Appendix D

Reenactments

Contents

Boxes

Figures

Tables

Appendix D
Reenactments

INTRODUCTION

In the context of this report, a "reenactment" is a repetition of a ballistic test that armor was, or might have been, subjected to. In particular, it is a test of:

1. armor worn by the victim of a shooting, who was hit on his or her armor by a bullet; or
2. a similar sample of armor, if the armor worn by the victim is unavailable or likely to have been damaged by the assault or subsequent testing.

In such a reenactment the armor is shot with one or more bullets of the same type used in the assault. Ideally, the bullets should impact the armor at the same velocity at which the bullet impacted in the assault.

However, other aspects of the reenactment—such as the angle of incidence at which the test bullets strike the armor-are not intended to replicate the conditions of the assault; they are intended to replicate the conditions of a test that might have been used to decide whether other samples of the armor tested had acceptable ballistic resistance. An example of such a test is a test for special-type ballistic resistance conducted in accordance with NIJ Standard 0101.03 using the weapon and ammunition used in the assault. It requires a wet sample and a dry sample of armor to be shot and the backface signature (crater depth) produced in clay behind the sample to be measured after the first fair impact on each sample. If either backface signature (BFS) exceeds 44 mm, the test is failed. By comparing the results of the reenactment to the effect of the shot on the victim, and by repeating this process for several victims, one may infer the risk associated with armor that passes the test, when worn by others for whom the victims are representative. That is, reenactments test the test, not the vest.

This appendix discusses some general considerations relating to the planning, conduct, and analysis of reenactments. It also analyzes the results of reenactments sponsored by E.I. du Pont de Neymours & Co., Inc., performed by H.P. White Laboratory, Inc., and observed by OTA in October 1991.

Why Reenact?

The "reenactment" of shootings of armor wearers is potentially a uniquely valuable procedure for characterizing the relationship between

1. the risk that a bullet stopped by armor in an actual assault will cause trauma to the wearer, and
2. the result of a ballistic test (e.g., backface signature measurement) used as an index of the risk of trauma.

The controlled shooting of armor on human wearers could provide more information faster but is considered unethical. The experimental shooting of armor on large mammals has provided the bulk of scientific knowledge about the correlation of ballistic measurements with risk of trauma in several species. Considering this information as well as the differences between animal and human anatomy and between laboratory and assault conditions, experts have predicted the risk of trauma in human wearers. However, the performance and analysis of reenactments is the only ethical means of testing such predictions.

Reenact What?

In this context, ''reenactment" refers not to the reenactment of an assault, but to the 'reenactment' of a ballistic test to which armor of the type involved was or might have been subjected. The purpose is to assess how reliably the ballistic test would have predicted the severity of any trauma caused by the stopped bullet.

For example, in one assault the front panel of a Point Blank model 15SR vest stopped2230.O-grain, .45-caliber, full-metal-jacketed bullets from Winchester Western cartridges fired by a Colt .45 ACP (semi-)automatic pistol with a 5-inch barrel 150 to 155 feet away. NIJ's *Body Armor Selection Guide [145]* cites .45 automatic as a II-A threat and the Point Blank model 15SR is NH-certified to have type 11-A ballistic resistance, but .45-caliber shots are not used in the NIJ-specified II-A test, nor in any of the other tests for *numbered types* of ballistic resistance. However, the 0101.03 standard provides for a test of 'special-type' ballistic resistance to any type of bullet at any impact velocity, to be specified by the customer.[1] Thus Point Blank (or a purchaser) could have had model 15SR vests tested for special-type ballistic resistance to 230.0-grain, .45-caliber, full-metal-jacketed bullets impacting at velocities typical of such bullets fired from a Colt .45 ACP automatic pistol with a 5-inch barrel at a range of 150 to 155 feet.

In some assaults, the impact velocity cannot be estimated with demonstrable accuracy and reliability after the fact. However, in some other cases both the weapon used in the assault and extra ammunition from the same box or lot as that used by the assailant are available. If the range from the weapon to the victim is known approximately, firing the left-over ammunition from the same weapon would produce approximately the same impact velocity at roughly the same range.

A difficulty arises because NIJ Standard 0101.03 specifies that soft armor must be shot at a range of 5 meters (about 16.4 feet) from the muzzle of the test weapon. This would usually preclude reenacting the range of the assault. There are several possible solutions: one is to ignore this rule and shoot the armor at the range at which it was hit in the assault. Another solution would be to (1) fire some of the leftover cartridges-not at the armor-and measure the bullet velocities at the range at which they impacted the armor in the assault, then (2) reload the remaining cartridges with a charge of powder judged likely to reproduce the recorded velocities at a range of 5 meters, and (3) fire them as specified in the .03 standard. This would complicate statistical analysis. One would want to calculate the statistical significance with which one could reject the hypothesis that the distribution of velocities of the bullets from the reloaded cartridges at a range of 5 meters differs from the distribution of velocities of the bullets from the factory-loaded cartridges at assault range.

In many cases, a third solution is reasonable: shoot the armor at a range of 5 meters and ignore the difference between the range in the assault and the range in the reenactment. Most shootings of police officers occur at very close range, and the momentum[2] of a bullet, on which BFS depends [7, 8], would change very little over the frost few meters of flight.[3] Except perhaps in the case of shotgun pellets, the muzzle velocity, the velocity of impact in an assault, and the velocity at the 16-foot range at which the test is conducted will be almost the same, because bullets slow down very little until they are far downrange. The same is true of shotgun slugs, but shotgun pellets slow more dramatically after they leave the muzzle and start to spread. Spreading depends on the design of the shot shell, the downrange distance, and the shotgun's choke.[4] As a load of shot travels downrange and spreads, its effectiveness as a penetrator or producer of backface signature is reduced so much that a test at a range of 16 feet may not indicate the likely result of a zero-range assault.

In the example at hand—reenactment of Colt .45 shots fired at a range of 150 to 155 feet (50 to 51.7 yards, 45.7 to 47.2 meters)-it is reasonable to shoot the armor at a range of 5 meters and ignore the difference between the range in the assault and the range in the reenactment. Federal, Remington, and Winchester Western cartridges propel their 230-grain, .45-caliber, full-metal-jacketed bullets to velocities of 835 to 850 ft/s at the muzzles of 5-inch test barrels; at such velocities, they lose about 4 to 5 percent of their velocity (and momentum) in the first 50 yards of flight. (See table D-l.) [85] The velocity loss is about 35 to 40 ft/s, which is within the 50 ft/s variation allowed in a .03 Special Type test.

In reality, there will be some shot-to-shot variation in velocity. A portion of this variation is systematic—for example, the first shot fired from a tight barrel at room temperature tends to be slightly slower, on the average, than subsequent shots fired in rapid succession from the same barrel, which has been heated by previous shots and has expanded. But most of the variation is unexplained (i.e., apparently random) and presumed to arise from cartridge-to-cartridge differences in the ammunition. Firing several shots to reenact each assault shot will allow subsequent statistical analysis (described below) to estimate risk in terms of BFS by averaging over the

1 As of December 10, 1991, this option had never been exercised.

2 The momentum of a projectile is its mass times its velocity.

3 However, the probability that a bullet will penetrate may vary significantly over the first few meters of flight; in particular, it may be greater near the muzzle if the bullet pitches or yaws as it exits the barrel, but the pitching and yawing maybe damped (i.e., may die out) within a few meters.

