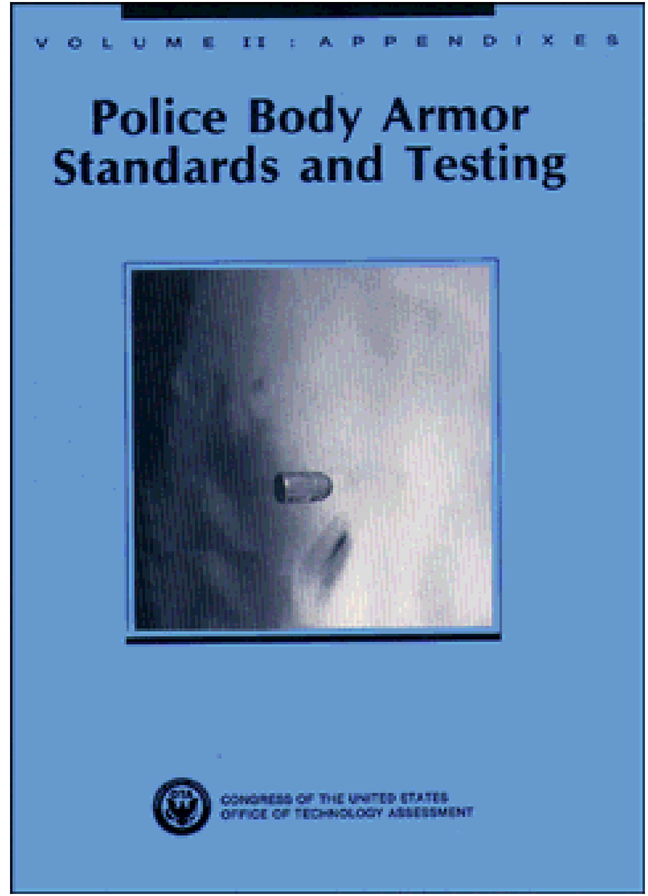
Police Body Armor Standards and Testing, Vol. II—Appendixes

August 1992

OTA-ISC-535
NTIS order #PB92-101731



Recommended Citation:

U.S. Congress, Office of Technology Assessment, *Police Body Armor Standards and Testing, Volume II: Appendices*, *OTA-ISC-535* (Washington, DC: U.S. Government Printing Office, September 1992).

For sale by the U.S.. Government Prlnting Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328
ISBN O-1 6-038074-x

Foreword

For two decades, the number of police officers shot to death each year has been declining while the number of officers shot has been increasing. The decrease in the lethality of shootings is partly attributable to the increase in wearing of bullet-resistant body armor, especially soft, inconspicuous armor designed to be worn full-time.

A prospective purchaser can see how much of the body an armor garment covers but cannot see whether it will stop a particular kind of bullet at a particular velocity and protect the wearer from the impact. To provide benchmarks for protection, the National Institute of Justice issued NIJ Standard 0101.03 in 1987. It specifies standard procedures for testing samples of armor. If samples of a model pass, the NIJ or the manfacturer may certify that the model has the type of ballistic resistance for which it was tested.

The standard has been controversial since it was issued. This report describes the origin of the standard, the rationale for particular provisions, and the main points of controversy, which concern acceptable risks, the validity and discrimination of the test, and the reproducibility of results. OTA finds that resolving these controversies will require specifying acceptable risks quantitatively, performing additional research to test validity (the correspondence of test results to performance in service), and implementing a quality-control program.

To date, all armor of NIJ-certified models has performed as rated in service-but uncertified armor, including armor that would fail the test specified by the standard, has also performed as advertised. This has provoked charges that the NIJ test is too stringent and fails to discriminate some safe armor from unsafe armor. The validity and discrimination of the test are technical issues that are susceptible to scientific analysis-if the NIJ specifies maximum acceptable risks quantitatively. The report describes illustrative specifications of acceptable risks and an experimental method for deciding whether the current test, or any proposed alternative, limits the risks as required.

NIJ does not inspector test marketed units of certified models to see whether they are like the samples that passed the model-certification test. Without a quality-control program, NIJ has no basis for assuring police that the garments they buy and wear are like the samples NIJ deemed adequate. Indeed, samples of some NIJ-certified models have failed retests and in some cases differed from the samples originally tested for certification. This report describes and compares several options for a quality-control program.

This assessment was requested by Senator Joseph R. Biden, Jr. (Chairman), Senator Strom Thurmond (Ranking Minority Member), Senator Dennis DeConcini, and Senator Edward M. Kennedy of the Senate Committee on the Judiciary; Congressman John Joseph Moakley, Chairman of the House Rules Committee; and Congressman Edward F. Feighan of the House Committee on the Judiciary and of its Subcommittees on Crime and on Economic and Commercial Law.

OTA's findings and analysis of options were reported in *Policy Body Armor Standards and Testing: Volume* **Z in** August 1992. This volume contains all appendices to the report.

JOHN H. GIBBONS
Director

Police Body Armor Standards and Testing Advisory Panel

Lester B. Lave, *Panel Chair*
James H. Higgins Professor of Economics
Graduate School of Industrial Administration
Carnegie-Mellon University

George N. Austin, Jr.
National Officer
Fraternal Order of Police

Lane Bishop
Statistician
Center for Applied Mathematics
Allied-Signal, Inc.

Alfred Blumstein
Dean and J. Erik Jonsson Professor of Urban
Systems and Operations Research
School of Urban and Public Affairs
Carnegie-Mellon University

Michael Bowman
Vice President and General Manager
Fibers Department
E.I. duPont de Nemours Co., Inc.

Milton Brand
President
The Brand Consulting Group

James T. Curran
Professor and Dean for Special Programs
John Jay College of Criminal Justice
City University of New York

Donald R. Dunn
President
H.P. White Laboratory, Inc.

Martin Fackler
President
International Wound Ballistics Association

Michael A. Goldfarb
General Surgeon
Monmouth Medical Center

David C. Hill
President
Fibers Division
Engineered Materials Sector
Allied-Signal, Inc.

Max Henrion
Member of the Technical Staff
Rockwell International Science Center

Alexander Jason
Ballistics Consultant
ANITE Group

Harlin R. McEwen
Chief
Ithaca Police Department

Isaac Papier
Managing Engineer
Burglary Detection and Signaling Dept.
Underwriters Laboratories, Inc.

Richard Stone
President
Point Blank Body Armor, Inc.

Dieter Wachter
Vice President of High-Performance Fabric
Clark-Schwebel Fiberglass Corp.

Robert Wantz
President
Personal Protective Armor Association

NOTE: OTA appreciates and is grateful for the valuable assistance and thoughtful critiques provided by the advisory panel members. The panel does not, however, necessarily approve, disapprove, or endorse this background paper. OTA assumes full responsibility for the background paper and the accuracy of its contents.

OTA Project Staff-Police Body Armor Standards and Testing

Lionel S. Johns, *Assistant Director, OTA*
Energy, Materials, and International Security Division

Akin Shaw, *International Security and Commerce Program Manager*

Michael B. Callaham, *Project Director*

Brian McCue, *Senior Analyst*

Jonathan Tucker, *Analyst* (through May 1991)

Administrative Staff

Jacqueline Robinson-Boykin
Office *Administrator*

Louise Staley
Administrative Secretary

Acknowledgements

OTA gratefully acknowledges the assistance of the following individuals and organizations for their help in supplying information or in reviewing drafts of portions of this report. The individuals and organizations listed do not necessarily approve, disapprove, or endorse this report; OTA assumes full responsibility for the report and the accuracy of its contents.

Allied-Signal, Inc.
Kevin McCarter
Sam White
Steven A. Young

Aspen Systems, Inc.
Marc H. Caplan
Wendy Howe
Candace McIlhemy

Canadian General Standards Board
Marian L. Gaucher

E.I. duPont de Nemours and Co., Inc.
Thomas E. Bachner, Jr.
William Brierly
Louis H. Miner
Helen A. Slavin

Elgin (IL) Police Department
Chief Charles A. Gruber

General Motors Research Laboratories
David C. Viano

Hartford (VT) Police Department
Chief Joseph G. Estey

Home Office Police Scientific Development Branch
Eric Brown

Jaba Associates (Ontario)
Alan Athey

Point Blank Body Armor, Inc.
Gaetan (Tom) J. Dragone

Second Chance Body Armor, Inc.
Clinton Davis
Pamela Hinz

Lester Shubin

U.S. Department of Commerce
National Institute of Standards and Technology
Keith Eberhardt
Lawrence K. Eliason
Daniel E. Frank
John Whidden
Patent and Trademark Office
Deborah L. Kyle

U.S. Department of Defense
Strategic Defense Initiative Organization
Nicholas Montanarelli
Department of the Army
Ballistics Research Laboratory
Russell Prather
Chemical Research, Development, and Engineering Center
Larry Sturdivan

U.S. Department of Justice
Bureau of Alcohol, Tobacco, and Firearms
Daniel Hartnett
Federal Bureau of Investigation
Bunny Morris
David Pisenti
Charles Barry Smith
National Institute of Justice
Paul Cascarano
Charles DeWitt
Paul Estaver

University of Maryland
Prof. Girish Grover
Ann Beth Jenkins
Ian Twilley

Frederick Peter Watkins

Contents

Appendix A

The Origin of and Rationale for the NIJ Standard

Contents

Contents—Continued

Boxes

Figures

Tables

Appendix A

The Origin of and Rationale for the NIJ Standard

INTRODUCTION TO NIJ BODY ARMOR STANDARDS

General

Four standards for body armor, numbered 0101.00 through 0101.03, have been successively promulgated by the U.S. Department of Justice's National Institute of Justice (ND) and its predecessor, the National Institute of Law Enforcement and Criminal Justice (NILECJ). Compliance with these standards has been voluntary--companies perceiving that benefit in the marketplace would accrue from their products' compliance with a Federal standard can submit their vests for certification according to the standard. Recognizing that different customers will feel different needs for protection, the Justice Department created standards that specify more than one level of protection: 0101.00 set standards for three types of armor, expanded to six in later standards.

The Justice Department recognized at the outset that there is no such thing as 100-percent safety. In particular, it stated that the blunt trauma (bruising of internal organs) caused by the impact from a nonpenetrating bullet on armor was to be survivable in 90 percent of cases. As will be shown below, implementors of the standard used conservative judgment at a number of stages, leading to a situation in which (as of this writing) nobody[2] wearing NIJ-certified armor has been killed by blunt trauma.

The question of technology-specific considerations-those based on current vest construction, not desired vest performance-arises repeatedly in the formulation of standards for police body armor. To date, the standards have specified performance, not construction: manufacturers are free to make a vest any way they want as long as it passes the test. However, some technology-specific considerations have crept into the standards here and there. The most obvious of these, introduced in the 0101.01 standard, is the requirement that the vest be tested wet as well as dry. This test was instituted in response to the finding that a certain vest material could be penetrated more readily when saturated with water than when dry. Granting that police officers' vests become wet and that wetness could make a difference to the ballistic performance of the vest,[3] testing under wet conditions clearly makes sense. Yet why not test the vests when they are cold, or hot, or covered with powdered sugar? The answer that vests do not, in normal use, become sufficiently cold, hot, or covered with powdered sugar to degrade their performance is at once a technology-specific consideration (somebody might someday come forward with a vest that proved highly sensitive to these conditions) and an invitation to argue about the conditions arising in normal use, including the level of wetness to which one can reasonably expect a vest to be subjected. We shall revisit the wetness issue in describing the 0101.01 standard-the purpose of raising it here is merely to show how technology-specific considerations can infiltrate a supposedly performance-oriented standard.

Overview of the Current Standard and the Controversy Surrounding It

The National Institute of Justice 0101.03 Standard for concealable body armor provides for the testing of four types of soft body armor and two types of rigid armor,[4] collectively offering protection from the full spectrum of small-arms threats. Compliance with the standard is voluntary: some companies choose to comply and some do not, presumably reflecting different assessments of the benefits of NIJ certification as compared to the costs of producing compliant vests. In a gray area, some companies assert that their vests comply with the standard, but have not submitted them for official

---

1 Or, perhaps, in the courtroom.

2 In one problematic case, a very heavy bullet fired from a rifle killed an officer without penetrating his vest. Some therefore call this a death by blunt trauma, while others point to the fact that the vest and the bullet both penetrated the officer, making the death more closely resemble a regular wound and not blunt trauma.

3 Though it need not—the material that performs poorly when wet can be waterproofed or encased in a waterproof cover and thereby retain its ballistic efficacy.

4 As well as for a generic test of "special type" armor, whose desired level of ballistic performance is left up to the user.

Table A-l-Summary of 0101.03 Armor Types According to the Ammunition Against Which They Are Tested (velocities compared to those of Federal brand)

Type	Cailber	Mass (grains)	Test velocity (ft/sec)	Federal velocity (ft/seo)
I	.22 LRHV	40	1,050 to 1,100	1,255
	.38 RNL	158	850 to 900	755
ii-A	.357 JSP	158	1,250 to 1,300	1,235
	9mm FMJ	124	1,090 to 1,140	1,120
ii	.357 JSP	158	1,395 to 1,445	1,235
	9mm FMJ	124	1,175 to 1,225	1,120
hi-A	44 Magnum	240	1,400 to 1,450	1,180
	9mm FMJ	124	1,400 to 1,450	1,120
III	7.62 mm FMJ	150	2,750 to 2,800	2,910
Iv	.30-06 AP	166	2,850 to 2,900	2,800

KEY: AP = armor piercing; FMJ = full metal jacket; JSP = jacketed soft-point; LRHV = long rifle high velocity; RNL = round-nose lead.

SOURCES: National Institute of Justice, NIJ Standard 0101.03, April 1987 [144], and William S. Jarrett, 1991 [85].

testing, while others advertise that their vests have been tested without stating the outcome of the test.

In general,[5] the armor must demonstrate an ability to stop, without the transmission of unduly concentrated blunt impact, two types of ammunition. (See table A-l.) It must do so when wet as well as when dry. The armor is shot while attached to a clay backing-the resulting dents in this backing provide a means of assessing the amount of impact that the vest would transmit to its wearer.

The velocities to be used in the test are representative of those found in commercial ammunition, with some exceptions. (See table A-l.) The most salient exceptions are the velocities specified for testing type III-A armor, which is not intended for daily wear and was created in response to the threat posed by terrorists, not common criminals. [145][6] Another exception is the velocity specified for the .357-caliber jacketed soft-point bullets used in type-II tests.

Four vests are consumed by the test[7]-one for each of the four combinations resulting from the two ammunition types and the two wet-dry conditions. Each vest has two panels, the front and the back. Each panel is shot 6 times, so that the vest model must endure 48 shots to pass. For soft body armor, the first shot on each panel is used in assessing the transmission of blunt impact.[8] For armor intended to protect the wearer against handgun bullets, two shots on each panel strike at an angle of 30 degrees away from head-on: the rest (including that used in the assessment of blunt impact) are head-on.

As of Oct. 31, 1991, 329 of the 555 models submitted for NIJ certification testing under the 0101.03 standard had passed, 221 had failed, and 5 tests were inconclusive. Penetration caused 166 failures, excessive backface signature (an index of blunt-trauma risk) caused 15, and 40 models failed because of both penetration and excessive backface signature.

Critics of the standard charge that its stringency and the variability of results force manufacturers to build unduly rugged armor, creating extra expense and discomfort for the consumer, and ultimately resulting in the perverse effect of officers dying because armor that meets the standard is so uncomfortable or expensive that it is not used. Critics point to the perfect record of armor in the field (no officer has died from a shot that his or her armor was supposed to be able to stop), much of it set by armor that has not passed-and, in many cases, could not pass-the NIJ test. In addition, they cite cases in which officers have been saved from shots that their armor was *not* rated to stop, and even cases in which subsequent "reenactment" of the shot under the laboratory conditions mandated by the NIJ standard

5 Some of what follows does not apply to the strongest of the rigid armors.

6 Numbers in brackets cite references in the bibliography in volume 1 of this report.

7 In practice, six vests are forwarded for testing, to allow for the possibility that a vest or two would be spoiled during the test process or otherwise tested inconclusively.

8 In the case of Type III armor, a verbal communication to the test facility mandates the use of the first two fair shots on each panel.

resulted in either a penetration of the vest or a backface deformation greater than that allowed by the NIJ.

Specifically, critics cite as unduly stringent the requirement that the vest retain its bullet-stopping ability even when wet. Although they have nothing against vests that perform well when wet and admit that some officers may need or desire such vests, they question a standard that makes wet-testing, and thus wet-strength, mandatory. While a variety of means to assure unimpaired performance when wet are available, all add at least a little cost, weight, and stiffness to the vest. Critics also decry the requirement that each panel endure six shots. Not only do they see six shots as an unrealistically high number in itself, but in addition they point out that the tendency of the vest to squirm about while under fire on the test fixture leads to delamination of the ballistic material and raises the probability of penetration on the later shots. They further maintain that this "bunching and balling' of the vest does not occur when the vest is on a human torso, so that the test does not give a true assessment of vest performance in the rare case of multiple impacts. Finally, some critics claim that the maximum allowable depth of the dentin the clay (44 mm) is too little, and has no basis in physical, clinical, or experiential reality.

Upon introduction of the 0101.03 standard, many vests that had passed the 0101.02 test failed a retest under the new standard. Critics asserted that the mass failure of vests previously deemed acceptable indicated that there was something wrong with the new standard or, considering the textual similarity between the two standards, with the implementation of the new standard by the test laboratory. Others have asserted that certain practices, such as poor recordkeeping and the mixing and matching of passed panels, created undue leniency in the 0101.02 era.

Defenders of the 0101.03 standard point out that a standard for a safety-related product *should* be somewhat conservative, it being far better to fail some adequate vests than to pass even a few inadequate ones. They defend the requirement that the vest should function while wet on the grounds that, while total immersion of an officer is a rare occurrence, perspiration is not, and could readily soak a vest. They point out that officers fortunate enough to have survived shootings their vests were not rated to stop may have survived more because of the obliquity of the shot than because of superior body armor. They defend the requirement that the vest withstand six shots per panel on the grounds that the weapons available today can fire many more shots than that. They see the claim that bunching and balling does not occur on the human torso as unsubstantiated at best, and perhaps even contradicted by videos featuring the president of a body armor company shooting himself in the vest. [121] Finally, they cite animal tests performed at the beginning of the body armor program as the basis for the 44 mm backface signature criterion.[9]

NILECJ STANDARD 0101.00

The NILECJ,[10] a part of the Law Enforcement Assistance Administration at the U.S. Department of Justice, promulgated NILECJ-STD 0101.00, *Ballistic Resistance of Police Body Armor, in* March of 1972.[11] The standard was formulated in conjunction with the Law Enforcement Standards Laboratory (LESL)[12] of the National Bureau of Standards.[13]

Sampling

Each "lot" of armor submitted for certification was to be sampled at random. The standard specified the number of vests constituting an adequate sample, with larger lots requiring larger samples. Alternatively, manufacturers could assure lot-to-lot quality through application of quality control procedures. Though the standard does not explicitly state as much, the reader is left to infer that certification of an initial lot and lot-to-lot consistency as documented by ''quality control charts' would permit the manufacturer to present later lots as 'certified." In practice, the term "lot" is more ambiguous than one might suppose, because body armor manufacturers buy the components of body armor from different vendors at different times. A set of, vests all made at once from the same shipment of ballistic material may contain waterproof coverings made from differ-

9 Like other parts of this controversy, the relationship between the animal tests and the 44 mm criterion is explored more deeply in a later section.

10 Now the National Institute of Justice (NIJ).

11 Facts in this section come from the standard itself [141] if no Other source is cited.

12 Now the Office of Law Enforcement Standards (OLES).

13 Now the National Institute of Standards and Technology (NIST).

ent shipments of waterproof materials, and the ballistic material itself may have been made from fibers spun at different times, or scoured with chemicals produced at different times.

Marking and Workmanship

The 0101.00 standard required that armor be "free of wrinkles, blisters, cracks, crazing, fabric tears, chipped or sharp corners, and other evidence of inferior workmanship, " and further specified that "Each armor part shall be clearly and durably marked with the manufacturer's name, brand name or logo, the model number, and the lot number. "

Penetration

The standard specified that each armor part was to withstand 5 "fair hits" by test bullets with no penetrations, except (1) "armor fronts" were to withstand 10 "fair hits" with no penetrations, and (2) armor parts-front or back—being tested for Type .30 AP (armor-piercing) ballistic resistance were required to withstand only 1 "fair hit" by a .30-06 AP test bullet with no penetration. A "fair hit" was a hit by a bullet with velocity of at least that required for the type, striking the armor at no more than 5 degrees away from norma1[14] incidence and no closer than 2 inches to the edge of the armor or to a prior hit.

Different set-ups were prescribed for the penetration test and the deformation test. The test set-up for penetration did not use the now-familiar clay backing, nor indeed any backing at all. Penetration was to be assessed with a' 'witness plate," mounted six inches behind the armor. A witness plate is a thin piece of sheet metal inspected for holes after the test by holding it up to a light. Passage of light through the witness plate signified a penetration of the vest and caused the vest to fail. In fact, "penetration by any fair hit, no matter what its velocity, shall cause rejection of the lot. "

Deformation

The set-up specified for the deformation test included a backing made of "nonhardening modeling clay.' A method for determining the depth of the deformation in the backing (the creation of a plaster cast) was given, but the maximimum acceptable depth of the dent in the clay behind the armor was explicitly cited as "not yet established." No mention was made of the possibility of a penetration occurring during a deformation test.

Types of Armor

The standard recognized three types of armor, known by the guns and ammunition against which they were to afford protection. (See table A-2.) These were Type .22 LR (long rifle)-.38 Special, Type .357 Magnum, and Type .30 AP. Type .22 LR-.38 Special was to be tested with the .22 caliber ammunition and, if it passed, then tested with the .38 Special ammunition. The Type .30 AP armor needed only to stop one bullet, not five.

Type .22 LR-.38 Special was to afford protection against the .22 caliber Long Rifle rounds freed from handguns and .38 Special Metal Point rounds against which it was to be tested as well as other .22, .25,.32, and .45 caliber rounds and 12-gauge#4 lead shot—protection against these latter rounds was taken for granted if the armor passed the test with .22 LR and .38 Special Metal Point.

Type .357 Magnum was to protect against the .357 Magnum rounds against which it would be tested as well as 9-mm Luger, 12-gauge #00 Buckshot, and all of the Type .22 LR-.38 Special threats-protection against these latter rounds was taken for granted if the armor passed the test with .357 Magnum ammunition.

Type .30 AP was to protect against the .30 caliber armor piercing rifle round against which it was to be tested as well as .41 and .44 Magnum handgun rounds, .30 caliber carbine rounds, 12-gauge rifled slugs, and all of the threats specified for the two other types of armor-protection against these latter rounds was taken for granted if the armor passed the test with .30 AP rifle ammunition. It was expected that Type .30 AP armor would stop the .30 caliber AP round with a ceramic material that might well be broken in the process-a nonceramic rear element was 'normally' to be made of Type .357 armor. The test of the Type .30 AP armor did not, however, include a test of the rear element.

The velocities lie towards the upper end of the range attainable by the firing of commercially available ammunition from commercially available

14 I.e, head-on,

Table A-2-Summary of 0101.00 Armor Types According to the Ammunition Against Which They Were Tested

Type	Caliber	Mass (grains)	Minimum velocity (ft/sec)
.22 LR-.38 Special . . .	.22	40	1,181
	.38	158	782
.357 Magnum	.357	158	1,261
.30 AP	.30	166	2,694

SOURCE: National Institute of Law Enforcement and Criminal Justice, NILECJ Standard 0101.00, March 1972.

guns.[15] The International Association of chiefs of Police (IACP) has published the research underlying these velocity selections.

Comments on Technology Specificity in the 0101.00 Standard

An important instance of technology specificity is the requirement that an armor part need only stop one .30-06 armor-piercing bullet in order to demonstrate Type .30 AP ballistic resistance, but it must stop 5 or 10 .357 Magnum bullets in order to demonstrate .357 Magnum Type ballistic resistance. The explicit reason for this is that Type .30 AP vests were expected to be ceramic, and thus only capable of reliably stopping a single bullet-ceramic vests absorb impact energy by shattering.

NILECJ STANDARD 0101.01

NILECJ-STD-O1O1.O1 was promulgated in December, 1978.[16] The first full-fledged U.S. standard for police body armor, it was formulated with the active participation of the Personal Protective Armor Association (PPAA). [150] After the release of 0101.00, NIJ had established the Technology Assessment Program Advisory Council (TAPAC), to advise NIJ about the direction of its Technology Assessment Program (TAP). TAPAC recommended that NIJ establish a testing program for law enforcement equipment, including body armor. The resulting test program was administered by the IACP. [150]

Reasons for Replacing the 0101.00 Standard

As indicated by its number, the 0101.00 standard was created in order to be replaced. Its writers anticipated the eventual articulation of an acceptable degree of backface deformation-they specified the test procedure, but left the allowable depth “not yet established. ” [141]

The 0101.01 standard set forth five levels of armor in place of the three specified by the 0101.00 standard. One new level was a second level for rigid armor, offering protection against a sporting, as opposed to military, rifle threat; the other was an intermediate level of protection against handguns.

The 0101.01 standard also introduced the testing of vests while wet, a reaction to the discovery that wetness could severely reduce the ballistic performance of the vest material then in most common use.[17]

Sampling

The 0101.01 standard specifies that “two complete armors, selected at random, shall constitute a test sample.” Two extra armors might be needed if the tester wanted to exercise the option not to test both types of ammunition on the same panels. The 0101.00 standard’s suggested sample sizes based on lot sizes and the use of a table of random numbers to attain random selection were dropped. Moreover, no reference to the “lot” concept appears; unlike 0101.00,0101.01 does not specify that ‘penetration by any fair hit, no matter what its velocity, shall cause rejection of the lot. In fact, the standard itself does not spell out the exact consequences of failure.

