Appendix C

Issues

Contents

Figures

Table

Issues

INTRODUCTION

This appendix discusses prominent policy issues and technical issues related to standardizing the assessment of protection provided by body armor, and in particular to the National Institute of Justice's Standard 0101.03, *Ballistic Resistance of Police Body Armor*. The policy issues relate to the scope and safety goals of such standardization; the technical issues concern whether provisions of the current NIJ standard achieve them, and whether proposed revisions would improve the standard.

POLICY ISSUES

The major policy issues in the current debate are

1. whether compliance with the Federal standard should be mandated;
2. whether the purpose of standardization is to inform, or to protect, consumers;
3. the threats from which protection is to be certified, and whether manufacturers, consumers, or the government should specify them;
4. the types of injuries to be prevented;
5. the maximurn acceptable probability of failing to prevent such injuries;
6. whether the purpose of standardization is to assure reliability of product performance or merely adequacy of design; and
7. whether the body armor test procedure ought to be within the technical capability of individual police departments ("a field test' '), or instead a lab test of whatever complexity is necessary to meet policy goals.

Issues 3, 4, and 5 are discussed in volume 1 and appendices B and D; they are not discussed further here.

Should the Standard Be Voluntary or Mandatory?

Compliance with NIJ-STD-0101.03 is voluntary: manufacturers may make and sell body armor without testing it for compliance with the standard—or even if it is tested and fails. But many customers value certification of compliance, so major manufacturers offer certified armor. Some offer uncertified armor as well, and it sells. The current regime of voluntary compliance allows purchasers who demand it to buy armor certified to comply with a governmental standard in which they have confidence, but it does not prevent customers who do not demand such certification from buying whatever they want.

The requirement that the vest perform properly while wet showcases this feature of voluntary-compliance tests. Manufacturers who believe they would benefit from a governmental "seal of approval" can participate in the NIJ's body armor program, while those who see the wet-testing requirement as unnecessary and onerous can (and do) sell vests that would not pass the wet test. If customers find these vests to be better in some other way (perhaps comfort), they can go ahead and buy them.

The voluntary system thus affords the manufacturer and the consumer alike considerable freedom, while allowing for a governmental role in the assessment of body armor. A shortcoming of the current regime is that it allows manufacturers to certify compliance without concomitant NIJ certification of compliance. Manufacturers can, for example, perform the test themselves, or have a test laboratory do so under contract. If the samples of a model of armor pass the test, the manufacturer can truthfully certify on the labels of other samples that they comply with NIJ Standard 0101.03, even if the NIJ's Technology Assessment Program Information Center (TAPIC) has never seen samples of the model before and does not list the model on its Consumer Product List of models it certifies to comply. Consumers may not understand the distinction between certification of compliance by a manufacturer and certification by TAPIC, which will not certify armor unless its testing complies not only with the standard but also with several additional conditions, which manufacturers are not obliged to observe.

Armor of models certified to comply with NIJ Standard 0101.02 but failing to comply with NIJ Standard 0101.03 are still offered for sale, their labels truthfully certifying compliance with NIJ standard 0101.02. A mandatory-testing regime with regulatory authority vested in a body such as the NIJ would clarify many of these gray areas. H.R. 322, the Police Protection Act of 1991, was introduced in the 102d Congress to provide such a regime, as was H.R. 4830/S. 2639 in the 101st Congress.

Choosing between voluntary and mandatory testing entails a great many value judgments. Some argue that testing and compliance with standards ought to be mandatory for body armor, just as it is for automobiles. On the other hand, there is also considerable sentiment against Federal regulation of equipment used by local law-enforcement agencies.

Selection of mandatory testing leads to a number of secondary issues involving enforcement-ought the regulatory body go out into the marketplace, buy random vests, and test them? What should be the reaction of the regulatory body when signs of false claims appear-how should the right of the manufacturer to due process be squared with the right of the consumer to be protected by the standard?

While selection of voluntary-compliance testing eliminates some enforcement issues, it renders others much more complex. Clearly a manufacturer ought not to make false claims regarding a product, and, if any armor manufacturer does, he could be prosecuted under fair-trade statutes,[1] and possibly for wire or mail fraud. Though compliance with the NIJ body armor standard is voluntary, the NIJ, through TAPIC, endeavors to ensure that compliance is not claimed falsely and has disseminated a few "Body Armor Safety Alerts" to local law-enforcement agencies nationwide over the National Law Enforcement Telecommunications System (NLETS) when it suspected that compliance was being claimed falsely.

NIJ-STD 0108.01, **a** voluntary standard for ballistic resistance of structures,[2] has attracted far less attention than 0101.03, despite great technical similarity. A contributing reason is that the NIJ, having established the standard, has had no further involvement. Manufacturers submit their products to a laboratory for testing, get the results, and use them in selling their product if they so desire; the laboratory confirms the results to potential customers who inquire, but there is no NIJ or TAPIC role.

Purpose: To Inform or To Protect?

An important consideration in deciding whether standardization ought to be voluntary or mandatory is deciding whether the purpose of standardization is to inform consumers so that they may make informed choices in an unregulated marketplace, or whether the purpose is to protect consumers: to protect some from making uninformed, misinformed, or irrational choices, and to protect others from particular risks they might knowingly and willingly accept. An answer to this question has implications not only for deciding between voluntary versus mandatory compliance, but also for the kind of testing the standard should specify and for the presentation of test results.

The question of whether the purpose of standardization is to inform or protect consumers has not been raised prominently in the current debate, but OTA believes that asking it might clarify decisionmaking on whether standardization ought to be mandatory and on the provisions of the standard and the form of certification.

Typically, standards intended to inform define several quality levels or categories and may (or may not) be voluntary, whereas those whose purpose is to protect are mandatory and have a pass-fail form. For example, eggs are graded so as to inform the shopper of their quality, whereas airplanes are inspected (and passed or rejected) so as to protect passengers and crews from the hazard of flying on unsafe airplanes.

The NIJ standard for concealable body armor combines informative and protective goals, resulting in pass-fail testing at a number of levels of protection. A standard whose purpose is to protect the body armor consumer would embody ballistics standards something like those in NIJ-STD-0101.03 and might well also specify the region of the body that the vest is supposed to cover. It might even go so far as to require particular ballistic qualities, eliminating the consumer's choice as to the level of protection. A standard whose purpose is to *inform, while* it would inform the consumer about the vest's ballistic qualities, would not specify the vest's coverage because the consumer can discern that by simply trying on the vest.

[1] E.g., the Federal Trade Commission Act; see 15 USCA 45.

[2] Such as "body bunkers," portable booths used in such tactical situations as drug busts.

The choice of which type of standard is appropriate involves value judgments, in particular a value judgment about the importance of free choice by consumers, as well as other judgments about who is better suited to select an officer's equipment, the officer or the Federal Government. NIJ's body armor selection guide, which is cited by its 0101.03 standard, provides aid in the process of selecting armor appropriate for local conditions rather than dictating these from the Federal level.

Field Test or Lab Test?

The test specified in NIJ 0101.00 and revisions thereof was originally conceived of as a field test that police departments could perform for themselves. The formulators of the test sought to avoid any specialized test equipment or procedures that would be beyond the typical police department's means or beyond most wearers' comprehension.[3]

Whether or not the rationale behind choosing a field test over something more complicated was a good one is, in part, a value judgment, but the trade-off should be made clear: afield testis simpler to perform and more realistic than a lab test, but the test conditions are less reproducible, so the results may be, too. A field test is intended to be easy to understand, but the uncertainties in the implications of the results areas hard to understand, and may well be greater, than the uncertainties implied by the results of a lab test.

The fact is that few, if any, police departments have undertaken to apply the NIJ test on their own. Perhaps two or three departments apply their own (roughly comparable) tests,[4] but most either send vests to the same laboratory as TAPIC, or apply the crudest of impromptu tests on their owns

Trade-Off: Test Cost Versus Reproducibility

One fundamental trade-off in vest-testing (or, indeed, in any testing) is that between cost and reproducibility. The result of any test is going to be an estimate of some kind, and further testing can always further refine the estimate. The more extensive (and costly) the test, the more refried the estimate, and the greater the likelihood that a second test would give a second estimate that was close to the first one. The question of how reproducible a result has to be in order to be ''reproducible enough" entails a value judgment regarding the desired level of reproducibility. This value judgment does, or ought to, take into account the cost of the testing and the reality that somebody—probably the customer or the taxpayer-must bear that cost.

A related test issue has a much more startling bottom line. Suppose we are presented with Vest 1, that has passed a test with 48 shots (in which a vest fails if even 1 shot penetrates, as in the NIJ 0101.03 test for concealable body armor), and a different-looking Vest 2, that has passed a test with just 1 shot, and that we have no other information regarding these vests. The test facility now proposes to test a second vest—Vest 1A, identical to Vest l—in the first test and a copy—Vest 2A-of Vest 2 in the second test. How surprised should we be if the A models pass the same tests that the originals did? Vest 1A is probably a tough vest, but it has to pass a tough test, and while Vest 2A remains a largely unknown vest because Vest 2 passed only the least stringent of tests, Vest 2A faces only the same easy test. Of course, extra information that we had obtained in some other way—for example, an expert's examination of the vests' construction—might tell us a great deal about the vests and how surprised we should be if they pass the retest, but *the mere fact that a vest has passed a test says very little about the probability that an identical vest will pass the same test, regardless of the details of the vests or the tests.*

Statisticians express their uncertainty about the statements they make in terms of "levels of confidence,' expressed in percentage terms. The idea is that, for example, 90 percent of statements made "at the 90-percent confidence level" are true,[6] though of

[3] For example, the NILECJ and the NIJ rejected V_{50} testing (discussed below) partly because it would do more than just test compliance with the standard at a specified level of ballistic resistance-it would result in a score that would indicate the margin by which certified armor exceeded minimum performance specifications. [145] However, the fact that the V_{50} is a statistical parameter, and the fact that V_{50} testing requires armor to be penetrated, which might diminish some prospective wearers' subjective confidence in its performance, were also considered.

4 OTA knows of only two; the State police Departments of Pennsylvania and California.

5 OTA interviewed an officer who "tested" a Type II-A vest by wrapping it around a knapsack and shooting at it with a .357 Magnum.

6 This concept differs from the related concept, generally rejected by statisticians, that each statement made at the 90-percent confidence level has, in itself, a 90-percent chance of being true. Classical statisticians stick to the idea that the statements have, individually, either a O-percent chance or a 100-percent" chance of being true, only one doesn't know which.

course there is no way of telling, a priori, which 90 percent. In these terms, given a vest that has been tested once and has passed, one can have 50-percent confidence that it has a 50-percent or better chance of passing a second test, and 90-percent confidence that it has a 10-percent or better chance of passing a second test—regardless of the type of test.

Quality Assurance

"Quality assurance" refers to inspection of products (sometimes only final inspection) and rejection of defective ones. "Quality control" and, especially, "statistical process control" refer to monitoring and adjusting the production process itself to reduce the fraction of defective items detected by final inspection; they follow the maxim, "quality cannot be inspected in-it must be built in. " Some body armor manufacturers have implemented sophisticated quality-control processes.

There is no known method for thoroughly testing body armor nondestructively. Ballistic testing of samples is considered necessary but weakens them in places, so that thereafter they cannot be considered as protective as a ''virgin" (unshot) vest of presumably similar manufacture. There are three ways of dealing with this problem. The first is to ignore it—to make no representations about the quality of units not actually tested. A second is to try to make sure production units are made in the same way as samples that were tested and deemed acceptable. A third is to infer the acceptability of units not tested on the basis of tests of randomly selected samples; this approach, sometimes called statistical quality control (SQC), provides assurances couched in statistical jargon. Statistical process control (SPC) combines the second and third approaches.

The present system of testing vests is really one of *design certification:* when the manufacturer presents a vest of new design and has it certified, it is really the design that is certified. Continuing quality control, and assurance that vest production continues to use the same methods and materials as were used in the test article, are entirely up to the manufacturer. For that matter, assurance that the same design will be used is almost entirely up to the manufacturer; TAPIC and the NIJ only compare the construction of vests offered for sale to the construction of those originally presented for testing in the rare event that some kind of accusation is made.

OTA has discovered that not all police officers are aware of this state of affairs. Some assume, for example, that NIJ testing is to be redone whenever a manufacturer switches to a new lot of fabric.[7]

NIJ could institute a program of ongoing quality control. This could be done in any of several ways (see app. E for details). One option that NIJ has considered is Classification of body armor, by Underwriters Laboratories Inc. (UL), as complying with the NIJ standard. UL now estimates that a minimum-cost program might cost about $3,000 for initial testing of a model (plus about $1,500 for each additional model from the same manufacturer tested at the same time) plus a recurring annual cost of little more than about $700 to $1,000 for the ongoing ''follow-up' inspection program. This option would not provide purchasers with quantitative estimates of risks of UL-Classified armor.

A different approach would be needed to calculate and advise purchasers and wearers of the quantitative limits on risk implied by test results. The procedure for lot certification described in appendix E is one example; it would rely on sampling and ballistic testing, not on inspection of the manufacturer's production process or auditing of the manufacturer's quality-control program. The inventorying of lots and selection of samples for testing could be performed for the NIJ (or a manufacturer) by a grantee or contractor; the ballistic testing could be performed by an independent ballistic-testing laboratory such as UL or H.P. White.