4 A shotgun's *choke* is a slight constriction at the muzzle. It controls the rapidity with which the shot spread after they depart the gun. Greater penetration and blunt impact (at the price of a smaller pattern) are to be expected from more strongly choked guns. Conversely, a strong choke will slow a slug, lessening its ability as a penetrator or blunt impactor (as well as causing possible damage to the gun).

Table D-l—Downrange Velocities of 230-grain, .45-caliber Bullets From Factory-Loaded Cartridges

Manufacturer	Velocity[a] (ft/s) at			
	Bbl.	Muzzle	25 yds	50 yds
Federal 45A.	5 in.	850	830	810
Remington R45AP[b].	5 in.	835	—	800
Winchester X45A1 P2...	5 in.	835	—	800

KEY: Bbl. = Barrel length; — = Not given.
[a]Nominal.
b "Metal case" (FMJ).

SOURCE: William S. Jarrett (cd.), Shooter's *Bible,* 1992 edition [85].

impact velocities representative of the impact velocity in the assault and over the corresponding BFSs, which, for any impact velocity, may vary with impact location or for other reasons, including unexplained randomness.

How many shots should constitute one reenactment? (This is distinct from the question of how many reenactments should be performed for each assault shot, which is considered below.) A special-type test of a ballistic element (e.g., a front panel) requires shooting two elements one wet, the other dry-and measuring the BFS caused by the first fair shot on each panel. This is the case for considering 1 reenactment to consist of 2 shots, 1 of which impacts armor that has been sprayed with water as prescribed by NIJ Standard 0101.03. However, some reenacted shots were stopped by armor not designed to resist penetration when wet. Should such vests be tested wet? If they are, the result would likely be a penetration, not a measurable BFS.

In choosing the number of shots, it is useful to consider the evolution of the NILECJ/NIJ standards and the origin of the 44-mm BFS limit, both of which are explained in appendix A. NILECJ Standard 0101.00 required the BFS to be measured on one dry sample of each element, but it required the BFS caused by each of 5 fair shots impacting the element (10 if a front panel) to be measured. Although the BFS was to be recorded, no BFS limit was specified; the standard itself indicated that it would be amended later to specify a limit when one was determined. This was done 6 years later, when NILECJ Standard 0101.01 introduced the 44-mm BFS limit, which was based on NILECJ-sponsored research performed by the Army. (See app. A.)

Documentation does not clearly indicate whether the Army intended the limit to apply to a l-shot test or to a test consisting of a greater number of shots, nor whether the Army or NILECJ appreciated that, for fixed risk, the BFS limit should increase as the number of BFS measurements (any of which could cause failure) increases. In any case, since it was introduced in NILECJ Standard 0101.01 in 1978, the 44-mm limit has applied to a 2-shot test—and partly for this reason may have been more conservative than originally intended.

NILECJ Standard 0101.01 also introduced the requirement for testing a wet sample as well as a dry one, hence for making only 2 BFS measurements per bullet type per element, instead of 5 or 10.

In light of all this, we consider the following approaches reasonable:

1. For purposes of correlating BFS with risk of trauma, one may consider 2 BFS measurements behind dry armor to constitute 1 reenactment of the BFS part of a test for special-type ballistic resistance conducted in accordance with NIJ Standard 0101.03. Had the ballistic element been enclosed in a thin waterproof cover (e.g., of polyurethane-coated ripstop nylon), this would have made little difference in the BFS (or penetration) and would have kept the ballistic element dry, had the cover been sprayed with water before one of the shots.
2. One could consider each BFS measurement behind dry armor to constitute 1 reenactment of a l-shot BFS test like that specified by NIJ Standard 0101.03 except for the number of shots. The probability that the armor would have failed a similar 2-shot test (i.e., failed on either or both of 2 shots) maybe estimated by statistical methods.

We will consider only the first of these approaches, because it is simpler.

Quite apart from the question of how many shots should constitute one reenactment, the expectation that BFS will vary from shot to shot under similar conditions makes it desirable to conduct as many reenactments as possible for each case. The analysis of results, which should include an analysis of uncertainty, will generally estimate less uncertainty if more reenactments are performed. Practical or economic constraints, such as the amount of left-over assailant's ammunition or unweakened area on the victim's armor, may limit the number of reenactments possible-perhaps to different numbers in different cases.

SELECTION OF CASES FOR REENACTMENT

To estimate the risk of injury associated with a particular BFS on the basis of an experiment in which the experimenter selected the numbers of reenactments of injurious and noninjurious shots to be performed, we use an analytical procedure called separate-sample logistic discrimination [9]. It is widely used for epidemiological case-control studies, in which, for example, 20 persons with a particular type of cancer ("cases") and 20 persons without the disease ("controls") are selected and interviewed to assess their exposure to a suspected carcinogen over the last 20 years. The procedure allows the risk of getting the cancer to be predicted as a function of degree or duration of exposure, based on such retrospective data. *It accounts for the fact that the number of cases and the number of controls were chosen by the experimenter,* not necessarily in proportion to the number of persons known to have the disease and the number of persons believed to not have it. In fact, it is particularly efficient when the disease of interest is rare; researchers may investigate the past exposures of all known cases but need only investigate the past exposures of a comparable number of controls chosen randomly from the much larger group of people believed to be free of the disease.

By analogy, the *cases we* consider are those who were killed or seriously[5] injured or incapacitated by the impact of a bullet (or slug, or shotgun blast) stopped by soft armor they were wearing. *Controls* should be representative of (e.g., chosen randomly from) the much larger group of people whose armor stopped a bullet, slug, or blast but who did not suffer death or serious injury or incapacitation. The *exposure* of interest is exposure to impact of a bullet stopped by armor; the *dose* (amount of exposure) is O or 1 depending on whether the 2-shot BFS test reenacted is passed (0) or failed (l). (For purposes of estimating the BFS limit that corresponds to a specified risk, the dose could be the BFS measured in a l-shot test.)

At the end of 1991, about 540 people were known to have been saved by body armor from death or serious injury by gunshot wound. About 90 percent of the incidents involved 1 impact on armor, and most of the rest involved 2 impacts, so about 594 shots were stopped with no death or serious injury resulting from the impact. Only 2 or 3 (maybe 4) people were known to have been killed or seriously injured by a bullet, slug, or shot stopped by armor. The number depends on where one draws the line between degrees of trauma severity. (See box D-l-Categories of Trauma and Incapacitation.) It is convenient to use the Abbreviated Injury Scale (AIS) to distinguish degrees of trauma severity [88]. On this scale, a rating of 6 denotes a fatality. One such injury has occurred; the anonymous victim was killed by a .45-70 bullet fried from a carbine. [133]

An AIS rating of 5 denotes a critical injury with survival uncertain; a rating of 4 denotes a life-threatening injury with survival probable. The injury sustained by Officer Bryan Power of the Mercedes (Texas) Police Department probably would be rated AIS 4 or 5; he was hit on his armor over his upper left chest by a 12-gauge slug, which made a 10-cm diameter open wound in his chest and bruised his lung underneath.[6]

A rating of 3 denotes a severe but not life-threatening injury, which describes the injury of Officer Steve Draper of the East Hempfield Township (Pennsylvania) Police Department. He was hit on his armor over his left chest by a 347.5-grain 16-gauge slug, which caused "penetration to chest cavity within 1-1.5 in of heart. "[7] This required sutures of muscle and skin.