Wet Testing

A separate set of armor was to be tested while wet, the wetness having been attained by a gentle spray of specified rate and duration. The most obvious consequence of this wet-testing was to oblige manufacturers to make their products impervious to water.

15 “. . . the approach taken was to use actual handguns and factory ammunition to conduct the ballistic tests. . . . the measured impact velocities for each type of test round were averaged, the standard deviation calculated, and testing velocities selected to be in the upper boundary of the standard deviation. . . to provide a margin of safety should an assailant utilize ammunition providing bullet velocities at the high end of the nominal range for these bullets.” [150]

16 Facts in this section come from the standard itself [142] if no other source is cited.

17 This reduction only lasts as long as the wetness does: once dry, a vest returns to its original level of ballistic performance. It is thought that the wetness lubricates the fibers, allowing them to slip against one another more easily and eliminating the net-like action by which the vest stops the bullet.

Marking and Workmanship

The 0101.01 standard again required that armor be free of specified evidence of inferior workmanship. The labeling requirements were enhanced to include size, type (according to the standard itself), month and year of manufacture, cleaning instructions, and strike face. (The strike face is the side of the armor panel intended to be hit by the bullets.)

Penetration

The 0101.01 standard eliminated the witness plate and required use of clay backing for penetration testing, relying on examination of the backing material and the armor itself to determine whether a penetration has occurred. The introduction of upper limits on velocity necessitated an additional clause in the definition of a fair hit-a hit was unfair if the bullet was going too fast, except in the case of a bullet that was going too fast and even so did not penetrate. Such a hit was a fair hit.

If the vest construction included any seams, a fair hit had to be administered to a seam. Because the standard did not specify that one of the first two fair hits (those used in measuring deformation) must fall on a seam, deformation of the backing material by a hit on a seam was not required to be measured.

Deformation

An expected innovation in the 0101.01 standard was the specification of a maximum allowable backface deformation. No backing material was specified, although the report stated that Roma Plastilina No. 1 modeling clay was "found to be suitable.

Conditioning of the material was specified, as was a test for consistency: measuring the depths of craters formed by dropping weights onto the clay. The clay was to be maintained at a temperature between 15 and30'C (59 and 86 'F). Deformability of Roma Plastilina No. 1 and similar modeling clays depends strongly on temperature.[18 19]

The standard specified that the dents resulting from the first two fair shots with each type of ammunition were to be no more than 44 mm deep. Hits were to be placed as far apart as possible, and the standard instructs the laboratory to "reposition the backing material (as required) to avoid any overlap of depressions. '

To be a fair hit for the purpose of measuring deformation, a bullet had to be within the allowable velocity bounds-for measuring deformation, no clause (analogous to the clause counting overspeed bullets as fair tests if they did not penetrate) allowed overspeed bullets to be considered fair if they did not create a disqualifying deformation.

Origin and Rationale of the 44-mm BFS Limit

Considerable confusion and controversy surround the genesis of the 44 mm backface signature (BFS) limit, in part because the rationale for it was never documented. There *is* a rationale for the limit, at least for Type I Kevlar armor. However, the experiments recognized as necessary to assess the validity of the criterion for higher energy bullets were never completed, for fiscal reasons.

OTA has reconstructed the following account based on Army reports on research performed for the NILECJ and interviews of individuals responsible for setting the limit or conducting the research on which the limit was to be based.

It appears that there were three thrusts to the body armor research performed by the Army. The earliest research [104] and some of the later biomedical research [74, 75, 101, 127] was aimed at predicting the injurious effects of particular types of bullets striking particular types of armor at specified velocities over particular parts of the torso. For this, goats wearing various types of armor were shot, sacrificed, and autopsied.

This work originated when the NILECJ's body armor program aspired only to develop armor against "common handguns''-in practice, against .22 LR and .38 Special rounds. Although assaults by other low-energy handgun rounds-e. g., .25- and .32-caliber-were common, the .22 LR was considered the most likely of then-common handgun rounds to penetrate armor, and the .38 Special was considered most likely to cause blunt trauma if stopped. Thus the early experiments mostly used .38 Special bullets impacting 7-ply Kevlar panels at about 800 ft/s.

18 See [8] for the dependence of Plastilina and [28] for that Of Plasticine.

19A difference in temperature ~@ explain the difference in backface deformations produced by two seemingly identical shots shown in the video *Second Chance v. Magnum Force* [121] to demonstrate to the viewer how deformation tests can bemanipulated.

Another thrust [35,[20] 114, 130] was the development of species-independent, parametric models of blunt-trauma lethality-for example, predicting lethality of shots on armor over the lung, in terms of properties of the projectile (mass, diameter, velocity), armor (mass per unit area), and victim (weight, body wall thickness). Such a model would allow data collected in previous experiments+. g., shootings of animals with tear-gas grenades-to be compared with the shootings of armored goats by bullets. This requires treating the bullet plus the portion of armor it pushes into the torso (without penetrating the skin) as a single, blunt projectile, moving slower than the bullet at impact. This blunt projectile would have the same momentum as the bullet; its effective diameter was considered to be the diameter of the depression made by the armor in the torso or, approximately, in gelatin or clay backing material. An advantage of this approach is that a parametric blunt-trauma lethality model could be used to predict the lethality of new projectile-armor combinations without shooting more animals; it would only require shooting the projectile of interest at the armor of interest on *a flesh-simulating* backing material. (See box A-l.)

A third thrust was to record the diameter and depth of the depression made by various armor struck by various bullets in gelatin [100] and clay [114] backing material. The gelatin data were to be correlated with the results of shooting the armored goats. The clay data were to be used in conjunction with the parametric blunt-trauma lethality models described above. But the Prather report [114] also compared the maximum momentary depth of indentation of gelatin by a blunt projectile with the maximum depth of indentation of clay, based on one shot per backing. This tenuous comparison allowed BFS in clay to be correlated with maximum deformation depth in gelatin, which had been correlated with ballistic parameters, which in turn had been related to nonlethality in goats and extrapolated to norilethtity in humans. This series of correlations provided the basis for the 44-mm BFS limit in NILECJ-Std. 0101.01. For this use the backing need not simulate the density or resiliency of tissue.

The Army's soft body armor medical assessment team, led by Dr. Michael Goldfarb, recommended a BFS limit of 44mm for 158-grain, .38-ca.liber bullets

Table A-3-Backface Signature Parameters .38-Caliber, 158-Grain Projectile Versus 7-Ply Kevlar-29, 400/2 Denier

Film no.	Striking velocity (m/s)	Maximum depth (cm)	Maximum base radius (cm)
30008	243.7	4.82	4.76
30177	253.9	4.99	4.12
30178	255.4	5.17	5.18
30179	249.6	5.00	4.61
30180	247.8	4.72	4.01
30181	249.3	4.88	4.99
30182	251.5	4.60	3.79
30183	249.0	4.64	4.60
30184	259.1	5.08	4.79
30185	254.8	5.20	4.62
30186	255.4	4.80	4.97
30187	254.5	3.98	4.50
30318	249.8	4.65	4.91
30319	246.8	4.71	3.99
30320	247.3	4.84	3.77
30321	245.9	4.14	3.84
30322	248.1	4.42	4.45
Mean	250.7	4.74	4.46
Standard deviation . . .	4.17	0.33	0.46

SOURCE: LeRoy W. Metker et al., 1975 [100], table 3.

striking 7-ply, 400/2-denier Kevlar-29 armor at about 800 ii/s. Their recommendation was based in part on the gelatin deformation data reprinted in table A-3. The third column shows the maximum depth of deformation of ballistic gelatin behind 7-ply, 400/2-denier Kevlar-29 armor struck by a 158-grain, .38-caliber bullet in each of 17 shots intended to simulate the shots at the 14 armored goats examined by Goldfarb et al. [74] The maximum depths of deformation averaged 4.74 cm, with a sample standard deviation of 0.33 cm. The goats examined by Goldfarb et al. all lived until they were sacrificed 24 hours after being shot, and none sustained serious injuries. According to Goldfarb, he and his medical assessment team reasoned that goats shot under the less stressful of the experimental conditions-which correlate with gelatin deformations 1 standard deviation less than the mean, or about 4.4 cm—would be very unlikely to sustain serious or lethal trauma. Their report concludes that humans would be even less likely to sustain serious or lethal trauma under similar conditions.

To complete the correlation of trauma with deformation in clay, the researchers compared defor-

[20] See also Victor R. Clare, James H. Lewis, Alexander. Mickiewicz, and Larry M. Sturdivan, *Body Armor—Blunt Trauma Data* (Washington, DC: U.S. Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice, May 1976).

297-923 0 - 92 - 2 : OL 3

Box A-l—Parametric Models for Estimating Probability of Blunt-Trauma Lethality

Under NILECJ sponsorship, the Army developed several mathematical formulas, or "parametric lethality models,' for estimating the probability of blunt-trauma lethality on the basis of numbers (''parameters' describing properties of an impacting bullet (mass and velocity), the armor (areal density, i.e., mass per unit area), and the wearer (body mass and, for some models, body-wall thickness). Most were developed just after the 44-mm BFS limit was recommended, but before issuance of NILECJ Std. -0101.01, the first standard to specify the limit. Some of the models were considered to provide a rough confirmation of the adequacy of the 44-mm limit, the medical rationale for which was limited to .38-Special bullets stopped by 7-ply Kevlar 29 armor, and especially for extending that limit to other threats and armors. In fact, the models suggest that it would be appropriate for the BFS limit to depend on the threat, the armor, and measurements of the wearer. The NILECJ opted for a simpler, conservative, uniform limit.

To use these models would require measuring the diameter of the crater made in the backing, instead of (or in addition to) its depth. It would also be necessary to measure the areal density of the armor at the point of impact, or to infer it from the other parameters.

The most highly developed predictive models developed for the NILECJ are two developed by Larry Sturdivan: one for estimating the probability of lethal blunt trauma resulting from impacts on the abdomen over the liver, the other-discussed here-for estimating the probability of lethality from impacts on the thorax over the heart or a lung. Both models predict probability of lethality based on the mass M, diameter D, and velocity V of the impacting, nonpenetrating projectile, and the body mass W and body-wall thickness T of the victim.[a] They are based on data obtained by shooting anesthetized goats and calves with blunt plastic cylinders or similar nonpenetrating projectiles used to simulate impacts of bullets stopped by armor. [130]

The model for lethality of thoracic blunt trauma is

$$P(L) = 1/(1+ \exp(34.13 -3.597 \ln(MV^2/W^{1/3}TD)))$$

or, equivalently,

$$P(L) = 1 /(1 + 6.645X10^{14}/ (MV^2MV//DW^{1/3}T)^{3.597})$$

where

- P(L) denotes the probability of lethality,
- exp() the exponential function,
- In() the natural (base-e) logarithm,
- M the projectile mass in grams,
- V the projectile velocity in meters per second,
- W the mass of the victim in kilograms,
- T the thickness of victim's body wall (skin, fascia, fat, muscle, bone) at impact point, in centimeters, and
- D the projectile diameter in centimeters.

To use the model, one must estimate the mass, diameter, and velocity (M, D, and V) of the blunt "projectile" formed by the bullet plus the portion of the armor that it pushes into the body. M, D, and V may be estimated by the method proposed by Prather et al., which requires knowing the areal density a_d of the armor at the point of impact: The diameter D of the blunt projectile formed by the bullet plus a portion of the armor is considered to be the diameter of the backface signature made in clay backing by the bullet-armor combination; its mass M is accordingly the bullet mass M_P plus the mass of armor over the crater:

$$M = M_P + 3.14\ (D/2)^2 a_d$$

with a_d in g/cm^2.

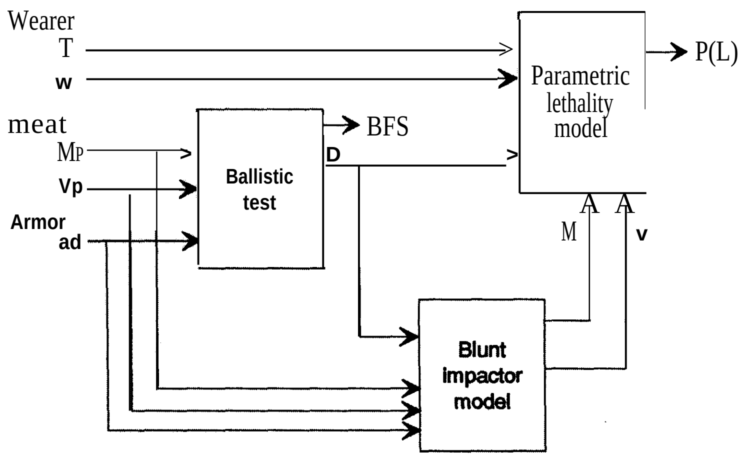
The velocity V of the blunt projectile is estimated from the velocity V_P of the bullet by noting that conservation of momentum, a basic physical law, requires the momentum MV of the blunt projectile to equal the momentum $M_P V_p$ of the bullet. Hence

$$V = (M_P/M)\ V_P$$

The figure illustrates the procedure for estimating the probability of lethality from the backface signature using the parametric lethality model. It is assumed that the model applies to humans as well as to the larger animals (calves) and smaller animals (goats) shot in the experiments that generated the data to which the model was fitted. However, these animals were shot by heavy, slow, blunt projectiles aimed at especially vulnerable locations. In extrapolating predictions to assault situations, allowance should be made for less deadly targeting.

[a] The body wall includes the skin, fat, muscle, bone (if any) and fascia covering the organs inside the abdominal or thoracic cavity. An individual's body-wall thickness varies with location.

Estimating the Probability of Blunt-Trauma Lethality Using a Parametric Lethality Model



ad= areal density of armor (mass per unit area)
BFS = depth of crater
D = diameter of crater
M = mass of projectile + portion of armor pushed into crater
Mp = mass of projectile (bullet)
P(L) = probability of lethality
T= thickness of body wall (skin, muscle, bone...) of wearer
V = velocity of projectile + portion of armor pushed into crater
Vp = velocity of projectile
W = "weight" (i.e., body mass) of wearer

SOURCE: Office of Technology Assessment, 1991.

mation depth of goat thorax with that in clay, gelatin, and other backing media.[21]

The medical team also considered the fatal "massive, contralateral right lung damage" produced in the one armored goat shot with a .45-caliber bullet [101], reenactments of which produced deformations of 5.2 cm in clay and 5.3 cm in 20-percent gelatin [114].

In another, unpublished, experiment, a goat (no. 21644) wearing a 5-ply Kevlar panel was struck by a ***.38*** caliber bullet. Although the vest stopped the bullet and produced only a superficial skin contusion, autopsy revealed that blunt trauma had produced a massive lung hemorrhage involving roughly 150 cubic centimeters of tissue. (See figure A-l.) When the average deformation depth of .38 caliber bullets against 5-ply Kevlar was later measured in 20-percent ("ballistic") gelatin, it was only 48.2 mm, with a standard deviation of 3.9 mm. [100] From this, Dr. Goldfarb concludes that the margin of safety provided by the NIJ backface deformation standard may amount to only about half a centimeter. He questions "whether it is really worth throwing out a proven standard because of difference of a few millimeters. "[22]

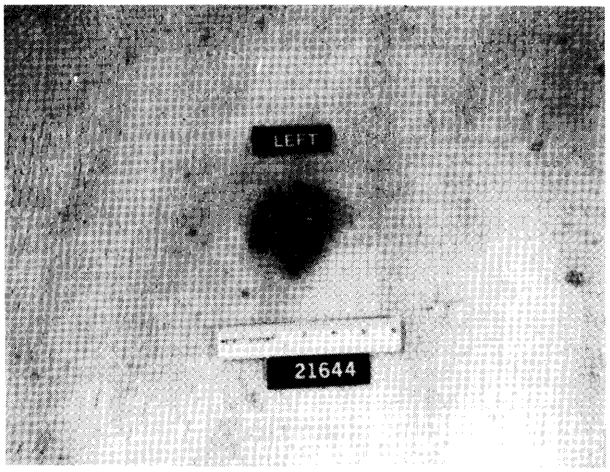
In addition, Goldfarb said that he and other medical team members were concerned that impacts that would not kill a man of large or medium build might kill a woman of medium or small build. Indeed, the parametric models suggest that a lighter person with a thinner body wall (skin, fat, muscle, bone, fascia) would not survive some impacts that a larger person would. The medical team was not asked to recommend a weight- or sex-dependent limit, so they wanted an extra margin of safety for adequate protection of small, typically female, officers.

Critics have recently noted [86, 87] what appears to be a discrepancy between the deformations listed in table 3 of [100] and the minimum, nominal, and maximum deformations shown in figure 5 of that report (reproduced in figure A-2). The discrepancy is only apparent: as we understand it, table 3 lists the maximum depth reached by any point of the indentation at any time, measured from the film. In particular, it lists four maximum depths equaling or exceeding 5.0 cm. The deformation envelopes shown in figure 5 bound the parabolic curves listed in table 1 of [100], which were obtained as fits to the (not necessarily parabolic) indentation profile read from the film frame exposed at the time of maximum indentation. The curve-fitting process generated approximating parabolas, some of which were not as

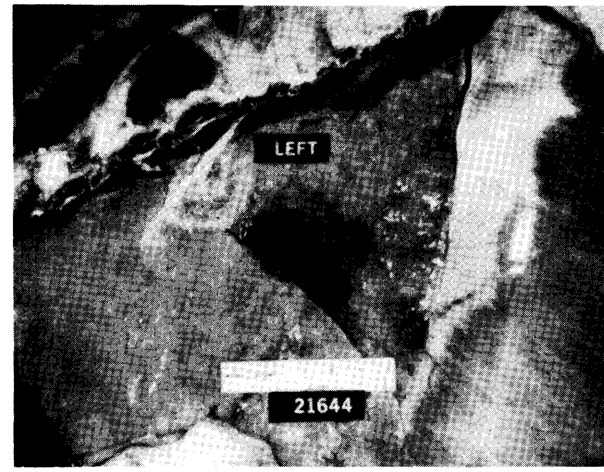
21 See table A.2 and figure B-2 of [114]; "BASELINE" refers to goat thorax.

22 Michael A. Goldfarb, M.D., personal communication, Apr. 25, 1991.

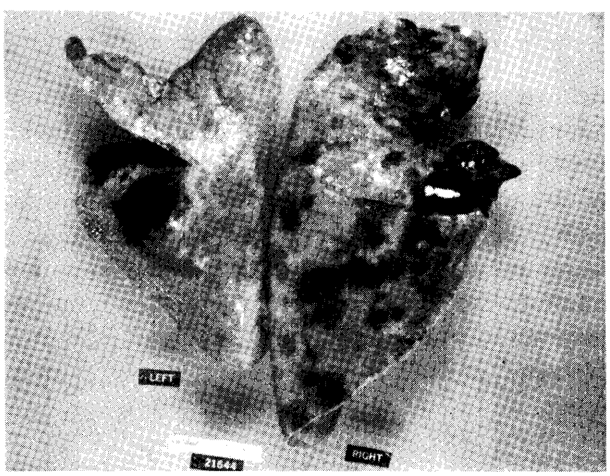
Figure A-l—Trauma to Goat Lung Caused by 158-Grain, .38-Caliber Bullet Stopped by 5-Ply Kevlar Armor



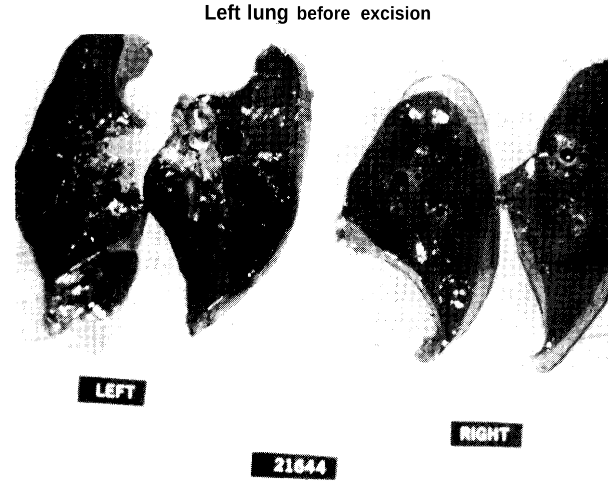
Superficial laceration



Left lung before excision



Left and right lungs after excision



Left and **right lungs after section**

SOURCE: Michael A. Goldfarb, M. D., 1991.

deep as the deepest part of the uneven surface they approximated.[23]

The NILECJ also funded similar Army experiments in which goats armored with Kevlar were shot with 9-mm and .357 Magnum bullets; however the studies were never completed (funding was stopped) and no report on them was published.[24] Mr. Lester Shubin, then the NILECJ's Director of Science and Technology, recently rationalized the specification of a 44-mm limit for all bullets and armor in NILECJ 0101.01 by noting that it was implausible that a greater BFS should be allowed for higher energy bullets, so if 44mm was appropriate for .38 Special, it was probably the maximum that should be allowed for higher energy threats. It might be that a smaller limit would be appropriate for higher energy threats, but there was no research to show what it should be.[25]

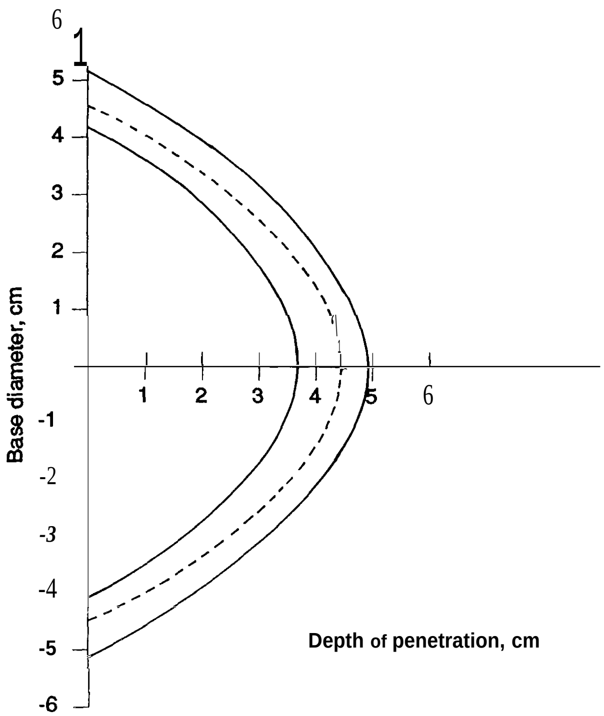
A different group of Army researchers working for the NILECJ provided additional support for a limit of about 44 mm in a 1977 report. [114] Figure B-10 of that report (reproduced herein figure A-3)

23 For example, table 3 lists the the maximum depth of the indentation shown in film no. 30178 as 5.17 cm, but table 1 shows the equation for the parabola fitted to the indentation shown in that film $y^2 = 26.94 - 5.6105\ x$, where y is the radius of the indentation and x is its depth. The maximum depth of this fitted parabola occurs along the centerline, where $Y = 0$, and is given by $0 = 26.94 - 5.6105\ x$, or $x = 26.94/5.6105 = 4.80$ cm.

24 Russel Prather, personal communication, Jan. 10, 1992.

25 Mr. Lester Shubin, personal communication, Nov. 13, 1991.

Figure A-2—The .38-Caliber Deformation Envelope in 20 Percent Ballistic Gelatin Backing 7-Ply, 400/2-Denier Kevlar 29 Armor Struck by 158-Grain, .38-Caliber Bullets



SOURCE: LeRoy W. Metker et al., 1975 [100].

plots a curve for probability of lethality ("PROB. LETH. ' as a function of 'LN DEFORMATION. ' The accompanying text (p. 10) indicates that "LN DEFORMATION" is the natural (base-e) logarithm of deformation in centimeters, so that, for example, a deformation of 5.0 cm (50 mm) corresponds to LN DEFORMATION = 1.61, for which the curve in figure B-10 predicts a probability of lethality of about 0.15, or 15 percent. A deformation of 4.4 cm (44 mm) corresponds to LN DEFORMATION = 1.48, for which the curve predicts a probability of lethality of about 0.06, or 6 percent.

The figure also plots circles with PROB. LETH. = O or 1, indicating survivals or fatalities, respectively, in experiments. The text indicates that the data are "the original blunt impactor data," for which [100] had been cited. However, the text does not specify which of the very numerous blunt impactor data in [100] were plotted. In separate interviews, Mr. Larry M. Sturdivan and Mr. Russell N. Prather told OTA that the data in figure B-10 are for shootings of unarmored goats by blunt impactors-rigid cylinders, some with a hemispherical nose-and that the deformations recorded are the maximum depths of indentation *of the animals' skin* momentarily produced by the projectiles.[26] They are not, as is sometimes assumed, [86, 87] deformations in clay produced by reenactments. The depths were measured, according to Sturdivan, from frames of high-speed films of the impacts; the projectiles were scored at intervals along their length to calibrate the readings. The report did compare deformation of goat skin (''Baseline' and clay by blunt impactors in its table A-2 and figure B-2. However, the comparison is for only one shot per backing; it gives no indication of variation to be expected under similar conditions or of the correlation to be expected at other impact velocities and momenta.