The cost would depend on the reliability and confidence in reliability demanded. Demanding more of either will require more testing and will cost more. However, only 2 tests would be needed to decide whether to certify a lot of arbitrary size with a consumers' risk no greater than 10 percent and a producer's risk no greater than 10 percent (see figure E-12), if consumers' risk is defined as the probability that a lot containing armor with a probability of passing lower than 8.53 percent[8] is accepted, and if producer's risk is defined as the probability that a lot containing armor with a probability of passing no

[7] Responses of police officers attending Body Armor seminar at University of Maryland, Department of Textiles and Consumer Economics, Comfort and Perception Research Laboratory, Apr. 23, 1991.

[8] This corresponds to a geometric-mean single-shot probability of 95 percent of stopping the bullet and, if appropriate, leaving an acceptable BFS.

lower than 95.31 percent[9] is rejected (see appendix E for other options and additional details).

The main policy choice for the NIJ is whether to undertake to assure purchasers and wearers that armor *of a model* certified to comply with the NIJ standard would itself pass a certification test. If so, the NIJ must decide whether to provide such assurance quantitatively-e. g., in terms of statistical confidence limits on the probability that a sample of armor of a certified model or lot would pass a certification (“re-”) test. If so, the NIJ must decide the minimum statistical confidence it will accept and the minimum passing probability to which it refers. Demanding higher reliability and confidence in reliability will require more testing and will cost more. “How much is enough?” is a policy choice (i.e., value judgment).

A current issue for Congress is whether to enact H.R. 322, the Police Protection Act of 1991, which would, inter alia, mandate an NIJ-supervised quality-control program and require manufacturers to submit “representative samples” periodically to the NIJ to be tested for compliance with the current or a future standard. Because NIJ has not specified in detail what it would do to implement the quality-control provisions of the Act, OTA cannot assess the effectiveness of the NIJ’s implementation.

The act has many other provisions that will be weighed along with its effect on quality control: it would authorize the director of the NIJ to establish procedures for recertification of body armor models. Moreover, it would prohibit the manufacture, sale, or distribution in commerce of armor not complying with the standard. This would curtail industry’s current freedom to produce and sell what the market demands. It would likewise curtail consumers’ current freedom to take certain risks (e.g., that armor will be soaked, shot, and penetrated in service) hoping to reduce others (e.g., that armor will not be worn). It suggests that some law enforcement officials cannot understand the risks they would take and would not accept them if they understood them: Congress finds that . . . the complexities of body armor and the diverse nature and abilities of law enforcement officials to purchase and test it result in unnecessary risk.

If H.R. 322 is not enacted, Congress could fired a voluntary quality-control program. The Department of Justice could propose one, or Congress could require the administration to propose one.

Enforcement

One can imagine means of violating the letter or the spirit of NIJ Standard-0101.03, TAPIC’S ‘Compliance Testing Procedure for Police Body Armor,”[10] or fair-trade laws; for example:

- Certifying on a label that armor is of a model that complies with NIJ Standard-0101.03, when in fact samples have never passed the test specified by the standard-anywhere. (This *could be* judged to violate the Federal Trade Commission Act.[11] However, the burden of proving that samples never passed the test specified by the standard, anywhere, would be the governments.)
- Repeatedly submitting for TAPIC-supervised testing samples of armor made identically but bearing a different model designation in each case, until one set of samples passes and is certified, and then manufacturing more such garments and offering them for sale labeled with the model designation of the samples that passed. (TAPIC *would* consider this a violation of its “Compliance Testing Procedure for Police Body Armor,” which specifies that ‘In the event that a body armor model fails to comply with the requirements of NIJ Standard-0101.03, the manufacturer must abandon that model designation. A noncomplying model cannot be submitted for retesting.” TAPIC would consider samples to be of the same model if only the model designations differed. However, this is not the only sensible interpretation of the ambiguous provision: any manufacturer found to engage in this practice could argue, in effect, “Samples of the armor I designated Model A did not comply, so I abandoned that model *designation,* produced more samples, designated them Model B, and submitted them to TAPIC. Model B was not known to be noncomplying.”)
- Submitting atypically good samples that were not selected randomly as the standard specifies.

9 This corresponds to a geometric-mean single-shot probability of 99.9 percent of stopping the bullet and, if appropriate, leaving an acceptable BFS.

10 Appendix B of [137].

11 See 15 USCA 45.

Making test samples from a stock of fabric of a particularly good lot reserved for such samples *would be an* egregious example, and difficult to detect.

- Submitting samples apparently larger than allowed by the standard. This practice is a clear example of non-random selection;[12] nevertheless, it is recognized and tolerated by TAPIC because it minimizes the chance of having to shoot spare samples because of unfair hits on the initial samples, and because it has not been proven that larger samples have abetter chance of passing. However, physical reasoning suggests that larger samples do have a better chance of passing, especially if the test shots are aimed to approximately maximize the minimum distance from any impact to any other impact or an edge.[13]

One may influence a model's chance of being certified without resorting to such expressly prohibited practices. The following are not expressly prohibited by the NIJ standard or TAPIC'S compliance testing procedure:

- Certifying on a label that armor is of a model that complies with NIJ Standard 0101.03, when the compliance testing was performed privately, not through TAPIC. (Private testing in accordance with the standard need not comply with a number of restrictive provisions that the NIJ has specified, in addition to those specified by the standard, for TAPIC-supervised testing.[14] For example, although TAPIC prohibits, and attempts to detect, submission of samples of a noncomplying model for retesting, a manufacturer may certify that a model complies with the standard even if samples did not pass on the frost attempt.)
- Submitting, at the same time, several sets of armor samples produced in the same way but labeled as different models, and then, if any set passes, manufacturing more such garments and offering them for sale labeled with the model designation of a set that passed. (If all sets are submitted before any has been tested and failed, this would not violate the letter of TAPIC'S Compliance Testing Procedure. Nevertheless, TAPIC has objected to one apparent attempt. [32])
- Labeling armor as ''tested for compliance with NIJ Standard 0101.03" without specifying whether samples of the model *passed the* test.
- Asking the operator performing the test to try (by adjusting the powder charge) to achieve bullet velocities slightly greater than the maximum velocities specified by the standard, so that nonpenetrations will count as fair shots while penetrations, if any occur, will count as unfair shots. (In the case of contoured vests such as those designed for female officers, one could further suggest to those performing the test that they ensure that one of the first six shots lands on a seam, obviating the need for a seventh fair shot.)
- Stipulating the loosest possible attachment of the armor to the backing, so as to raise the probability that the armor will fall off the backing and be replaced, providing-in effect—for a smoothing of the armor between shots.
- Availing oneself of the option to have the second type of ammunition tested on the same panels as the first type. (The panels are inverted so that the prescribed impact sites are on relatively fresh armor; see app. A). In the event of a failure when testing against a second type of ammunition in this fashion, the standard provides for a restart of the test using a fresh panel and the second ammunition type. Thus the manufacturer who specifies the use of this option (which was intended to conserve vests) gives the model two chances to pass the second-ammunition part of the test instead of one. Possible degradation of the armor by the first-ammunition part of the test makes the first

12 In defense of this practice, it might also be noted that it is impossible to comply with the provision of the standard that requires random selection of samples foresting, unless the sampling is done after all units of a model have been produced but before any unit has been sold. Volume 1 and appendix E of this volume discuss sampling in greater detail.

13 A large area of a panel is stretched momentarily when a shot impacts; this allows the panel@ absorb the ballet's energy without being penetrated. The larger the panel, the more energy it can absorb, until the radius exceeds the strain-wave velocity of the material times the duration of deceleration of the bullet by the armor. The panel maybe stretched permanently, penetrated partially, or otherwise weakened near the impacted area, and a subsequent shot may be more likely to penetrate if it impacts such a weakened area. To prevent such interactions, the NIJ standard requires a minimum separation of 2 inches between shots; however, high-energy projectiles may weaken the armor over a greater radius. It is plausible that probability of penetration decreases with increasing separation between shots, although only slightly at large separations.

14 For example, TAPIC's "Compliance 'resting Procedure for Police Body Armor" governs only *TAPIC's* certification Of Compliance; it does not govern a manufacturer's certification of compliance, which the standard itself provides for.

of these chances worse than the second, but still better than none at all.

- Fastening the vest to the fixture with copious wrapped layers of adhesive tape, virtually mummifying the vest-fixture combination. This set-up reduces or eliminates ply separation. While not necessarily any less accurate than the standard set-up consisting of straps, and while clearly allowable under the letter of the 0101.03 standard, this set-up gives results that are not comparable with those obtained in the usual way, in which plies can separate.
- If it is desired that a given vest *fail,* directing or encouraging the test operator to aim the later shots in the test sequence at a region of ply separation (if any) created by earlier shots in the test sequence. Such a region would be visible as a "hill" on the bunched-up vest. This may increase the probability of penetration for some combinations of bullet type and armor type. (If the armor is contoured-e.g., a model designed for female officers-one could further suggest that the operator ensure that none of the frost six shots hits a seam; this would necessitate a seventh fair shot—an additional opportunity to fail.)

Revising the standard to specify the test protocol in greater detail could prohibit those practices that seek to influence a model's chance of certification by exploiting a laxity or vagueness in the wording of the standard, such as intentional over-velocity shooting or loose attachment.

Teaching potential purchasers the distinction between a manufacturer's certification of compliance (on the label) and TAPIC'S certification of compliance (by listing a model on TAPIC'S Consumer Product List) might deter manufacturers from certifying compliance without concomitant TAPIC certification, or at least alert consumers to the *possible* insubstantiality of such certification. Revising the standard to specify that the test it specifies must be passed on the first attempt would clarify the intent of the standard. Revising the standard to apply to lot certification, as described in vol. 1 and app. E, would go even further and provide quantitative estimates of maximum risk.

Such measures would not suffice to prevent or detect deliberate fraud, such as labeling noncomplying armor with a model designation listed on TAPIC'S Consumer Product List. Nor would enactment of H.R. 322, The Police Protection Act of 1991. Detecting such fraud reliably probably will require purchasing samples of armor in the marketplace covertly (e.g., in concert with consumers) and inspecting and testing them. This would not be foolproof, because noncomplying counterfeit armor could resemble certified armor visibly, and samples of a certified model might fail the ballistic test because of poor quality control or bad luck. A market-surveillance program would be most effective in concert with a government-supervised lot-certification program, including government-supervised inventorying and tagging of units of certified lots, as described in app. E.

Although NIJ currently lacks enforcement authority, the Federal Trade Commission (FTC) has jurisdiction to enforce fair trade practices; it can prosecute a manufacturer it believes to be misrepresenting a product's compliance with NIJ's voluntary standard, and has done so. Conceivably, a manufacturer could be prosecuted for one of the practices mentioned above that, although not prohibited expressly by the NIJ standard, is not disclosed to prospective purchasers and, if disclosed, would influence their decisions of whether to purchase:

> [F]ailure to disclose by mark or label material facts concerning merchandise, which, if known to prospective purchasers, would influence their decisions of whether to purchase, is an unfair trade practice violative of this chapter [of the Federal Trade Commission Act].[15]

How far this protection extends will remain unclear until clarified by case law.

There is considerable difference of opinion as to the course of action NIJ should take on receiving word that its certification is being improperly claimed for noncomplying vests: ought it to warn police departments immediately (and thus risk irreparable damage to the reputation of a manufacturer not yet proven guilty), or ought it to investigate the accusation fully before saying anything (and thus risk the death of an officer, killed while wearing a vest believed to falsely claim NIJ certification)?

[15] 15 USCA 45, n. 93.

297-923 0 - 92 - 4 : QL 3

Regulation of Trade in or Wearing of Armor

A related but larger issue for Congress is whether to ban trade in armor not certified by NIJ, as H.R. 322 would if enacted. An issue for OSHA is whether to exercise its existing authority to mandate wear and issue applicable standards and regulations.

Style *Certification*

While there is a technical side to the complicated issue of style certification (see below), there is a policy issue involved as well: to what degree is it acceptable to certify diverse vests without ballistic testing, on the grounds that they are merely variant styles of a basic model that has already passed ballistic testing? One could seemingly do away with this issue by insisting that all vests be tested, but—given that the test destroys the vest, so the customer will always be buying something on the grounds that it is "just like" something else that passed the test—where should the line be drawn? Even before NIJ procedures included the style certification concept, only one size and color of vest needed to be tested: vests of other sizes and colors were sold on the grounds that they were ballistically identical to the vest that had been tested.[16]

Once one admits that not all sizes and colors of a vest need to be tested, one is opening the door to the "style certification" concept.[17] the only question is where the line between "style" and "model' is to be drawn. A value judgment enters into the determination of how much testing and confidence are enough, given that everything costs money.

TECHNICAL ISSUES

Trade-Offs in Body Armor Testing

Most people would probably feel that a test of a product such as police body armor ought to be conservative (i.e., stringent), realistic, and reproducible. It should be conservative, so that undetected flaws in test formulation or post-test variation in the product would not make the difference between a safe product and an unsafe one. It should be realistic, so that test conditions accurately reproduce the circumstances under which the product will be used in the field. It should be reproducible, so that an item that passes the test one time will not fail if retested.

The trouble with these criteria is that they are mutually contradictory. In particular, realism is at odds with conservativism and reproducibility. Realism requires that test conditions be the same as those in the field; conservativism requires that they be more stringent. The conditions found in the real world are anything but reproducible; no two actual shooting incidents will be identical.

For these reasons, some realism is often sacrificed when a test is formulated. To criticize a test such as the NIJ test for police body armor purely on the grounds that it is unrealistic is a value judgment, as was the trade-off selected in designing the test. While it is easy to charge that the testis flawed on the grounds that 'the bad guys won't always use that kind of ammunition' or "most people don't get hit 6 times in the chest," it is important to realize that certain artificialities have to be introduced in order to make the test conservative and reproducible.[18]

There is also a tradeoff between stringency and reproducibility, at freed cost. More generally, there is a tradeoff among stringency, validity, reproducibility, cost, and other valued attributes, such as simplicity. Threats are multidimensional (i.e., vary in many ways: bullet types, velocities, angles of incidence, and impacts per panel) and pose different risks of penetration. If reproducibility were the only concern, the test wouldn't use bullets at all: it would use fragment simulators. They are machined, not cast, and hence yield highly reproducible results, but they cost 100 times as much. They also penetrate better than typical bullets of similar energy, so the test results would have to be calibrated to penetrations in service.