All other stopped bullets known to us produced injuries rated lower than 3. The most serious of these

5 We will consider various degrees of seriousness.

6 Christopher H. Hansen, M.D, medical report, July 4, 1984. There was no "gross escape of air," pneumothorax (air in the chest cavity), or evidence of injury to the heart.

7 Questionnaire completed by DuPont based on telephone interview of victim.

Box D-1-Categories of Trauma and Incapacitation

In any attempt to correlate BFS or any other measurement with the incidence or severity of trauma, one must decide where to draw the line between categories (types or degrees of severity) of trauma or incapacitation. The NIJ has not defined such categories precisely, aIthough theNILECJ attempted to. However, the NILECJJ's specification left many ambiguities that complicate attempts to assess what BFS limit is appropriate.

The NILECJ specified that "protective garment . . . *should prevent penetration by the bullet* into the chest, abdomen, or back" and that "Any blunt trauma effects requiring surgical repair should have a mortality risk of 10% or less." in addition, "A man wearing the garment should be able to walk from the site of a shooting after being hit in the chest or abdomen by a bullet of specified caliber or weight and velocity." It was assumed that "the patient will receive *medical attention at a hospital within one hour." [104]*[a]

The statement about mortality risk, interpreted literally, does not specify an upperbound on the acceptable risk of mortality from nonpenetrating impacts that do not require surgery, including impacts that kill before surgery can be attempted and impacts that produce penetrating wounds, rather than blunt trauma even though they do not penetrate the armor. An example-the only lethal one we know of—is the case of an officer who was hit on his armor over his right upper anterior thorax [chest] by a bullet from a .45-70 carbine, which penetrated his metal nameplate before encountering the armor. His armor stopped the bullet but penetrated his skin and right lung to a depth of about 4 inches, breaking a rib. The medical examiner attributed the cause of death not to the penetration, per se, but to "The shock wave created by the missile," which "lacerated the aorta, the pulmonary artery, and the vena cava immediately adjacent to the heart, resulting in death by insanguination into the thoracic cavities. ''[130][b]

OTA interviewed three individuals involved in the formulation of the NILECJ goals (Michael Goldfarb, Nicholas Montanarelli, and Lester Shubin), and all three agreed that the goals were not intended to exclude such cases; they agreed that a more accurate rendition of the intent might be: "A bullet stopped by armor certified to withstand it should have no more than a 10-percent chance of causing trauma that is lethal, requires surgery, or renders the wearer unable to walk from the site of the shooting." OTA did not ask them whether they would distinguish between minor surgery (e.g., sutures in skin) from major surgery, but others have proposed such a distinction.

Police officers and chiefs have also expressed a desire for protection against incapacitation, particularly against being rendered unable to return fire. In his first test of his company's nylon body armor, Richard Davis made a point of demonstrating that he could shoot at targets immediately after shooting himself in the region protected by his vest. [121] The NILECJ considered this but decided not to incorporate it explicitly into the safety criterion: "Consideration had to be given to such things as . . . whether the wearer should be able to pursue his duties, returning fire if necessary after being shot. The criterion adopted by the Institute was that a man wearing the garment should be able to walk from the site of a shooting after being hit in the chest, back, or abdomen" [104] The ability to walk away was used as a proxy for other abilities, some of which-such as the ability to return fire-are more difficult to assess after the fact. It is not always necessary or appropriate to return fire, so it is problematic to determine the extent to which this goal had been achieved. But it is usually necessary or appropriate to walk from the site of a shooting (in some cases, to return fire), so it is easier to determine the extent to which this goal had been achieved.

---

a Cf. reference [1411.

b Although the speed of sound may be so low in lung tissue that the pressure wave maybe supersonic—hence a shockwave-there, [166] the pressure wave was probably subsonic (not a shock wave) in the aorta, the pulmonary artery, and the vena cava. However, even a subsonic pressure wave, if sufficiently strong, could cause the damage noted.

injuries is probably that suffered by Officer Torben Beith of the Long Beach (California) Police Department, who was hit on his armor over his upper right chest by a l-ounce, 12-gauge slug, which caused laceration requiring 8 sutures.

Thus 2 of 596 shots stopped by armor caused injuries rated AIS 4-6, and the rest did not. Let P_1' = 594/596, the proportion of all shots stopped by armor that caused no injury rated 4 or higher, and let P_z' = 2/596, the proportion of all shots-stopped by armor that caused injuries rated AIS 4-6.

P_2' is the maximum-likelihood estimate of the probability that a shot stopped by armor would cause injury rated AIS 4-6, regardless of whether the armor passed, or would pass, any test. This is called the *unconditional* probability (per shot) of injury rated

AIS 4-6; it is the historical probability of such injury averaged over all types of armor worn, wearers, and threats. Armor that passed, or would pass, a BFS test should have a lower probability than P_2' of allowing a shot that it stops to cause injury rated AIS 4-6, and armor that failed, or would fail, the same test should have a higher probability than P_2' of allowing such an injury. The purpose of separate-sample logistic discrimination is to estimate these *conditional* probabilities.

If the model that results is used to predict future risks, the confidence limits on prediction errors would be as estimated (see discussion below), if future threats and armor are statistically like past ones. If not, the prediction errors could be greater. If there is particular concern that the future may differ significantly from the past, either of two statistical methods may be used to address the problem directly. One is to use time-series methods<. g., to estimate probability of injury as a function both of BFS and year. This would be a relatively simple elaboration of the analysis presented here. Another option would be to use Bayesian inference based in part on subjective estimates [11].

We consider first the problem of estimating, based on reenactment results, the probability that a shot stopped by armor would cause injury rated AIS 4-6, given that the armor (or armor of the same model) passed (after the fact) a 2-shot BFS test using bullets of the type the vest stopped in the assault impacting at the velocity at which the bullet impacted in the assault. Estimating the probability of injury in some other range of severity may be done in the same reamer.

Let n_1 be the number of (2-shot) *tests* conducted in the lab to reenact the shots that caused no injury rated AIS 4 or higher; n_1 is the number of *controls.*

- The shots to be reenacted could be chosen exhaustively-i.e., one test could be performed in the lab to reenact each shot that was stopped by armor and caused no injury rated AIS 4 or higher. There are 594 such shots, so exhaustive sampling would require as many tests (n_1 = 594), a total of 1,188 shots.
- Alternatively, the shots to be reenacted could be selected randomly, so that each of the 594 shots stopped by armor without causing injury rated AIS 4 or higher has the same probability of being selected a priori. One could choose n_1 in advance, perhaps based on the budget for reenactment, and continue the random sampling, with or without replacement, until a program of n_1 tests is obtained. If the sampling is done with replacement, 2 or more of the n_1 tests might reenact the same shot stopped by armor. This is not redundant, because the BFSs may differ, and reenacting a shot several times tends to average out such variation.

Similarly, let n_2 be the number of tests conducted to reenact the shots that caused injury rated AIS 4 or higher; n_2 is the number of *cases. The* shots to be reenacted could be chosen exhaustively or randomly.

Because only 2 shots caused injury rated AIS 4 or higher, it is feasible and desirable to conduct more than 2 tests; the shots to be reenacted could be selected randomly with replacement. [Alternatively, each shot that caused such injury could be reenacted the same number of times.]