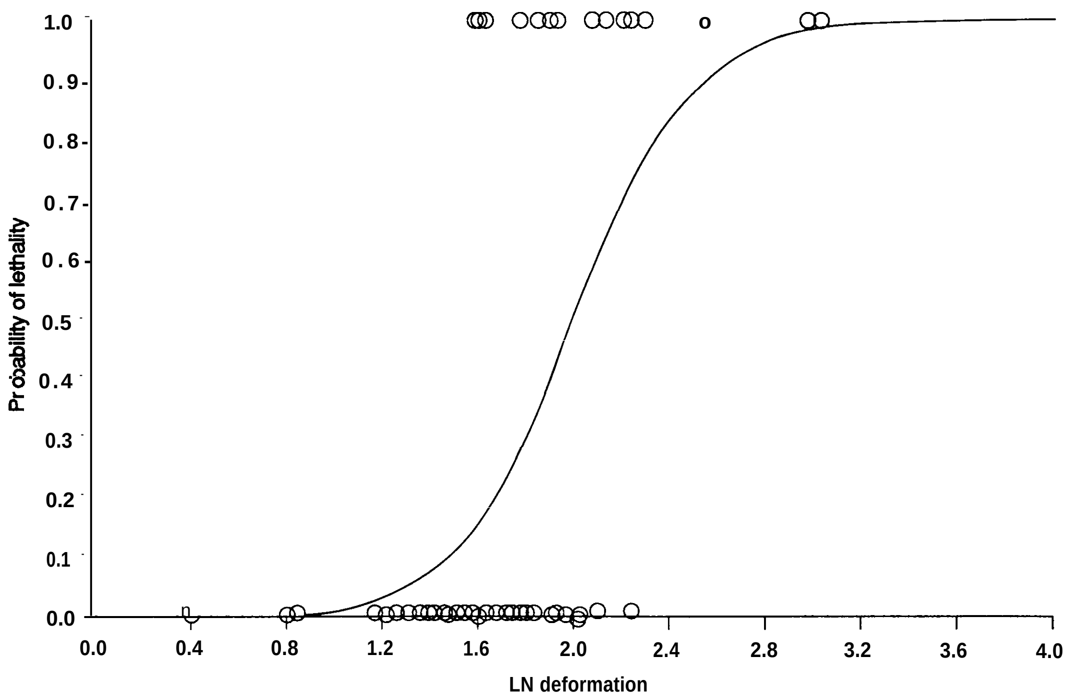
The blunt impactors, simulating the impact of bullet plus armor, were targeted at particularly vulnerable areas. There was no adjustment (as there was in the study by Goldfarb et al.) for goat-human differences or for the imperfect targeting in actual assaults. There was no adjustment for goat-human differences because the model was intended to be species-independent; similar but more complicated parametric lethality models developed by the Army sought to explain differences in lethality on the basis of biometric indices such as weight and body-wall thickness rather than species per se. However, in order to compare figure B-10 to lethality data from actual assaults and deformation data from ballistic reenactments, the deformation data should be adjusted for clay-skin differences and the lethality data

26 In response to OTA's request for the data to which figure B-10 had been fit, Russell Prather noted that he was "unsuccessful in locating the exact data set used to generate figure B-10 from report ARCSL-TR-77055," but "managed to locate much of the basic raw data from the blunt impactor program," which he provided to OTA [115]. He noted that a logistic model he fitted to the data he located was slightly more conservative (i.e., pessimistic) than figure B-10 at a deformation of 5 cm, predicting a probability of lethality of 0.20, compared to 0.15 or 0.16 for figure B-10; the former value was quoted in [114]; the latter by Prather in his letter of 18 Aprii 1991. The difference is insignificant and difficult to measure from the figure or discern by eye.

To fit the model, Prather used the Waker-Duncan method of logistic regression, which requires an initial estimate, which influences the fitted model [164]. OTA fit a model to the data using a Newton-Raphson procedure [91], which also requires an initial estimate, but it does net influence the fitted model. OTA found that it predicted a probability of lethality of 0.154 at a deformation of 5 cm, in agreement with figure B-10.

When OTA included a separate, non-overlapping set of data (provided by Larry Sturdivan) on blunt-impactor shots at goats, targeted over the liver the resulting model predicted a probability of lethality of 0.157 at a deformation of 5 cm, also in agreement with figure B-10.

Figure A-3—Correlation of Probability of Lethality With Deformation Depth



SOURCE: Russell N. Prather et al. 1977 [1 14]. Redrawn by the Office of Technology Assessment, 1992.

should be adjusted for imperfect targeting in actual assaults. Both adjustments would result in prediction of lower lethalities for assaults on armored humans than are indicated in figure B-10.

Types of Armor

The standard specified five levels of armor: I, II-A, II, Ill, and IV. Types I, II, and IV corresponded to the three types defined in 0101.00: II-A offers protection against an intermediate handgun threat and III offers protection against a rifle threat less than that of IV, the old .30 AP type. The velocity requirements changed somewhat, and a plus-or-minus tolerance was introduced in place of the previous no-slower-than specification of velocities. (See table A-4.) presumably manufacturers were concerned that the no-slower-than specification would leave any vest vulnerable to penetration if tested by a sufficiently fast bullet.

The 0101.01 standard also provides for "special type" armor; armor whose ballistic protection is specified by the manufacturer in terms of the exact test rounds it will withstand.

Results of Testing Under 0101.01

Nearly half the armor submitted on the promulgation of the 0101.01 standard failed. Manufacturers responded by improving their armor, and 87 models of armor were eventually certified according to this standard. [148, 150]

An important consequence of the wet-testing protocol is often overlooked. Not only does it require that vests withstand bullets when wet, it doubles the number of shots fired during a test. Separate vests take the damage, so there is no issue of cumulative damage on a given vest. Nevertheless, there is an issue of cumulative probability that the vest will fail on some shot or other. Vest samples that have a 95-percent chance of passing the dry shots would have only a 90-percent chance of passing both the wet shots and the dry shots, even if they performed exactly as well wet as they did dry.[27]

Comments on Technology Specificity in the 0101.01 Standard

With textbook avoidance of technology specificity in their standard, the formulators reacted to the

27 Because, in effect, they have to pass two tests, which they can do with 95-percent probability each, for an overall probability of $0.95 \times 0.95 = 0.9025$.

Table A-4-Summary of 0101.01 Armor Types According to the Ammunition Against Which They Were Tested

Type	Caliber	Mass (grains)	Velocity (ft/sec)	Tolerance (ft/sec)
I	.22 LRHV	40	1,050	+/-40
	.38 RNL	158	850	+/–50
II-A	.357 JSP	158	1,250	+/–50
	9-mm FMJ	124	1,090	+/–50
II	.357JSP	158	1,395	+/-50
	9-mm FMJ	124	1,175	+/-50
III	7.62-mm FMJ	150	2,863	+/-1 51
Iv	.30-06 AP	166	2,750	+/–50

KEY: AP = armor piercing; FMJ = full metal jacket; JSP = jacketed soft-point; LRHV = long rifle high velocity; RNL = round-nose lead.

SOURCE: National Institute of Law Enforcement and Criminal Justice, NILECJ Standard 0101.01, December 1978.

Table A-5-Summary of 0101.02 Armor Types According to the Ammunition Against Which They Were Tested

Type	Caliber	Mass (grains)	Velocity (ft/sec)	Tolerance (ft/sec)
I	.22 LRHV	40	1,050	+/-40
	.38 RNL	158	850	+/–50
II-A	.357JSP	158	1,250	+/-50
	9-mm FMJ	124	1,090	+/-50
II	.357JSP	158	1,395	+/-50
	9-mm FMJ	124	1,175	+/-40
III-A	44 Magnum	240	1,400	+/-50
	9-mm FMJ	124	1,400	+/-50
III	7.62-mm FMJ	150	2,750	+/–50
Iv	.30-06 AP	166	2,850	+/–50

KEY: AP - armor piercing; FMJ = full metal jacket; JSP = jacketed soft-point; LRHV = long rifle high velocity; RNL = round-nose lead.

SOURCE: National Institute of Justice, NIJ Standard 0101.02, 1985.

finding that ballistic material in common use fails when wet by requiring that the armor stop bullets when wet, not that it be waterproofed. Most manufacturers complied by using a waterproofing agent, while others placed the ballistic material in a waterproof carrier. Eventually, an alternative, non-woven, material would prove impervious to water and come into use.

The requirement that the strike face be specified stemmed from an incident in which a particular piece of armor was easily penetrated when mistakenly shot at from the wrong side.

NIJ STANDARD 0101.02

The 0101.02 standard was promulgated in March, 1985 by NIJ's Technology Assessment Program.

Reasons for Replacing the 0101.01 Standard

Researchers had become aware that, whereas a head-on shot is considered the most stressful case for rigid armor, woven armor could actually be more penetrable from an oblique angle than head-on. [150] The exact mechanics of this vulnerability evidently depend on the geometries of the weave and the bullet: a new fabric introduced in the late 1980s seemed particularly vulnerable to angle shots. [150] In particular, 9-mm bullets hitting loosely-woven Kevlar fabric penetrated best when hitting at an angle of about 30 degrees away from head-on. For soft body armor, the 0101.02 test added two shots at 30 degree angles, removing one other shot from the test so that each panel had to withstand six fair shots.

The 0101.02 standard introduced a new category of ballistic resistance, type III-A, for armor intended to withstand. the high energy handgun bullets fired by .44 Magnum handguns and 9-mm submachine guns[28 29] (See table A-5.) Some say that type III-A was introduced as a result of the increased threat to police officers on the street. Heretofore the multiplicity of the shots against a single test item armor (except for type III armor, which only receives one shot) was apparently seen only as a means of collecting an adequate amount of data. With the increased prominence of autoloading pistols and even submachine guns, however, the ability of the armor to stop more than one shot became a requirement in itself. For this reason, the placement of the shots on the vest was considered with a view to providing a basis for the evaluation of the vest's ability to stop multiple shots. [150]

The 0101.02 standard also introduced a higher level of specificity as to the placement of shots. Diagrams showed where, on a typical panel, fair shots ought to fall.

Sampling

The 0101.02 standard again requires that two to four complete sets of armor be 'selected at random' from some unspecified set. In a new stipulation, these armors are to be sized for a 46"-48" chest. The

n A submachine gun is a selective-fire weapon that fires pistol ammunition.

29 Facts in this section come from the standard itself [143] if 110 other source is cited.

rather large size lowers the likelihood of shots being deemed unfair because they are too close together. In the case of vests designed for female officers, it is difficult to believe that enough vests of such large size would be made to permit selection of four test articles "at random. "

In a new section entitled "Acceptance Criteria," 0101.02 articulates that a "model" of a vest meets the standard if it meets the workrnanship, labeling, penetration, and deformation requirements. This concept represents a departure from the ''lot' concept.

Marking and Workmanship

The 0101.02 standard reiterated the marking and workmanship requirements of the 0101.01 standard. The labeling requirements were enhanced to include a requirement that the type specification explicitly state the type and the standard according to which it was categorized. Thus a label could declare a vest to be "Type II-A under NIJ Standard 0101.02." For armor of types I through III-A, 0101.02 required a warning printed in large type declaring that the vest was not intended to protect against rifle fire or attacks from edged or pointed weapons. Curiously, the labeling portion of the standard also required a label certifying compliance with the standard-presumably a manufacturer could not affix such a label prior to certification, and yet presence of the label was declared to be a requirement for certification.

Penetration

The 0101.02 standard contained the first specific reference to vests contoured for female officers: in the case of such vests, at least one of the 30-degree angled shots had to fall on a bust cup. (The backing material under the vest was to be contoured so as to fill the bust cups.) Because the 30-degree shots are numbers 4 and 5, the resulting deformation is not measured. If the cup contains a seam, the shot must land on a seam. Though the 30-degree incidence is in principle measured between the line of fire and the tangent plane of the vest, the departure of the bust cup from the main plane of the vest makes these shots' angles of incidence questionable, and very probably less than 30 degrees.

In practice, the requirement that a seam be hit can necessitate a seventh fair shot in the case of female vests.

Deformation

Backface deformation was measured only on the first fair shot under the 0101.02 standard, rather than on the frost two fair shots as under the 0101.01 standard. During the transition to the 0101.03 standard, Justice Department officials investigated rumors that the clay block used in 0101.02 testing had had a plywood backing, lessening deformation. This backing, not mandated by the 0101.02 standard, did exist but was not used for 0101.02 testing. [148]

Types of Armor

The 0101.02 standard introduced the Type III-A armor, a soft armor capable of stopping .44 Magnum bullets. This armor type was created at the behest of another Federal Department, whose employees sometimes needed such protection. Some in the NIJ rue the inclusion of III-A armor in the standard, because of the implication that it is appropriate for daily use by law enforcement officers. They feel that local police departments will, acting through understandable and laudable concern for the welfare of their employees, obtain III-A vests without realizing that they are far more robust, expensive, and uncomfortable than is appropriate for police use. In that case, the probable outcome would be that the vests would go unworn. The NIJ's Body Armor Selection Guide cites Type III-A armor as ''generally considered unsuitable for routine wear. However, individuals confronted with a terrorist weapon threat may often be willing to tolerate the weight and bulk of such armor while on duty." [145][30]

Results of Testing Under 0101.02

As was the case with the addition of wet testing in the transition from 0101.00 to 0101.01, the addition of an extra shot in 0101.02 made the test harder to pass. Not only did the total number of opportunities to fail increase (albeit by 20 percent instead of 100 percent), but the number of fair shots per vest actually increased, increasing cumulative damage to the vest.

Because of administrative disarray at the IACP during the 0101.02 period, it is not clear how many vests, or which ones, were tested under the 0101.02 standard. Sixty-two models were certified as having passed. [148]

---

30 Nevertheless, some wearers report that discomfort varies only slightly with vest thickness.

Comments on Technology Specificity in the 0101.02 Standard

Angled shots against armor of types I through III-A were introduced in response to the discovery that 9 mm bullets penetrated Kevlar more readily at that angle than they did at normal incidence. This modification of the test represents the technology-specific consideration that these vests, but not the rigid vests of types 111 and IV, were likely to be made out of Kevlar and thus subject to the angle penetration.

NIJ STANDARD 0101.03

The 0101.03 standard was promulgated in April 1987 by NIJ's Technology Assessment Program.[31] Clarifications and modifications of the test procedure have been made since.

Reasons for Replacing the 0101.02 Standard

As mentioned above, it is not clear how many vests were tested under the 0101.02 standard. Worse, samples of certified models were not retained in an orderly way, so that there was no way for the NIJ to determine if the construction of a given model offered for sale was the same as the construction of the model of the same name that had passed the 0101.02 test. These circumstances were brought about by administrative disarray at the IACP. The NIJ reassigned the Technology Assessment Program Information Center (TAPIC) function of the IACP to a new grantee (Aspen Systems), but some information on body armor tested under 0101.02 could not be recovered [150] and a rationale for beginning anew with Aspen Systems was needed.

Retesting and recertification appeared to be the only recourse. The NIJ offered to pay for retesting if the manufacturers would supply the vests, but the manufacturers balked, fearing the consequences if a previously certified model should happen to fail the retest. In such a case, what would be the status of the vests of that model that had already been sold? The NIJ and the manufacturers agreed to let the results of the 0101.02 period stand, but to create for the retest anew standard, 0101.03, that would be substantially the same as 0101.02. The purpose of 0101.03 was simply that it would be a different standard, so that if a vest that had passed under 0101.02 failed the retest, it would not create an anomaly in which vest had passed and then failed the same test. [150] Even today, vests are sold on the strength of their 0101.02 compliance test.

Table A-6-Summary of 0101.03 Armor Types According to the Ammunition Against Which They Were Tested

Type	Caliber	Mass (grains)	Velocity (ft/sec)	Tolerance (ft/sec)
I	.22 LRHV	40	1,050	+50
	.38 RNL	158	850	+50
II-A	.357JSP	158	1,250	+50
	9-mm FMJ	124	1,090	+50
II	.357 JSP	158	1,395	+50
	9-mm FMJ	124	1,175	+50
III-A	44 Magnum	240	1,400	+50
	9-mm FMJ	124	1,400	+50
111	7.62-mm FMJ	150	2,750	+50
Iv	.30-06 AP	166	2,850	+50

KEY: AP = armor piercing; FMJ = full metal jacket; JSP = jacketed soft-point; LRHV = long rifle high velocity; RNL = round-nose lead.

SOURCE: National Institute of Justice, NIJ Standard 0101.03, 1987.

Minor changes in 0101.03 as compared to 0101.02 included the elimination of the negative side of the plus-or-minus standard, so that the nominal velocity figure could be cited as a minimum. (See table A-6.) Records of tests performed under the 0101.02 standard revealed that the majority of shots fell in the plus side of the standard anyway, so that this change was not viewed as significant. [150]

In a more major change in the test protocol, the 0101.03 standard clarified the point that vests were not to be smoothed out or repositioned between shots.

Perhaps because of difficulties in determining which vests had been tested under 0101.02 and which had not, the 0101.03 standard introduced the distinction between a *model* and a style: several styles of the same model vest could all be certified by the same test, inasmuch as they were ballistically identical and only superficially different.

Sampling

The 0101.03 standard takes for granted that a full set of four armors will be needed, though there is still a tester's option to test the same panel with two types of ammunition. 0101.03 says that a "style" (not a ''model,' as in 0101 .02) of a vest meets the standard if it meets the workmanship, labeling, penetration, and deformation requirements. An administrative procedure issued by TAPIC clarifies the course of

[31] Facts in this section come from the standard itself [144], if no other source is cited.

action to be taken if a model fails-the manufacturer must abandon that model. [150] Not only must the manufacturer abandon the model designation, he or she may not submit a noncomplying model for retesting. [146]

Vests tested under the 0101.03 standard are archived by TAPIC for later reference. Under this system, any question about whether a given vest is the of same model as was tested can be resolved by direct comparison of the test item and the vest in question.

Marking and Workmanship

The 0101.03 standard departs from the marking and workmanship requirements of the 0101.02 standard in that a distinction is drawn between ballistic panels and the carriers in which they are used. Standard 0101.03 recognizes that some armor consists of a carrier and removable panels, whereas other armor consists of a carrier containing nonremovable panels. The 'panel' labeling requirements generally follow the 'armor' labeling requirements of the 0101.02 standard, enhanced to include a serial number and model or style designation uniquely identifying the panel for purchasing purposes. Under 0101.03, care instructions have to conform to part 423 of the Federal Trade Commission Regulation Rule. Carriers with nonremovable panels must, "in addition to the label required for the ballistic element, have a label on the carrier that is in conformance with the requirements for the ballistic panels,' unless the label on the panel is not covered by the carrier. Carriers with removable panels must be labeled with an identification of the manufacturer, ''a statement telling the user to look at the ballistic panels to determine the protection provialed, ' the size, date, and model name of the carrier, care instructions, and certification of compliance with NIJ Standard 0101.03.

Penetration

A clarification issued March 18, 1988 addressed the question of vests that may have been weakened by unfair hits. If a panel that has already received two or more unfair hits fails owing to penetration, the test is deemed inconclusive and another panel is tested.

A modification issued May 11, 1989 defined penetration to include "perforation of the last layer of fabric to the extent that the projectile breaks threads in that layer and protrudes from the inside surface of the layer. " [82]

Deformation

The 0101.03 standard eased a special requirement formerly placed on the first shot on each panel, the one that is used in the assessment of backface signature. Under 0101.02, the velocity of this shot had to be in the upper 32.8 ft/s (10 m/s) of the allowable range of velocities.[32] In the context of its elimination of the bottom 50 ft/s of the allowable range, 0101.03 permitted the velocity of the first shot to be anywhere in the remaining 50 ft/s, not restricting it to the upper 32.8 ft/s. In this respect, 0101.03 relaxed the backface deformation standard by allowing shots of slightly lower velocity.

On October 10, 1989, H.P. White Laboratories proposed a modification under which backface deformation would be measured for all normal-incidence shots, not just the first on each panel. The measurements would be made after all of the shots were fired, so as to avoid any rearrangement of the vest between shots. (Under the current practice, the measurement of the BFS of the first shot is made right after the shot, in effect Wowing for a rearrangement of the vest.) Any deformation in excess of 44 mm would constitute a failure of the vest. The NIJ has not accepted this modification. [82]

Types of Armor

The 0101.03 standard did not introduce any new armor types, nor any new shots. Subsequent modifications to the standard moved the sites of the fourth, fifth, and sixth shots slightly, to avoid placing any shot directly on threads weakened by a previous shot. [82]

Results of Testing Under 0101.03

Manufacturers and government officials alike expected that some vests certified under 0101.02 would fail the 0101.03 retest purely through the operation of chance alone: as described above, this expectation was a principal reason for the creation of the 0101.03 standard in the first place. However, far more vests failed than anybody expected: 50 out of

32 NIJ STD 0101.02 [143], page 10, first paragraph. The operative sentence can be seen as ambiguous: H.P. White Laboratory personnel explained its interpretation to the OTA staff.

Table A-7—Results of 0101.03 Compliance Retests

Ballistic resistance level	I	II-A	II	II I-A	III	Iv	III/Iv	Total
Tested	12	21	29	17	2	2	1	84
Certified	11	9	11	1	1	0	1	34
Failed	1	12	18	16	1	2	0	50

SOURCE: National Institute of Standards and Technology, undated [1 371.

84,[33] The results of these tests are shown in table A-7.

Experts differ as to whether the slight increase in velocity caused by abandonment of the negative side of the velocity tolerance could, statistically, explain so many failures in a group of 84 vests that had previously passed.[34] However, a variety of other causes have been suggested.

As mentioned above, the 0101.02 standard provided for testing of the second ammunition type on the same vest as had been used for the first ammunition type. If a failure occurred with the second ammunition type, the successful passage with the first ammunition type was allowed to stand and the test with the second ammunition type was restarted on a fresh panel. The purpose of this protocol was to save money by consuming the minimum number of panels possible. An important consequence, however, was that the vest could have two chances to pass the second part of the test. The majority of 0101.02 testing was done in this fashion. [137]

Existing records of successful tests under 0101.02 cite some reports as “revised,” without further explanation. Unsupported allegations exist that individual panels were submitted to substitute for ones that failed, until a complete set of eight passes was garnered. [137] This practice could perhaps be seen as having been fostered by the protocol allowing a restart of a second-ammunition test upon failure.

It seems possible to OTA that the large number of failures could be attributed to the 0101.03 standard’s heightened strictures against smoothing down or repositioning the vests between shots. Allegations are also sometimes made to the effect that, under 0101.02, vests were intentionally strapped to the test fixture so weakly that they would fall off after a shot, producing a free rearrangement of the vest as it was reattached to the test fixture. Regardless of any change in intent, the 0101.03 standard provided (at the time of the retest) for 4 straps attaching the vest to the test future rather than the 2 used under the 0101.02 standard. Presently, the 0101.03 standard provides for 5 straps, an extra strap having been mandated by the NIJ in a procedural modification.

Because 0101.02 testing was coordinated directly between the manufacturer and the test lab, it is possible that failures existed and were not reported to the IACP. It is also possible, given the recordkeeping difficulties experienced by the IACP during the 0101.02 era, that records of failures were received but not preserved in an accessible manner.

While no single difference between the 0101.02 and 0101.03 revisions, or the procedures associated with them, can satisfactorily explain the large number of failures during the 0101.03 retest, the above factors, working in concert, may have exerted a cumulative effect greater than any individual effect.

Hundreds more vests have been tested since the retest program. The results of this testing are shown in table A-8.

The deformation standard has occasioned a debate out of proportion to the number of failures attributable to deformation alone. [150] Manufacturers and

33 Because of the recordkeeping anomalies prevalent during the 0101.02 era and because some manufacturers took the precaution Of renaming vest models before submitting them for the retest, the NIJ and NIST—though possessing evidence that some vests that failed in the retest had passed 0101 .02-cannot fully document all such cases with confidence. However, submission of the vest for a retest, as such, constituted an implicit statement that the vest had passed 0101.02 and was being retested as part of the pact that the government made with the industry when introducing 0101.03 In addition, OTA has received confirmation from members of the body armor industry that many vests that failed the retest had passed under 0101.02. No party has contested the figure of 50 out of 84, which appears in [137], page 31. This source also says, on the same page, that 62 vests passed 0101.02—it is not clear where the other 22 (i.e. 84- 62) 0101.02-compliant vests came from.

34 Based on a review of extant records of 0101.02 testing, [137] makes a strong case (on pp. 31-33) that most shots fired in 0101.02 testing lay within the velocity window specified later for 0101.03, concluding that “the test results for at most 25 percent of the armor could be influenced to some extent by the elimination of the negative velocity tolerance.”

Table A-8-Results of 0101.03 Certification Tests (as of June 1991)

Ballistic resistance level	I	II-A	II	II I-A	III	Iv	Total
Certified	29	70	102	57	15	8	281
Failed (penetration)	2	46	68	14	7	8	145
Failed (deformation)	1	1	1	10	2	0	15
Failed (both)							
(subtotal)	(4)	(54)	(78)	(44)	(9)	(8)	(197)
Inconclusive							
Tested	33	125	182	103	24	16	483

SOURCE: H.P. White Laboratory, Inc. June 7, 1991.

others have made statements such as "more than 50 percent of current vests fail the NIJ/NIST test procedure despite their perfect performance in the field" [87], because the rate at which vests fail the test (40 percent) greatly exceeds the rate at which they fail in the field (said to be O percent, on the grounds than no officer has ever been killed through being hit on the protected area by a bullet the vest was rated to stop). [150] Manufacturers say this discrepancy stems from over-conservatism in the standard. A more obvious reason is that vests that fail are not (presumably) presented in the market-place for sale. Other possible reasons include the fact that vests see use against all threat levels whereas they are only tested against the most threatening level they could hope to withstand. The fact that manufacturers feel an incentive to build close to the limit so as to avoid the extra weight, bulk, heat retention, and expense incurred by having more ballistic protection than is necessary may explain why the success rate of vests has not improved despite claims of technological progress by the manufacturers.