If armor having a mean single-shot penetration probability lower than a specified value is defined as ' 'good," and armor having a mean single-shot penetration probability higher than another specified value is defined as "bad," then it is possible to devise a test that ensures the probability of certifying bad armor (''the consumer's risk' is no greater than a specified maximum while the probability of

16 In the case of vests of differing size, this is almost certainly not true—the size makes a difference. On occasion, small sizes of a vest that had passed the NIJ test have been known to fail an NIJ-like test.

17 Currently, color is considered so obviously irrelevant that vests differing only in color are in fact deemed to be the same style.

18 The NIJ standard, the Police Protective Armor Association (PPAA) standard, the State of California SW&@ and various foreign standards all seek to ensure that armor is far better than it is necessary to withstand typical assaults.

rejecting good armor (''the producer's risk' is also no greater than a specified maximum. [60][19] [20] Lowering both ceilings would reduce risks to both consumer and producer but would increase the amount and cost of testing required; producers might bear part of this cost but would probably pass the balance of the cost along to consumers by increasing prices. Lowering both ceilings might also require permitting some penetrations. This may reduce the simplicity, and perhaps the understandability, of the test; it may have other risks and benefits discussed at the end of this appendix, in V_{50} Testing.

The high failure rate in tests is often contrasted to "the perfect record of vest performance in the field." Some of the discrepancy is attributable to conservativism (critics charge that it is attributable to over-conservativism) in the formulation of the test and some to the fact that a test will always operate near the limits of the vest[21] whereas field use spans the full spectrum of conditions. Insofar as the discrepancy is attributable to overconservativism, the correct course of action is not clear. "What should we do," asked one expert, "back off on the standard until somebody gets killed?" On the other hand, overly conservative standards could lead to overly uncomfortable and expensive vests, and thus to officers getting shot while wearing no armor at all.

Defintion of "Style"

The purpose of *style certification* is to allow certification of more than one style of the same *model* vest without incurring the additional cost[22] of testing each style.

For example, suppose that a vest manufacturer's Vest A has passed the 0101.03 test and has been duly certified. Vest A consists of two ballistic panels placed in a cotton carrier. It has sold well, including several sales to large police departments. A small-town department has examined Vest A closely and would like to buy 50, but wants the neck of the vest to be shaped slightly differently-they want a V-neck rather than a crew-neck on the front panel, because the crew-neck would show in the open collar worn in the summer by officers on the street. The manufacturer would certainly like to sell 50 vests, but not if doing so would require ballistic testing of a new Vest B that differs from Vest A only in the shape of the collar. The test would consume most or all of the profit to be made from a 50-vest sale, and if Vest B failed, the many purchasers of Vest A might lose confidence in their vests. The manufacturer needs a means of declaring that the new vests are really examples of Vest A, only with a different-shape collar. To respond to this need, NIJ has instituted a procedure for style *certification:* a vest is sent in to TAPIC with the request that it be certified as being a new style of a previously certified model of vest.

Because the certification of a new style is inherited from the certification of the original test article, stylistic differences are defined as those that do not affect the ballistic performance of the vest. The collar is such a difference: other such differences include flaps on the sides of the panels to increase coverage of the wearer's sides. Enlarging these is a style change only-decreasing their size would also be a style change if the shot pattern of the certification test would fit on the new vest without any shot being nearer than 3' to the edge of the modified vest. Changes in the color of the carrier are so immaterial that they are not even considered to be style changes.

A proposed change goes beyond being a style change if it involves changes in the ballistic material used, the number of layers of the material, or the stitching of the material.[23] In the past, some conflicts have arisen over what constitutes a mere stylistic difference and at what point two vests become so different that they are different models, not merely different styles. Formulation of a fool-proof definition of style remains an important technical issue.

19 The notions Of "COnsumer'srisk" and "producer' srisk" were originally introduced by the statistician (and inventor of quality control as we know it today) Edward Deming.

20 We note that the certification of bad armor also poses a liability risk to producers and, perhaps unfairly, a credibility risk to the certification process.

21 If it doesn't, the manufacturer may make the vest lighter and cheaper until it does.

22 And risk of failure.

23 OTA has encountered supporters of widely differing views regarding the effect that stitching has on ballistic performance *apart from resistance to bunching and balling. Nearly everybody* agrees that extra stitching lessens the tendancy towards bunching and balling, albeit at the price of increased stiffness.

Choice of Backing

The NIJ test, as well as the Police Protective Armor Association (PPAA) test, uses nonhardening, oil-based modeling clay as a backing for test samples of armor.[24] This clay has the virtue of being reusable, so that the (moderate) expense of creating the 2-foot square, 4-inch thick block of backing material can be amortized over many tests. By virtue of its lack of elasticity, it affords an easy means of measuring backface deformation, which in turn can be related to the probability that the wearer will be injured or killed by the impact of a bullet, slug, or shot stopped by the armor.

Some object to the clay backing on the grounds that it "does not realistically simulate human tissue." [87] In particular, objections allege, the hardness of the clay causes more penetrations, the inelasticity of the clay leads to bunching-up of the vest during testing, and the deformation of the clay has not been related to deformation or injury in humans.

Penetration

One's intuition suggests that attachment to a firm backing will make the vest more penetrable than no backing at all. Attachment to a backing influences penetration in two ways—attachment prevents the whole vest from moving out of the way, and the backing allows part of the vest to be pinched between the backing and the bullet. These effects are separable: Some experiments have used attachment without backing, and others have used backing without attachment. For example, in military and other V_{50} testing of armor fabric, a panel of fabric is attached to a frame with only air backing. (See below for a full description of V_{50} testing.)

How similar is the clay backing to the human body in terms of the ability to hold the vest in place and to create pinching between the bullet and the body or the backing? Clay backing prevents bulk movement of the vest away from the shot. Contrary to the impressions possibly fostered by Hollywood, so does the human body: the impact of a gunshot, even of a shotgun blast, is no more likely to knock over the target than the recoil from the same shot is to knock over the shooter. The clay *is* harder than some parts of the human body, and a bullet may have a greater chance of penetrating the vest on a clay backing than it would on a human's ventral region.[25] The human sternum, by contrast, is harder than the clay.

Bunching Up

Although one can argue that the clay is harder than some parts of the human body and softer than others, it is undeniably less elastic than any. Indeed, inelasticity-the quality of not springing back after having been deformed-was a quality sought after in the clay, for it is this quality that makes possible the measurement of backface signature (BFS) without high-speed photography or other elaborate, expensive means.

Some, however, see the inelasticity of the clay as fostering the readily observable bunching-up of some pieces of armor. After repeated shots against a clay backing, some armor is so bunched up as to give the appearance of having been wadded into a tight ball. On the inside of the armor, this bunching and balling causes the plies of ballistic fabric to separate, making them more easily penetrable. In the worst case, it can even lead to folding of the armor panel within its cover, so that a site marked for a shot no longer has the armor panel beneath it, resulting in a sure penetration and failure of the item.

Critics of the use of clay as a backing argue that the bunching and balling of the armor on clay does not reflect its true behavior on the human body and that therefore failures attributable to bunching and balling do not indicate unsafe armor.

A bullet's impact upon the soft armor protected body causes a momentary indentation that rebounds several times due to body tissue elasticity. The elastic body wall rebounding against the armor tends to smooth it and return any layers separated by the bullet's impact toward their original positions. This self-smoothing and repositioning of layers cannot occur when the armor is pushed into non-elastic clay. This effect makes it easier for subsequent bullets hitting the vest to penetrate completely. [87]

24 The NIJ test originally used air as a backing for penetration shots.

25 Prather, Swann, and Hawkins [114] reported V_{50}s for 7-ply Kevlar of 1079 and 1088 fps on clay in two tests (the second used clay that had been stored unwrapped), 1096 fps on [euthanized] goat abdomen, 1115 fps on [euthanized] goat thorax, and 1109 fps on gelatin. These values are ordered as the conventional wisdom would have them, but are not markedly disparate.

It is widely claimed that body armor manufacturers now construct armor with extra stiffness, e.g., by quilting, so as to minimize bunching and balling during the NIJ test. [87] *Insofar as the* conditions of the test are artificial, this extra stiffness (which carries a penalty in comfort and cost) is a needless burden on the manufacturers and wearers of the vests.

Alternative backing materials include ballistic gelatin, polyurethane foam, and solid synthetic rubber. Ballistic gelatin is an unflavored version of the food item and, like the food item, consists of a solution of water in animal protein. It owes its flabby texture to the fact that it is a solution of a liquid in a solid, rather than the more usual solution of a solid in a liquid. Some controversy exists over whether ballistic gelatin ought to be 10 percent protein or 20 percent protein. The latter more closely represents the density of human flesh, while the former better mimics its mechanical qualities.[26] Both approximate the density of human flesh better than does modeling clay, but neither can simulate the effect of bones or other rigid tissue. Polyurethane foam of the type used in "foam rubber" mattresses has been used in demonstrations staged by body armor companies. Slabs of foam are packed in a nylon cover or bag, to which the vest is strapped. RTV silicone synthetic rubber has also been used experimentally as a resilient backing for the testing of body armor. [160] The elasticity of all these materials would preclude later measurement of the backface deformation, though the gelatin is transparent and high-speed photography could be used to capture an image of the deformed backing.[27] This procedure, and the nonreusable nature of the gelatin, would add greatly to the expense of the test. Other techniques for recording deformation of resilient backing are described in appendix E; they would also require costly apparatus.

If films exist of the animal shootings the Army performed to correlate any blunt trauma produced with the maximum deformation of gelatin behind similar armor, they could be examined for signs of bunching and balling in the armor on the animals. Films were made of the deformation of gelatin behind armor; locating and analyzing the films might also provide information. (Some experts say that similar tests conducted elsewhere produced more bunching and balling on gelatin than on clay.)

An important piece of physical evidence-for both sides-is a videotape [121] of Richard C. Davis, founder and President of Second Chance Body Armor, Inc., shooting himself in the abdomen while wearing body armor of his own design in Walled Lake, Michigan in 1972[28]. The critics, including Davis himself, argue that the video shows no bunching; other viewers contend that it does. OTA staff judge that it does but have seen greater bunching, on occasion, in NIJ-type testing.

We know of no evidence that the hypothesized "self-smoothing and repositioning" goes beyond the return of the chest or abdomen to its pre-impact position (unless the stopped bullet fractures a bone or the armor penetrates the skin). A biomechanical model [90] of the adult male torso (see figure C-1) fitted to measurements made on cadavers, [109, 11 1] which had been correlated with measurements on live volunteers, predicts that the sternum-spine separation will not oscillate after an impact (see figure C-2). The change in sternum-spine separation begins to return to O after about 2 milliseconds (ins) and approaches O very gradually thereafter, taking 48 ms to subside to 37 percent of the maximum change and 100 ms to subside to 14 percent (see figure C-2). Engineers call such a response "over-damped" and call the time required for the response to subside to 37 percent of its maximum value the damping time; the damping time predicted by the biomechanical model, 48 ms, is roughly the period between successive impacts at 1,200 rounds per minute (rpm), the cyclic rate of fire of an Ingram MAC-11 submachinegun. Thus, the biomechanical model predicts only a fraction of an outward pat—never exceeding the preimpact position—between successive impacts at 1,200 rpm.[29]

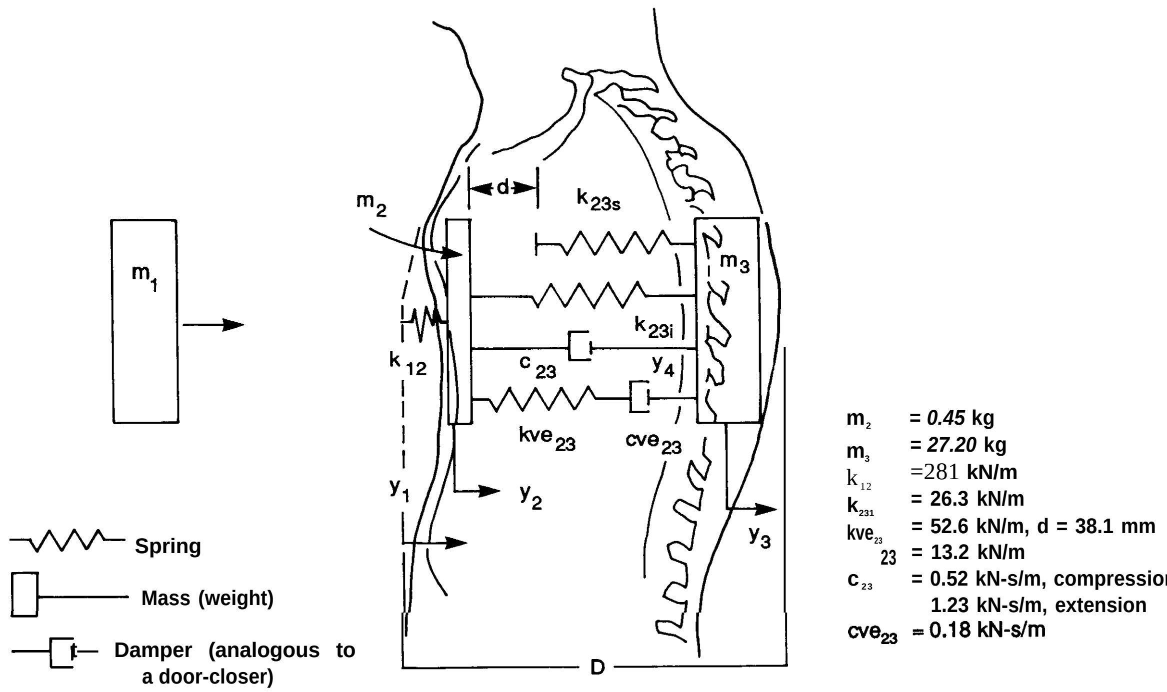
26 Dessert recipes lead to a concentration of about 10 percent.