In contrast, 594 shots caused no injury rated AIS 4 or higher. Because of the cost, it may not be desirable to perform 594 tests (1,188 shots) in reenactment—nor is it necessary. The number of controls, n_1, may be chosen to be comparable to the number of cases, n_2, although this is not necessary. If n_1 and n_2 are not both greater than O, there can be no statistical confidence in some of the resulting estimates.

METHOD OF ANALYSIS

"Whoever, therefore, deals with the problem of modern armor will go far astray if he does not consider on generous lines the index of probability."
— Bashford Dean, 1920. [53]

This section describes the constrained maximum-likelihood estimation (defined below) of values for the parameters of a logistic model that could be used to estimate the conditional probability of injury —viz., the probability that a shot stopped by armor would cause injury rated AIS 4-6, given that the armor (or armor of the same model) passed, or would pass, a BFS test using bullets of the type the vest stopped in the assault impacting at the velocity at which the bullet impacted in the assault.

Let $n = n_1 + n_2$ be the total number of reenactments. Let $P_1 = n_1/n$, the proportion of reenactments that reenact shots stopped by armor that caused no injury rated 4 or higher, and let $P_2 = n_2/n$, the proportion of reenactments that reenact shots stopped by armor that caused injury rated 4 or higher.

Of the n_1 tests reenacting the shots causing no injury rated AIS 4 or higher, let $n_{1(0)}$ be the number that result in a pass (viz., BFS no greater than 44 mm on both shots) and let $n_{1(1)}$ be the number that result in a failure. Of the n_2 tests reenacting the shots causing injury rated AIS 4 or higher, let $n_{2(0)}$ be the number that result in a pass, and let $n_{2(1)}$ be the number that result in a failure. Let $n_{(0)}$ be the total number of reenactments that result in a pass, and let $n_{(1)}$ be the total number that result in a failure, i.e., $n_{(0)} = n_{1(0)} + n_{2(0)}$ and $n_{(1)} = n_{1(1)} + n_{2(1)}$.

Finally, let $p_{1(0)}$ be the probability that a stopped shot would cause no injury rated AIS 4-6, given that the armor passed the BFS test, $p_{1(1)}$ the probability that a stopped shot would cause no injury rated AIS 4-6, given that the armor failed the BFS test, $p_{2(0)}'$ be the probability that a stopped shot would cause injury rated AIS 4-6, given that the armor passed the BFS test, and $p_{2(1)}'$ the probability that a stopped shot would cause injury rated AIS 4-6, given that the armor failed the BFS test.

Let $p_{1(0)}'$ $p_{1(1)}'$ $p_{2(0)}$ and $p_{2(1)}$ be defined similarly EXCEPT they apply only to the shots (and corresponding armor and victims) selected for reenactment.[8] We call these the within-sample conditional probabilities, and we call $p_{1(0)}'$, $P_{1(1)}'$, $p_{2(0)}'$, and $p_{2(1)}'$ the corresponding population probabilities, because they refer to the entire "population" of shots stopped by armor.

The estimate of $p_{1(0)}$ is simply $n_{1(0)}/n_{(0)}$, the fraction of the $n_{(0)}$ reenactments that resulted in a pass that reenacted shots that caused no injury rated AIS 4-6. Similarly, the estimate of $P_{1(1)}$ is simply These are called constrained maximum-likelihood estimates, because they retie the likelihood that the (reenactment) results actually observed would occur, subject to the constraint that, given $n_{(0)}$ passes and $n_{(1)}$ failures, the expected proportion of shots causing no injury rated AIS 4-6 must be P_1 (i.e., n_1/n). We use *italics* to denote constrained maximum-likelihood estimates of probabilities (or odds). Thus $p_{1(0)}$ is the constrained maximum-likelihood estimate of pi(o), and equals $n_{1(0)}/n_{(0)}$. Similarly, $P_{2(0)} = n_{2(0)}/n_{(0)}$ and $P_{2(1)} = n_{2(1)}/n_{(1)}$.

To adjust estimated within-sample risk to apply to the population, it is convenient to use odds instead of probabilities. Let O
odds for injury rated AIS 4-6, given the BFS test is passed, i.e.,

$$O_{2(0)} = P_{2(0)}/(1-P_{2(0)}) = p_{2(0)}/ /P_{2(0)}) = P_{2(0)} / P_{1(0)}"$$

Similarly, let $O_{2(1)}$ denote the within-sample odds for injury rated AIS 4-6, given the BFS testis failed; let $O_{2(0)}$ and $O_{2(1)}$ denote the the constrained maximum-likelihood estimates of these odds; and let $O_{2(0)}'$ and $O_{2(1)}'$ denote the the constrained maximum-likelihood estimates of the odds $O_{2(0)}'$ and $O_{2(1)}'$ for injury to the population. $O_{2(0)}'$ and $O_{2(1)}$ are calculated from $O_{2(0)}$ and $O_{2(1)}$ using the formulae

$$O_{2(0)}' = O_{2(0)} = p_1 p_2'/(p_1 p_2$$
$$O_{2(1)}' = O_{2(1)} P_1 P_2' / (P_1' P_2)$$

The estimated probabilities $p_{2(0)}'$ and $p_{2(1)}'$ may be calculated from the estimated odds $O_{2(0)}'$ and $O_{2(1)}'$ using the formulae

$$P_{2(0)}' = O_{2(0)}'/(1 + O_{2(0)}')$$
$$P_{2(1)}' = O_{2(1)}'/(1 + O_{2(1)}')$$

These estimates could be very inaccurate, so it is important to calculate confidence limits on possible values of $p_{2(0)}'$ and $p_{2(1)}'$. In general, confidence limits on $p_{2(0)}$ will depend on $p_{2(1)}$,' and vice versa. For example, if none of the n_2 reenactments of injurious shots results in a pass (i.e., if $n_{2(0)} = 0$), then the reenactments would provide 100C-percent confidence that $p_{2(0)}'$ is no greater than the confidence limit

$$CL = P_2' P_{2(1)} [1 - (1 - C)_1/n_2]/[p_{2(1)}' - p_2' (1 - C) 1/n2]$$

which is called the upper 100C-percent confidence limit on $p_{2(0)}$. The confidence level C is the minimum probability with which the reenactment results ($n_{2(0)}$ and n(o)) would have led to a larger estimate $p_{2(0)}' = n_{2(0)}/n_{(0)}$ than the one obtained $P_{2(0)}'$ = o), if $p_{2(0)}$ were as large as CL, or larger.[9]

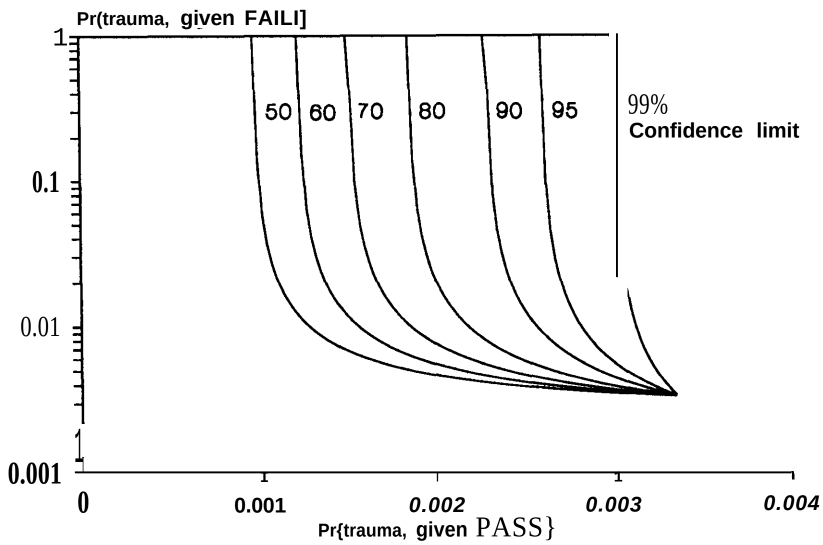
As an example, figure D-1 shows the upper 50-, 60-, 70-, 80-, 90-, 95-, and 99-percent confidence limits on $p_{2(0)}'$ ("Pr{trauma, given PASS}") for a range of values of $p_{2(1)}$(' 'Pr{trauma, given FAIL} "), for the case $n_2 = 2$ and $n_{2(0)} = 0$.[10]

[8] OTA is indebted to Keith Eberhardt of NIST for pointing out the importance of this distinction.