The tendency of the test armor, if untouched, to bunch up on the clay during testing has previously been mentioned as a possible cause of failure. Tests conducted under the 0101.00,0101.01, and 0101.02 standards resulted in a large number of truncated trials because, to save money, testing stopped immediately upon a failure. The procedure of the 0101.03 test, unlike that of its predecessors, mandates continued shooting even after a failure, so that complete data are available. These data can be examined for signs pointing to bunching as a significant cause of failures.

Appendix B

The Utility of Police Body Armor

Contents

Box

Figures

Tables

The Utility of Police Body Armor

SUMMARY

Every year, about 60 police officers are killed by gunfire-the majority by handguns. Concealable body armor offers several levels of protection encompassing the full spectrum of the handgun threat. In addition, some vests protect against shotguns and certain rifles. Every year, vests save one or two dozen officers from death by gunfire. If every officer wore a vest, the number of officers saved from death by gunfire might be doubled.

THE FIREARMS THREAT TO POLICE OFFICERS

Police Confiscate More Powerful Firearms, Perceive Increasing Threat

Jurisdictions all across America report an upswing, during the last few years, in the confiscation of especially sophisticated and deadly firearms. These include "assault rifes" and high-powered automatic pistols. Police officers feel they are more threatened by these guns than they were in the past. [102][1] Some blame the increase on the affluence of criminals involved in the drug trade; others see it as an unfortunate outcome of the move to ban the cheap handguns known as "Saturday night specials."

One incontrovertible increase in the threat to police officers is the officers' own guns. Many departments, responding to the heightened firearms threat on the street, invested in more powerful guns themselves, typically replacing .38 Specials with .357s, 9-mm "automatics," or even larger guns. Because 20 percent of officers who get shot are shot with their own or their partners' guns, [140] an upgrade of the officers' weaponry increases the threat they face.

One response to the perception of a growing threat to police officers is the wearing of soft, concealable body armor. Such a protective garment has a soft, padded feel, fits under the officer's shirt, and is intended to be worn at all times. It is not a "flak jacket" or bomb squad outfit, worn outwardly and only at times of great threat. Nor does it include rigid metal plates, though many examples include a large pocket into which a rigid plate (perhaps carried in the squad car's glove compartment) can be placed if a greater-than-expected threat arises. Many officers feel that they owe their lives to the practice of day-to-day wear of soft body armor,[2] but shooting deaths of officers continue.

The Guns That Kill Police Officers

There is considerable evidence that the perceived threat to police officers posed by high-powered guns is exaggerated. Some of the perception is doubtless founded in newspaper headlines and departmental scuttlebutt, sources that disproportionately report interesting cases and thus overstate the threat from exotic weaponry.

Some officers may, more objectively, base their threat estimate on the statistics of weapons confiscated by their department or nationwide. Even this would exaggerate the threat. For example, the mix of firearms confiscated by the Bureau of Alcohol, Tobacco, and Firearms (see figure B-1), which is presumably representative of those confiscated by local law-enforcement agencies nationwide, is far richer in powerful weapons than is the mix of firearms used in fatal assaults on police officers (see figure B-2), according to information collected systematically from local police departments and Federal agencies by the Federal Bureau of Investigation (FBI), which publishes it. [140] An estimate based on departmental confiscations might be more representative of the threat in a particular jurisdiction but would be ''noisy'—prone to error—because of the small sample size.

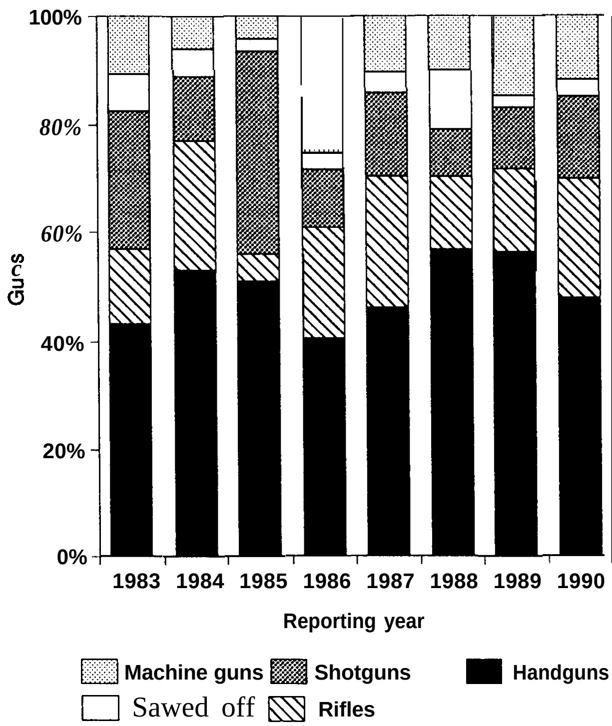
It is plausible that the mix of guns used in all assaults on police might have an even smaller proportion of powerful guns than does the mix of guns used in fatal assaults on police. However, the FBI does not collect comprehensive data on types of guns used in nonfatal shootings of law-enforcement

[1] Numbers in brackets cite references in the bibliography in volume 1 of this report.

[2] The IACP/DuPont Kevlar Survivors' Club (S. M.) includes about 1,400 members, over 500 of whom credit soft body armor with having saved them in shooting incidents.

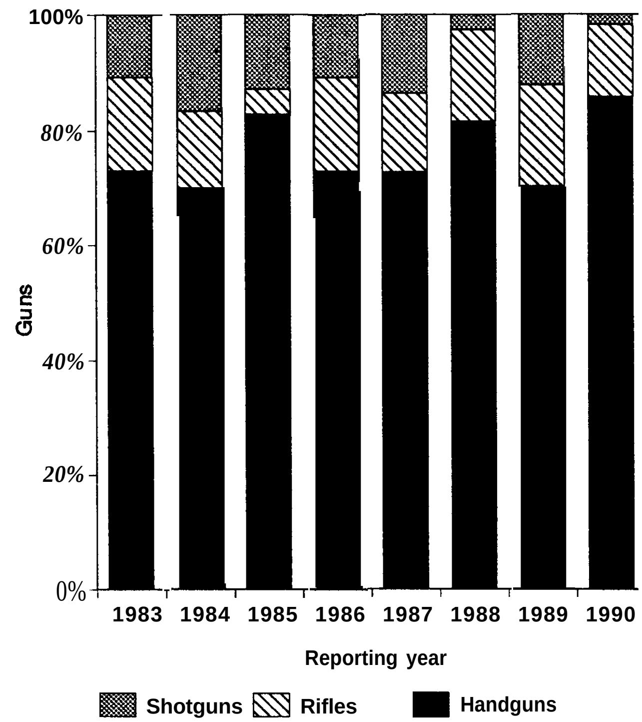
[3] Indeed, the National Institute of Justice commends the use of confiscated weapons as an indicator of what vest to buy.

Figure B-l—Mix of Guns Confiscated by the Bureau of Alc:ohol, Tobacco, and Firearms Is



SOURCE: Bureau of Alcohol, Tobacco, and Firearms, 1991.

Figure B-2—Types of Guns That Killed Police Officers



SOURCE: Federal Bureau of Investigation, 1984-1991.

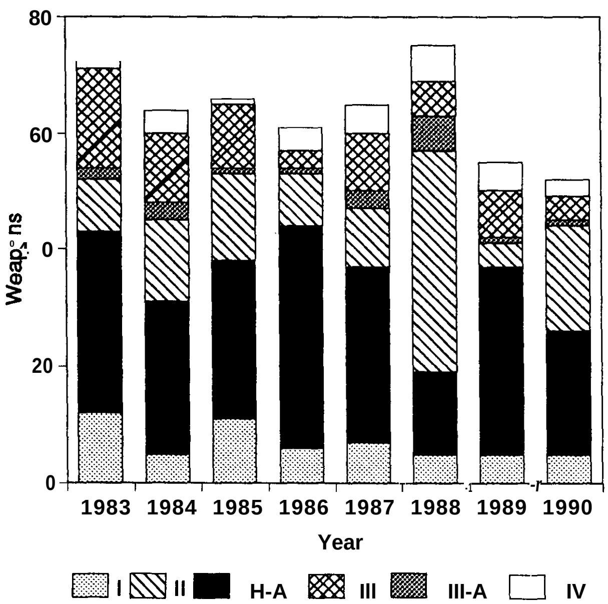
officers. The Bureau plans to expand its data-collection program to collect such data, if resources permit. [108] Currently, the FBI's annual report, Law *Enforcement Officers Killed and Assaulted,* tabulates reported assaults on law-enforcement officers by type of weapon used but lumps all types of firearms together in a single category. Moreover, the tabulation includes assaults without battery, so the assaults with firearms include incidents in which guns were used only to threaten officers or were fired without hitting them.

Figure B-3 shows the mix of guns used to kill police officers in the United States in recent years, categorized (by OTA) according to the minimum level of ballistic resistance the National Institute of Justice (NIJ) has recommended for protection from the threat. The National Institute of Justice categorizes body armor into levels of ballistic resistance in terms of the gunfire threats it is supposed to withstand (see table B-l). Each level of armor is expected to offer protection against the threat associated with it and with all lower numbered levels of armor. For threats, such as birdshot and buckshot, that are not specifically mentioned by NIJ Standard 0101.03 or NIJ Guide 100-87, OTA used the guidelines in National Institute of Law Enforcement and Criminal Justice Standard 0101.01 (1978).

The data reflect only fatal attacks: because an officer is more likely to survive an injury from a lower level threat than from a higher level one, one would expect that the data on killings understate the incidence of low-level shootings. Especially in this light, the continued prominence of threat-level I and II-A killings is worthy of note: anecdotal evidence, surveys based on officer's opinions, and perhaps even tabulations of weapons confiscated from criminals, would have one believe that the threat to the police officer is swinging dramatically towards the high end of the spectrum. The FBI data, however, do not particularly bear this impression out.

Felonious gunfire kills about 60 officers per year; a handful of officers are feloniously killed each year by other weapons, or without weapons. About the same number of officers are killed accidentally as are killed feloniously (see figure B-4). The majority of the accidental deaths involve motor vehicles.

Figure B-3-Types of Guns Used to Kill Officers (sorted by lowest level of armor expected to stop projectile)



SOURCE: Office of Technology Assessment, 1992, using data from the Federal Bureau of Investigation, 1983-1990.

These data include only sworn law enforcement officers; deaths of other possible civilian users of body armor, such as security guards, do not appear.

THE BUYING AND WEARING OF BODY ARMOR

Estimating Actual Wear Rate

Exact data on body armor sales are treated as proprietary by the manufacturers, but we can make a rough estimate of the number of vests extant in the United States. The concealable body armor industry grosses about $40 million per year in sales for U.S. civilian[4] use. [14, 129, 150] Assuming that a vest costs $400 and lasts for 5 or more years,[5] 100,000 vests are sold yearly and 500,000 or more are in

Table B-I—Levels of Protection

Level	Threat
I	.22, .25 and .32-caliber handguns, .38 Special lead
II-A	.38 Special high velocity, .45s, low velocity .357 Magnum & 9-mm, and .22 rifles
II	Higher velocity .357 Magnum and 9-mm
III-A	.44 Magnum and submachine gun 9-mm
III	High-power rifle: 5.56-mm, 7.62-mm full metal jacket, .30-caliber carbine, .30-06 pointed soft point, 12-gauge rifled slug
Iv	Armor-piercing, .30-caliber rifle bullets

SOURCE: National Institute of Justice, NIJ Standard 0101.03, 1987, and NIJ Guide 100-87, 1989.

useable condition at any one time. Considering that there are about a half a million police officers (not counting other potential users of concealable body armor such as security guards), [78] the industry can supply most of those who could benefit from concealable body armor. These estimates arguably *understate the* number of vests produced because-especially with recent price competition-the average price of a vest may be lower than $400. They arguably *overstate the* number of vests in use because the business has grown to the $40 million figure in recent years[6] and because some vests are replaced before they wear out, owing to a perception that they are insufficient to meet the present threat.

Naturally, some officers are more at risk than others-some work in peaceful small towns and others in the more violent environment of today's big cities. Departments or individual officers in the more dangerous settings could be expected to be more likely to buy and wear body armor, so we might expect to find more wearers of body armor among those officers who get shot than among the population of officers as a whole.

This expectation is borne out by the FBI Uniform Crime Reports (UCR) data. As noted above, there is no systematic collection of the specifics of shootings not leading to the death of an officer. The FBI does report, in conjunction with the locations of officers'

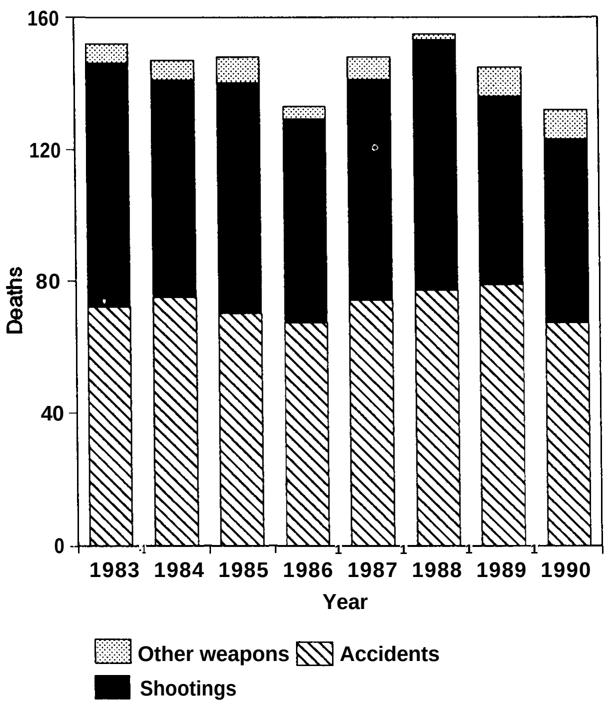
4 I.e., nonmilitary. U.S. civilian users of body armor include sworn law enforcement officers, security guards, and others. A few vests are known to have made their way into the criminal world.

5 These figures, while chosen for convenience, are roughly correct. Armor prices vary widely according to size, level, and style. Some authorities advocate a "rational replacement policy" that begins to consider a vest for replacement after 5 years of use but recognizes that, with proper care, a vest can last twice that long. [145] Anecdotal evidence suggests that many vests receive improper care.

6 NIJ has estimated that the body armor industry grosses $50 or $60 million per yearand that vests cost $200. These data would lead to the conclusion that 250,000 to 300,000 vests are manufactured per year, enough to supply every officer with a new vest every 3 or 4 years once the entire fores had been outfitted. NIJ's figures certainly understate the average cost of a vest and arguably overstate the industry's output of concealable body armor destined for domestic use: foreign sales account for part of the industry's gross earnings, and manufacturers of concealable body armor also produce body armor of types III and IV as well as a variety of other products such as helmets and helicopter seat cushions.

297-923 0 - 92 - 3 : QL 3

Figure B-4-Law Enforcement Officers Killed



SOURCE: Federal Bureau of Investigation, 1984-1991.

fatal gunshot wounds, whether or not the officers were wearing vests. [140] (See table B-2.) As one might expect, few officers wearing armor are killed by shots to the upper torso; to date, no officer has been killed when struck on the protected area of a vest by around that his or her vest was rated to stop. The proportion of officers wearing body armor when they get shot can be estimated from the proportion of officers wearing body armor when they died of gunshot wounds in locations *other than the upper torso.*[7] *This* proportion initially increased as body armor penetrated the market but has fluctuated between 30 and 40 percent for several years. The sample size introduces some uncertainty, but even the region spanned by 95-percent confidence intervals shows some fluctuation (see figure B-5).

The estimate that 30 to 40 percent of officers wear body armor is low by comparison to survey data. It has been suggested that officers who get caught in gunfights might be the sort who don't tend to wear their armor. However, one could equally well argue the opposite, that the wear rate of officers who come under fire is greater than that of the population of officers as a whole, either because of some knowledge that a shooting was in the offing, or simply because, as mentioned earlier, officers serving in dangerous areas may tend to wear their vests more than do other officers. In any case, the wear rate estimated from other-than-upper-torso deaths may be termed an "under-free wear rate," to distinguish it from the true average wear rate.

Factors Influencing the Wearing of Armor

Many officers who possess armor do not always wear it. Because armor is rarely shared,[8] the proportion of officers who wear armor would not be expected to exceed the proportion who posssess it.

Comfort

Concealable body armor can be somewhat uncomfortable to wear. Even though some officers claim, in responses to a recent survey, that they want a vest that protects and do not care if it is uncomfortable, [102][9] officers who own vests often find reasons not to wear a vest on a particular day. Most of these reasons center on comfort. Wearers (and, especially, nonwearers) commonly cite the armor as "hot," "heavy, " "stiff," "chafing, ' ' and the like. Complaints about chafing, and to some degree about stiffness and the impression of great weight, can often be traced to a bad fit, or simply to the armor being strapped on too tightly. Armor should be the right size-the front panel should just reach the navel if the officer is to be comfortable when seated. Female officers can expect particular difficulty in getting armor to fit: one body armor manufacturer expressed the view that custom fitting was the only way to guarantee a female officer that her armor would be comfortable.

The complaint that armor is heavy strikes some as minor because the weight is well-distributed (a backpack that weighed only a few pounds would hardly be considered a load at all) and because police officers already carry a number of other heavy items,

7 Because armor protects the upper torso, officers who wear armor are under-represented among those who die of upper torso wounds and thus among "officers killed" as a whole. For this reason it is inappropriate to estimate wear rate from the total population of officers killed. [144] Wearers may be slightly over-represented among those who die of non-upper-torso wounds, inasmuch as the armor may block one or more upper torso shots prior to a fatal shot elsewhere, e.g., the criminal keeps shooting until he hits the head.

8 "Armor is underwear, " as one company phrases its admonition against armor-sharing.

9 Cf. reference [23].

Table B-2—Location of Officers' Fatal Gunshot Wounds

	Wound location					
	Head		Upper torso		Lower torso	
	Total victims	Armored victims	Total victims	Armored victims	Total victims	Armored victims
1981	36	6	47	5	3	0
1982	24	5	56	7	2	1
1983	29	10	42	8	3	2
1984	33	13	32	4	1	0
1985	27	8	43	3	0	0
1986	26	6	33	6	3	2
1987	31	13	32	3	4	0
1988	37	15	36	3	3	2
1989	27	9	24	6	6	3
1990	31	11	22	2	3	2
Total	301	96	367	47	28	12

SOURCE: Federal Bureau of investigation, 1987and 1990.

notably their guns. On the other hand, one could argue that the weight of the vest, taken on top of the weight of all the other things an officer is expected to carry or wear while on duty, is a significant burden. The most salient aspect of complaints about the weight of the vest is that the weight (unlike chafing and even, as we will see below, heat) is directly related to the ballistic performance of the vest. A thicker vest will weigh more and will offer protection against a broader region of the threat spectrum. Insofar as weight lessens comfort, there exists a true comfort-v. -protection tradeoff. However, a pioneering study of influences on wear rate, by the Aerospace Corp., found that wear rate was independent of the areal density (weight per unit area) of armor with an areal density less than about 4.5 kilograms per square meter, but decreased slightly with increasing areal density above 4.5 kg/m^2 (see figure 7 of vol. 1). Officers' complaints that armor makes them feel hot cannot be attributed to improper fit. Not only is commonly used armor material a good insulator, but also the thickness of the armor blocks the evaporation needed for the body's normal perspirative cooling. Just six plies of fabric, waterproofed or not, are enough to block the evaporation of sweat, so any vest—regardless of level[10] or waterproofing-an block perspiration. Some officers find that they can lessen the blocking **effect** of the vest by wearing a purpose-made ribbed undergarment, whose vertical ribs hold the vest away from the body and allow circulation of air under the vest.[11]

Though the added weight of the vest is not much compared to the other clothing and equipment worn by a police officer, the subtracted perspirative area is significant compared to the total area of the officer's skin. The vest imposes a true cost to the officer in terms of his body's ability to cool itself and can be viewed as a "legitimate complaint" about body armor. The Aerospace Corp. found that the strongest influence on wear rate, of those considered, was the Temperature-Humidity Index (THI) defined by the U.S. Weather Bureau. Reported wear rate was higher at times and locations with lower values of the THI (see figure 5 of vol. 1)-e. g., in winter (see figure 6 of vol. 1). [8][12] The correlation of wear rate with THI was -0.75. Manufacturers presumably feel an incentive to make their products more acceptable in this regard, so vests may eventually improve in their ability to let the wearer keep cool.

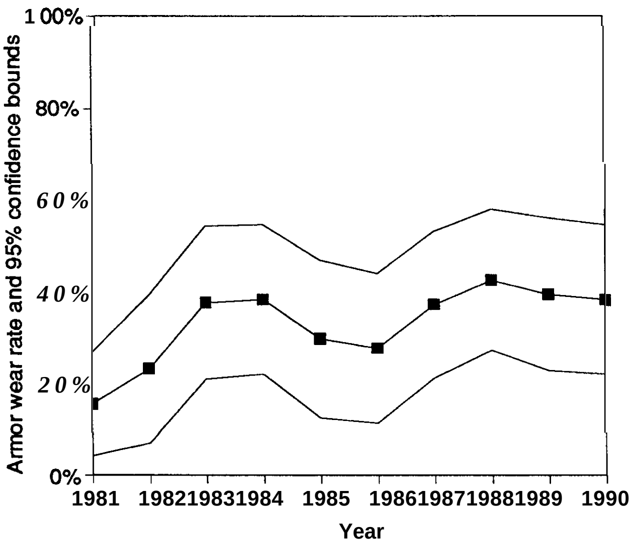
The Aerospace Corp. found the second strongest influence on wear rate was the officer's weight:

10 The NIJ-prescribed &S@ for a level I vest specified seven layers.

11 Additionally, these garments are made so as to wick perspiration away from the body and evaporate it from the garment's ribs. This effect increases cooling and eliminates the uncomfortable feeling of sweat dripping down one's body underneath the vest.

12 A more recent study by Strategy Polling Corp. and the John Jay College of Criminal Justice [102] found that self-reported wear rates by front-line officers were lowest in the Northeast (52 percent) and highest in the West (83 percent), with the South (66 percent) and North Central States (69 percent) inbetween. Wear rates by police management—an indicator of management support for wearing armor-were lower but followed the same geographical pattern, supporting earlier findings by the Brand Consulting Group [22, 23] that management support, including exemplary wearing, would increase wearing by front-line officers.

Figure B-5—Armor Wear Rate and 9&Percent Confidence Bounds



SOURCE: Office of Technology Assessment, 1992.

heavier officers tended to wear their armor less than lighter officers did (correlation with weight: -0.49). The third strongest influence on wear rate was the officer's age: older officers tended to wear their armor more than younger officers did (correlation with age: + 0.39). In contrast, the Brand Consulting Group reported, after surveying smaller samples of officers, that older officers wore armor less than younger officers did. [21, 22, 23] These results may not be inconsistent, because the Aerospace Corp. adjusted for weight in correlating wear with age, which is presumably positively correlated with weight. That is, the Aerospace Corp. found that lighter officers wore their armor more frequently than did heavier officers, but within each weight category, the older officers wore their armor more frequently than did younger officers.

Factors Other Than Comfort

Many factors other than comfort can influence an officer's decision as to whether to wear body armor on a particular shift. These include the perceived level of danger, orders to wear the armor, potential impact on disability or death benefits if it is found that armor was not being worn during an incident, and management support for armor wear.

Notoriously, harm seems to come when one least expects it. Many officers saved by their vests report that they had no particular feeling of danger when dressing for duty on the day they were shot. [121] In the larger sense, however, the officers and departments that have acquired body armor have done so for a reason: the perception that theirs is a dangerous jurisdiction. Similarly, officers assigned to particular parts of town, to particular shifts, or to duty on particular days of the week, might be more likely than others to wear their armor, even in the absence of any particular knowledge, foreboding, or premonition of danger.

Department-wide standing orders to wear armor are not unheard of. In some ways, it is surprising that mandatory wear is not more widespread: construction workers have to wear their hardhats, and even the National Hockey League has now adopted a helmet rule. It is difficult to assess how fully standing vest-wear orders are obeyed, but one would certainly expect them to have a positive influence on wear rate.

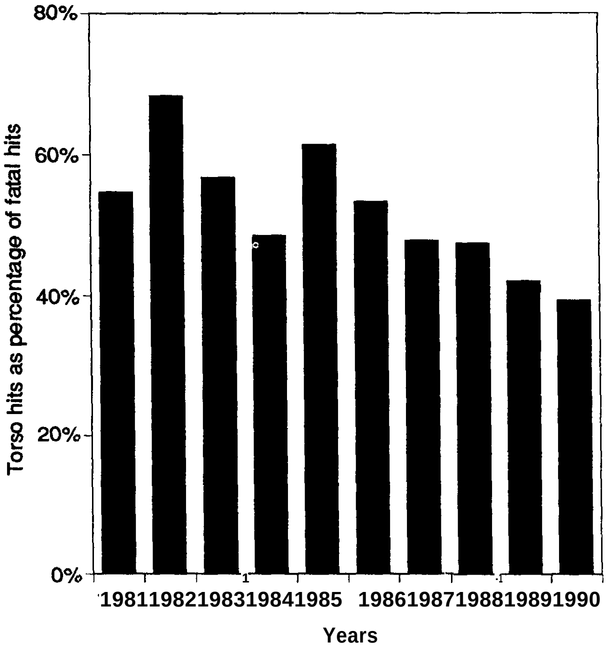
While the nonwearing of a vest, in contravention of standing orders, could be dealt with as a minor uniform infraction, the real sanction for an officer not in compliance with a mandatory-wear policy would be the potential loss of his or her survivors' benefits should he or she come to harm.