27 In a seldom-noted effect, the sides of the gelatin bow out and act as lenses, complicating the measurement of dimensions photographed through the gelatin.

28 Mr. Davis has shot himself on many other occasions to demonstrate the capabilities of his company's armor, but he typically inserts a thick telephone book between his abdomen and the front armor panel which he shoots, so such shootings are not a realistic simulation of the effect of an actual assault, at least for purposes of simulating ply separation.

29 Sellier and Wehner have stated that the resonant frequency of the chest cavity is about 10 Hz, [123] but they did not note the damping, or the source of the information.

Figure C-l—A Biomechanical Model of the Human Torso



SOURCE: Albert 1. King and David C. Viano, 1986. [90]

Patting Down

Those who believe that the clay backing causes unrealistic bunching and balling of the ballistic fabric but who also feel that the practicality of clay (in terms of cost, reusability, and measurement of backface deformation) makes it an otherwise preferable backing advocate patting the vest down between shots, smoothing out the bunches of fabric. Others see the patting or smoothing as unrealistic, on the grounds that police officers do not smooth out their vests during gunfights. Advocates of smoothing the vests between shots agree that police officers do not deliberately readjust their clothing after each hit, but cite the "self-patting" effect, by which "the multiple rebounds of the elastic body wall" "return the body armor layers (which are separated to some degree by bullet impact) to their original positions.' [86, 87]

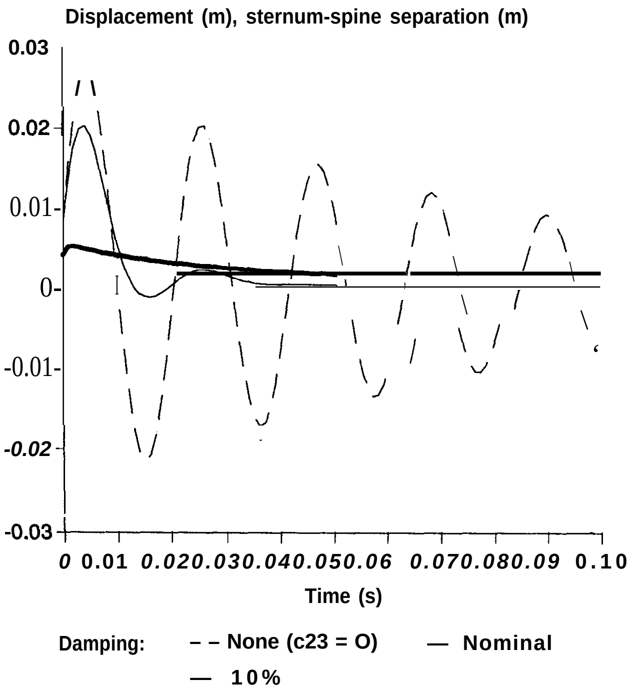
Strictures against patting the vest down have become stronger with each successive edition of the NIJ standard. (See appendix A, Origins of and Rationale for the NIJ Standard.) This issue could be revisited yet again, especially if compelling evidence of the self-patting effect were developed.

Deformation

Because of the desire not to disturb the armor, BFS is measured only after the first fair hit; measurement after each shot would create the opportunity for smoothing the vest, and in fact would probably render such smoothing unavoidable. Shot #l is a head-on shot, so BFS is measured only for a head-on shot. This shot is unlikely to be on a seam: in normal vest construction practice, the only vests with seams in the ballistic material are those constructed for female officers. The seams in these vests are nowhere near the site of shot no. 1, and are likely to be hit only by the one angle shot required to hit a seam and a bust cup.

One drawback of clay as a means of measuring deformation is that its deformability depends on temperature and preparation. The test protocol specifies how the backing material is to be prepared and specifies a temperature range within which it must be maintained for sometime before the test and during the test, as well as a more limited ambient temperature range to be maintained. Three drop tests are required to establish that the deformability of the backing is within acceptable limits. However, in current practice, clay used to fill in dents in the

Figure C-2—Movement of Sternum Relative to Spine After an Impact (predicted by biomechanical model with parameters for adult male torso)



Note that the response is overdamped-i.e., the sternum-spine separation does not oscillate. Also shown are the responses predicted if the damping parameter c_{23} were O percent and 10 percent of the nominal value. There is oscillation at about 50 Hz (cydes per second) in both cases, although 10 percent of nominal damping is apparently near critical damping, at which oscillation does not occur.

SOURCE: Office of Technology Assessment, 1992.

backing material may be drawn from a supply kept at a different temperature than that of the block of backing, which may warm or cool as testing proceeds. The drop tests are typically done only at the beginning of a test and provide no check on possible changes in the consistency of the clay later in the test. Clay used to fill in the bust cups of vests contoured for female officers is to be conditioned in the same way as that in the main body of the backing material, but the standard specifies that no drop test need be performed on the bust cup clay.

The drop test does not assure that backface signatures produced in different backing materials behind similar armors by similar bullets impacting at similar velocities will be the same. Different modeling clays conditioned to pass the drop test yield different backface signatures at the much higher deformation velocities typical of a ballistic test conducted in accordance with NIJ Std. 0101.03. In tests conducted by the British Police Scientific Development Branch (PSDB), under otherwise similar conditions the nominal backface signatures produced in U.S.-made Plastilina and U.K.-made Plasticize were similar at impact velocities of 485 m/s but differed by about 4.4 mm for each 100 m/s above or below 485 m/s.[30]

Other backing materials not yet tested by NIJ or NIST, and potentially usable by a tester attempting to certify compliance of armor that would fail the deformation test on Roma Plastalina No. 1, could differ more dramatically. Specification of a backing material would eliminate this potential source of variation in-or operator influence on—test conditions.[31] [32]

Shape of Test Fixture

The usual test fixture is a rectangular frame containing a 24" x 24" x 5" block of clay backing material-the exterior dimensions of the frame might typically be 26" x 26" x 5", because the 24" x24" front and back surfaces of the clay are exposed. At the request of the armor manufacturer, the clay backing may itself have a plywood backing.[33] The armor is spread flat on the frame and strapped thereto with large elastic straps.[34] (The 0101.03 standard

30 See equation 2 of [28]. The fitted backface signatures (figure 8 of [28]) differed by about 5.4 mm for each 100 m/s above or below 487 @; the greatest difference was observed at the lowest velocities-about 260 m/s. An updated nominal model, [29] based on additional data but apparently excluding the low-velocity impacts on Plastilina, predicts BFSS will differ by 4.4 mm for each 100 m/s above or below 350 m/s. The corresponding fitted model predicts BFSS will differ by 5.1 mm for each 100 m/s above or below 336 m/s. Iremonger and Bell [84] reported yet another model based on the same research program but also apparently excluding the low-velocity impacts on Plastilina.

31 Although clay composition demonstrably affects the results of the deformation test (for protection from nonpenetrating bullets), it is not certain that it affects the results of the penetration test. More research would be needed to find out whether it does.

32 The important question of allowable backface signature will be deferred to a later section-the purpose of this section is only to discuss issues of deformation as they relate to the choice of backing material and the issue of repositioning the armor.

33 This option is noted in communications of test results when the manufacturer chooses it. It appears that manufacturers so choose more often than not.

34 The number of straps used is not specified in the NIJ 0101 series of standards, but has, in practice, increased over the years. Five straps are now used.

specifies that inelastic "tape" can also be used, but this option is rarely, if ever, chosen by a manufacturer.) Armor contoured for female officers typically will not lie flat, and additional clay is built up so as to fill the vest.

The NIJ has considered alternatives to the fixture specified in the .03 standard and has had NIST/ OLES conduct tests [149] of three alternative fixtures: (1) the flat clay block specified by NIJ Std. -0101.03, (2) a mannequin as specified by PPAA STD-1989-05, [113] and (3) an experimental curvilinear test fixture (already known as "the curv") consisting of a rectangular frame holding a clay block but with semicylindrical sides facilitating the attachment of a complete armor by means of its own strapping and fasteners (see figure C-3).

The flat shape of the test fixture facilitates determination of the angles (O or 30 degrees) at which the bullet strikes the armor. It also facilitates measurement of the BFS, which is defined as the displacement of the clay below the plane in which it originally lay. This measurement is established by using a metal straightedge to shave off the upwelling of clay around the crater, and then using a measuring device whose three legs rest on the clay surrounding the crater and whose plunger measures the distance from that plane to the bottom of the crater.

However, the shape and size of the test fixture preclude the attachment of the armor to the fixture by its own straps. Those who cite bunching and balling as an artificiality of the NIJ test sometimes point to this fact as a secondary cause of bunching and balling: they maintain that if the vest were held taut by its own straps, rather than swaddled to the backing by other straps, it would be less prone to bunching and balling. In practice, such an attachment would probably hold the vest more tautly than would an actual officer, who would adjust his or her vest for a looser and therefore more comfortable fit.

The obvious alternative is a mannequin. The mannequins constructed for the PPPA and other test protocols typically consist of a head and upper torso made of hard plastic, with a cavity hollowed out in the middle of the torso to receive the backing material. The examples seen by OTA staff used oil-based nonhardening modeling clay as a backing material. The vest can be strapped to the mannequin just as it would be to a police officer. The front surface of the clay can be shaped as the true torso would be or sheared off flat to facilitate measurement of backface signature.

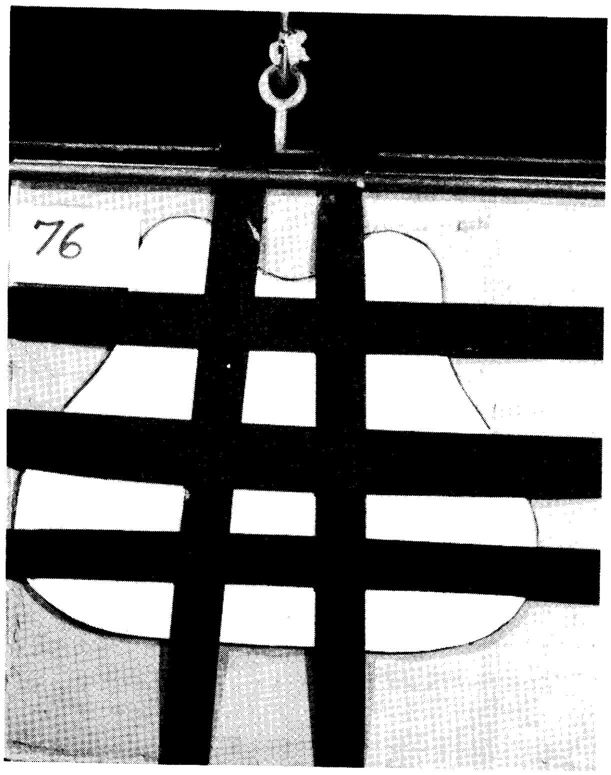
The mannequin test could be further refined by suspending the mamequin as if in a swing, rather than firmly anchoring it to the floor as is generally done with the clay block in compliance testing at H.P. White Laboratory, Inc. (although neither anchoring nor use of a frame for the backing material is required by NIJ Standard 0101.03). The suspended mamequin would thus be free to swing back when hit, transforming some of the energy of the bullet into the energy of the swinging motion and thus lessening the energy deposited in the vest and the clay—as would happen in an actual shooting incident. If the mannequin weighed as much as a vest-wearer, this set-up would more accurately capture the dynamics of the victim-bullet collision. (Some have objected that the officer might be running; the officer's-or the mannequin' s-initial velocity affects the amount of bullet energy absorbed by changing that velocity.) However, inasmuch as the backward motion imparted to an actual shooting victim is slight (as mentioned above, it is comparable to that imparted to the shooter by the recoil of the gun), this refinement would add very little accuracy and might not be worth the trouble. The portion of initial kinetic energy available to permanently deform the backing and possibly the armor is the change in total kinetic energy; it is proportional to the square of the difference between the initial velocities of the bullet and the backing. This would vary by at most a fraction of a percent even if the backing were initially moving at 10 m/s.[35]

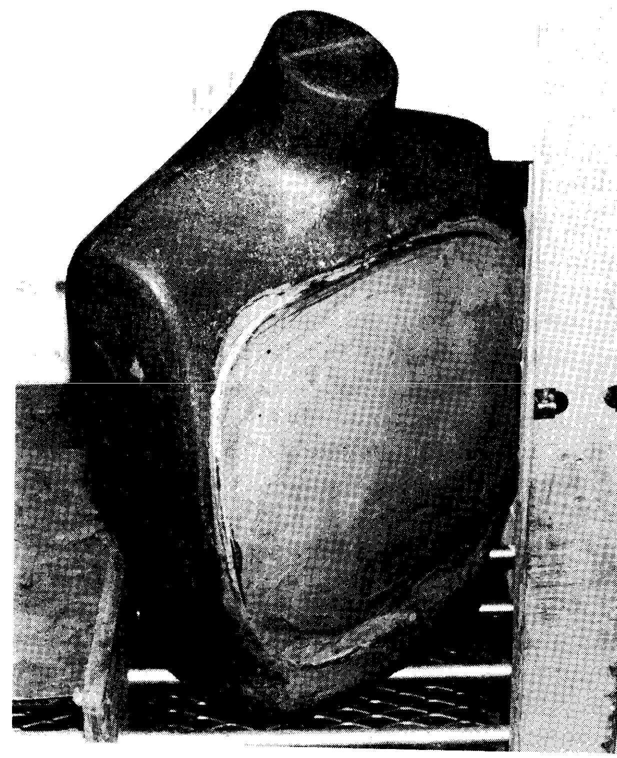
Using the flat block, one panel of a vest is tested and then it is replaced with the other panel. The use of a vest-wearing mannequin without provision for patting down or adjusting the vest between shots would raise the question of whether the vest could be adjusted between the test of the front panel and the test of the back panel.