[9] The upper 100C-percent confidence limit CL on $p_{2(0)}'$ may be obtained for any value of $n_{2(0)}$ by solving the equation obtained by letting C equal the binomial cumulative distribution function of parameters n_2 and $p = (CL/P_2')(p_{2(1)}' - P_2')/(p_{2(1)}' - CL)$, evaluated at argument $n_{2(0)}$.

[10] We assume $p_{2(1)}'$ does not exceed $P_{2(0)}'$.

Figure D-l-Confidence Limits on the Probability of Death or Life-Threatening Injury, Given the BFS Test is Passed



For the case $n_2 = 2$, $n_{2(0)} = O$ (see text).
SOURCE: Office of Technology Assessment, 1992.

For the same case ($n_2 = 2$, $n_{2(0)} = O$), figure D-2, which is based on the 90-percent confidence curve of figure D-1, shows an exact 90-percent confidence region for $p_{2(o)}$' and $P_{2(1)}$'. That is, the data provide 90-percent confidence that $P_{2(0)}$' and $p_{2(1\sim}$ are in the shaded region shown. If they were at the upper left-hand corner of the region, the test would have perfect discrimination; if they were at the lower right-hand corner of the region, the test would have no discrimination.

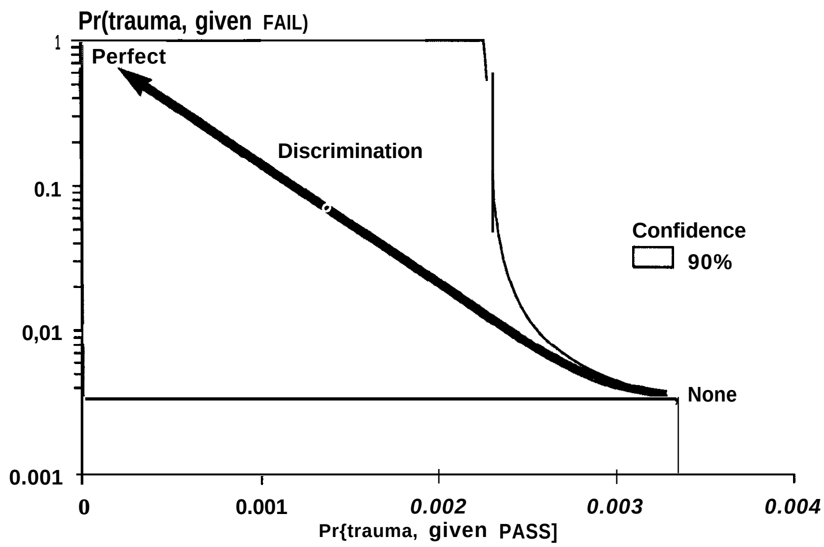
In some cases, separate-sample logistic discrimination may be used to estimate the probability that a stopped shot would cause injury, as a function of the backface signature measured in a l-shot test. The procedure is more complicated and may not always be applicable, but if it is, it allows the estimated probability of injury to be plotted versus backface signature; see Logistic Model for Probability of Injury versus BFS, below.

RESULTS OF DUPONT-SPONSORED REENACTMENTS

In October 1991, reenactments of 22 assaults were performed by H.P. White Laboratory, Inc. They were sponsored by the E.I. duPont de Neymours Co., Inc. (hereinafter "DuPont") and observed by OTA at DuPont's invitation. Dupont sought to reenact all known assaults (described above) in which death or serious injury was caused by a stopped bullet. In addition, DuPont wanted to reenact ''magnum saves—shootings in which the victim was saved by armor from penetration and from death or serious injury by the stopped shot, and in which the assailants weapon and ammunition and the victim's armor are believed likely to cause a large BFS in a reenactment. The backface signatures and penetrations produced in the reenactments are summarized in table D-2.[11]

11 The table includes the backface signatures obtained in the reenactment of the shot from a Winchester Model 37 shotgun with a sawed-off 14-inch-long barrel that struck Mr. Herman Joyner at 6-inch range, but it excludes backface signatures of 20,22, and 28 mm produced at longer range (5 m, as specified by the .03 standard) using the same gun, and backface signatures of 31,34,34, and 37 mm produced at a range of 5 meters using a testbarreland PPAAtest ammunition and velocities. OTAdoubts that the impact velocities in the excluded reenactments approximate the impact velocity of the slug that hit Mr. Joyner. DuPont directed HPWLI to try the different ranges, barrels, and ammunition in an attempt to strike a balance between the desire to recreate the impact velocity and the desire to measure it. OTA considers the recreation of the impact velocity most important.

Figure D-2-Confidence Regions for the Probability of Death or Life-Threatening Injury, Given the BFS Test is Passed, versus the Probability of Death or Life-Threatening Injury, Given the BFS Testis Failed



For the case $n_2 = 2$, $n_{2(0)} = 0$ (see text).

SOURCE: Office of Technology Assessment, 1992.

Almost a quarter of the shots reenacting shots that armor stopped in service penetrated the armor in the reenactments. This may be partly attributable to the fact that in the reenactments the shots impacted at normal incidence, at which penetration probability is expected to be greatest; in the assaults the shots generally did not impact at normal incidence. There may be other reasons. (See box D-2—Penetrations in BFS Testing.) We score any penetration as a failure.

Because the magnum saves were selected neither randomly nor exhaustively from all the saves, they cannot be used for separate-sample logistic discrimination.[12] 13 However, if all other saves were reenacted, the results could be combined with the results of the magnum saves to produce an exhaustive set of reenactments of saves, which could be used, and we expect that the results of the magnum save reenactments would be the most influential of the results.[14] (See box D-3-Magnum Saves.)

All (i.e., both) shots producing trauma rated AIS 4 to 6 were reenacted, but the one producing AIS 4 to 5 was reenacted thrice (6 shots total), while the one producing AIS 6 was reenacted only once (2 shots total). The different numbers of reenactments per injurious shot did not result from sampling with replacement. If we discard 2 of the 3 reenactments of the shot producing trauma rated AIS 4 to 5,[15] the remaining reenactment, together with the reenactment of the shot that produced trauma rated AIS 6,

12 OTA is indebted to Lane Bishop of Allied-Signal for pointing this out.

13 DuPont and OTA believed that the results of reenactments of "magnum saves" would be particularly informative and should have particular influence on the conclusions. Indeed, they should, if the magnum-save cases were among cases selected for reenactment by random or exhaustive sampling. However, OTA staff had identified separate-sample logistic discrimination as an appropriate method of statistical analysis only a few days before the reenactments began, and did not until later appreciate the importance of randomly selecting the cases to be reenacted.

