 184.
56 Hemenway, *supra* note 3, at 1444-45.
57 *Id.* at 1444.
58 *See infra* notes 62-66 and accompanying text.
59 Hemenway, *supra* note 3, at 1436-37.
60 *Id.* at 1435-37.
61 *Id.* at 1434-35.

hand, he says nothing about evidence concerning the validity of responses to questions requiring Rs 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 H addresses. We will correct this conspicuous omission.

A large body of empirical evidence indicates that when asked sensitive questions about illegal behavior, survey Rs, 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 illegal drug use. Unlike other illegal behaviors, there is a strong external criterion that analysts can use to judge the validity of 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 Rs 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.[62] For example, 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.[63] Among patients at a walk-in clinic who had positive urine tests for illicit drug use, only 28% had reported the use in earlier interviews.[64] Actual use was thus at least 3.6 times higher (100/28=3.6) than reported use. Among a

---

62 *See generally* Z. Amsel et al., *Reliability and Validity of Self-Reported Illegal Activites and Drug Use Collected from Narcotics Addicts*, 11 INT'L J. OF THE ADDICTIONS 325 (1976); W.A. Baumgartner et al., *Hair Analysis for the Detection of Drug Use in Pretrial/Probation/Parle Populations, in Summary Report to the Nat'l. Institute of Justice* 1 (1990); I.H. Cisin & H.L. Parry, *Sensitivity of Survey Techniques In Measuring Illicit Drug Use, in* DEVELOPMENTAL PAPERS: ATTEMPTS TO IMPROSE THE MEASUREMENT OF HERION IN THE NATIONAL SURVEY (J.D. Rittenhouse ed., 1971); R. Dembo et al., *Urine Testing of Detained Juveniles to Identify High-Risk Youth*, NAT'L. INSTITUTE OF JUSTICE RESEARCH (1990); R. Falck et al., *The Validity of Injection Drug Users' Self Reported Use of Opiates and Cocaine*, 22 J. OF DRUG ISSUES 823 (1992); Stephen Magura et al., *The Validity of Methadone Clients' Self-Reported Drug Use*, 22 INT'L J. OF THE ADDICTIONS 727 (1987); Tom Mieczkowski et al., *Concordance of Three Measures of Cocaine Use in an Arrestee Population: Hair, Urine and Self-Report*, J. OF PSYCHOACTIVE DRUGS 241, 246 (1991); Tom Mieczkowski, *The Accuracy of Self-Reported Drug Use: An Evaluation and Analysis of New Data, in* DRUGS AND CRIME 275 (James Q. Wilson & Michael Tonry eds., 1990); Eric D. Wish & Bernard Cropper, *Drug Testing by the Criminal Justice System in* DRUGS AND CRIME 391 (James Q. Wilson & Michael Tonry eds., 1990); Eric D. Wish, *Drug Use Forecasting: New York 1984 to 1986*, NAT'L. INSTITUTE OF JUSTICE (Feb. 1987).

63 COOK ET AL., *supra* note 22, at 403.

64 Sally E. McNagny & Ruth M. Parker, *High Prevalence of Recent Cocaine Use and the Unreliability of Patient Self-report in an Inner-city Walk-in Clinic*, 267 JAMA 1106 (1992).

group of juvenile arrestees, hair analysis indicated 56.8% had used cocaine, but only 7.4% self-reported it in interviews.[65] Thus actual use levels were at least 7.7 times higher than self-reports indicated. In a group of youthful jail releasees, 67% tested positive for cocaine with hair analysis, but only 23% self-reported cocaine use in the preceding 90 days, and only 36% reported ever using it.[66]

Some studies have reported counts 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).[67] Likewise, among 86 subjects studied by Baumgartner, 16 falsely denied cocaine use by self-report, but only one reported drug use without a hair assay confirming it.[68] This again indicates that false negatives are common and false positives close to nonexistent.

These examples could be multiplied, but to no purpose.[69] The research record is clear enough: 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. Self-report surveys therefore underestimate illegal behavior. To use H'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).