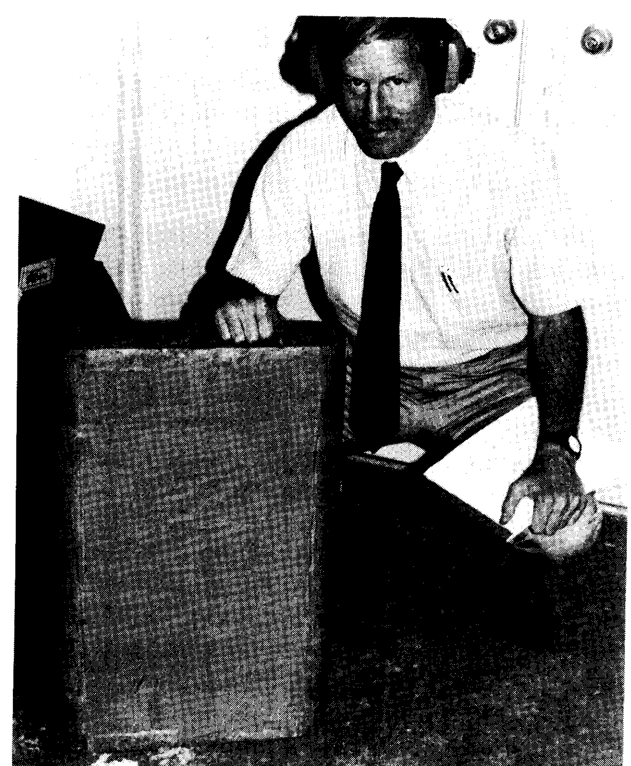
A compromise test fixture could consist of a flat block of clay contained in a fixture to which the vest could be attached with its own straps-such a fixture is termed a "curv.' The NIJ found the curv to be

35 However, it *can* make a difference, and did so in shootings performed by DuPont. In the course of a program of reenactments (see below), DuPont used the PPAA test set-up, but performed a pre-test with armor of the same style as the victim's mounted on an *unanchored* frame containing an NIJ-like clay block. In one reenactment, the armor on the mannequin was penetrated even though the corresponding panel on the clay block was not.

Figure C-3-Mounting Fixtures for Ballistic Tests of Body Armor







Top left: Clay block specified by NIJ Std. 0101.03 in rectangular frame.

SOURCE: El. du Pont de Neymours & Co., Inc., 1992.

Top right: Clay-filled mannequin specified by PPAA STD-1989-05.

SOURCE: Office of Technology Assessment, 1992.

Bottom right: Clay block in curvilinear frame tested at HPWLI by NIST/OLES for NIJ.

SOURCE: Office of Technology Assessment, 1992.

superior to the .03 block, partly on the grounds that ballistic tests of identical armor showed greater consistency on the curv than on the block. OTA concurs. We have not assessed the statistical signficance at which the data support this conclusion, but the data do not contradict it, and the greater realism of testing whole armor attached by its own strapping and fasteners is a strong argument for the curv relative to the block.

NIJ also found the curv to be superior to the PPAA mannequin. In these tests, the face of the clay in the mamequin's box was planed to facilitate accurate measurement of the backface signature. When armor was mounted on the clay, it arched over the clay in the box, and was not 'in intimate contact' with the clay as required by both NIJ Std. 0101.03 and PPAA STD-1989-05. This arching may have contributed to ply separation, penetration, and variance in results. OTA does not believe that ND's test of the mannequin was consistent with provisions in NIJ Std. 0101.03 for testing of Type IV or female models, which, like PPAA STD-1989-05, allow—and, in fact, require-clay to be mounded behind the armor panel to assure that the panel is in intimate contact with the clay. We acknowledge that measurement of BFS would be most accurate if the crater were made in an initially flat part of the clay, but this need not include the whole face of the clay block. We see no reason why the PPAA mamequin is necessarily inferior to the curv, and some armor might fit a mannequin better than the curv.

Test Ammunition and Velocities

Test ammunition has been critiqued both for inconsistency and for outlandishness. As critics point out, the standard's specification of bullet weight, caliber, and construction (e.g., 158 grain .357 jacketed soft point) allows for considerable variation: "Bullets of identical weight and caliber are made by many different manufacturers, each with its own particular bullet design and metal/alloy formulation.' On the other hand, the 0101.03 standard states that "The test ammunition specified in this standard represent common threats to law enforcement officers. ' [144] For this reason, a test facility was asked to cease using a brand of particularly effective bullets on the grounds that they were available only as ingredients for hand-loading (not in ready-to-fire cartridges) and thus did not represent a "common threat" to law enforcement officers.

The ranges of specified test velocities lie towards the upper end of the velocities obtainable with commercially available ammunition and guns, consistent with the principle that the test should be conservative. Some argue that the velocity of the .357 bullet used in testing Type II armor (1,395-1,445 ft/s) is beyond what would credibly be encountered in real life.[36] If so, Type II armor is being overstressed by the test and could be made lighter, more comfortable, and cheaper while still protecting against a realistic .357 threat. The question of the distribution of speeds at which this round hits armor in assaults is a technical issue that can be revisited by the NIJ.

It is not quite the case that test bullets with velocities outside the allowable range are ignored: for obvious reasons, a bullet that goes too slowly but penetrates the vest anyway suffices to fail the vest, and a bullet that goes too fast but is stopped by the vest counts as a fair hit. Only underspeed nonfailures and overspeed failures are counted as unfair hits. These rules led one manufacturer to request that his vests be shot with slightly overspeed bullets: any penetrations would be unfair hits, whereas non-penetrations would count towards passage of the vest.

Backface Signature Limit

The rationale for the 44-mm backface signature limit is described in appendix A, Origins of and Rationale for the NIJ Standard.

Critics of the 44-mm backface signature limit cite a variety of alleged defects in the way it was derived, including:

- the use (in some tests) of blunt, heavy, and slow test projectiles instead of small, fast bullets;
- the lack of any armor on the animals shot with the blunt impactors;
- the use of a type of armor fabric never commercially used for body armor in those tests that were done with armored animals;
- the lack of variety in the momenta of the bullets shot at armored animals;

[36] The Pennsylvania State Police used several .357 revolvers in an attempt to find one that could shoot the specified round at 1,395 ft/s or faster. They found that one cylinder of one of the guns could do so.

- the reliance on kinetic energy as an explanatory variable in mathematical models of wound causation;
- the dependence of backface signature depth upon momentum;
- the use of gelatin as a tissue simulant for purposes of assessing backface signature when such gelatin is, at best, representative of tissue only for purposes of penetration;
- the use of 20 percent gelatin, which behaves differently from tissue even with respect to penetration;
- the use of goats as test animals, despite their overall small size compared to humans and, in particular, the thinness of their body walls; and so on.

These critics also generally acknowledge that the researchers were doing the best they could with the resources available to them, and that the backface signature limit at which they arrived was probably reasonable at the time. [87] However, they argue that we are now in a position to improve on the original set of conclusions.

Defenders of the 44-mm backface signature limit can adduce a variety of rebuttals to the above allegations. They rebut the objections about bullets, armor, and blunt impactors by explaining that the blunt, heavy, and slow impactors were meant to simulate the effect of the bullet-backed armor thudding into the victim's torso. To simulate the impact of bullets of the same momentum (mass times velocity), they had a heavier mass and slower velocity. Being heavier, they could be wider (i.e., blunter), to distribute the pressure over an area comparable to the diameter of the depression made in gelatin or clay by armor stopping a bullet. They could also be longer, which allowed the maximum momentary indentation produced in an animal's skin (or gelatin) to be recorded by high-speed cinematography and later measured. They excuse the goat-human dissimilarities on the grounds that goats are conservative models of humans, in the sense that if a goat survived a certain impact, a human would be able to survive it at least as well. (The experimenters aspired to later shootings of primates, but lack of funding and a changing attitude towards such experiments left this hope unrealized.)

Some defenders of the 44-mm backface signature limit also cite the 25-mm British (PSDB) limit and an alleged 20-mm German BFS limit as evidence that it is reasonable to have a BFS limit even smaller that 44 mm. It may be, but the argument cannot rest on the British and German BFS limits, because 25 and 20 mm are not the respective limits for lightweight concealable armor and were not derived using the same backing material normally used for NIJ certification tests. Consequently, the risk they allow may differ from the risk a similar BFS limit would allow in a test otherwise similar to a NIJ certification test. In any case, the appropriateness of a BFS limit for the NIJ test cannot be decided until the NIJ makes explicit the maximum risks that it will accept and the minimum confidence with which it wishes the validity of the test to be demonstrated.

The 25-mm PSDB limit applies only to heavy armor having an areal density greater than 7 kg/m^2; this is equivalent to more than 25 plies of 1,000-denier, 31x31 Kevlar 29® fabric and heavier than most concealable armor worn in the United States. The limit was based on early PSDB tests using Plasticize (a modeling clay made in the United Kingdom) as backing material. Recent tests showed that under otherwise similar conditions (except temperature) and with both backing materials conditioned and warmed to pass the NIJ drop test, the BFS produced in U.S.-made Roma Plastilina No. 1 was greater than the BFS produced in Plasticize (almost double, at low velocities). In consideration of the results, the PSDB expressed ''some unhappiness with the (probably) conservative PSDB figure of 25 mm indentation" and "would welcome discussion on the need to revise this figure upwards." [28]

The (September 1988) German BFS limit for concealable armor is confidential,[37] but it is not 20 mm.[38] In any case, the developer of the German trauma-protection criteria observed that the BFSs produced behind 12-layer protective vests tested by the Bundeskriminalamt (BKA) were smaller by a factor of 1.8 than those obtained under roughly similar conditions in research sponsored by the NILECJ (now the NIJ). He conjectured that the

[37] The Management Board of the Technical Commission of the Police Management Academy, Research and Development Institute for Police Technology [Box 480 353,4400 Muenster, tel. (02501) 806-1] does not allow the September 1988 Technical Guideline for Bulletproof Vests [Richlinie Schutzwesten] to be quoted without its written permission.

[38] Ref. [28] errs on this point.

difference was most likely caused by some difference in the properties of the backing materials used, but note that another possible explanation was that the vests tested by the BKA had "foils" between the layers of ballistic fabric, which may have had an important effect on their stiffness. [122, 123] With this adjustment, a 44-mm BFS in a NIJ test was presumed to correspond to a 24-mm BFS under conditions of the BKA. OTA cannot endorse this adjustment, because the difference may have been due to the foils, but the important point is that the developer of the German BFS criteria believed that this *was* a proper adjustment.

Reenactments

Critics of the 44-mm backface signature limit often point to experience in the field, where no deaths due to blunt trauma caused by a nonpenetrating bullet are known. Some have gone so far as to attempt to reenact the circumstances of selected ‘‘saves’ from death or serious injury by shooting so as to see what the backface signature may have been. These reenactments use armor, weapons, and ammunition identical, or as nearly so as possible, to the armor, weapons, and ammunition involved in the original saves. In each reenactment, a shot is fried at the vest while it is mounted on a clay backing. The backface signature of the shot is measured in the backing. Those responsible for creating the reenactments point out that great accuracy in range is not needed because projectiles slow down only slightly as they move downrange. They recognize that the incidence angle in an assault, which may not be known accurately, may influence lethality significantly, but, in the reenactment, they shoot the vest at normal incidence for comparison with the NIJ (or PPAA) deformation test, justifying normal incidence as the "worst case." [87]

The use of normal incidence in a reenactment is the worst case, but it is the worst case for the NIJ standard, not for the victim officer. Suppose that the victim receives a shotgun blast at some random angle of incidence and lives. A reenactment done at zero degree (i.e., perpendicular to the plane of the armor, or 'normal") incidence for the sake of being the "worst case" will almost certainly subject the clay to a greater impact than that received by the shooting victim. Because almost no shootings occur at exactly normal incidence, a normal-incidence reenactment would be almost guaranteed to stress the vest more than did the original shooting, creating a backface signature corresponding to a greater blunt trauma than the one originally received by the victim officer, or even penetrating the vest outright. In fact, in some "saves,' it can be argued a priori that the angle of incidence was nonzero, on the basis that a head-on shot would have penetrated the vest.[39]

However, a cogent argument for the use of normal incidence in reenactments can be made on grounds other than that it is the worst case. The purpose of the reenactment is to ‘‘test the test, not the vest' we know (in some sense) about the vest already because we know the condition of the victim officer after the shooting.[40] The reenactment tells US if the test is a good one. Especially because many of the shootings involve guns and ammunition (in particular, shotguns and .45s) not used in the Type I, II-A, II, III-A, or III tests, it is worth thinking of the reenactment as a test of Special Type[41] armor made to stop the ammunition in question. As an NIJ test, then, not as a reenactment, the shot should be fired at normal incidence.