14 When estimating probability of trauma as a function of BFS depth (as distinct from BFS category), it may be desirable to exclude from the analysis, at some point, results (BFSs) that lead the model being fitted to predict odds, the natural logarithm of which is less than -3 or greater than 3; see [9], p. 31. This is equivalent to excluding BFSs that lead the model being fitted to predict probabilities smaller than 0.05 or greater than 0.95. If this is not done, the estimates of the regression coefficients from which the estimated probabilities are calculated may be unreliable. This does not necessarily make the estimated probabilities inaccurate, but it complicates the assessment of their accuracy and reliability.

Saves from bullets of lower energy than the maximum for which the armor is rated are likely to produce relatively small BFSs that would be excluded by this criterion; results from magnum saves would be retained and would be influential.

15 Because the results are all the same—failure—it does not matter which result is retained.

Table D-2—Backface Signatures and Penetrations Produced in Reenactments

AIS[a]	Victim	BFS(s) [mm]	Penetrations
6	Anonymous		2
4-5	Power		6
3	Draper	64	3
0-2	Beith	54	6
"	Bartlett	39,42,41,41	
"	Beijin	29,32,32,34	4
"	Bennetts	35,37,28,28	
"	Bohne	36,30,32,33,34,30,32, 35	
"	Gazeik	25,27,32,38	
"	Hyatt	42,44,39,38,32,40,34	
"	Joyner	71,78,76,72,80	
"	Knight	33,28,31,30	
"	Martin	31,34,34,37	
"	Mulata	44,42,39,38,39	14
"	Norris	38,33,34,37,41	
"	Page	35,33,20,24	
"	Perez	22,35,34,24	
"	Seward		4
"	Solheim	22,23,24,26	
"	Stewart	30,28,28,29	
"	Wengert	41,39,38,43	
"	Yearick	49,53,53,56,47,55,49, 47	

NOTE: The total number of shots fired to reenact each felonious shot differs from case to case. This table lists all shots except for seven shots fired to reenact the assault on Mr. Joyner, which OTA estimates did not have an impact velocity comparable to that in the assault. (See fn. 1 1.)

a Abbreviated Injury scale:
- 6: fatal
- 5: critical-survival uncertain
- 4: severe, life-threatening-survival probable
- 3: severe, not life-threatening
- O-2: not severe

forms a set of 2 reenactments of shots selected by exhaustive sampling from the results available. This set ($n_2 = 2$) can be used for separate-sample logistic discrimination.

Similarly, if we discard 1 of the 2 reenactments of the shot that caused trauma rated AIS 3, the remaining reenactment, together with the reenactments of the shot that produced trauma rated AIS 4 to 6, would form a set of 3 reenactments of shots selected by exhaustive sampling, which could be used for separate-sample logistic discrimination to estimate the risk of injury rated AIS 3 to 6.

Table D-3 (top) is a statistical summary of the results in table D-2, by BFS category. Table D-3 (bottom) shows the subset of results we deem usable for separate-sample logistic discrimination, counting each 2 shots as one reenactment. To estimate the risk of trauma rated AIS 4 to 6, we use only the top 2 rows, which include a total of 2 reenactments ($n_2 = 2$), both failures ($n_{2(0)} = O$, $n_{2(1)} = 2$). To estimate the risk of trauma rated AIS 3 to 6, we would use all 3 rows: $n_2 = 3$, $n_{2(0)} = O$, $n_{2(1)} = 3$.

Table D-3-Summary of Results

AIS	Shots	BFS O-44 mm	BFS 44+ mm
All results			
6	2	0	2
4-5	6	0	6
3	4	0	4
0-2	111	69	42

AIS	Reenactments	BFS test result: Pass	BFS test result: Fail
Results used for analysis			
6	1	0	1
4-5	1	0	1
3	1	0	1

NOTE: Injuries requiring only skin sutures are rated AIS O-2. "44+ mm BFS" includes penetrations. Each reenactment consists of 2 shots.

SOURCE: Office of Technology Assessment, 1992.

ANALYSES

Risk Associated With the Current 44-mm BFS Limit

The within-sample probability $p_{2(1)}$ may be estimated from the data in table D-3 (bottom): $P_{2(1)}$ = %(@(l) = 1/1 = 1. However, $p_{2(0)}$ may not be estimated as long as $n_{(0)} = O$. Calculating a constrained maximum-likelihood estimate $p_{2(0)}$ will require more data (i.e., more reenactments\; so will adjusting the estimate $p_{2(1)}$ to apply to the population.

One may nevertheless calculate confidence limits on $p_{2(0)}$'; they depend on $p_{2(1)}$' as well as the data n_2 and ~(0). Because all results of the n_2 reenactments of injurious shots were failures (~$_{(0)}$= O), the upper confidence limits on $p_{2(0)}$' are those shown in figures D-1 and D-2. They indicate that $p_{2(0)}$' is less than about 0.0025 unless the test has little discrimination.

Figure D-3 shows the 90-percent confidence region analogous to that of figure D-2, but in this case for the probability of death (AIS 6). This would be of interest to those who consider death to be the only unacceptable category of trauma.

The method of constrained maximum-likelihood estimation used here could be elaborated to estimate the risk of excessive trauma or incapacitation for each of several categories of wearers (e.g., men and women), given the backface signature measured in a ballistic test. However, such additional stratification would degrade the statistical significance with

Box D-2-Penetrations in BFS Testing

The reenactments of three assaults (Anonymous, Power, and Seward) produced only penetrations. It is reasonable to ask what factors might explain the penetration of armor on clay and the nonpenetration of the same or similar armor on the human victim. One possibility to be considered is that armor is more easily penetrated on clay than on a human torso under conditions of wear. This begs a related question: is armor more easily penetrated on some areas of a human torso than on others?

It is difficult to settle these questions at present, partly because of the limited data available, and partly because other factors may have influenced the results. For example, the bullet that killed the anonymous officer without penetrating his armor was first slowed, and perhaps deformed, by his metal nameplate, which it shattered. In the reenactment, no nameplate was used, and the bullet penetrated.

Some speculate that the probable ballistic limit (V_{50}) of armor on a human torso (especially the abdomen)might be comparable to that measured in tests with gelatin backing and between that measured in tests with clay backing (which is denser and less resilient than soft tissue) and that measured in tests with air backing (which is less dense than soft tissue).

In research sponsored by the NILECJ, the Army's Chemical Systems Laboratory found the V_{50} for .22-caliber bullets impacting 7-ply Kevlar-29 armor to be 1,096 ft/s on goat abdomen, 1,115 ft/s on goat thorax, 1,109 ft/s on 20-percent ballistic gelatin,[a] and 1,079 and 1,088 ft/s on 2 samples of Roma Plastina No. 1 modeling clay that had been stored under different conditions. [1 12] The V_{50} for gelatin backing was between the values for goat abdomen and thorax, and V_{50} for the clay samples were slower than those for goat abdomen and thorax, i.e., the armor was more likely to be penetrated on clay than on goat abdomen or thorax, The Army concluded that agreement was good enough to recommend the use of clay as a backing for armor testing, citing its availability and ease of use compared to gelatin.