Finally, the value of management support for armor wear should not be under-rated. While exhortations, poster campaigns, and the like can sometimes seem "hokey" to those involved, management support for armor wear need not be limited to purchase of the armor. In the long run, and certainly after a "save,' a properly managed program of management support for the wearing of body armor will be seen as a meaningful expression of concern for the men and women on the force.

One would expect that, since the introduction of vests in the mid-1970s, the proportion of officers killed by wounds to the upper torso would have gone down. It has, but only very slightly: the small size of the decline can be attributed to the dilution of the vests' effect on upper torso hits owing to the FBI's expansive definition of "upper torso," which includes the arms and part of the neck.[13] A significant decrease has occurred since 1982 (see figure B-6).

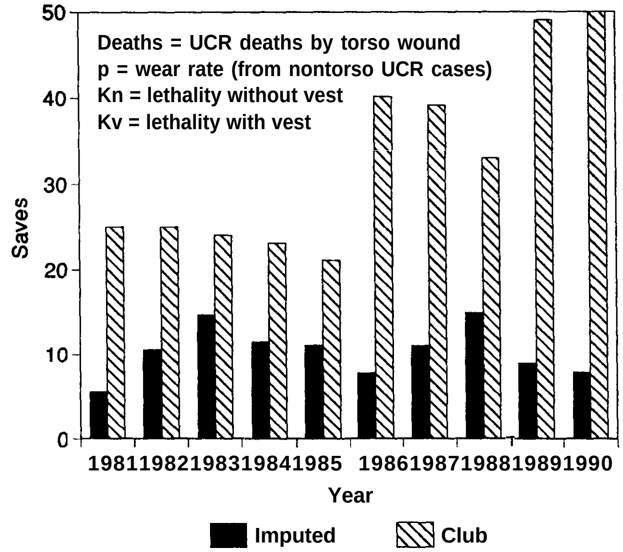
13 See reference [140], 1986, p. 14.

Figure B-6-Decline in Torso-Wound Share of Deaths



SOURCE: Office of Technology Assessment, 1992, based on data from the Federal Bureau of Investigation, 1982 to 1991.

Figure B-7-Saves Recorded by the IACP/DuPont Kevlar Survivors' Club (S.M.) v. Saves Estimated by OTA



SOURCE: IACP/DuPont Kevlar Survivors' Club (S.M.), 1991, and Office of Technology Assessment, 1992.

Officers Saved By Armor From Death by Gunfire

Based on body armor's effect in reducing torso wounds, one could estimate that body armor saves about 10 officers per year from death by gunfire.[14] Firms involved in the body armor business collect and publish data on the number of ''firearms saves—instances in which an officer *probably* would have died by gunshot wound were it not for body armor-and report numbers considerably greater than 10 per year (see figure B-7).[15] These numbers exceed OTA's estimate of saves from death by armor partly because some wearers saved from *probable* death would not have certainly died had they not been wearing armor.[16] In the aggregate, therefore, the set of people counted as saves will be slightly larger than the set of people who would have died had they received the same hits without any vests on.

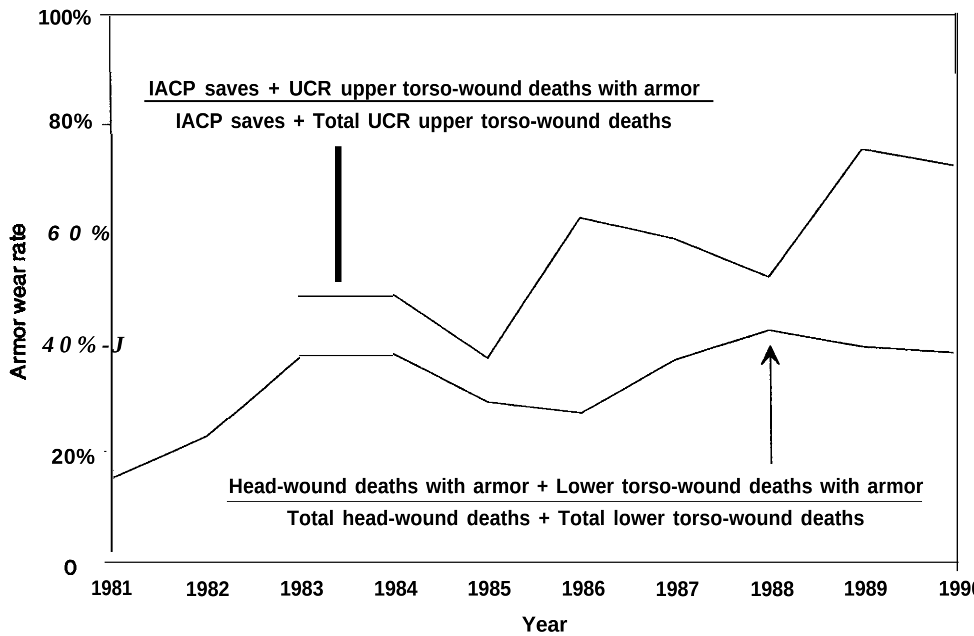
One way to check the validity of the 'saves" data reported by industry is to see what wear rate it implies. Those officers saved were hit on the torso; the FBI reports the number of officers killed by hits on the torso (including some additional armor-wearers), and we may make a second estimate of

14 This estimate is derived is subject to some statistical uncertainty, resting as it does on estimates of K_V and K_N, the probabilities that a torso hit is fatal with and without (respectively) a vest on. In the absence of a break-down, by wound site, of nonfatal hits corresponding to the breakdown of fatal hits provided by the FBI's *Uniform Crime Reports [140], these quantities* must be estimated from the available data (fatal wounds and their sites, and aggregate nonfatal attacks) through the use of various reasonable but not guaranteed assumptions. The principal assumptions are that vest wear acts only to lower the probability that a torso hit will be fatal, does not affect the probability that other hits are fatal, and does not affect the probability that a torso hit occurs in the first place. The resulting K_V and K_N (0.11 and 0.43) seem plausible in light of military studies of wounding and death, although those studies are not strictly comparable because of the different weaponry and projectiles used.

15 Though manufacturers recognize that demsnd for their vests stems from the firearms threat, and supply separate data on firearms saves alone, they report all instances in which body armor arguably saved an officer from death or serious injury. Recent yearly totals amount to over 100 saves per year: one tally records a total of over 1,350 saves to date. About two-thirds of these saves, however, are not of officers attacked with firearms: they include officers involved in serious auto accidents and officers attacked with all manner of other weapons, including knives. Makers of concealable body armor emphasize that their product is not intended to, and cannot be expected to, offer protection against slashing or stabbing weapons. NIJ calls attention to a death from a stab wound incurred in the course of an ill-advised armor demonstration. [145] However, such armor has deflected such attacks in many instances.

16 Even a doctor's statement that death would probably resulted had not the victim been wearing a vest allows fOr some chance that the victim would have lived anyway.

Figure B-8-Armor Wear Rate Estimates Compared



SOURCE: Office of Technology Assessment, 1992.

under-fire wear rate from this figure. It is somewhat higher than the estimate made from the non-upper-torso wounds, which is not surprising in that the industry's estimate of "saves" inevitably includes some officers who would have lived anyway (see figure B-8). In 1990, it is consistent with the 1990 wear-rate survey data.

Saves and Fatal Wounds Per Shooting

Another way of looking at the effect of body armor is to consider the number of saves and fatal wounds per shooting. Using FBI data and either the OTA estimate of saves or that provided by the body armor industry, one may calculate the chances that a shooting incident will result in a save, a fatal shot to the head, a fatal shot to the upper torso, or a fatal shot to the lower torso (see figure B-9).

The saves as estimated by OTA are defined as saves from gunshot wounds that *would have been fatal* and therefore displace fatal upper torso wounds. The saves recorded by DuPont are saves from gunshot wounds that *probably would have been fatal* and therefore more than displace fatal upper torso wounds. A save or a fatality occurs in roughly 10 to 15 percent of shooting incidents, a save or a fatal upper torso wound occurring about 10 percent of the time: years with more saves have correspondingly fewer fatal upper torso wounds. There is no particular indication that widespread use of body armor is leading criminals to adopt a policy of shooting at officers' heads. Indeed, such a policy would probably not be productive, from the criminals' standpoint [even assuming he or she will not later be held to account for the shooting], in that aiming for the head would increase the percentage of shots that miss the target altogether.

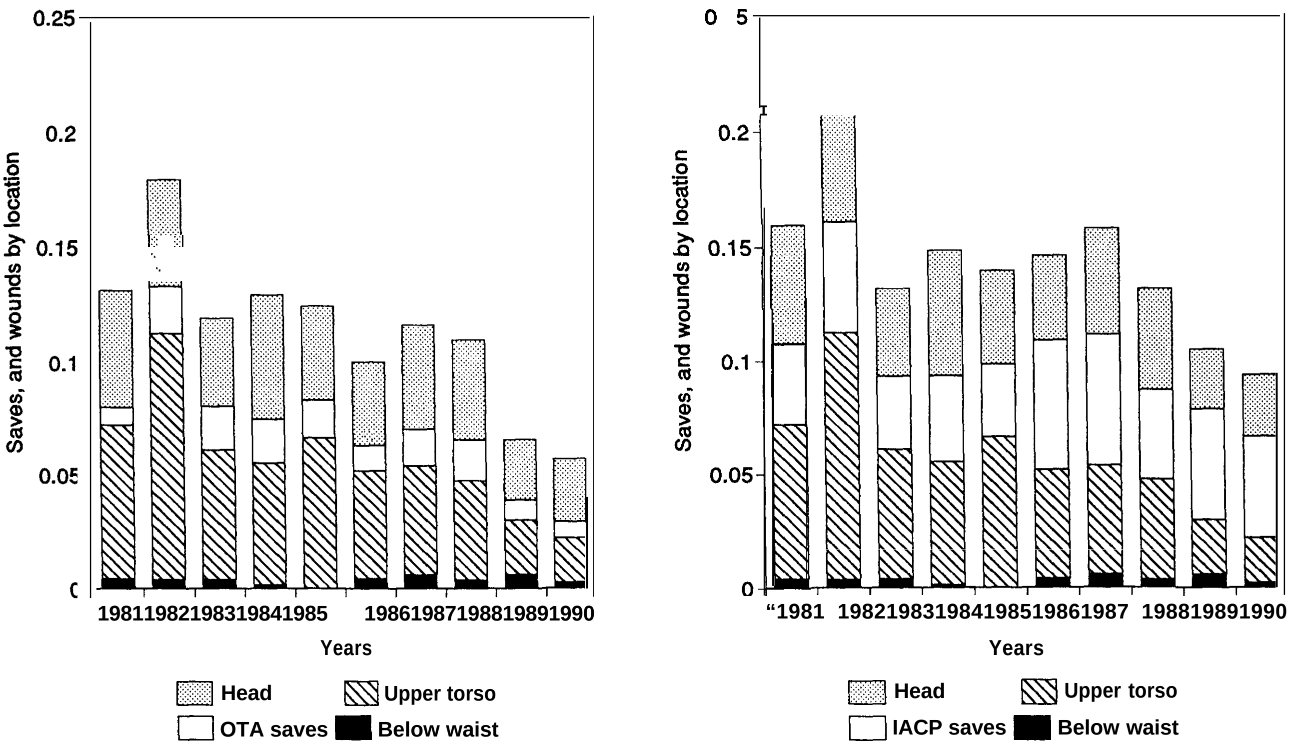
Is Body Armor a Good Buy?

Certainly an officer and his or her family will retrospectively consider a vest to have been a good buy after it has accomplished a save. But is body armor a wise choice for every officer, or for society as a whole? The preceding sections show that body armor costs society $40 million each year. What is the return on this investment, in economic terms?

Currently, the wearing of armor saves 10 to 20 officers per year from death by gunshot wound. It is problematical in principle to estimate the value to society, in monetary terms, of each life saved (or anything else[17]). It is simpler to estimate the cost of each death. [76]

17 See Kenneth Joseph Arrow, *Social Choice and Individual Values* (New York, NY: Wiley, 1963).

Figure B-9-Saves and Fatal Wounds Per Shooting



Left: Saves as **estimated by** OTA, 1992.

SOURCE: Office of Technology Assessment, 1992.

Right: **Saves as** recorded **by** the IACP/DuPont Kevlar Survivors' Club (S.M.), 1991.

The tangible cost to society of a police officer's death may be in the neighborhood of $1 million or more. The average officer killed has about 10 years of service, [140] suggesting that she is about 30 years old at the time of his or her death. A death benefit of $100,000 is paid by the Justice Department. Local jurisdictions may also pay substantial benefits. Many officers leave young widows who receive their husbands' pensions for decades. A woman receiving her late husband's salary of $25,000 per year for 40 years receives a million dollars, though the annuity cost to the department is perhaps half that figure. In addition, some survivors sue departments for damages, alleging wrongful death. [145]

The direct and indirect costs of the memorial service for an officer slain in the line of duty are considerable. They sometimes include a day's pay for officers who attend as an official duty; this alone may exceed a million dollars. For example, the funeral of slain New York Police Department Officer Hector Fontanez was attended by 9,000 officers from as far away as Washington, DC.[18]

The training of a new officer costs another $25 to 50 thousand and produces only a rookie; another 10 years' salary must be paid to produce a seasoned officer with 10 years' experience.

Spending $40 million to save 10 to 20 officers therefore seems like a reasonable choice for society to make purely on the basis of dollars saved, let alone lives saved (see box B-l). In addition, another 20 or so officers escape serious injury (these are the officers logged as saves even though their wounds would not have been fatal-though we can estimate their number statistically we cannot say which vest-wearing victims of shooting they were) and thus avoid thousands of dollars in hospital payments

[18] **Sarah Krulwich, "New York Buries Two** *Killed in the Line* of Duty," New ***York Times,*** Sept. 17, 1991, p. Al, **and Sara Lyall,** "2 **Funerals:** Day of Pain And Anger," New ***York Times,*** Sept. 17, 1991, p. B3.

Box B-l—Spending Money, Saving Lives

The statement of policy issues such as those surrounding police body armor often evokes the response, “No amount of money is too great to spend when lives are at stake--this is what we pay taxes for.” Though no Administration would set an explicit ceiling on the expenditure allowable to save a single human life, two important facts combine to create implicit ceilings:

1. Almost any endeavor to save lives by spending money faces increased costs with each successive life saved.[a] In the case of body armor testing, for example, increased accuracy and reproduceability could always be gained by spending more time and money.
2. Other means of saving lives are competing for the same dollars.

Taken together, these facts lead to a situation in which the further pursuit of a particular life-saving endeavor will cost more per life saved than does some other endeavor. At that point the government would ideally stop trying to save lives the expensive way and shift the unspent dollars over to the program that saves lives the cheap way: more lives will be saved for the same dollars. In this way, competing means of saving lives through government programs create implicit ceilings on the size of any one way of doing so. In practice, the number of lives saved per dollar is difficult to compute, so the suggested calculation is only done in the most approximate of senses.

a Indeed, almost any endeavor to do anything faces increasing costs as it grows, or what economists call “diminishing returns to scale.”

as well as a great deal of pain and suffering. Finally, the wearing of vests saves some officers from death by nonfelonious, nonballistic threats (chiefly automobile accidents)-upwards of 50 officers per year by one count. [16, 17, 18] These calculations suggest that, even in a strict cost-accounting sense that assigns no cost to human suffering, loss of life, or bereavement, the purchase of concealable body armor for police officers is a “good buy” for the officers, the departments, and for society as a whole.

Armor might have been an even better buy than the foregoing analysis indicates, if armor has, or attains, an average service life greater than the 5 years assumed and is properly cared for during its service life. In this case, the annual benefits estimated might be obtained in the future at a lower annual cost than the recent annual cost. A continued decline in the prices of the least expensive models would further reduce the annual cost to society for reaping the current annual benefits.

Although spending $40 million per year saves 10 to ***20*** officers per year from being shot to death, and may save at least as many more from other hazards, doubling the annual expenditure for armor would not double the saves, because most officers in large jurisdictions (including the most dangerous ones) report that they already own armor. [102] Buying each officer two vests would not double the reported ownership rate (nor the reported wear rate), and those who don’t own armor may be those least at risk.

However, if the wear rate is 30 to 40 percent, it could be at least doubled and possibly tripled, in principle. This would not increase saves in proportion, because those who wear armor least may be those least at risk. It is unrealistic to expect, and perhaps unwise to desire, universal wearing of armor,[19] Nevertheless, there is a clear potential for increasing wear rate and, thereby, saves.

[19] Undercover officers, whose deaths account for a considerable fraction of the total, cannot be expected to wear armor; if the armor were detected, it would expose them.

Appendix C

Issues

Contents

Figures

Table

Appendix C
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.

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's risk” and “producer' s risk” 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]

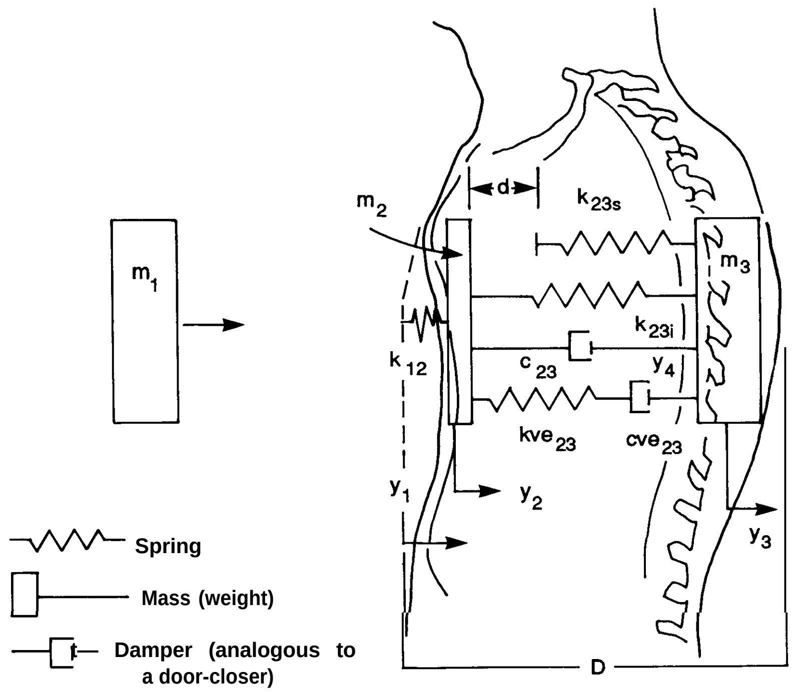
26 Dessert recipes lead to a concentration of about 10 percent.

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



m_2 = ***0.45*** **kg**
m_3 = ***27.20*** **kg**
k_{12} =281 **kN/m**
k_{231} = 26.3 kN/m
kve_{23} = 52.6 kN/m, d = 38.1 mm
23 = 13.2 kN/m
c_{23} = 0.52 kN-s/m, compression
1.23 kN-s/m, extension
cve_{23} = 0.18 kN-s/m

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]

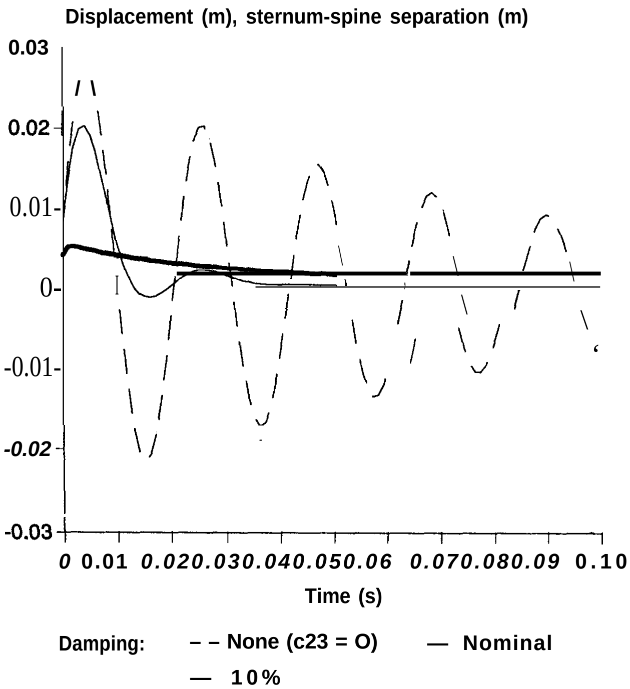
Strictures against patting the vest down have become stronger with each successive edition of the NIJ standard. (See appendix A, Origins of and Rationale for the NIJ Standard.) This issue could be revisited yet again, especially if compelling evidence of the self-patting effect were developed.

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.

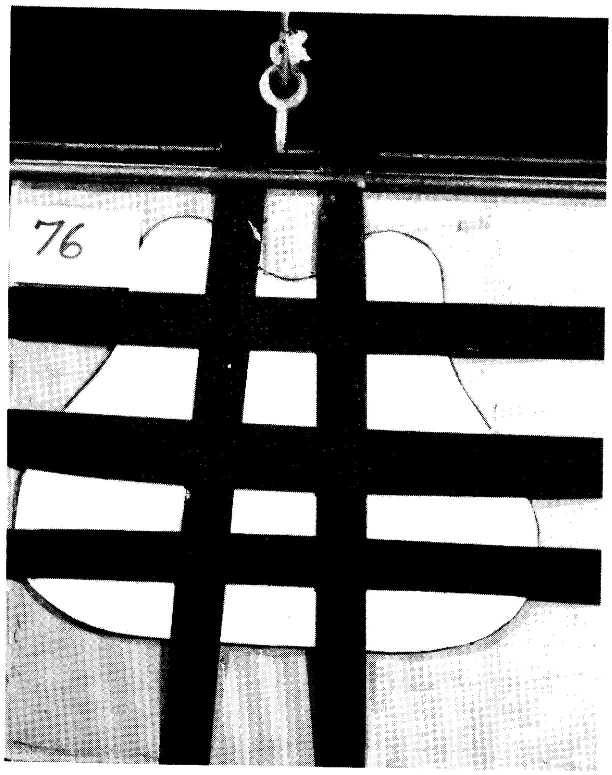
The mannequin test could be further refined by suspending the mamequin as if in a swing, rather than firmly anchoring it to the floor as is generally done with the clay block in compliance testing at H.P. White Laboratory, Inc. (although neither anchoring nor use of a frame for the backing material is 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]

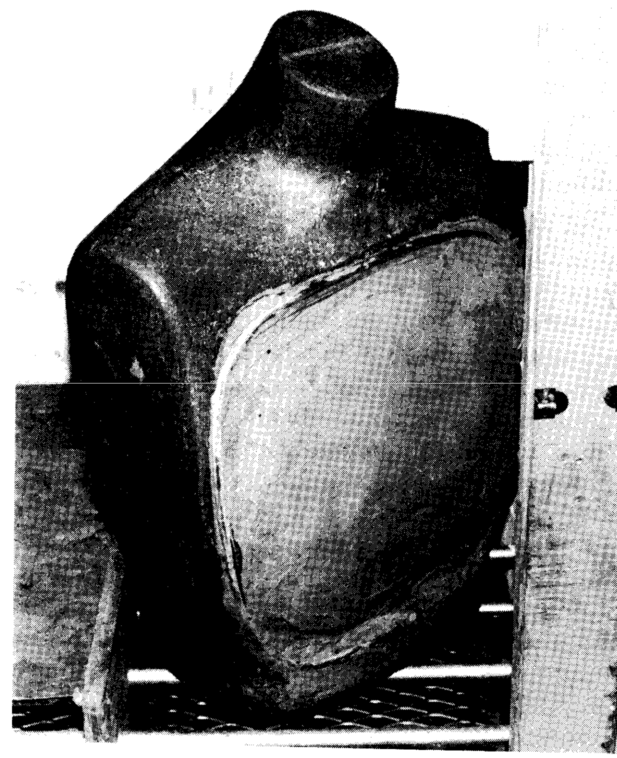
Using the flat block, one panel of a vest is tested and then it is replaced with the other panel. The use of a vest-wearing mannequin without provision for patting down or adjusting the vest between shots would raise the question of whether the vest could be adjusted between the test of the front panel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

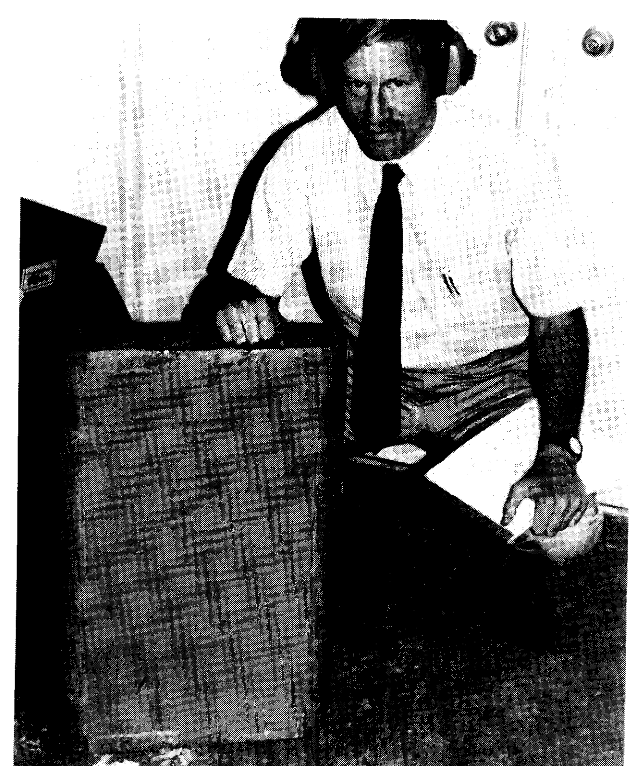
A compromise test fixture could consist of a flat block of clay contained in a fixture to which the vest could be attached with its own straps-such a fixture is termed a "curv.' The NIJ found the curv to be

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[R] 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]

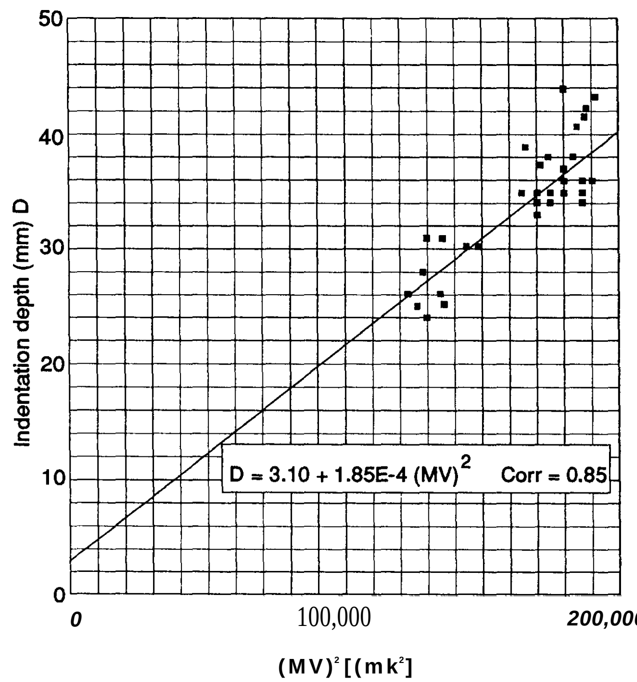
One must assume that the real-life variables artificially held constant in the test-range, angle, the chance that one shot could land near where an earlier one had, etc.—are subsumed into the 10 percent, along with the fact that real ammunition may be shot at velocities different from those specified 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

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.