The 44-mm criterion for BFS is (one must assume) chosen so that passing it in a normal-incidence test shot indicates that the vest is adequate to protect the victim officer from blunt trauma. Following the goals enunciated by the NILECJ, we interpret "adequate" to mean that a person hit on armor by one nonpenetrating bullet at a velocity that would produce a BFS greater than 44 mm in an NIJ test would have a 10-percent or greater probability of suffering blunt trauma serious enough to

1. kill him or her (even if medical attention is available within an hour),
2. indicate corrective or diagnostic surgery, or

39 See, e.g., [123], p.13.

40 In many of the reenactments performed to date, the condition of the victim officer is the only indicator of vest quality because the vest was not NIJ-certified.

41 See appendix A.

42 The NILECJ refers to the victim officer as a ‘‘man’’ in stating this part of the requirement. It is not clear to OTA whether NILECJ meant to state the standard in terms of the effect on males or was merely conforming to the nongender-neutral language standards still in use at the time.

43 The Army's body armor medical assessment team assumed this also meant "regardless of the wearer's weight, sex, or body-wall thickness." Accordingly, one should not be surprised if far fewer than 10 percent of all shots producing a BFS of about 44 mm are lethal.

3. render him[42] unable to walk away from the scene of the shooting.[43]

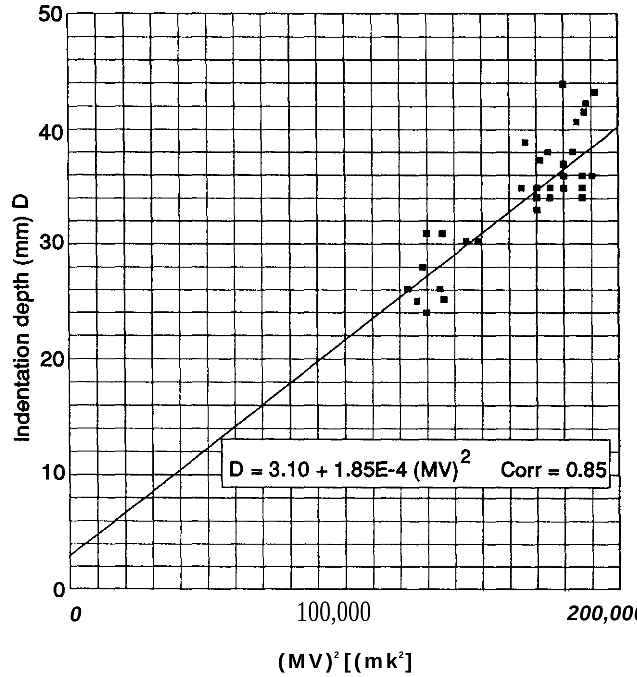
One must assume that the real-life variables artificially held constant in the test-range, angle, the chance that one shot could land near where an earlier one had, etc.—are subsumed into the 10 percent, along with the fact that real ammunition may be shot at velocities different from those specified for the test ammunition. Figure C-4 shows some examples of variations in muzzle velocities and backface signatures produced under similar conditions.[44] To account for this variation accurately, it is valuable to perform several reenactments of each shot and to consider the distribution of backface signatures corresponding to each reenacted shot.

How To Interpret the Results of Reenactments

Because reenactments are done retrospectively and, inevitably, with some amount of selection, they are in no way a random sample of past shootings. Therefore their results require a special form of interpretation, more complicated than the freshman statistics that suffice for interpreting simpler test data. The rationale for prospective inferences based on retrospective tests is explained fully in appendix D. An important conclusion is that if the test is to have any statistical significance, it will be necessary to reenact at least one shot that caused excessive trauma as well as shots that caused acceptable trauma. Otherwise, the fact that the measured backface signatures would be associated with only acceptable trauma would have no statistical significance; it would be the only possible outcome that could result from such an experiment. Put another way, the test cannot meaningfully find 44 mm to be too little if the cases are selected so that it cannot find some amount that is too much.

Because the interpretation of the results will take into account the fact that the cases are selected retrospectively, there is no reason to make the sample in any sense representative. A nonrepresentative sample, such as one with a more even mixture of acceptable and unacceptable outcomes than is present in real life, can be even desirable on the grounds that it will shed the most light on what level of BFS best represents the dividing line between vests that will transmit unacceptable blunt trauma and those that won't. Indeed, there is no reason not to recycle the few unacceptable events, re-enacting each several times so as to provide this even mixture. Again, the price that is paid for these freedoms is the need to perform the specialized and relatively complicated statistical analysis described in appendix D.

Figure C-4-Variations in Muzzle Velocities and Back-face Signatures Produced Under Similar Conditions



NOTE: Mark 22 9-mm bullets fired from a Thompson Contender with a 10-inch Barrel at Panels of 20-ply, 1,000denier, Kevlar-29 on Roma Plastilina No. 1 Modeling Clay Conditioned in Accordance with NIJ Standard 0101.08.

SOURCE: M.J. Iremonger and S.J. Bell, 1991 [84]. Redrawn by the Office of Technology Assessment, 1992.

It is important to note that death is not the only outcome deemed unacceptable by the NIJ: the need for surgery or the inability to walk from the scene of the assault also qualify as unacceptable results of blunt trauma. Even so, there are few—if any-cases of lethal, operable, or incapacitating blunt trauma caused by ballistic impacts on armor. The number of cases depends on the definition of blunt trauma. For example, one officer was killed by a rifle bullet that his soft armor stopped, but the armor, pushed by the bullet, penetrated into his chest cavity, killing him. [133] Some argue that this was a penetrating wound-not blunt trauma-even though the bullet

44 Others may be inferred from clay cavity data published in [8].

did not penetrate the vest. Others argue that some laceration—i.e., superficial penetration of the skin—often accompanies blunt trauma, and the depth of penetration is a question of degree, not kind.

Whatever definition one adopts, it is clear that the intent of the BFS limit was to limit to 10 percent the risk of death, operable injury, or incapacitation resulting from a stopped bullet of the type and velocity against which the armor is certified to have ballistic resistance. This is a very real risk that should not be underestimated. In particular, even if up to 90 percent of shots that would produce backface signatures deeper than 44 mm in clay did not produce serious blunt trauma, there would be no reason to change the limit. But if reenactments show that an even greater percentage of shots that would produce backface signatures deeper than 44 mm in clay did not produce serious blunt trauma, it would make a case that the BFS limit could be greater than 44 mm without exposing wearers to greater risk than that allowed by the original NILECJ safety goals, which have not been revised (or explicitly endorsed) by the NIJ.

Another goals-related issue remains, one about which it is harder to divine the intent of the NILECJ: the desired probability of acceptable armor passing the test. This issue is perhaps more salient if it is recast as the NILECJ's tolerance of cases in which acceptable armor would fail the test. No explicit statement of this level was made, and yet it is a key parameter: a testing program that did not aspire to any particular ability to approve acceptable items could (like some movie reviewers) avoid ever approving a defective item by the simple expedient of rejecting everything. (See also app. E.)

Some Reenactments Have Been Done

Recognizing limitations of the few "scouting test' reenactments performed in 1990, DuPont contracted with H.P. White Laboratories to perform a larger number of reenactments on October 23-25, 1991. DuPont invited OTA to send observers. OTA sent one observer to witness the reenactments.

One question immediately raised by the reenactments is how one is to treat cases in which the reenactment shot penetrates the vest, especially those in which the vest in the original event was not penetrated. The simplest answer to this question, based on the precept that the purpose of the reenactment is to see how a vest would have performed in test, is to count a penetration as an infinitely deep BFS failure. More subtly, one can analyze the reenactment data in such a way as to arrive at a BFS equivalent in danger to a penetration. (See also appendices D and E.)

Reenactment is the only approach that can permit models of human lethality to be tested scientifically. (In most cases, experimental shootings of armored humans would be unethical.[45] The suggestion has been made that one could establish some limit on BFS through a series of shootings that approached the unacceptable from below, starting with a very mild impact and working upwards until the volunteer subject stated that he or she had had enough.) Such data could be used to develop or improve, as well as test, lethality models, as described in appendices D and E.

Importance of the Backface Signature Limit

The stakes in the controversy over the backface signature limit have been lower than those in the controversy over penetration testing. Whatever its validity, the BFS limit has not been nearly so demanding as the nonpenetration criterion: Of the 550 models of armor submitted for certification testing to the .03 standard through Oct. 31, 1991, only 15 failed the BFS test alone (1 each at levels I, II-A, and II; 10 at level III-A, and 2 at level III), while 166 failed because of penetration only and 40 failed because of both penetration and excessive backface deformation. The number of BFS failures is somewhat deflated by the fact that no BFS measurement is made in the event of a penetration failure on the frost shot. [55, 56,57, 58]

These statistics, and the rarity of serious blunt trauma injuries in the field, have led some to suggest that the idea of danger from blunt trauma is a red herring and that the BFS limit could be abandoned altogether. Not only would such a course of action render moot the difficult question of finding the correct BFS threshold, it would also open the way to using a backing material other than clay. After all, clay was chosen because its inelasticity afforded the opportunity to measure BFS. Some believe that a

45 The Federal Aviation Administration test to assure that an airplane can be evacuated quickly is performed with paid volunteers. Injuries can occur as these people all try to get out of the darkened airplane in 90 seconds. Participants are warned in advance that people have been hurt before in such tests. [6]

more elastic and flesh-like material-such as ballistic gelatin-would eliminate the test armor's bunching and balling; they would see elimination of the BFS criterion as paving the way for a switch away from clay.

Others deny that the rarity of failures due to BFS alone indicates that armor passing the penetration test alone would provide adequate protection from blunt trauma. They point out that armor tested in the past was at least designed in the hope of passing the BFS part of the test, and claim that in the absence of any BFS criterion, whatsoever, radical and dangerous new armor designs could arise, designed solely to prevent penetration of bullets and with the possibility of transmitting enormous blunt impact to the wearer. For example, armor made of aramid felt or knit (as opposed to woven) aramid fiber could stop bullets and could even be very flexible, light, and cool, but would have enormous proboscis-shaped backface signatures.

Another reason to have a standard for protection from blunt trauma is that a typical reaction to a suggestion to buy or wear flexible body armor is to question whether the impact of a stopped bullet would not be dangerous or fatal.[46] When this question arises, the answer, "It can be, but there is a Federal standard for protection from blunt trauma, and my armor meets it," may be more credible and persuasive than the answer, "No, blunt trauma isn't really much of a problem, so the armor isn't tested for its ability to withstand it.'

Number of Shots

As explained in appendix A, the rationale for NIJ standard's multiplicity of shots against a single panel gradually evolved from economy to replication of a perceived multishot threat.

Police officers certainly do face a multishot threat. The introduction of 9-mm and .380 caliber handguns with magazines holding over a dozen rounds has increased the number of shots a criminal can fire. FBI statistics do not, however, show an increase in the average number of shots impacting on the upper torsos of victim officers-this number has hovered around 1.5 for the last 10 years, showing no definite trend. Nor has the maximum number of shots on the vest-protectable area increased: if anything, it decreased from 5 to 4 during the 1980s. The majority of multiple-shot cases are two-shot cases, and in some of these the impacts are divided between the front and back panels, so that neither panel sustains a multiple hit attack even though the officer wearing the vest does.

Perhaps because of recent attention to advanced weapons in the hands of criminals, or perhaps simply because of attention to the 35 percent or so of cases in which more than one shot impacts the upper torso, body armor customers want to be assured of protection from multiple shots [102] and the NIJ wants to test vests accordingly. (See also app. A of this volume.) The 0101.03 standard's test protocol, in which two angle shots (no. 4 and no. 5) are followed by a head-on shot (no. 6), is designed to test the resistance of the vest to multiple shots.

Especially because the angled shots push the edges of the vest towards the middle, rather than away from it, the last two shots are likely to hit a thoroughly bunched-up vest. Opponents of the current test see this effect as an artificiality: proponents see it as a useful feature of the test, assessing the multiple-shot resistance of the vest in an admittedly stressful manner. One option would be to shoot these shots across the vest, so that they stretch the vest rather than push it together.

Variation or "Inconsistency" of Test Results

Critics of NIJ testing have pointed out variation or ''inconsistency " in the test results, citing instances in which a particular model of vest passed the test and later failed it or vice versa, instances in which one panel of a vest passed the test when the opposite panel failed, and the disparity between the percentage of shots that result in failures and the percentage of vests that fail. In a widely cited sample, [65] 2.6 percent of the shots penetrated, 13 percent of the panels failed, 51 percent of the vests failed, and 72 percent of the panels that failed had opposite panels that passed.

If the behavior of vests were completely deterministic, and if the vests and tests were identical, there would be no occurrences such as those described above: a model of vest would either be capable of passing the test and would do so all of the time, or it would be incapable of passing the test and would

46 This common reaction is not unfounded. Individuals have been killed by a batted ball, or even a punch, landing on the chest. [70, 90, 154, 155, 159]

experience the same test history on the front panel as on the back, with six failures per panel for certain types of ammunition and/or conditions of wetness and zero failures for the rest.

The behavior of body armor is not completely deterministic, however. This fact alone explains some of the variation in test results. If, for example, samples of a certain model of armor are 99-percent certain to stop the test bullets, then the percentage of 48-shot tests the model should be expected to pass is[47]

$$0.99^{48} \times 100 \text{ percent} = 62 \text{ percent.}$$

Thus, there should be a large disparity between the percentage of shots that result in failure (1 percent) and the percentage of vests that fail (48 percent). Under the same conditions,

$$0.99^{6} \times 100 \text{ percent} = 94 \text{ percent}$$

of the panels will pass, so that 94 percent of the panels that fail will have an opposite panel that passes.