In other research sponsored by the NILECJ, the Aerospace Corp. compared V_{50}s measured using clay and air backing. They found V_{50}s slower with clay backing than with air backing--i.e., other conditions being identical, the armor was more likely to be penetrated on clay than with no backing. [8)

More recently, NIST has conducted ballistic tests for the NIJ to measure the V_{50}s of armor test panels made of various numbers of plies of treated or untreated Kevlar-129, or Spectra Shield, impacted by 9-mm Full Metal Jacketed or .357 Magnum Jacketed Soft Point bullets on clay or air backing. For one bullet-armor combination, the clay-backed and air-backed V_{50}s were essentially identical for each panel thickness tested For another bullet-armor combination (9-mm FMJ v. untreated Kevlar-129), the clay-backed V_{50}s exceeded the air-backed V_{50}s at each panel thickness tested; the difference would be almost 200 ft/s for 7-ply untreated Kevlar-129, based on interpolation. For 2 other Imllet-armor combinations, the air-backed V_{50}s exceeded the clay~backed V_{50}s at each panel thickness tested.[b] These results may indicate that whether armor is more easily penetrated on day or air depends on bullet-armor-backing interactions not yet understood another possibility is that the apparent dependence on backing is not statistically significant. However, it does seem consistent across samples of different thicknesses, although varying with bullet-armor-backing combination. NIST is still analyzing the results.

Angle of incidence-which is O degrees in each reenactment-may also affect penetration. Under laboratory conditions, increasing the angle of incidence (as defined in the .03 standard) decreases the probability of penetration for most, but not all, bullet-armor combinations tested. Officer Power estimated that the slug that struck his armor had an "angle of impact" of approximately 30 degrees.[c] This probably decreased the probability of penetration in the assault, compared to that in the reenactments, in which all six shots penetrated

Recent tests conducted by H.P. White Laboratory, Inc., for Allied-Signal illustrate how the fraction of slugs that penetrate fabric armor decreases as the angle of incidence increases, under otherwise similar conditions, In these tests, 437.5-grain 12-gauge slugs impacted shootpacks (test panels) made of 31 plies of Kevlar 129[R] fabric, style 704, at about 1,600 ft/s. Of the 9 slugs impacting 1 shootpack at 0 degrees, 6 (67 percent) penetrated. Of the 12 slugs impacting 2 other shootpacks at 45 degrees, only half (50 percent) penetrated. Of the 6 slugs impacting another shootpack at 60 degrees, none (() percent) penetrated.[d]

[a] That is, the dry gelatin used constituted 20 percent Of the weight of the gelled product.

[b] Daniel E. Frank, NIST/OLES, personal communication, Sep. 17, 1991.

[c] Questionnaire completed by DuPont based on telephone interview of victim.

[d] Bishop, Allied-Signal, personal communication, Mar. 13, 1992.

Box D-3—"Magnum Saves"

Representatives of DuPont have indicated that their selection of cases to reenact emphasized so-called "Magnum Saves," i.e. those in which the assailant's weapon was a .357,.41, or .44 Magnum pistol. These cases are dramatic instances of vest performance, especially because most of them involve vests not certified to stop such high-energy rounds. In fact, many of the DuPont re-enactments feature vests that were not certified at all. Most of these were non-waterproofed fabric vests that would almost certainly fail the wet test. It is claimed that these vests account for more than their share of saves-the proportion of saves involving such vests exceeds the proportion of such vests in the population of extant vests.[a]

Saves of officers assaulted by shotguns are also of particular credit to the vest, especially in cases in which the shotgun fired a slug, or in which the range was so short that the pellets had not significantly spread out before they hit the vest.

Reenactments of "magnum saves" are likely to have particular influence on the conclusions drawn from an analysis of reenactments, because they are likely to result in backface signatures associated (by the regression procedure) with probabilities not smaller than 0.05 nor larger than 0.95 and if so would not be discarded by the regression procedure. Reenactments of shootings by low-energy bullets that caused death or serious injury (if there were any) would likewise be particularly influential. lt would seem to be economical to select these cases for reenactment and not attempt to reenact the many more shots from which officers were saved, the data from which would likely be discarded by the the regression procedure at some stage, Unfortunately, some means is needed to estimate what proportion of such shots are represented by those causing large backface signatures but little injury. The simplest approach is to select shots to be reenacted randomly (with replacement) from each trauma category-and ignore most of the results later. Further research might devise other techniques that could use data, including previously collected data, more economically.

[a] This suggests to some that the non-waterproof vests have a higher wear rate than waterproof vests, suggesting in turn that they are more comfortable. Other interpretations are possible. For example, a vest truthfully advertised to have passed the dry test during manufacturer-sponsored testing may not have passed it on the first try, whereas NIJ tests each model only once. Thus vests intended for NIJ testing may be more conservatively *designed.*

which risk could be inferred from a limited set of data. Additional stratification is a logical next step to be undertaken when additional reenactments have been performed. (See box D-4-Control for What?.)

Logistic Model for Probability of Injury versus BFS

The risk associated with any BFS limit could be estimated by the procedure used above, if the reenactment results (BFSs and penetrations) are resorted into redefined categories of passing and failing, based on the hypothetical BFS limit. Estimates of the probability of various degrees of injury, and confidence limits on these, could then be calculated as above. However, this would require many tables and figures to display the estimates and confidence regions for many alternative BFS limits.

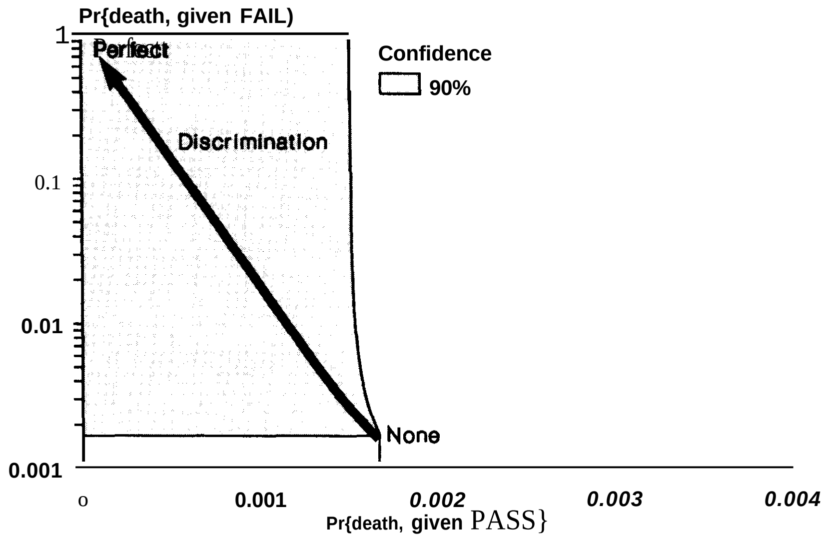
It maybe more convenient to use separate-sample logistic discrimination to obtain a logistic model that estimates the probability of injury associated with any BFS. (This would be called separate-sample logistic regression.) The model would befitted to the results (BFS or penetration) of l-shot reenactments (that is, each shot performed in reenactment would be considered a separate reenactment).

This approach will not work, however, if a condition called "complete separation of sample points" occurs. [9] This would occur, for example, if all reenactments of shots that caused injury (of the severity of interest) produced only penetrations, or penetrations and BFSs larger than any produced by reenactments of shots that did not cause such injury. This was the case with the reenactments described above; it necessitated the more complicated categorical analysis described above, which is applicable when the sample size (number of reenactments) is small.