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.

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.

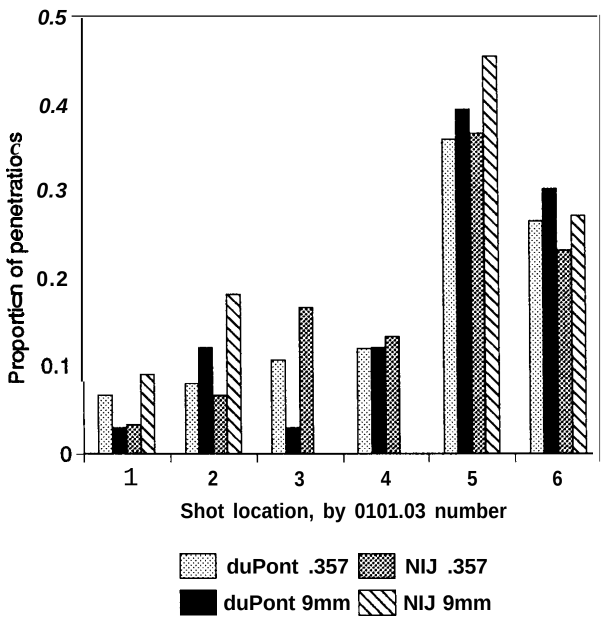
49 Because $(1-0.026)^6 = 0.85 = 1-0.15$.

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; anuntreatedgarrnentof the style that absorbed 22 percent perspiration in the FBI Academy testabsorbed26.2 percent water in an Army test using a copper mannequin. Even this may not represent complete saturation, but OTAknows 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 °F, but only about 7.5 percent as the temperature increased from 80 to 160 °F. [106]

In contrast, the extended-chain polyethylene (ECPE) plastic from which Spectra™ 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 1-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.)

V50 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_{10}, it is best to adjust the velocities used in the test to be near what one expects the V_{10} to be. A procedure analogous to the DOD V_{50} procedure could be developed for finding the V_{10}.

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

68 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 The 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.

Appendix D

Reenactments

Contents

Boxes

Figures

Tables

Appendix D
Reenactments

INTRODUCTION

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

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

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

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

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

Why Reenact?

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

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

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

Reenact What?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

SELECTION OF CASES FOR REENACTMENT

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

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

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

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

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

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

5 We will consider various degrees of seriousness.

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

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

Box D-1-Categories of Trauma and Incapacitation

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

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

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

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

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

---

a Cf. reference [1411.

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

---

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

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

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

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

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

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

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

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

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

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

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

METHOD OF ANALYSIS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

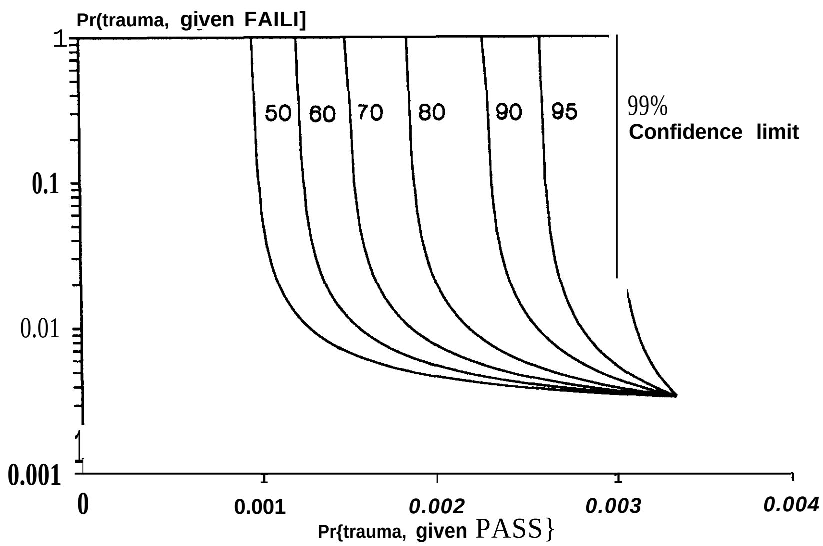
As an example, figure D-1 shows the upper 50-, 60-, 70-, 80-, 90-, 95-, and 99-percent confidence limits on $p_{2(0)}'$ ("Pr{trauma, given PASS}") for a range of values of $p_{2(1)}$(' 'Pr{trauma, given FAIL} "), for the case $n_2 = 2$ and $n_{2(0)} = O$.[10]

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

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

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

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



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

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

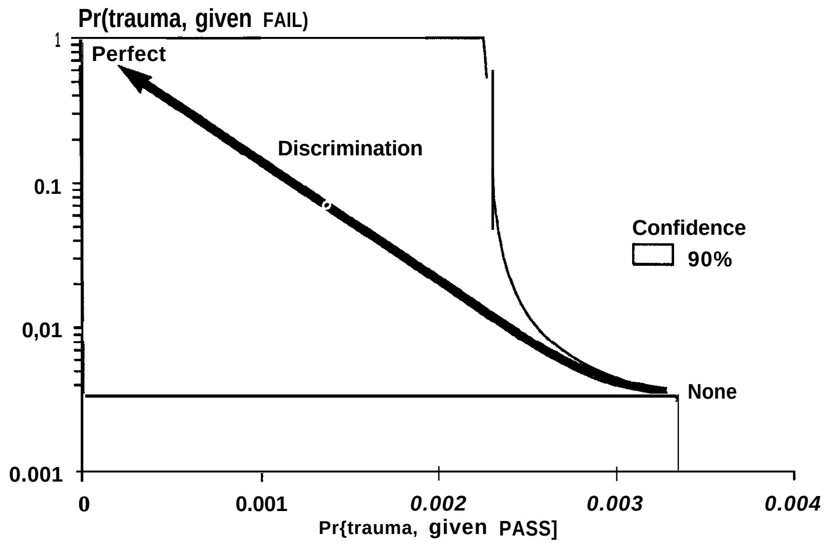
In some cases, separate-sample logistic discrimination may be used to estimate the probability that a stopped shot would cause injury, as a function of the backface signature measured in a l-shot test. The procedure is more complicated and may not always be applicable, but if it is, it allows the estimated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

RESULTS OF DUPONT-SPONSORED REENACTMENTS

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

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

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



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

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

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

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

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

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

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

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

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

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

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

NOTE: The total number of shots fired to reenact each felonious shot differs from case to case. This table lists all shots except for seven shots fired to reenact the assault on Mr. Joyner, which OTA estimates did not have an impact velocity comparable to that in the assault. (See fn. 11.)

a Abbreviated Injury scale:
- 6: fatal
- 5: critical-survival uncertain
- 4: severe, life-threatening-survival probable
- 3: severe, not life-threatening
- 0-2: not severe

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

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

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

Table D-3-Summary of Results

AIS	Shots	BFS: 0-44 mm	BFS: 44+ mm
All results			
6	2	0	2
4-5	6	0	6
3	4	0	4
0-2	111	69	42
Results used for analysis			
AIS	**Reenactments**	**BFS test result: Pass**	**BFS test result: Fail**
6	1	0	1
4-5	1	0	1
3	1	0	1

NOTE: Injuries requiring only skin sutures are rated AIS 0-2. "44+ mm BFS" includes penetrations. Each reenactment consists of 2 shots.

SOURCE: Office of Technology Assessment, 1992.

ANALYSES

Risk Associated With the Current 44-mm BFS Limit

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

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

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

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

Box D-2-Penetrations in BFS Testing

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

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

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

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

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

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

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

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

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

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

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

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

Box D-3—"Magnum Saves"

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

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

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

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

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

Logistic Model for Probability of Injury versus BFS

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

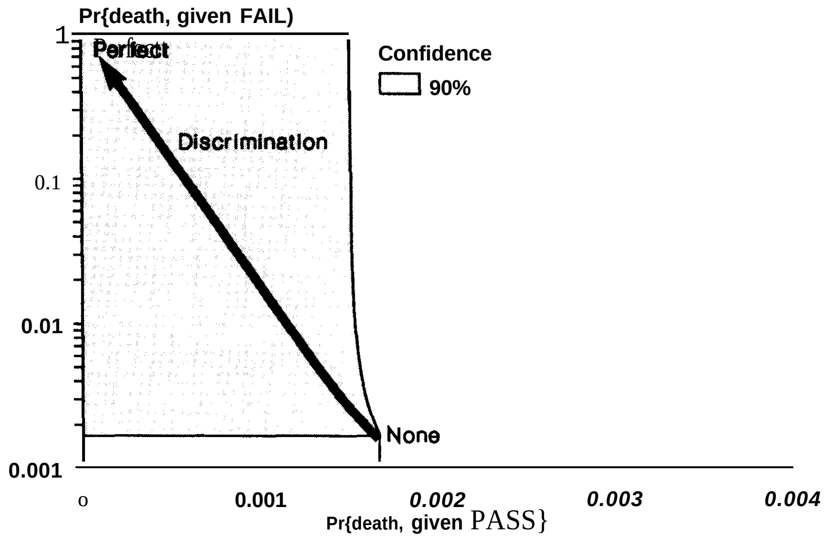
It maybe more convenient to use separate-sample logistic discrimination to obtain a logistic model that estimates the probability of injury associated with any BFS. (This would be called separate-sample logistic regression.) The model would befitted to the results (BFS or penetration) of l-shot reenactments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

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

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

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

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



SOURCE: Office of Technology Assessment, 1992.

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

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

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

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

Sensitivity Analysis

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

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

Box D-4-Control for What?

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

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

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

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

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

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

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

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

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

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

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

Appendix E

Options for the Department of Justice

Contents

Box

Figures

Table

Appendix E
Options for the Department of Justice

GENERAL

This appendix describes and assesses several options that the Department of Justice could exercise to revise NIJ Standard 0101.03 and/or the process by which compliance with it is certified, in order to

- limit the variance in test conditions,
- provide more information on ballistic resistance of certified armor (including uncertainties and limits of ballistic resistance, dependence on wearer, etc.),
- decrease producers' financial risks as well as consumers' safety risks, and
- assure consumers that certified armor offered for sale is as good as the samples tested for certification.

Some of the options could be undertaken by the National Institute of Justice (NIJ) without additional authority or funding. Others-research and quality-assurance programs—would require substantially increased funding.

Status Quo

One option is to postpone any change to NIJ Std. 0101.03 and the current method of certifying compliance with it. The argument for this is that armor of styles certified to comply with NIJ Std. 0101.03 has saved many lives (see app. B) and is not known to have failed, in actual assaults, to stop any bullet of a type that it was certified to resist, nor to prevent lethal blunt trauma. Yet the criterion for protection from blunt trauma is not so strict that many models fail it: as of Oct. 31, 1991, of the 555 models submitted for testing for NIJ certification of compliance with the .03 standard, only 15 failed solely because of excessive backface signature (BFS), the test's index of risk of blunt trauma.

The vast majority of the failures were caused by penetration, alone (166) or in combination with excessive BFS (40). Most of the dissatisfaction of some parties with the current standard stems from these failures, or from penetrations in retests. Complaints charge that the test is "a crap shoot" (i.e., not reproducible) or too stringent. These and other arguments against the status quo were summarized in appendices A and B.

Arguments for the alternative options discussed in the remainder of this appendix are also arguments against the status quo.

Make Conventional Practices Mandatory

On several occasions since NIJ Std. 0101.03 was issued, NIJ has instructed H.P. White Laboratory, Inc., (HPWLI) in letters, telephone calls, or meetings, to perform certain test procedures in certain ways consistent with the standard. In effect, these instructions rule out other ways of performing test procedures that could reasonably be considered consistent with the printed standard. Sometimes this was done to clarify a portion of the standard; in other cases it was done with the intent of reducing variability of results that might be attributable to variability of test procedures. For example, in 1988 an official at NIST directed that the test facility use only 124-grain, FMJ 9-mm bullets made by Remington. [82]

On other occasions, HPWLI has informed NIJ that, unless instructed otherwise, it would henceforth perform certain test procedures only in certain ways but not in other ways consistent with the printed standard. Again the intent was to reduce variability. Sometimes NIJ would indicate its concurrence; sometimes NIJ would object, proposing a different procedure. For example, on March 28, 1988, HPWLI informed NTJ-in response to a modification made earlier in the month by TAPIC that the locations of shots 4 and 5 be altered slightly to ensure nonalignment with each other and the new location of shot 6-that shot 5 be raised 1 inch and shot 4 left unchanged. In May of the same year, TAPIC responded with a letter approving the new shot locations. [82]

On still other occasions, HPWLI has proposed to change certain test procedures in a way that actually departs somewhat from those specified in the printed standard, but is clearly justifiable on technical grounds. Such proposed changes are not implemented until approval is received. For example, on October 10, 1989, HPWLI proposed that the 30-degree obliquity of the fourth and fifth shots be rotated so as to be combination of horizontal and vertical obliquity, as opposed to the present situation in which all shots lie in a horizontal plane with respect to the vertical vest. [82]

In at least one instance, NIJ has presented a major procedural change-the replacement of the flat-faced block of clay with a curved, abstractly torso-like fixture (containing a smaller flat-faced block of clay) on which the vest is mounted by its own straps as if worn by an officer-as a possible modification to the 0101.03 standard. This possible change highlights issues always present, albeit perhaps to a lesser degree, when the test procedure is changed:

1. Does the change make the test harder or easier to pass? Either way, vests already tested might experience a different outcome if tested again. Manufacturers of vests that failed the earlier test will want a repeat opportunity, while those whose vests passed will seek to avoid further testing.
2. Does the change confer a particular advantage on certain manufacturers?
3. What artificialities have been introduced? While it would be naive to suppose that any test or test procedure could avoid all artificialities, it is wise to consider these artificialties before they are introduced. In the case of the curvilinear test fixture, one might well ask what will happen to a vest if its straps give way during the test. Will it be picked up off the floor and reattached? If so, vest manufacturers will strive for the most tenuous possible attachment so that their vests can be picked up and smoothed out as many times as possible, reducing or even eliminating bunching and balling. If not, does the vest fail if its straps come undone? What if one strap breaks and the vest droops, obscuring the next shot's line of fire? What if an unfair shot penetrates the vest and hits the opposite panel, arguably weakening it?

The underlying point is that procedural changes have become de facto parts of the standard. NIJ should consider incorporating them into the next version of NIJ Std. 0101. Of course, some of these instructions and practices may become obsolete if the current standard is changed in other respects. It would be especially important to incorporate the applicable instructions and practices into the standard if NIJ should authorize a different laboratory to test armor for certification (or quality assurance).

Specify Backing Material

A simple but possibly helpful change would be to specify the backing material to be used. In practice, only one backing material, Roma Plastilina No. 1 modeling clay, is used by HPWLI for NIJ certification tests. However, NIJ Standard 0101.03 does not require it; it *defines* "backing material" as "a block of nonhardening, oil-base modeling clay placed in contact with the back of the test specimen during ballistic testing." This is confusing, because a variety of materials other than modeling clay are often used as backing in tests for other purposes than NIJ certification. Examples include 10-percent ballistic gelatin, 20-percent ballistic gelatin, rigid foamed polystyrene (Styrofoam), foamed polyurethane rubber, RTV silicone rubber, soap, plywood, human and animal cadavers, and live animals. Of these, only Styrofoam and soap are sufficiently inelastic for use for deformation measurement in an NIJ-like test (i.e., without high-speed cinematography or other expensive techniques).

The definition is also confusing because, although clay is *placed in* contact with the back of the test specimen at the *beginning* of ballistic testing according to NIJ Standard 0101.03, the standard prohibits ''disturbing the relationship between the armor and the backing material' to assure that the clay *remains in* contact with the back of the test specimen *during* ballistic testing (or for any other purpose). Amending the definition of backing material in section 3 (Definitions) of the standard would improve clarity, whether or not a particular backing material is specified in section 4 (Requirements) or section 5 (Test Methods).

Laboratories in England, France, and Germany have used other types of modeling clay as backing material and found that deformation is affected by choice of material. For example, researchers in England have calibrated deformation in Plastilina to deformations in Plasticize and PP2 as a function of bullet velocity. In these comparisons all three backings were conditioned so as to pass the drop test specified in NIJ Std. 0101.03. This required heating Plasticize to temperatures higher than the maximum allowed by NIJ Std. 0101.03. [28, 29, 84] As noted above, some experts consider backing temperature unimportant provided the drop test is satisfied. However, strict adherence to all provisions of NIJ Std. 0101.03, including allowable temperature, would exclude use of Plasticize and perhaps some other backings sometimes used. This has not been an issue in NIJ certification testing; H.P. White Laboratory uses only Roma Plastilina No. 1.

Even if different backing materials can pass the drop test at temperatures within the allowed range, specifying only one of them might improve reproducibility. It is possible that the consistency (flowability) of candidate backing materials might depend strongly, but differently, on the rate of deformation.[1]

Some backing materials conditioned to produce comparable drop-test results yield different backface signatures at the much higher deformation velocities typical of a ballistic test conducted in accordance with NIJ Std. 0101.03. For example, in tests conducted by the British Police Scientific Development Branch, under otherwise similar conditions the average (viz., fitted) backface signatures produced in U.S.-made Plastilina and U.K.-made Plasticize were similar at impact velocities of 350 m/s but differed by about 4.4 mm for each 100 m/s above or below 350 m/s. [29; cf. 28] Thus, 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. Some materials are known to yield different results; others, not yet tested by NIJ or NIST, could diner more dramatically. Specification of a backing material would eliminate this potential source of variation in-or operator influence on—test conditions.

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.

Reduce Allowable Range of Backing Material Temperature

One way to reduce or at least limit the variability of test conditions is to reduce the range of acceptable temperatures of the backing material. Currently, the clay's temperature can be anywhere between 15 and 30 'C, i.e. 59 and 86 'F. Tightening this tolerance up, however, might make little real difference because the backing material must also pass a drop test, in which a special weight is dropped 2 meters and the resulting dent must be between 22 and 28 millimeters in depth. Some experts consider backing temperature unimportant provided the drop test is satisfied. [69, 29] The standard does not require use of Roma Plastilina No. 1, but does point out that this nonhardening modeling clay fulfills the requirements of the test.

Research by the Aerospace Corp. indicated that the volume (especially) and surface area of the crater produced in Roma Plastilina No. 1 by the drop test is very sensitive to temperature, and the Aerospace Corp. recommended that the temperature of this backing material be maintained in the range 68 to 72 OF. [8] The Aerospace Corp. calculated crater volume and surface area from depth and diameter measurements, assuming the crater to be a right circular cone. Using the same approximation, OTA has reconstructed the unrecorded depth and diameter measurements and found that crater depth is less sensitive to temperature than is crater volume (see figure E-1).[2]

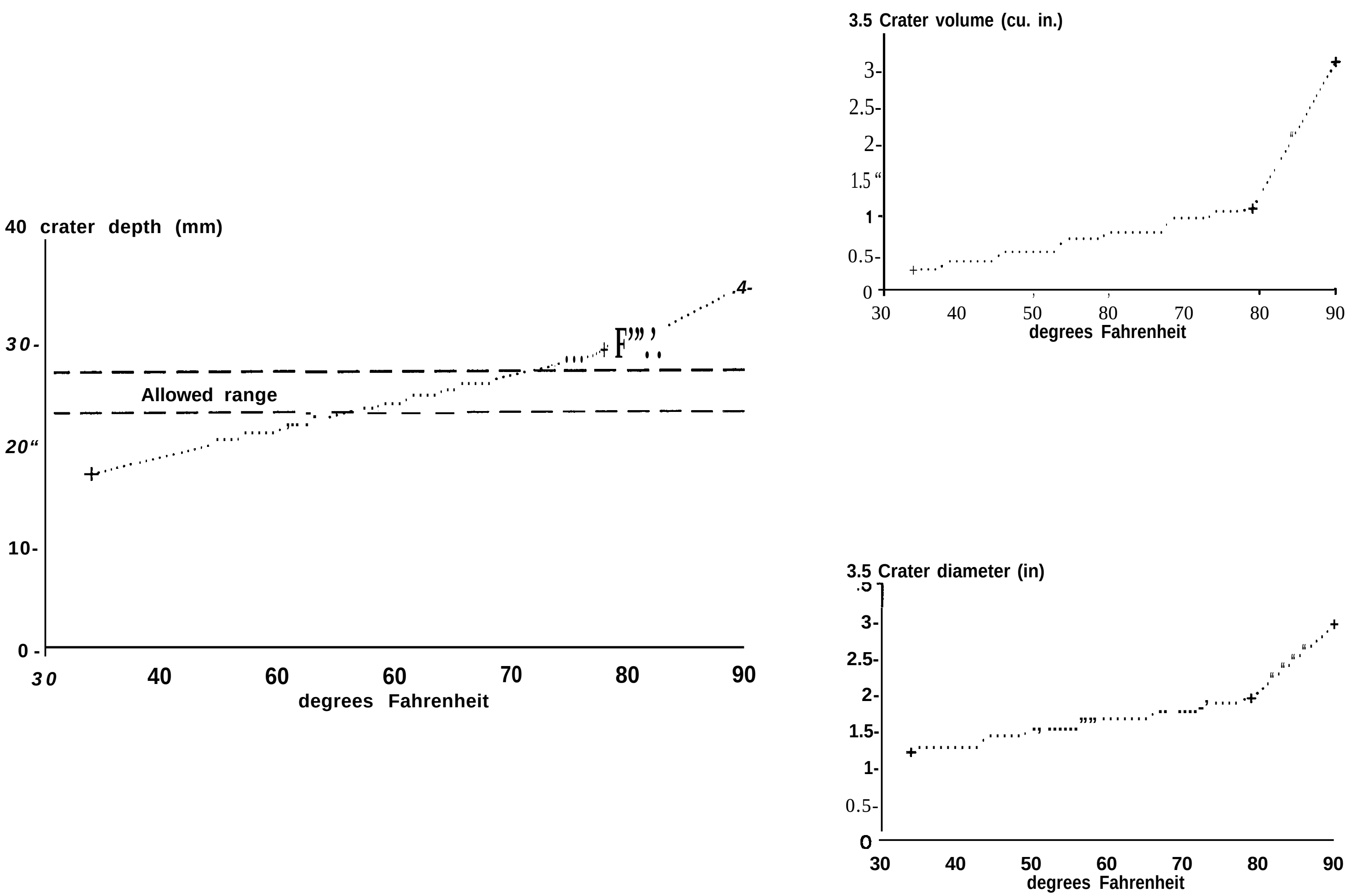
The drop test, if performed at the beginning and at the end of a test, would standardize the consistency to some extent, but it is doubtful that it is an adequate substitute for temperature control. For example, if the clay block were left for many hours in an area colder than 59 'F, then brought into an area maintained at 59 'F and kept there for 3 hours, the surface of the clay block might warm enough so that the drop test could be passed, indenting the clay only about 25 mm. But in subsequent testing, a shot might push the armor into a deeper, colder, stiffer layer of the clay-e.g., to the BFS limit. Were the clay at that depth warmer, as required by the standard, the BFS test would be failed. But in practice, in testing observed by OTA, clay temperature is not measured during testing, nor is the drop test performed after the beginning of a test.[3] Thus in practice temperature may not be controlled to within specified tolerances, which would allow considerable inadvertent or operator-controlled variation in test conditions.

1 The consistency of Silly Putty®, a familiar toy item, illustrates strong strain-rate-dependence.

2 Damon the temperature sensitivity of Plasticize are shown in [281.

3 In testing observed by OTA at H.P. White Laboratory, Inc., clay was conditioned and ambient temperature was maintained within the tolerances allowed by NIJ Standard 0101.03. Moreover, clay was routinely stored at the temperature used for conditioning, even before the 3-hour conditioning period prescribed by the standard. The conditioning temperature was warmer than the ambient temperature, and the face of the clay block cooled during testing. To recondition the block as prescribed by the standard, warmer, softer clay was taken from storage and used to fill the craters made by previous shots in the test sequence.

Figure E-l—Variation of Drop-Test Crater Dimensions with Temperature



SOURCE: Office of Technology Assessment, 1992.