Viewing the results of NIJ testing in this light can be instructive. If the 2.6 percent per shot chance of failure[48] were evenly distributed among the panels, 15 percent would fai1[49]-the fact that only 13 percent do is indicative of some amount of shot-to-shot consistency, in that failures were more concentrated in certain panels than would be expected by chance alone. If the 13-percent per panel chance of failure were evenly distributed among vests, 68 percent of the vests would fai1[50]-the fact that only 51 percent do is indicative of some amount of panel-to-panel consistency. Similarly, the fact that 72 percent of the panels that failed had opposite panels that passed indicates some level of panel-to-panel consistency, inasmuch as if the 13 percent of panels that were bad were evenly distributed, a full 87 percent of the panels that failed would have opposite panels that passed. In other words, a gambler who placed bets about the performance of back panels on the basis of the corresponding front panels' performance would make money: a back panel whose front panel failed is more than twice as likely to fail as one whose front panel passed.[51]

While it is reassuring to know that the results of NIJ testing display some consistency, one might well wonder how much of the remaining randomness or inconsistency is attributable to the test and how much is inherent in the performance of soft body armor when operating near its limits of performance.

The bunching and balling effects described above have been cited as a source of randomness in test results.[52] One means of assessing their contribution is to examine the distribution of penetrations for signs that penetrations tend to occur on shots in the latter portion of the test sequence. Figure C-5, Locations of Level-II Penetrations, shows that shot 6 results in far more penetrations than do the other head-on shots and shot 5 results in more penetrations than shot 4. (Shots 1,2, and 3 impact head-on; shots 4 and 5 impact from directions 30 degrees right and left, respectively, of the perpendicular to the plane of the armor panel; shot 6 impacts head-on between shots 4 and 5.) These data suggest that the number of previous shots has a strong bearing on whether or not a given shot will penetrate. One possible explanation for this effect is that the bunching and balling, which increases with every shot, may cause amounting probability of failure. Alternatively, the vest may be weakened by repeated hits. In either case, one would not expect the number of penetrations to be lower on shot 6 than on shot 5, but it was (though not to a statistically significant degree). The difference may be because, other things being equal, penetration probability of some ammunition is lower at normal incidence than at a 30-degree angle. (Recall that the angled shot was instituted in response to the finding that 9-mm ammunition penetrated some weaves of armor better at an angle than it did at normal incidence.)

There is a statistically significant difference-at better than 95-percent confidence-between the penetration probabilities of shots 4 and 5, that of shot 5 being greater. [59] The explanation could be ply separation, overall weakening, or both. One way to decide between these alternatives is to look at results of tests in which the vests were smoothed out

47 Neglecting any failures on account of BFS.

4S This is actually the sample mean; the erect probability of penetration cannot be measured but only estimated.

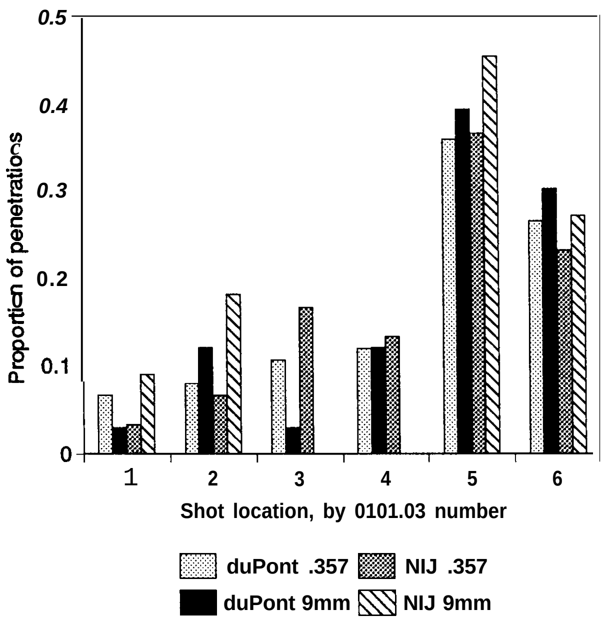
49 Because $(1-0.026)^6 = 0.85 = 1-0.15$.

50 Because (1-0,13)8 = 0.32= 1- ().68.

51 The conclusion that front-panel failures are not independent of ba&-panel failures is also supported by a chi-squared test of independence; see [59].

52 "It's a crapshoot," in the words of more than one expert interviewed for this study.

Figure C-5-Locations of Level-n Penetrations



SOURCES: Office of Technology Assessment, 1992, based on data provided by El. duPont de Nemours & Co., Inc., and National Institute of Justice.

between shots but otherwise tested according to NIJ 0101.03. Some tests of this type have been performed, and seemingly create a more even distribution of failures, but the testing was too limited (and the failures too rare) for firm statistics to be deduced. One could also examine the results of PPAA testing, in which the armor is smoothed between shots. Another avenue of investigation would be to consider all shots, not just the fair ones tallied above: some unfair hits would cause at least as much bunching and weakening as fair ones. Still another possibility, as yet unexplored, would be to shoot the six locations on each panel in inverse order. However, discovering the cause is not nearly so important as discovering whether ply separation is realistic—i.e., if it occurs frequently in actual assaults with several shots impacting on a panel.

Although ply separation, weakening, and other factors may cause shot-to-shot variations, a major joint cause of variation in passing retests is the variation in the ballistic resistance of armor submitted for certification testing and the stringency of the test, which fails about half the models submitted. It happens that the variance in outcomes of repeated testing is greatest when the probability of passing is one half. If the test were made less stringent (for example, by requiring fewer shots) so that it passed 99 percent of the models submitted, those that passed would pass a frost retest with a probability at least that high and would consistently (but not invariably) pass subsequent retests, but that would offer little evidence of their ballistic resistance. If the test were made more stringent so that it failed 99 percent of the models submitted, the few that passed would probably have greater ballistic resistance than most on the market today but would fail a first retest with a high probability, and would be very consistent in their failures of repeated retests.

A striking way of looking at the relationship between inherent statistical uncertainty and reproducibility is to consider that if a model passes a 48-shot test with no penetrations, one would have only 50-percent confidence in a (geometric-) mean stopping probability high enough for the model to pass a retest with a probability of 50 percent. One would have only 10-percent confidence in a mean stopping probability high enough for the model to pass a retest with a probability of 90 percent. *These bounds do not depend on the actual mean stopping probability or probability of passing the test; if the model were completely bulletproof, the inherent uncertainties of statistical inference would still be this great. In* particular, they would occur even if panels were patted down between shots, and so forth.

These bounds are also independent of the number of shots required by the test and the number of penetrations allowed. Increasing the stringency of the test (for example, by requiring more shots without changing the number of penetrations allowed) will increase confidence that any model that passes it will have some minimum mean stopping probability, such as 99 percent, but it will also reduce the probability that a model with a mean stopping probability of 99 percent will pass a retest. These opposite effects cancel one another exactly!

However, increasing the stringency of the test will allow it to show how good a good model really is, at a fixed level of confidence. Appendix E discusses some options for increasing reproducibility of test results without drastically increasing or decreasing consistency.

Temperature and Moisture During Actual Wear

Questions of ballistic materials' flammability, penetrability under conditions of heat or cold, and the observably increased penetrability of woven armor fabric when wet in turn raise questions about conditions of temperature and moisture during actual use. In the case of concealable body armor, which is worn on the torso, under clothing, and near the skin, the temperature is unlikely to depart from the 60°-1000 range within which the armor is tested. Some have questioned the need for wet testing on the grounds that officers' vests do not, in real life, get soaked.[53] Others point to the profuse sweating that can accompany vest wear in hot weather, as well as to a 1990 incident in which an officer was in fact shot twice by an assailant who had just held him underwater in an unsuccessful attempt to drown him.[54]

There is no doubt that fabric armor not treated for water-repellency or encapsulated in a waterproof cover loses some ballistic resistance *while* **wet** but recovers it after drying. For example, tests conducted by NIST's Law Enforcement Standards Laboratory showed that the V_{50} (the velocity at which bullets have a 50-percent chance of penetrating) for 20-ply Kevlar® panels struck by 124-grain, 9-mm, full-metal-jacketed bullets decreased from 1,406 ft/s for a dry panel to 1,222 ft/s for a panel that had gained 10.6 percent weight from soaking, to 930 ft/s for a panel that had gained 20.4 percent weight from soaking, to 828 ft/s for a panel that had gained an estimated 35 percent weight from soaking. For 12-ply Kevlar panels, the V_{50}s were 1,093 ft/s for a dry panel, 831 ft/s for a panel that had gained 15.6 percent weight from soaking, 781 ft/s for a panel that had gained 20.6 percent weight from soaking, and 721 ft/s for a panel that had gained an estimated 32 percent weight from soaking (see figure 11 of vol. 1 of this report) [62].[55] [56]

To pass a NIJ-like test for ballistic resistance, the V_{50} would have to be faster than the velocities specified for the test bullets. If wetting caused the V_{50} to approach the nominal test velocity, the probability of penetration per shot would approach or exceed 50 percent, and the armor would almost certainly fail the test. To estimate the risk of this happening in service, it would be desirable to collect statistics on moisture pickup by the armor when worn by the intended wearer; but that can't be done before the armor is purchased and worn! Second-best would be collecting statistics on moisture pickup by similar armor worn by other officers, ideally of a similar physique, performing similar duties in a similar climate. This could be done by any interested department; no survey of national scope has collected such data.

The feasibility and importance of weighing armor to measure its water uptake is illustrated by an experiment conducted at the FBI Academy, in which two instructors wearing 7-ply Kevlar® armor-one treated, the other untreated-exercised vigorously on a hot, humid day, playing handball 2 hours, eating lunch, teaching class, and then playing handball another half hour just before removing their armor to have it weighed to measure water uptake and shot to detect any degradation of ballistic resistance. The treated armor picked up 12 percent water (by weight); the untreated armor picked up 22 percent.[57] [58] Similar untreated armor worn by another

53 One manufacturer's promotional booklet [120] states that "'r'here is a 40-percent loss of stopping power when the [ballistic material] is 100-percent wet. Once the vest is dry, it is back to full stopping strength, [. . .] Even when totally soaked, [our II-A vest] will stop the commonly encountered .22's through .38's as well as buckshot and .45's. In other words, if someone can hold you underwater for 5-10 minutes, and then shoot you with a magnum, you are in trouble! Our experienced opinion is that waterproofing causes more trouble than it's worth because it gives the wearer a rubber-sheet effect, making the body armor too uncomfortable to wear.' [120]

54 See [150], p. 53. OTA could not determine whether the bullets impacted a wet portion of the armor.

55 The exact mechanism by which water degrades the performance of body armor fabric is not well understood. Experts consulted by OTA variously cited lubrication of the bullet's passage through the fabric, hydrostatic shock, and lubrication of the fibers themselves (making the fabric act like a safety net made with slipknots) as possible explanations. Conversely, one vest manufacturer's promotional material says that water makes the fibers swell, eliminating their ability to catch the bullet gracefully. All agree that performance is recovered when the fabric dries out.

56 When saturated, Spectra™ fabric holds less water than does saturated Kevlar™ fabric.

57 That is, the weights of the garments (the ballistic panels of which were not removable) increased by 12 and 22 percent of their initial (dry) weights as a result of absorption and retention of perspiration.

58 The higher figure does not represent complete saturation; an untreated garment of the style that absorbed 22 percent perspiration in the FBI Academy test absorbed 26.2 percent water in an Army test using a copper mannequin. Even this may not represent complete saturation, but OTA knows of no higher value measured for a similar garment. Water pickup in the NIST tests described above was for removable ballistic elements.

subject who spent his shift in a car picked up 5 Percent.[59] [8] The differences in percentage weight gained from absorption of perspiration may be attributed to the differences in treatment and type of duty.

The water absorption measured by the FBI, when compared to the NIST data on V_{50} versus water content, suggests that

1. prolonged exertion can cause untreated armor to lose a significant amount of ballistic resistance,
2. treatment decreases the loss, and
3. untreated armor may lose little ballistic resistance during sedentary duty.

However, there is too little information to assess, on a national scale, the effect on risk of making wet-testing optional or certifying wet and dry ballistic resistance separately.

Officers may also face exposure to blistering heat—for example, running through a puddle of flaming gasoline. Apparently such incidents are rare: the IACP/DuPont Survivors' Club attributes less than 2 percent of its more than 1,300 saves to protection afforded from flame, heat, or explosion by body armor. Rarer still is being shot under such a condition; we know of no such case in police use. Of course, it could happen, and protection may be desired.

Polyaramid fiber such as Kevlar® and Twaron® is inherently flame-resistant. It does not melt but does char at temperatures above 800 'F; it is self-extinguishing when the flame source is removed. The tensile strength of Kevlar 29 decreases about 45 percent as its temperature is increased from 80 to 560 OF, but only about 7.5 percent as the temperature increased from 80 to 160 'F. [106]

In contrast ,the extended-chainpolyethylene (ECPE) plastic from which Spectra[20] fabric and Spectra Shield® are made melts at about 300'F (150 'C), but Spectra™ fabric retains 94 percent of its room-temperature ballistic resistance[60] at a temperature of 160 °F (about 71 °C).[61] Armor that hot would be excruciatingly painful and would burn skin in less than a second. [128]

Spectra™ fabric and Spectra Shield® can be ignited but are less flammable than are cotton, nylon, or polyester fabrics commonly used for police uniforms.