X. Libelling our Interviewers

In discussing an alleged "limitation" of our survey, H writes: "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."[70] The unmistakable innuendo is that some of our interviewers faked or altered interviews to create phony accounts of "DGUs." To our knowledge, none of our interviewers knew anything about Kleck's views on DGU or what results he expected. H does not claim to have communicated with even one of the interviewers, to find out what they knew prior to interviewing. Therefore, as far as we can tell, he had no basis whatsoever for this outrageous charge.

---

65 Thomas E. Feucht et al., *Drug Use Among Juvenile Arrestees: A Comparison of Self Report, Urinalysis and Hair Assay*, 24 J. of Drug Issues 99, 103, 109, 111 (1994).

66 Stephen Magura et al., *Measuring Cocaine Use by Hair Analysis Among Criminally-Involved Youth*, 25 J. of Drug Issues 683, 691 (1995).

67 Mieczkowski et al., *supra* note 62, at 246.

68 Bamgartner et al., *supra* note 62, at 1.

69 See studies cited *supra* note 62.

70 Hemenway, *supra* note 3, at 1433.

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 the answer sheet would not generate a false positive. Furthermore, as we stated in our article, every single interview in which a DGU was alleged was validated by a call-back by a supervisor.[71] An interviewer-faked incident therefore could not have survived the quality control procedures unless a supervisor colluded.

XI. The Survey Hemenway Chose Not to Mention

Our estimates recently have been 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.[72] We can be certain that H 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.[73] 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 H and Cook.

H does not mention the results of this survey, perhaps for an understandable reason: It almost exactly confirms our results. We estimated 2.55 million annual DGUs, using a person-based one-year estimate.[74] The most comparable estimate generated by this survey was 2.45 million, well within sampling error of our estimate.[75] Many of the other estimates were even higher.[76] H 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 their findings by external measures."[77] In light of H's claim that "all checks for external validity of the Kleck-Gertz finding confirm that their estimate is highly exaggerated,"[78] what could possibly justify H's calculated decision to withhold the results of the Police Foundation survey, when it almost exactly confirmed our estimates?

We doubt that anything we can say will dissuade H from his re-

71 Kleck & Gertz, *supra* note 6, at 161.
72 Philip Cook & Jens Ludwig, Guns in America (1997).
73 *Id.*
74 Kleck & Gertz, *supra* note 6, at 184.
75 Cook & Ludwig, *supra* note 72, at 62.
76 *Id.*
77 Hemenway, *supra* note 3, at 1431.
78 *Id.*

markable theory that all surveys inevitably overestimate rare events, so he presumably would justify his decision to not mention this survey by asserting that all surveys are now irrelevant to the issue. Nevertheless, since the Police Foundation project yielded estimates almost identical to those of our earlier survey, one cannot plausibly argue that our estimates were an artifact of flaws peculiar to our survey.

This is a point that H has effectively conceded elsewhere,[79] raising the question: What was the point of all of his unsupported speculations about flaws supposedly afflicting our survey in particular,[80] if H knew that they were not responsible for our 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.

CONCLUSIONS

Hemenway has failed to cast even mild doubt on the accuracy of our estimates. The claim that there are huge 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 guns-violence research area. Given H's purposes, however, it is politically inconsequential that we can easily rebut all of his claims. We can be confident that ideologues will cite his series of one-sided speculations as authoritative proof that our estimates have been "discredited," while pro-control academics who fancy themselves moderates will conclude that although maybe H was wrong on some points, he has nevertheless somehow "cast doubt" on the estimates or "raised serious questions" about them. Left unmentioned will be one simple fact: In all of H's commentary, he does not once cite the one thing that could legitimately cast doubt on our estimates—better empirical evidence.

79 Philip J. Cook et al., *The Gun Debates's New Mythical Number: How Many Defensive Uses Per Year?*, 16 J. POL'Y ANALYSIS & MGMT. 463, 465 (1997).

80 Hemenway, *supra* note 3, at 1433-34.