Specifying that backing temperature be measured at several depths and locations and that the drop test be performed both before and after (and perhaps during) ballistic testing would insure that backing temperature is controlled to the current standard, and reducing the allowed temperature range further (e.g., to68to72‘F) would further improve control of test conditions and possibly the reproducibility of test results.

Another reason for doubting that the drop test is an adequate substitute for temperature control is the fact that deformation depends in a nonlinear way on the momentum of the dropweight or, in testing, the bullet. [8, 122, 123] These are quite different: the l-kilogram dropweight has a calculated momentum of 0.64 kg-m/son impact; but an 8-gram 9-mm bullet at 332 m/s (the nominal type II-A velocity) would have a calculated momentum of 2.7 kg-m/s. Deformation also varies nonlinearly with temperature, as shown in figure E-1, so the variation with (i.e., the sensitivity to) temperature at the momenta of bullets probably differs from that at the momentum of the dropweight on impact. However, we have no data characterizing the sensitivity to temperature at the momenta of bullets.

Although the drop test was developed to test the consistency of backing material for the purpose of standardizing the deformation test (for protection from nonpenetrating bullets), variation of consistency such as that shown in figure E-1 may also affect the results of the penetration test. Research would be needed to find out whether it does.

Certify Wet and Dry Ballistic Resistance Separately

The wet test could be mandatory or optional. The case for certifying dry ballistic resistance even if armor does not have, or is not tested for, wet ballistic resistance is that because of cost or comfort, many purchasers and wearers prefer armor with inadequate or untested wet ballistic resistance. They may suspect that the risk of its becoming dangerously wet is so low that they would accept it.

However, to learn what the risk is, they would have to weigh their armor regularly to measure and record water retention and analyze the records to calculate frequency with which retention exceeds dangerous levels. There is a risk that some may err in this, or not attempt it.

Even if it is done correctly, so that purchasers and wearers make an informed choice to accept the risk, it will be a higher risk than they would be exposed to if they bought *and wore* wet-certified armor. But in compensation, wear rate might be increased among those who find armor with inadequate wet ballistic resistance more affordable or comfortable but who also value NIJ’s certification.

Officers could weigh their armor panels at the beginning and end of each shift to measure moisture pickup, which they could record. However, this would indicate moisture *content,* which affects ballistic resistance, only if the armor were completely dry at the beginning of the shift. Some officers complain (to us) that their armor does not dry completely between shifts. Some officers may require two or more garments each in order to have a dry garment to wear while others are drying.

Even if officers measure and record the wetness of their armor, predicting the risk of future wetness and the uncertainty in the risk would be complicated, beyond the abilities of most officers and many departments. Aids in the form of worksheets or computer software would be required, along with training. The frequency with which dangerous wetness has occurred in the past is a reasonable (viz., a maximum-likelihood) estimate of the risk of dangerous wetness in the future, under similar conditions (e.g., season and duty). However, because the occurrence of dangerous wetness is apparently rare, there would be a substantial chance that the estimated risk would be inaccurate. To assess this risk, purchasers or wearers would have to calculate confidence limits on the estimated risk.

Subjecting armor only to the dry testing specified in the NIJ standard would reduce the stringency of the test, even for armor that performs as well wet as dry. For example, armor that is unaffected by moisture and has a 97-percent mean probability of stopping a bullet would have a 70 percent probability of passing a 12-shot dry test and would probably pass it; but if subjected to a wet-dry test (or a double dry test) of 24 shots, the same armor would more likely than not have failed (52 percent probability). If NIJ wished to compensate for this and maintain the stringency of the test, it could offer a choice of the current wet-dry test or a double-dry test with the same number of fair shots required.

To halve the cost of testing, one industry source has proposed testing and certifying dry ballistic resistance *or* wet ballistic resistance, but not requiring both tests. This is based on the premise that no conceivable type of armor has less ballistic resistance when dry than when wet. This is plausible, but even if true, armor would have a higher probability of passing a wet-only test than a wet-dry test with twice as many shots.

ASSESSING RESISTANCE TO PENETRATION

Smooth Armor Between Shots

[This topic was discussed in vol. 1.]

Use a Torso-Shaped Test Fixture

Appendix C notes that one of the technical issues surrounding the .03 standard is its requirement that armor be tested by removing its ballistic panels, strapping each to a flat block of clay, and shooting. This deprives the armor, in such testing, of any benefit (e.g. against bunching) it might derive from its own strapping or the carrier garment itself. A torso-shaped test fixture, be it a mannequin or a ‘‘curV,’ ‘ would lessen or eliminate these problems.

Use *Resilient Backing for Penetration Test*

NILECJ-STD-0101.00 , issued in 1972, specified the use of ‘a block of nonhardening modeling clay” as backing for the ballistic deformation test it described but *not* for the ballistic penetration test, which was to be air-backed. Three reasons were later given for the choice of air backing:

> First, excluding the backing material greatly simplifies the . . . projectile-fabric interaction; not only is the overall experimental scatter [variation in results] reduced, but the test results may be directly related to projectile-fabric interaction [alone].
>
> Second, exit velocities of the projectiles were desired; . . .
>
> Last, high-speed photography is much simpler without a backing material. [7]

However, the frost advantage cited was offset by the fact that there was little data relating air-backed test results to the projectile-fabric interaction on a torso, human or otherwise. Moreover, high-speed photography and measurement of exit velocities, although useful in research, are unnecessary in a test of resistance to penetration, and indeed NILECJ-STD-0101.00 did not require them. Accordingly,NILECJ-STD-O1O1.O1, which was issued in 1978, specified the use of a nonresilient backing material for testing both deformation and penetration. Like the current NIJ standard, it noted that Roma Plastilina No. 1 modeling clay was “found to be suitable” as a backing material but did not require its use, although it did specify a drop test to be performed to check the consistency of backing material.

As noted in appendix C, some critics of the current NIJ standard contend that the best technical option would be to use an inelastic backing such as clay for the blunt trauma test and an elastic backing for the penetration test.[4]

Other ballistic measurement techniques using costly apparatus might be adapted to measure deformation of resilient backing. Examples include multiflash photography, which has been used to measure deformation versus time in air backing; [39] multiflash x-radiography, which has been used to measure penetration (hence deformation) versus time in composite armor;[5] and Doppler radar, which has been used to measure velocity versus range of small projectiles impacting and penetrating media transparent to microwaves.[6] As of late 1990, the range resolution of the radar was 6.25 cm—too coarse to measure backface deformations with the accuracy needed for predicting blunt trauma. A planned improvement in signal processing was expected to improve (decrease) the range resolution tenfold, to 0.625 cm—still too coarse. Higher frequency, and costlier, millimeter-wave radar would probably be needed to provide the range resolution needed for predicting blunt trauma. Such apparatus could conceivably be afforded and used by a major ballistic testing facility such as H.P. White. However, specification of a backing that would require their use would void a major objective of the NIJ test procedure-to be reproducible at ballistic facilities typical of those used by many police departments, with no equipment more costly than a ballistic chronograph.

---

4 Dr. Martin Fackler proposed this at the NIJ Body Armor Users Workshop in Reston, Virginia, on June 6, 1990: ‘‘Maybe we need the springiness for the repeated testing, for the repeated shots; and for the backface deformation the clay. Maybe we need both of them.” See transcript p. 244, ll. 14-17.

5 M.S. Stephenson “A Flash X-Ray Study of the Penetration of Ceramic Faced Composite Armours,” pp. 143-159 in [134].

6 J.L.M.J. van Bree and E.J.M. van Riet, ‘‘Use of a Doppler Radar for Velocity Monitoring of Small-Calibre Projectiles,” pp. 261-269 in [134].

Standardize Test Bullets

The probability with which a commercially available bullet of specified mass and caliber will penetrate armor at a specified velocity depends sensitively on details of the bullet's construction and composition, which determine the hardness of the bullet and, more generally, its tendency to deform or fragment when impacting on armor. [28] A bullet that deforms may be stopped by relatively few layers of armor; many more layers may be needed to stop sharp fragments of a hard or steel-jacketed bullet.

Uncommon projectiles, ranging from the Teflon™ Thunderzap [121] to fragment-simulating projectiles, with a variety of so-called "cop-killer bullets" in between, span a greater range of penetration probabilities. This wide array of threats has led the PPAA [113] and the U.K. Home Office [28, 29] to specify test bullets more specifically than does the NIJ standard. In fact, even nominally identical bullets display considerable variation, sometimes even between different bullets in the same box of 50. Some years ago, the U.K. Home Office, noticing the variation in performance of 9-mm bullets of similar mass and velocity, purchased a large lot of one type of 9-mm round and has used it exclusively for the past 10 years, [29] even though variations in it have been noted since 1983. [28]

NIJ could follow this example and specify test bullets more strictly. This would probably increase reproducibility of test results, but it would decrease realism-it would not simulate the diversity of the threat faced by police officers.

Require a Full-Auto Test

According to a major survey, [102] police officers and chiefs are very interested in securing protection from automatic weapons, increasing numbers of which have been confiscated in recent years. However, to date they have been used in a very, very, small fraction of assaults on police officers, and in most of these no more than a very few shots hit the region covered by any one armor panel. As a risk to police officers, such assaults rank far below many others-head shots, for example. Nevertheless, assessment of ballistic resistance to automatic fire may be demanded. Providing it will require special equipment and will be costly.

One argument for the need for such a testis that in an actual assault with an automatic weapon, bunching and balling (ply separation) might occur and, if it does, might be patted down from *inside the armor* by "the dynamic, elastic human torso." However, this abdominal or thoracic undulation might not smooth the armor as completely as manual patting on clay backing would. One approach to assessing armor under such conditions would be to mount it on a resilient backing and expose it to automatic fire in a manner considered to be representative.

Before undertaking such an effort, one should critically examine the plausibility of the postulated biomechanical dynamics, an issue discussed in appendix C.

The Police Scientific Development Branch of the U.K. Home Office has developed a test fixture to expose armor to automatic fire in a predetermined pattern but has had difficulty achieving a reproducible shot pattern. [29]

Require a Ballistic Limit Test

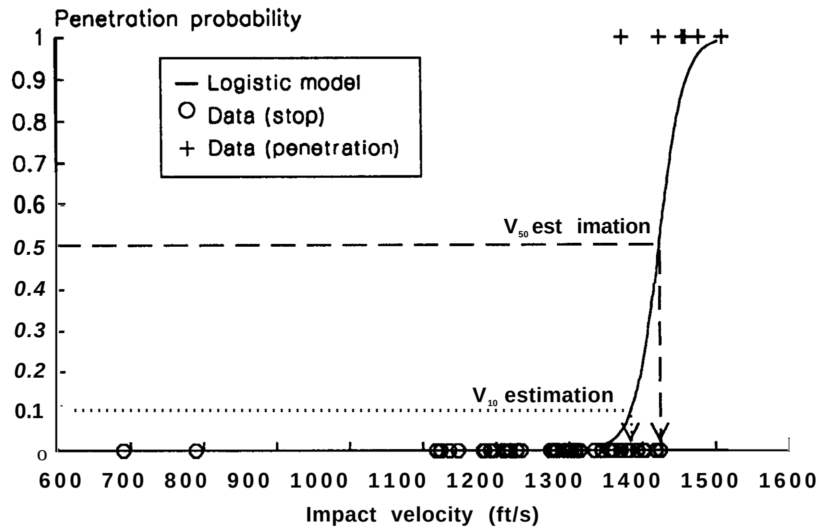
Armor could be subjected to a test to estimate its V_{50} ballistic limit-the velocity at which it has a 50-percent chance of being penetrated by the test projectile. [7] A model could be certified to have a specified type or level of ballistic resistance if the V_{50} estimated for each type of test bullet equals or exceeds a specified minimum value, and if samples also pass a test for protection from blunt trauma. But in addition, the model would be rated by the V_{50} estimate to let purchasers know the margin by which the model exceeds minimum NIJ standards.

The widely used test specified by the Department of Defense's Military Standard MIL-STD-662D [138] could be used. It uses air as the backing material (as did NILECJ 0101.00), but NIJ could specify that clay or some other backing material be used. Regardless of the material used, calibration of penetration probability in the test with penetration probability in assaults would be an issue.

An alternative score is the V_{10}, the estimated velocity at which test bullets have a 10-percent chance of penetrating-i.e., at which the armor stops a bullet with 90-percent reliability. The V_{10} could be estimated by logistic regression [91] using the

7 The DoD test uses air as the backing material, but the NIJ could specify that clay or some other backing material be used. Regardless of the material used, calibration of penetration probability in the test with penetration probability in assaults would be an issue.

Figure E-2—Estimates of V_{50} and V_{10} Obtained by Logistic Regression



.357 magnum at 30-deg incidence v. test panel on clay, smoothed between shots

SOURCE: Office of Technology Assessment, 1992.

results of a DOD-like test. (See figure E-2.) For purchasers who demand 90-percent, rather than 50-percent, reliability in stopping, the V_{10} would be more appropriate for comparing to typical or conservative threat velocities (e.g., the minimum velocity specified for bullets in the .03 test) than would the V_{50}.. By the same token, certification could be based on the estimated V_{05} or V_{01}, but estimating these velocities, which correspond to small probabilities of penetration, would require more shots than to estimate the V_{10} with the same accuracy, which in turn would require more shots than to estimate the V_{50}.

Increase Total Shots and Allow Penetrations

If a very large number of apparently identical armors of the same model and style are subjected to apparently identical tests as specified by NIJ Std. 0101.03, some of the armors would pass and some would fail. Some of the variation in test results might be caused by subtle variations in the armors; another component of the variation might be caused by slight variations in procedure from one test to another. Some of the variation in test results would remain unexplained at any stage in the scientific understanding of the process. Some of the variation—perhaps a small fraction-would be caused by fundamentally random quantum-mechanical processes.

Revising the standard or using a different one could alter-increase or decrease-the variation in test results. However, there will always be a random influence on test outcomes. As a result, an armor of a model that had passed 99 development tests conducted in accordance with NIJ Std. 0101.03 could fail the one NIJ certification test it is allowed. It is likewise possible for an armor of a model not subjected to development tests conducted in accordance with NIJ Std. 0101.03 to pass an NIJ certification and subsequently fail 99 acceptance tests or quality-assurance tests conducted in accordance with the standard.

Clearly these possibilities pose risks-different kinds, to be sure-to manufacturers, purchasers, wearers, and standard-setting authorities. Manufacturers want assurance that good armor does not fail certification testing because of chance variation (‘‘a crap shoot’ ‘), and purchasers and wearers want assurance that bad armor is not certified by a fluke. Certification of bad armor poses a safety risk to wearers as well as a liability risk to manufacturers and departmental purchasers. Any indication that good armor has flunked or that bad armor has been certified, even if not statistically significant, may provoke a challenge to the credibility of the testing and certifiication procedure.

There is a way to decrease the probability of certifying bad armor while at the same time decreasing the probability of flunking good armor. Reducing the consumers' risk requires more testing-g., repetitions of the protocol specified in NIJ Std. 0101.03. Extra testing will, of course, increase cost. Reducing the producer's risk requires allowing some penetrations. The following illustration of tradeoffs between several options is modeled after an analysis Keith Eberhardt of NIST prepared for NIJ in April 1991. [60] OTA performed all calculations used here.

To simplify presentation, we will consider as options only repetitions of the test prescribed by NIJ Std. 0101.03, and we will neglect the possibility of BFS failures. In this context, the phrase "mean stopping probability" means the geometric mean of the stopping probabilities of the 48 fair shots required by the protocol; individual stopping probabilities may vary with shot location and order, test bullet, and panel-front or back, wet or dry. The fraction of fair shots stopped in a particular test or series of tests is *not the* mean stopping probability; it is the mean stopping probability plus an unknown "sampling error.'

We define "good armor" and "bad armor" in terms of mean stopping probability.[8] This is a policy choice; it should be decided by NIJ if NIJ elects to use this approach. For illustration only, we define "good armor" as armor having a mean stopping probability of at least 0.999, and "bad armor" as armor having a mean stopping probability of no greater than 0.95.

Second, we define the options for testing and certification. For illustration, we consider only two:

Option 1: Subject panels of the model to the test prescribed by NIJ Std. 0101.03 (at a specified ballistic-resistance level), and certify it if and only if no fair shots penetrate.

Option 2: Subject panels of the model to *three repetition* of the test prescribed by NIJ Std. 0101.03 (at the specified ballistic-resistance level), and certify it if and only if no more than one fair shot penetrates.

Under Option 1, the model is subjected to 48 fair shots and certified if none penetrate. Under Option 2, the model is subjected to 144 fair shots and certified if no more than one penetrates.

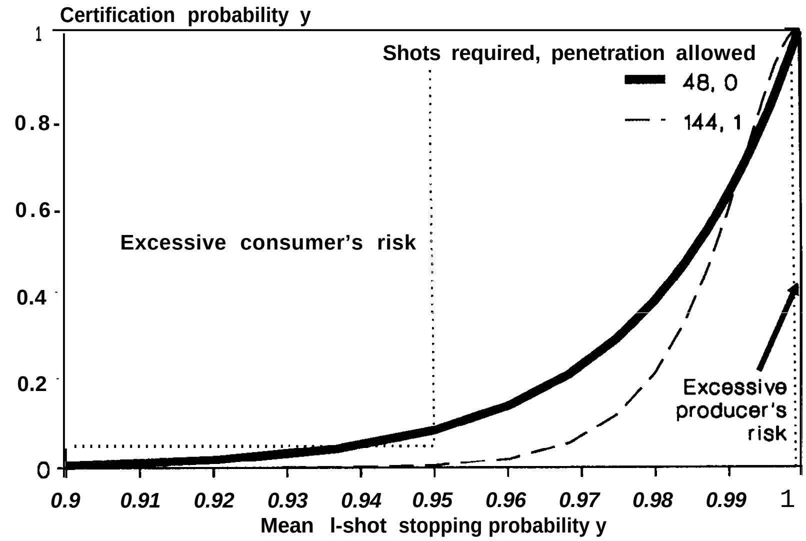
Third, we define "producer's risk" as the probability that good armor, as defined above, fails to be certified. We define ''consumers' risk' as the probability that bad armor, as defined above, is certified, recognizing that this also poses a financial risk to the producer.

Figure E-3 shows how the certification probability would vary with the mean stopping probability under each option. Note that the maximum consumers' risk of Option 2 (0.5 percent) is only about 1/17 that of Option 1 (8.5 percent), and its maximum producer's risk (0.9 percent) is also lower-only about 1/5 that of Option 1 (4.7 percent). However, Option 2 requires three times as many shots and would cost about three times as much as Option 1.

There are, of course, many other options. Even if one restricts consideration to repetitions of the .03 test sequence, one could require 1 repetition and allow O, 1,2, or up to 48 penetrations, or one could require 2 repetitions and allow O, 1, 2, or up to 96 penetrations, and so on. If upper bounds on consumers' risk and producer's risk are specified by policy, then it is a (solvable) technical problem to find the minimum number of repetitions required and the number of penetrations that must be allowed. In figure E-3, the upper left rectangular region labeled "Excessive Consumers' Risk" illustrates an upper bound of 0.05 (5 percent) on the consumers' risk, and the lower right rectangular region labeled "Excessive Producer's Risk" illustrates an upper bound of 0.05(5 percent) on the producer's risk. The graph (called an operating characteristic) for Option 1 passes through the region labeled "Excessive Consumers' Risk" and hence violates one of the bounds (hypothetically) set by policy. The operating characteristic for Option 2 avoids both prohibited regions and would be acceptable, but is not optimal, because the operating characteristic for an option not shown-two repetitions of the .03 sequence, allowi-ng one penetration-also avoids both prohibited regions but requires fewer repetitions. However, Option 2 would be optimal if consumers' risk and

[8] More generally, **one could define "good armor" and "bad armor" in terms of mean single-shot passing probability, which we define as the probability of stopping the [fair] shot and also leaving an acceptable BFS,** if it is a shot after which **BFS** is measured.

Figure E-3—Certification Probability y Versus Mean Stopping Probability for Two Certification Criteria



SOURCE: Office of Technology Assessment, 1992.

producer's risk were both prohibited from exceeding 1 percent.

Figure E-4 plots producer's risk versus consumers' risk for several options; it helps identify the minimum-cost certification criterion meeting the constraints on consumers' risk and producer's risk. Each curve corresponds to a certain number of repetitions of the 48-shot .03 test sequence and is therefore a curve of constant cost. Each break-point on it corresponds to the maximum number of penetrations allowed; the uppermost point on each curve-the one with greatest producer's risk—corresponds to allowing O penetrations, the next lower point to allowing 1 penetration, and so on. Options outside the rectangular region at lower left have excessive producer's risk, excessive consumers' risk, or both; bounds of 5 percent on producer's risk and consumers' risk are illustrated.

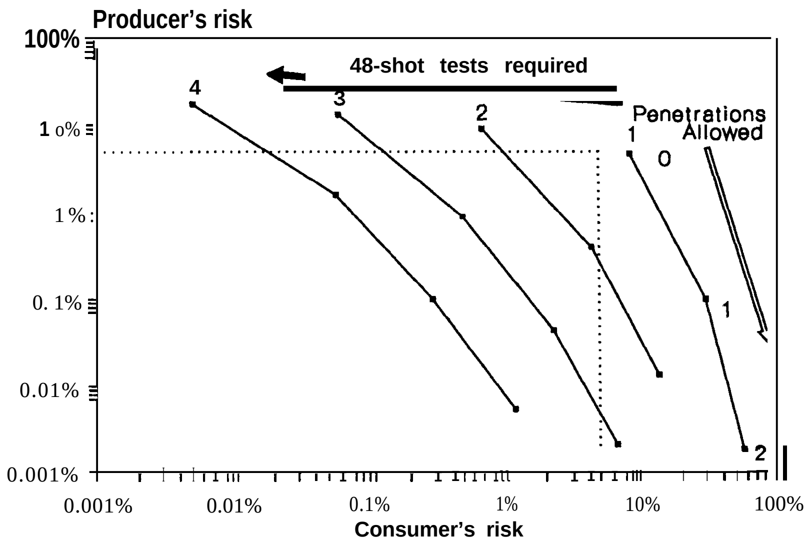
To identify acceptable minimum-cost criteria, one first examines the 1-test (48-shot) curve, and discovers that all points on it lie outside the acceptable region (only the first few points on it, including Option 1, are plotted). One next examines the 2-test (96-shot) curve, and discovers that only 1 point on it—the one corresponding to allowing 1 penetration—lies inside the acceptable region. This, then, is the unique minimum-cost criterion satisfying the constraints on producer's risk and consumers' risk.

In some cases there may be more than one minimum-cost criterion, requiring a choice between one criterion that minimizes producer's risk and another that minimizes consumers' risk. For example, if the bound on producer's risk is 10 percent and the bound on consumers' risk is 5 percent, then two repetitions would suffice, but one may allow one penetration or none. If 1 penetration were allowed, the producer's risk would be 0.4 percent and the consumers' risk 4.4 percent; if O penetrations were allowed, the producer's risk would be 9.2 percent and the consumers' risk 0.7 percent.

An alternative would be sequential testing with a stopping rule that allows testing to stop as soon as it demonstrates that both producer's and consumers' risks are acceptable. The number of tests required would not be fixed but would depend on the number of penetrations that occur as testing proceeds.

For example, a model could be certified if it withstood 96 shots with O penetrations, but if 1 penetration occurred in the first 96 shots, the armor could still be certified if it withstood 48 more shots with no more penetrations (i.e., if it withstood a total of 144 shots with no more than 1 penetration). This test would have a consumers' risk of 1 percent (slightly higher than that of the 96-shot test with no

Figure E-4-Consumers' Risk Versus Producer's Risk for Several Certification Criteria



SOURCE: Office of Technology Assessment, 1992.

penetrations allowed) and a producer's risk of 0.8 percent (much lower than that of the 96-shot test with no penetrations allowed). In some cases, it would require more testing and hence would cost more than the 96-shot test, but bad armor would have at most a 3.7-percent chance of needing more than 96 shots, and good armor would have at most a 8.7-percent chance of needing more than 96 shots.

A test *requiring 144* shots and allowing 1 penetration would have a slightly higher producer's risk (0.9 percent) but only half the consumers' risk (0.5 percent). Of course, it would cost more, on the average.

What effect would requiring more shots and allowing more penetrations have on reproducibility? It's a matter of definition. As noted in appendix C, neither these changes nor any others could provide more statistical confidence that the mean stopping probability is high enough for the model to pass a retest identical to the certification test with a specified probability. If this is the desired improvement in reproducibility, it is simply unattainable. However, if a retest is defined as, say, a 48-shot test with no penetrations allowed, then requiring several such tests for certification would reduce the probability that certified armor will fail such a retest (as distinct from the entire sequence of tests required for certification).

As noted in appendix C, the expected variance in outcomes of repeated testing is greatest when the probability of passing is one half. It approaches zero as the probability of passing approaches zero or one. Reducing the producer's risk can increase the probability that good armor will pass to as close to 1 as one desires; this will reduce the variance in outcomes of repeated testing of good armor to as small a value as maybe desired. Independently, the consumers' risk may be reduced, reducing the probability that bad armor will pass to as close to O as one desires and is willing to pay for; this will reduce the variance in outcomes of repeated testing of bad armor to as small a value as may be desired. The variance in outcomes of repeated testing of questionable armor-that having a stopping probability between that of good armor and that of bad armor-could still be high, but at least it could be argued that the variance in repeated testing of *good* armor would be low. This is qualitatively true of NIJ Std. 0101.03 and others standards such **as** PPAA STD-1989-05, but there are differences among these, and they do not define "good armor" quantitatively.