Armor made from Spectra Shield® has been tested for flammability by Southwest Research Institute (SwRI) under simulated conditions of police wear (on a mannequin standing in a pool of flaming gasoline from a Molotov cocktail) and by the Naval Air Development Center (NADC) under simulated conditions of military wear (running for 3 seconds over a pool of flaming JP-4 jet fuel). [98] The essence of the conclusions of both studies was that Spectra Shield® would *protect the* part of the body it covered from flame and blistering heat until well after other clothing had caught fire and other parts of the body had been subjected to blistering heat. These tests were sponsored by Allied-Signal. We note that the NADC test used military-style armor covered with flame-resistant Nomex™ fabric, which is not used on most models of police armor. The SwRI test used a police model covered with flame-retardant cotton/polyester fabric.

DuPont has also tested Spectra Shield® and Kevlar® armor for flammability and produced a videotape comparing the results. In these tests, the armor was placed on a mannequin outside of a flame-resistant Nomex™ coverall in which the mannequin was dressed. This, too, does not represent normal police use.

In general, the risk of flammability an armored officer faces depends not only on the ballistic material used in the armor but also on the material used for its cover and carrier garment, the material used for the officer's uniform or other clothing, and whether the armor is worn over or under such clothing. We judge that, in the case of armor undergarments, the ballistic material used in the armor is the least important of these factors.

[59] There were no penetrations of the untreated armor that picked up 5 percent weight or the treated armor that picked Up 12 percent weight, but the untreated armor that picked up 22 percent weight was penetrated by 9 of 10 .22-caliber bullets fired at the front panel. However, this difference in ballistic resistance cannot be attributed to differences in treatment or water uptake, because the velocities of all 10 shots fired at the panel that was penetrated were greater than the velocities of all 20 shots fired at the panels that were not penetrated. The probability that such a difference in velocities would occur by chance alone (i.e., under identical conditions) is less than 0.0001 (based on a l-sided Wilcoxon test).

[60] Viz., V_{50} measured per MIL-STD-662D using a .22-cal., 17-gr fragment-simulating projectile.

[61] These and other high-temperature tests were conducted by HPWLI for Allied-Signal, Inc.

There are other rare conditions (e.g. bleaching) to which ECPE is more resistant than is polyaramid. Manufacturers of Kevlar[R] fiber and armor caution wearers not to bleach it, as does the NIJ, but cases of bleaching Kevlar[R] armor have been reported, and degradation is irreversible. Future armor made from new materials may have different vulnerabilities to environmental conditions that cannot now be enumerated but exposure to which would be rare. For example, armor made from synthetic spider dragline silk might be degraded by exposure to lemon juice, vinegar, or battery acid.

Philosophy of Testing and Design

Conservativism

Only a tough vest can pass a tough test, so conservativism in testing engenders conservativism in design. For example, the bunching and balling described earlier occurs in tests and is not patted down because of the conservative assumption that it might occur in the field as well. Thus, the stiffening introduced by manufacturers[62] to mitigate bunching and balling is an expression of conservativism in their designs: while it helps pass the test, it may or may not help in the field.

Other examples of conservativism are readily found-the allowable amount of backface signature, the number of shots per panel, the velocities at which the bullets are shot, and so on, all reflect considerable conservativism.[63] These all translate into conservative designs for vests.

Few would argue with the idea that vest testing and design ought to include some element of conservativism: nobody would want a vest labeled "Guaranteed by the U.S. Government to pretty much protect the wearer most of time from average ammunition. However, some feel that the NIJ standard contains too much conservativism, and results in vests that are needlessly expensive and uncomfortable. Proponents of this view argue that the NIJ standard therefore lowers the number of officers in vests, ultimately leading to officer deaths that could have been avoided by promulgation of a less conservative standard. [87]

Officials of the NIJ respond to charges that the standard is overly conservative by citing the standard's several levels of armor, saying:

> Some argue that changing the standard will permit a lighter and more flexible vest, thus increasing the likelihood that the armor will be worn routinely. However, NIJ feels that the officer already has a range of choices—the classification of threat levels by which armor is already rated. [151]

and,

> An officer who feels uncomfortable with a vest at a given threat level can always chose to wear a vest complying to a lower threat level. However, in this circumstance, the officer knows that the lighter vest has less ballistic resistance. [151]

Presumably an officer who felt that the standard was too conservative and the resulting vests were too heavy and expensive could opt for a lower level vest and hope that, because of its conservative design, it would stop higher level threats. Actual experience shows that such a hope would be well-founded: many "saves" have involved lower level vests stopping higher level bullets. However, some vested deaths have involved lower level vests failing to stop higher level bullets: an individual officer could decide to take this chance, but how could a department make such a choice for its officers, or defend such a choice in a court case brought by a slain officer's surviving spouse?

"Go, No-Go" Testing

An NIJ certification test has only two possible outcomes-certification of the vest model, or failure. In this respect, it is like many tests faced by people. Presumably the person who fails and subsequently retakes a driving test learns more about driving in the time between the original test and the retest. Unlike people who fail driving tests, a vest model cannot improve, so it cannot retake the test: it must be abandoned by the manufacturer, who can then learn more about vest-making and submit a better model of vest next time.

62 By using extra stitching or by the use of Stiffer fabric.

63 The backface signature is one standard deviation less than the mean found to be safe for animals; the number of shots per panel is far mOre than the average number of hits per panel in a gunfight; the velocities are one standard deviation more tbanthemeanfound by testing commercial ammunition. (See also app. B, this volume.)

V_{50} Testing

Other things being equal, the probability that a nondeformable projectile will penetrate a piece of armor increases with the speed of the projectile: it is zero for stationary projectiles and is generally considered to be 100 percent for some suitably high speed, with a‘ ‘zone of mixed results’ in between.[64] Velocities in the zone of mixed results correspond to penetration probabilities between zero and one. The V_{50} is defined as that velocity at which a given projectile has a 50-percent chance of penetrating a given armor.

Being a statistical construct, V_{50} is estimated, not measured. There are two principal means of estimating it in use: a Department of Defense (DOD) protocol [138] and regression techniques for fitting a logistic [91] or probit[65] model (i.e., formula) for dependence of penetration probability on velocity.

In the DOD protocol, one seeks to develop a set of at least N shots such that there are an even number of shots, equally divided between penetration and nonpenetrations, and the velocities of the shots all lie within a 125-foot/second range. N is typically 6 or 10. The V_{50} is the mean[66] of the velocities in this set of shots.

In the regression methods, V_{50} is found by assuming that a certain functional form applies to the penetration probability as a function of velocity, regressing to find the parameter values that best explain the outcomes of test shots (in the sense of minimizing the mean squared error or maximizing the predicted likelihood of the outcomes), and then interpolating or extrapolating to find V_{50}.

For example, the data in table C-1 show the performance of a Type II-A vest against .44 Magnum ammunition.[67] The vest was shot on an NIJ-style clay block, but was smoothed after each shot. These data lead to a V_{50} of 1,327 feet per second by the logistic regression method. Because the DOD method actively “hunts” for the V_{50} by lowering the bullet velocity after a penetration and raising it after a stop, that method cannot be retrospectively applied to a given series of shots.

Table C-l—Example of Penetration Data

Velocitvy (ft/s)	Penetration
1,229	no
1,273	no
1,278	yes
1,292	no
1,369	no
1,382	yes
1,394	yes
1,403	yes
1,404	yes
1,414	yes
1,422	yes
1,422	yes
1,426	yes
1,429	no
1,429	yes
1,433	yes
1,436	yes
1,438	yes
1,449	yes

SOURCE: DuPont Co., 1991 (reenactments).

The V_{50} is of interest because it provides an alternative to the “go, no-go” format of the NIJ standard: It provides a quantitative index of ballistic resistance, but it can also be used for a‘ ‘go, no-go’ test by specifying a minimum acceptable V_{50}. Some body armor companies already use V_{50} tests of multi-ply sample panels of fabric to decide whether the fabric is acceptable for use in their body armor.

The V_{50} could be used in a variety of ways in the testing of body armor. One way would be to test the *design* of the vest with something resembling the present NIJ test, and measure the V_{50} as well. Subsequent lots of the same model would be given V_{50} tests to see if they are of the same quality as the original vest used in the design certification. The V_{50} provides a more sensitive measure of quality than does the NIJ test’s simple pass-fail grading, and has

64 In the case of deformable projectiles, increased speed can increase the flattening of the projectile and thus actually lower the probability of penetration. Even more extreme cases can be found-one expert told of firing a ball bearing at a speed measured in *miles per second* at a block of ballistic gelatin, only to have the ball bearing shatter and the block of gelatin remain unpenetrated! Conversely, there are some indications that very slow .22 caliber bullets can penetrate vests because of their shape and lack of deformation at low impact velocities.

65 See J.R. Ashford, ‘‘Quantal Response Analysis,’’ pp. 402-408 in Samuel Kotz & Norman L. Johnson, eds.), vol. 7 (New York, NY: John Wiley & Sons, 1986. Francis S. Mascianica, “Ballistic Testing Methodology,” pp. 60-61 of [93], describes an application to ballistics (without using the term “probit”).

66 Not, somewhat surprisingly, the median.

67 Shot at H.P. White Laboratory, Oct. 24, 1991. (The vest being a II-A, it is rated to stop 158-grain .357 bullets at 1,2501,300 ft/s and 124-grain 9-mm bullets at 1,090-1,140 ft/s.) The backface signatures resulting from the nonpenetrations were of 44 mm or less.

the advantage that there is no risk of failing an already-certified design, as there is under the present system.

One objection to the use of V_{50} tests is that "the average police officer won't understand them. " A related objection is that estimated V_{50}s would be viewed as scores, perhaps leading manufacturers to compete with one another in offering armor with the highest score—far in excess of what is needed to provide the level of ballistic resistance demanded and leading to increased manufacturing cost and reduced comfort and wearing. Another objection is that neither officers nor manufacturers want to deal with any concept that requires and demonstrates the penetration of vests, even if by much faster bullets than the vest is designed or certified to stop.

These concerns are understandable and have some validity. Nevertheless, other standards that involve the failure of a product do not appear to suffer from undue customer incomprehension, or revulsion at the idea that the product could fail. Fishing line, for example, is rated in terms of its breaking strength; lightbulbs and automobile batteries are rated in terms of their expected lifetimes; antifreeze comes with a table on the side of the container showing that the same product can fail two different ways, boiling and freezing. There is also value in reminding manufacturers, purchasers, and wearers that vests *can* be penetrated by sufficiently energetic rounds. This would underscore the NIJ's warnings that "there is no such thing as a *bulletproof* vest" [144] and, more generally, that "there is no such thing as 'bulletproof' armor." [145] Finally, the V_{50} test could be done (as it is by some manufacturers in their quality-assurance programs) with a non-bullet projectile, lessening the negative feeling arising from the penetration of the vest by a bullet.[68] As for the fear of competition in V_{50} scores, manufacturers[69] have already competed in matters such as liability coverage, backface signature, and the ability to stop very large numbers of shots or shots at very high velocity.

An advantage of estimating V_{50} by **regression** (instead of the DOD method) is that it provides a formalism for also estimating the velocity, V_{10},[70] at which the penetration probability is predicted to be 10 percent. Similarly one could use the same data to estimate the velocity at which the penetration probability is predicted to be 1 percent or any other value. There is a great deal of complex theory on the validity of such extrapolations,[71] but it boils down to this: one should be cautious of extrapolation, especially to extremes. In fact, simple logistic models and probit models are absurd at low velocities: they predict a nonzero penetration probability at zero velocity. More complicated logistic models that depend on certain nonlinear functions of velocity do not have this defect,[72] but even so, one must be cautious about using them to predict penetration probabilities at velocities substantially different than those of the projectiles fired in the tests to which the model was fitted.

If one is interested primarily in the V_{50}, it is best to adjust the velocities used in the test to be near what one expects the V_{50} to be, although one need not adhere to the DOD protocol for doing this. If, on the other hand, one is interested primarily in the V_{lo}, it is best to adjust the velocities used in the test to be near what one expects the V_{lo} to be. A procedure analogous to the DOD V_{50} procedure could be developed for finding the V_{lo}.

For comparable accuracy and statistical confi dence, more shots would be needed to estimate an extreme fractile (e.g., V_{10} or V_{90}) than to estimate the V_{50}. Partly for this reason, the V_{50} is of interest as an indicator of variation in the manufacturing (or testing) process. A more appropriate indicator of *quality* would be the fractile corresponding to the maximum acceptable penetration probability (if any) established by policy. For example, if the NIJ

6S The manufacturers are not seeking to avoid feeling bad when they use "fragment simulators" instead of bullets. Fragment simulators are made with much greater item-to-item uniformity than is available in any line of bullets; they are made of machined steel.

69 According to sources familiar with competitive practices in the industry.

70'r'he V_{10} for penetration is the V_{90} for stopping.

71 In the II-A example above, V_{01} and V_{10} only turn out to be slightly slower than V_{50}; about 1,370 feet per second for each, despite the almost 200 ft/s span of the zone of mixed results. Supporters of the idea that current vests are over-designed will point out that III-A vests-two levels higher-are tested against .44 Magnum rounds traveling at 1,400-1,450 feet per second.

72 They can also predict nonmonotonic behavior such as that described above: e.g., a decreasing of penetration probability with increasing velocity up to a point, then an increasing of penetration probability, then a decreasing of penetration probability with increasing velocity at extremely high velocity. In such a case there could be three distinct V_{50}s!

were to state a goal of no more than 10-percent probability of single-shot penetration (analogous the NILECJ's stated goal of no more than 10-percent probability of blunt-trauma lethality), then one would be interested in estimating the V_{10} and should fire shots at roughly the expected V_{10}, or at the minimum acceptable V_{10}. Actually, policy should not specify a minimum acceptable V_{10}, because the true V_{10} cannot be measured; it can only be estimated. A rational policy should therefore specify a lower confidence bound on the actual V_{10} and a level of statistical confidence to be demonstrated.