If complete separation of sample points does not occur, logistic regression could be used to obtain a 2-parameter logistic model that estimates probability of injury based on (1) whether penetration occurs in the BFS test, and (2) the BFS, if penetration does not occur.

It could also be used to obtain a l-parameter logistic model that estimates probability of injury based on the *effective BFS,* which we define as the

Figure D-3-Confidence Regions for the Probability of Death, Given the BFS Test is Passed, versus the Probability of Death, Given the BFS Test is Failed



SOURCE: Office of Technology Assessment, 1992.

measured BFS, if penetration does not occur, or a BFS equivalent (in risk of injury) to penetration, if penetration does occur. The BFS equivalent to penetration would be determined from a 2-parameter logistic model as described above; it would be the BFS for which the model predicts the same probability of injury (by a stopped bullet) that it predicts if a penetration occurs in the test.

A l-parameter logistic model could be used to determine a BFS limit (i.e., a limit on effective BFS) consistent with a specified estimated probability of injury. Moreover, confidence limits on the probability of injury as a function of BFS maybe calculated from the estimated dispersion (variance and covariance) of the errors in the estimates of the regression coefficients that determine the logistic model. Such confidence limits could be used to calculate the largest BFS limit that would limit probability of injury to a specified maximum acceptable value with a specified minimum acceptable statistical confidence.

Actually, logistic regression estimates the asymptotic dispersion—the limit that the dispersion would approach if the number of samples (i.e., reenactments) increased without bound, in which case the probability distribution of the errors in the estimates of the regression coefficients would approach a normal (i.e., gaussian) distribution. Unfortunately, there is no generally-accepted criterion for the number of samples required for the actual distribution to be acceptably asymptotic. A widely used rule of thumb is that 30 or more samples should suffice, but many more samples may be necessary if one demands high confidence in a small upper confidence limit on probability of injury.

If desired, confidence bounds on the BFS corresponding to any specified probability of injury (e.g., the maximum acceptable risk) may be calculated, using Fieller's theorem, [117] from the estimated regression coefficients and the estimated asymptotic dispersion of errors in their estimates. Such confidence limits on the explanatory variable (BFS in this case) are valid only in the limitofa''large' number of samples, but have been used (in other applications) when only a few tens of samples are available.

Sensitivity Analysis

The fact that only a very small fraction of shots stopped by armor have produced serious injury (regardless of whether the armor "passed" a reenactment) indicates that there is little risk that a bullet, slug, or shot stopped by armor will cause serious injury-unless new armor is distinctly different (ballistically) from the variety of past armor or unless the spectrum of weapons and ammunition *used against police officers* changes dramatically.

There is more uncertainty about how much selection based on passing a BFS test reduces the

Box D-4-Control for What?

It is possible to estimate a probability of death or injury that depends not only on the backface signature produced in a test, but also size and sex of the wearer,[a] the angle of impact, and other factors. This is called "controlling for" these factors in the analysis of the reenactments. It is done by stratification-i.e., grouping the data into categories, called "strata, ''in each of which the factors are similar, and estimating the risk in each stratum as a function of BFS.

Although this may be useful for some purposes, stratification reduces the data that can be used to estimate the risk in each stratum, so the resulting estimates may have greater uncertainty than the estimate that depends only on BFS, averaging over all victims, armor, and assault conditions. In any case, this estimate of averaged risk as a function of BFS will probably be the relevant one for assessing the validity of a BFS test, because, on legal and political grounds, it is doubtful that a statement of safety goals (the criteria for validity) would accept a greater risk for women than for men, or vice versa, or a greater risk for small wearers than for large wearers, or vice versa-although there is no scientific reason to avoid stating such goals.

There is, however, a scientific reason to avoid stating a safety goal applicable only to the "worst-case" situation. For one thing, the hypothetical worst-case combination of factors is not known with certainty. Reenactments could help predict them "in principle," but the prediction might be absurd, For example, other things being equal, predicted risk may increase with decreasing body weight of the wearer. The worst case would be a wearer who weighs nothing! Similarly, even if everyone agreed that, other things being equal, a bullet impacting at normal incidence is worse than a sirnilar bullet impacting at an angle, we doubt that there would ever be an assault (which could be reenacted to validate the estimate) in which the bullet could be proven to have impacted precisely at normal incidence. This leads to the main scientific objection to a safety goal applicable only to a "worst-case" situation: no test could ever be proven to be a valid guarantor of such safety (in scientific terms: no test could be logically positive), because there is zero probability that a case suitable for reenactment would ever occur.

Regardless of how the strata are defined, it is important that in each stratum the cases selected for reenactment be representative of all the cases that have occurred; random sampling of cases would do this, on the average, and exhaustive sampling would do it with certainty. [9]

However, there are practical obstacles to achieving this goal and sometimes a reason to deviate from it. Some censoring of cases maybe necessity because, for example, data or resources (e.g., similar ammunition) necessary for reenactment are lacking. Aside from this, it maybe desirable to exclude from the analysis (at some point) results (e.g., BFS) that lead the model being fitted to predict probabilities smaller than 0,05 or greater than 0.95. If this is not done, the estimates of the regression coefficients from which the estimated probabilities are calculated maybe unreliable. [9] This does not necessarily make the estimated probabilities inaccurate, but it complicates the assessment of their accuracy and reliability.

Other types of sampling could be used, if a statistical test shows that the results are representative. Such an approach would be valuable if it allowed use of data from reenactments already performed of cases that were not selected randomly. The problem is deciding in what ways the samples should resemble the population from which they are drawn and arguing persuasively that representativeness in other respects is irrelevant.

[a] As a practical matter, controlling for gender would be difficult, because only 5 incidents are known in which a female officer was shot and hit on the vest. T.E. Backer, duPont, personal commune ation, Mar. 13, 1992.

risk of injury by a stopped bullet of the type and energy used for the test, compared to the risk if armor is not subjected to such selection. What we know about the correlation of BFS with lethality or life-threatening injury to humans is based on fewer than a handful of cases-and the fact that hundreds of victims did not end up as ''cases. " This analysis of reenactments provides initial estimates of risks and their uncertainties; reenactment of additional cases would narrow the confidence intervals derived here and possibly change the maximum-likelihood estimates significantly.

It would also be valuable to identify and reenact some assaults on female officers and small officers, to determine whether the results depend significantly on the sex and size (or weight) of the wearer. However, assaults on female officers are rare: as of March 13, 1992, only 4 female officers are recorded in the IACP/DuPont Kevlar Survivors' Club fries as having been saved by armor from gunfire. Another female officer, recorded in DuPont's Casualty Reduction Analysis files, was killed by a head wound moments after her armor stopped a rifle bullet. We know of no female officer killed by a bullet stopped by her armor.

One might expect that victims of the two or so assaults that killed or seriously injured officers would include a disproportionate number of small officers. In fact, they were not unusually large: the anonymous officer who was killed was a 25-year-old 6'0" 160-pound male; Officer Bryan Power was 20 years old and slender.[16] Although this sample is also small, it is representative; there were no other cases of such severe trauma by nonpenetrating bullets to reenact.

16 His weight was not recorded in a medical report prepared after initial surgery, which included a temporary colostomy. He was described as "slender" in a medical report dated almost 2 months later. See also photo as "Save No. 329" in [120].