Some may object to allowing penetrations in a certification test in the belief that many, if not most, law-enforcement officers would not understand the statistical rationale and, in particular, might not trust or buy armor of a style that had been penetrated by a round of a type it is certified to stop, even if it stopped 99.9 percent of such rounds. This is a valid concern for NIJ to weigh in deciding whether to allow penetrations. However, NIJ should also weigh a related danger-that allowing no penetrations allows purchasers and wearers of armor who are so inclined to believe, unscientifically, that certified armor will certainly stop, in testing and in use, all rounds for which it is rated. Although NIJ Standard 0101.03 and NIJ Guide 100-87, *Selection and Application Guide to Police Body Armor*, caution purchasers and wearers that there is no such thing as bullet-proof armor, neither specifies the statistical confidence with which the probability of stopping rated rounds can be said to be at least 90, 95, or 99 percent on the basis of certification. Purchasers and wearers should know that neither the NIJ test nor any other provides more than O percent confidence that the probability of stopping a specified round is 100 percent.

ASSESSING RISK OF TRAUMA FROM STOPPED BULLETS

Several changes could be adopted to improve the validity, accuracy, and reproducibility of the current test for acceptable risk of blunt trauma, which consists of shooting the test armor on an unspecified but calibrated inelastic backing material, measuring the. depths of craters made in the backing, and failing the model if any crater is deeper than 44 mm.

For example, specifying the backing material to be used and reducing the currently allowed tolerance on its temperature might improve reproducibility, but perhaps not significantly. Reproducibility could also be improved (in the limited sense defined in the discussion of penetration resistance) by measuring more backface signatures and optionally, allowing some to exceed the specified limit [9] Reproducibility might also be improved by options for improving validity, such as those described below.

To improve the validity[10] Of the current test, NIJ could elect any of several options. If NIJ retains the current type of deformation test with a single BFS limit applicable to all bullets, velocities, types of armor, and wearers, there is evidence (see app. D, that the BFS limit corresponding to 90-percent safety exceeds 44-mm, with 95-percent confidence. NIJ could increase the BFS limit and still provide 90-percent safety with 90-percent confidence while reducing producers' risk.

Alternatively, NIJ could undertake to assess risk of blunt trauma based on the diameter(s), and perhaps also the depth, of backface signatures using a parametric lethality model similar to those proposed by Army researchers in the 1970s. A model appropriate for use does not exist today, but one could be developed, partly on the basis of reenactments and, if desired, partly on the basis of expert opinion informed by analogous animal experiments such as those performed for the NILECJ by the Army in the 1970s. Such a criterion might lead to different BFS limits for wearers of different sizes or weights and for armors of different areal density (i.e., mass per unit area); there could also be different BFS limits for portions of armor covering different parts of the body. This would increase complexity, but could be more accurate, hence more valid.

(A simpler and more conservative-hence less accurate-alternative would be to certify armor only in sizes greater than some minimum size that depends on test results or for wearers heavier than a minimum weight that depends on test results.)

There is also the option of using tests that would require additional instrumentation than that currently used (primarily, a ballistic chronograph, a thermometer, and rulers). Possibilities include measuring pressure in the backing during impact, or measuring velocity and deformation simultaneously, to use in predicting lethal trauma according to a ' 'viscous Criterion. '

The same procedures used to establish maximum allowable depths or other limits for each ballistic-resistance class could be used thereafter to revise those limits on the basis of new data on experiments with animals or assaults on humans.

Determine BFS Limits Based on Animal Experiments

As noted in appendix A, the current 44-mm BFS limit was originally derived specifically for the case of .38 Special round-nose lead bullets impacting on

---

9 By averaging, PPAA STD-1989.05 [113] allows more than half the measured backface signatures to exceed the specified limit of 44 mm.

10 For purposes of this discussion, we say a test is a 'valid' test if there is scientific evidence that the test accomplishes the purpose for which it was designed-in this context, the NILECJ's safety criterion, until it is superseded by a new NIJ safety criterion.

7-ply, 400/2-denier, Kevlar-29 armor at about 800 feet per second. The animal testing that would have been required to derive BFS limits for other threats and armors was begun but not completed.[11] Nevertheless, NILECJ-Std.-0101.01 and its successors, including NIJ-Std.-0101.03, specify a 44-mm BFS limit for all classes ("levels") of ballistic-resistance, for all types of armor.

No rationale for this generalization was documented. It was proposed by Lester Shubin, then Director of Science and Technology at the NILECJ, who in 1991 explained the rationale as the combination of

1. his judgment that it might be unsafe to allow higher energy bullets to produce a deeper BFS than the maximum deemed safe for .38 Special bullets impacting 7-ply Kevlar-29 at 800 ft/s;[12]
2. the absence of data showing that the BFS limit for higher energy bullets should be less than44 mm; and
3. the urgency of the need, inasmuch as armor was then being certified (under NILECJ-Std.-0101 .00) and worn without any test for protection from stopped bullets.

One option for improving the validity of the current test would be to conduct

1. additional experiments on animals, similar to those performed by Goldfarb et al.; [74] and
2. corresponding ballistic tests, analogous to *those performed.* by Prather et al., *[114]* to determine the backface signatures (or ether ballistic measurements) on clay backing (or whatever backing may be specified) that correlate with the the various degrees Of injury observed in the animal experiments.

One set of animal and ballistic experiments would be neededl for each combination of threat bullet and velocity) and and for which a BFS limit is to be determined. Umpublished records of the NILECJ-funded Army shootings of armored goats with .357 Magnum and 9-mm bullets, which remain in Ballistics Research Laboratory files, could supply some of the data needed to determine BFS limits appropriate for these bullets impacting the particular types of armor used in those experiments.

In principle, this approach has the potential to predict the probability of lethality from blunt trauma more accurately than can approaches that rely on (simple) parametric lethality models (described below). However, there are several disadvantages to this approach:

1. It would be expensive and time-consumin g to perform the large number of experiments that would needed just to determine BFS limits for the threat-armor combinations already tested under PTL_LStd.-OlOl.O3.
2. There would be a delay: until such experiments are performed and their results analyzed, there would be no explicit rationale for certifying armor (other than 7-ply, 400/2-denier, Kevlar-29 armor) as reducing the risk of blunt trauma (from threats other than .38 Special round-nose lead bullets at about 800 feet per second) to an acceptable level.
3. There would likewise be a barrier to technological innovation: armor not of the generic types tested in the experiments could not be certified. Developers of novel armor material—for example, synthetic spider silk-would have to fund experiments to estimate the deformation-trauma correlation in armor made from their material, or else lobby for Federal funding for such. experiments, and convince NIJ of the validity of the results before they couldi have any hope of having their product incorporated in NIJ-certified armor.
4. Extrapolation of the experimental results from animals to humans would be judgmental, as it was in the study by Goldfarb et al.

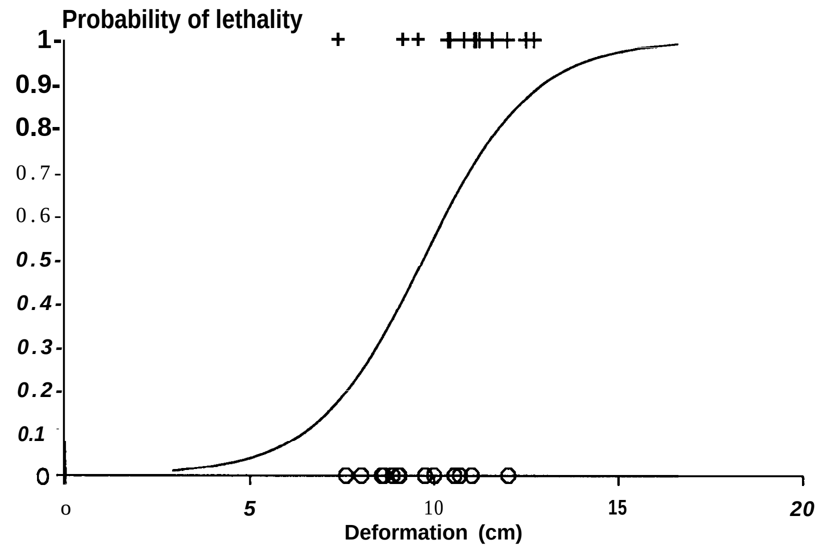
Determine BFS Limits Based on Parametric Lethality Models

Another option for improving the validity of the current test would be to base BFS limits on parametric lethality models of the type described in appendix A. An advantage of this approach, relative to the one @t described, is that extending it to additional threats or types of armor does not require additional biomedical tests (read: shooting large manmals, and killing some); it requires only additional ballistic tests: shooting the armor of interest

[11] The NILECJ funded Army experiments in which armored goats were shot with .357 Magnum and 9-mm bullets, as was armor on clay backing, but the research was not completed or published.

[12] Shubin worried, and some others still worry, that a BFS limit less than 44 mm might be appropriate for higher energy bullets, especially rifle bullets.

Figure E-5—Lethality Versus Prediction Based on Deformation



SOURCE: Office of Technology Assessment, 1992.

with bullets of interest at velocities of interest, using a backing such as clay.

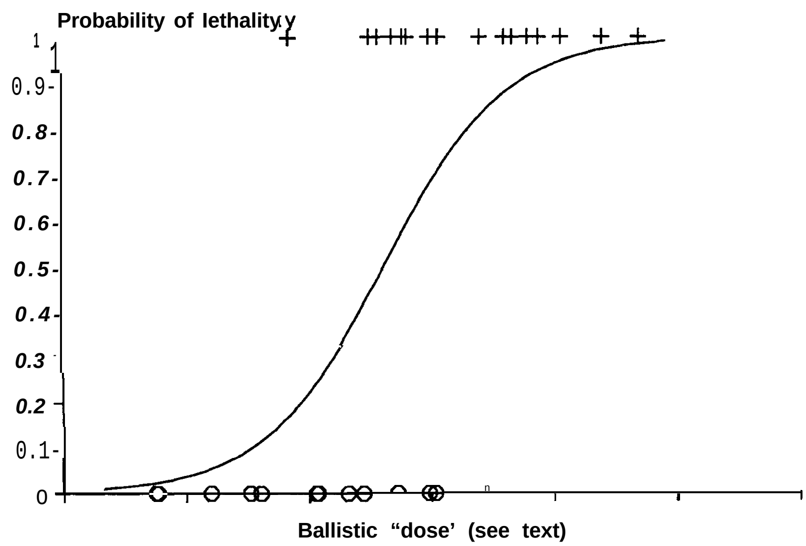
A simple parametric lethality model is a mathematical formula or graph that predicts the probability that a single shot would cause lethal blunt trauma, based on the value of a single parameter, such as BFS. Such a model could be used to derive a maximum acceptable BFS from the maximum acceptable probability of lethality specified by policy. Similarly, models of lethality or serious injury may also be developed and used (see app. D).

More complicated models, such as those proposed by Clare et al. [35] and by Sturdivan, [130] predict probability of lethality based on the values of several parameters, some describing the wearer (e.g., body mass and body-wall thickness), some describing the threat (e.g., bullet mass and velocity), some describing ballistic test results (e.g., the diameter of the crater made in flesh-simulating backing by the armor when hit by a bullet), and some describing properties of the armor (e.g., areal density: mass per unit area). In general, using more parameters provides more information and may improve the model, at the expense of the cost of making the additional measurements and the increased complexity of calculating the predicted lethality from them.

For example, figure E-5 shows the results—deaths (+) and survivals (o)-of shooting 29 goats over the liver with blunt, nonpenetrating projectiles simulating nonpenetrating ballets hitting armor, and the probability of death predicted on the basis of the maximum momentary deformation of each goat's abdomen, which is comparable to the depth of the crater the projectile would make in clay; table E-1 shows the data. Figure E-6 shows the same results, but with the probability of death predicted (by OTA) on the basis of a ballistic "dose" that depends on maximum deformation and five other parameters (the projectile's mass, diameter, and velocity, and the goat's weight and body-wall thickness). It is apparent that ordering the results by the ballistic dose as in figure E-6 separates the deaths from the survivals better than ordering them by deformation as in figure E-5. A vertical line can be drawn in figure E-6 to separate deaths from survivals with only 5 misclassifications; a similar line in figure E-5 would produce 9 misclassifications. Moreover, the model (i.e., the estimated probability of lethality) in figure E-6 predicts the results (deaths and survivals) with 67 times the likelihood predicted by the model in figure E-5.

A model (prediction) similar to that used in figure E-6 could be used for certifying acceptable protection from the impact of stopped bullets on the basis of multiple measurements.

Figure E-6-Lethality Versus Prediction Based on Multiple Measurements



SOURCE: Office of Technology Assessment, 1992.

Figure E-7 shows another example-a logistic discriminant model developed by OTA that discriminates perfectly the survivals from the fatalities of goats shot on the chest with blunt, nonpenetrating projectiles as reported by Clare et al. [35] Each shot is described in terms of a ''victim size parameter," which depends on the subject's weight, squared weight, and body-wall thickness,[13] and a "bullet-armor parameter,' which depends on the mass, speed, and diameter of the blunt projectile[14] used to simulate a combination of bullet, velocity, and armor. The model describes a straight line that separates shots that were survived from those that resulted in fatalities.

Both examples illustrate the general principle that the use of more parameters allows a model to better fit the data to which it is fitted, and may allow it to predict lethality with greater reliability. However, using more parameters may decrease the statistical confidence with which one can accept (i.e., not reject) the model. By using enough parameters, a model can be made to fit perfectly the data to which it is fitted, but this provides no confidence that the model would have been rejected had the data been different.

Appendix A described a method proposed by Prather, et al., for treating a bullet stopped by an-nor as a blunt projectile, and a multiparameter lethality model developed by Sturdivan [130] to estimate the probability that such a nonpenetrating projectile will cause lethal blunt trauma to the *thorax.* Here we will discuss how the model could be used to assess the acceptability of protection from lethal blunt trauma. Assessment of the acceptability of protection from lethal or critical trauma using a different parametric model-e.g., one based on data from reenactments—would proceed in a similar manner.

Sturdivan's model for probability of lethality, P(L), is

$$P(L) = 1/(1 + \exp(34.13 - 3.597 \ln(MW^2/W^{1/3}TD))),$$

where M denotes the projectile mass (g), V the projectile velocity (m/s), W the victim's body mass (kg), T the victim's body-wall thickness (cm), and D the projectile diameter (cm). D is estimated as the diameter of the crater made in clay backing, which is measured in a ballistic test. M and V are estimated from D, the bullet mass, M_P, and velocity, VP, and

13 The victim-size arameter is given (approximately) by the expression 86.89 W -0.9996 W^2 + 185.5 T, where W is the victim's weight (kg) and T is the victim's body-wall thickness (cm).

14 The bullet-armor parameter is given (approximately) by the expression 1.8434 M + 11.77 V -0.5788 D, where M is the mass of the blunt projectile (g), D is its diameter (mm), and V its velocity @/s).

Table E-l—Lethality of Blunt Trauma to Liver v. Characteristics of Projectile and Victim

M (g)	V (m/s)	D (mm)	W (kg)	T (cm)	Depth(cm)	Survival?
300	49.2	74	47.31	2.7	7.60	yes
300	38.6	74	55.88	2.1	11.68	no
300	41.2	74	48.46	2.4	9.75	yes
300	46.1	74	52.60	2.3	10.51	no
300	43.0	74	55.64	2.6	8.87	yes
300	39.4	74	54.67	1.9	8.58	yes
300	38.6	74	52.20	2.3	9.24	no
300	56.2	74	57.86	4.1	8.63	yes
300	49.1	74	58.19	2.9	7.46	no
300	45.0	74	56.75	2.6	9.24	no
300	36.0	74	56.02	2.1	9.86	no
300	48.5	74	55.65	3.3	9.03	yes
300	31.2	74	43.63	1.6	11.34	no
300	40.9	74	55.35	2.7	9.06	yes
300	41.2	74	50.23	2.6	10.54	no
430	42.0	100	53.15	1.8	11.25	no
430	38.2	100	44.02	2.4	10.70	yes
430	37.0	100	37.21	2.5	8.02	yes
430	32.5	100	40.98	1.5	10.00	yes
430	33.4	100	57.90	1.8	12.81	no
430	58.3	100	62.06	4.2	11.02	yes
430	34.3	100	48.96	1.6	12.01	yes
430	38.2	100	59.30	3.6	11.20	no
430	32.4	100	56.06	2.1	12.09	no
430	29.9	100	41.80	1.8	9.69	no
430	30.4	100	46.04	1.4	12.60	no
430	33.5	100	48.62	2.0	10.49	no
430	27.0	100	43.21	1.7	10.92	no
430	58.4	100	63.31	3.2	10.55	yes

Legend: M: projectile mass.
V: projectile speed.
D: projectile diameter.
W: weight of victim (goat).
T:thickness of victim's bodywall (skin, fat, muscle) at impact point.
Depth: maximum momentary depth of depression of victim's (goat's) skin by projectile.

SOURCE: U.S. Army Chemical Research, Development, and Engineering Center, April 4,1991 [131].

the areal density of the armor, a_d (g/cm*), using the formulas

$$M = \quad + 3.14\ (D/2)^2 a_d$$
$$V = (Mp/M) \quad V_p.$$

To assess risk of blunt trauma to a particular wearer, the wearer's body mass W and body-wall thickness Tare measured and used, along with D, M, and V in the formula for P(L). This procedure would be reversed in a certification test: P(L) would be set equal to the maximum acceptable probability of lethality, P(L)_, and the equation

$$P(L)_{max} = 1/(1 + \exp(34.13 - 3.597 \ln(MV^2/W^{1/3}TD)))$$

would be solved for $W^{1/3}T$. The value obtained would be the minimum allowable value:

$$(W^{1/3}T)_{min} = [\exp(-34.13)\ (1 - P(L)_{max})/P(L)_{max}]^{1/3.597}\ MV^2/D$$

That is, the armor could be certified to provide acceptable protection from lethal blunt trauma to wearers having a body mass W and body-wall thickness T large enough so that $W^{1/3}T$ equals or exceeds a value, $(W^{1/3}T)$min, derived from the specified threat Mp, Vp), the ballistic test result (D), and the areal density of the armor (a_d). Figure E-8 illustrates the process.

For a maximum acceptable probability of lethality of 10 percent ($P(L)^{max}$ = O. 1), $W^{1/3}T$ must equal or exceed 0.0001395 MV^2/D.

If it is not desired to certify armor for wearers having at least a specified value of $W^{1/3}T$, a conservative alternative would be to certify armor unconditionally if its calculated value of $(W^{1/3}T)$min exceeds the the value corresponding to a small fractile of officers, perhaps $(W^{1/3}T)$min = 7.6, for W= 55 kg and T = 2 cm. Of course, conservatism has its risks-of decreasing wear rate and increasing producer's risk unnecessarily. The option of certifying armor only for wearers having at least a specified value of $W^{1/3}T$, and variations of this, are discussed in greater detail below.

We will illustrate the calculation of $(W^{1/3}T)_{ti}$, assuming $P(L)_{max}$ =0.1, for a test in which a .38-cal., 158-grain (10.2-gram) lead round-nose bullet was fired at an armor panel made from 7-ply, 1,000-denier Kevlar 29. The impact velocity was 833 fps (V_p = 254 m/s), and the BFS was a crater 3.4 cm deep, with a roughly elliptical base measuring 6.2 cm x 5.5 cm. [114] The geometric mean of these major and minor axes (5.8 cm, the square root of 6.2 cm x 5.5 cm) should be used as the diameter D in calculating M. The nominal areal density of 1,000-denier, 31x31 Kevlar 29 fabric is 8.3 ounces per square yard (0.028 g/cm^2) per ply, so the areal density a_d of the 7-ply panel would be about 0.20 g/cm^2. Hence

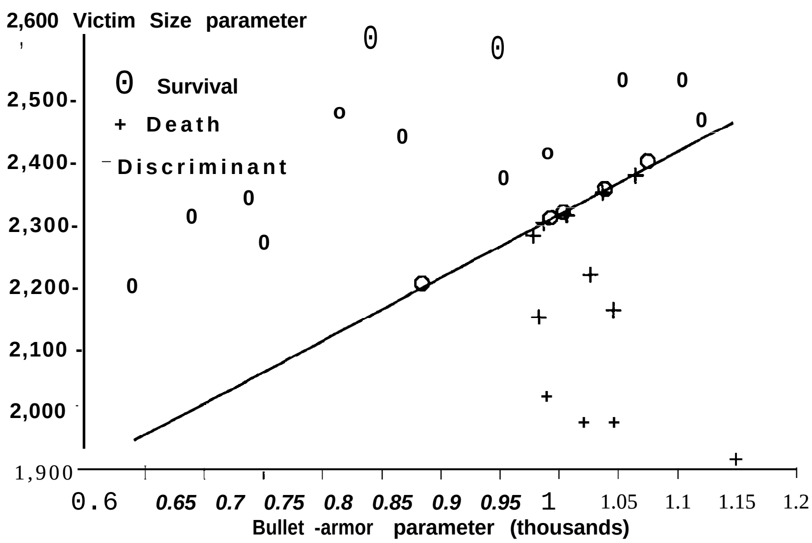
$$M = Mp + 3.14\ (D/2)^2 ad$$
$$= 10.2 + 3.14\ (6.2/2)^2\ 0.20$$
$$= 16 \text{ g}$$
$$V = (Mp/M)\ V_p$$
$$= (10.2/16)\ 254$$
$$= 162 \text{ m/S}$$

~/D= 67726, and

$$(W^{1/3}T)_{min} = 9.448 \text{ kg}^{1/3}\text{-cm}$$

Measurement of the armor's areal density over the crater presents a problem: Should the portion of armor over the crater be excised, cleaned of bullet, fragments, and backing, and weighed? This may degrade the value of the armor as an archival

Figure E-7—Discriminant Model for Assessing Protection From Lethal Trauma by a Stopped Bullet



SOURCE: Office of Technology Assessment, 1992.

standard for quality-assurance. For some armor, there is an alternative: the areal density of armor made from 1000-denier, 31x31 Kevlar 29 fabric could be inferred from bullet momentum and crater depth and diameter, using a clay-cavity model published by the Aerospace Corp. [7] This procedure is illustrated in figure E-9. One could attempt to develop similar models and procedures for other armor materials, but this may be costly (although less costly than animal experiments) and may pose a barrier to innovation.

Before putting these procedures into practice, it would be advisable to adjust the lethality predicted by Sturdivan's models, or others fitted to data obtained by targeting vulnerable organs, to account for the less accurate marksmanship typical of assaults. The adjustment process would weigh the blunt-trauma lethality predicted for each vulnerable organ by an organ-specific model according to the probability that a shot on armor (or on the upper torso) would impact over that organ, as was done in the medical assessment by Goldfarb et al.

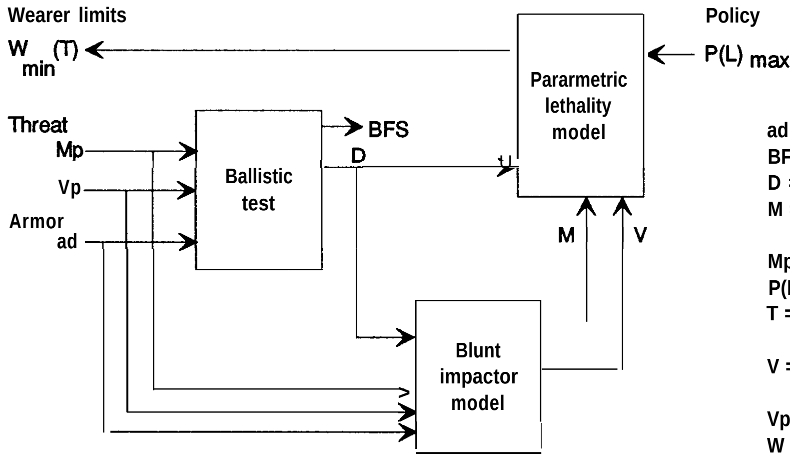
The extrapolation of predictions based on animal data to humans would be necessarily judgmental, as it was in the original body armor medical assessment sponsored by the NILECJ. Different experts, considering the animal data, might estimate different probabilities of death or trauma in humans under the same conditions. There is a procedure for combining these estimates, [95] and if this is done for c conditions, a c-parameter logistic model (counting the "dummy regressor") could be fit to the c combined estimates. Advantages of a logistic model include its great generality and the ability to update it easily on the basis of additional data [164] from reenactments of assaults.

Specify Size-Dependent BFS Limits

As noted in appendix A, the body armor medical assessment team that recommended the current 44-mm BFS limit did so to guarantee protection to light, female wearers with a thin body wall; they expected that heavier male wearers with a thicker body wall would face a lower probability of surgery or death if shot by a round that would cause a 44-mm BFS behind their armor. The parametric lethality models discussed above also support this expectation. These considerations provide a rationale for allowing a deeper BFS behind armor sized for large males, or certified only for male or female wearers heavier than specified minimum (perhaps sex-dependent) weights.

As examples of how this could be done, consider the 0.20 g/cm^2 vest mentioned above that stopped a 10.2-gram bullet that impacted at 254 m/s and made a crater measuring 6.2 cm x 5.5 cm in diameter. The calculated value of $(W^{1/3}T)_{min}$ was 9.448 $kg_{1/3}$-cm.

Figure E-8-Assessing Acceptability of Protection From Lethal Blunt Trauma Using a Parametric Lethality Model



ad = areal density of armor (mass per unit area)
BFS = depth of crater
D = diameter of crater
M = mass of projectile + portion of armor pushed into crater
Mp = mass of projectile (bullet)
P(L) = probability of lethality
T = thickness of wearer's body wall (skin, muscle, bone...)
V = velocity of projectile + portion of armor pushed into crater
Vp = velocity of projectile
W = "weight" (i.e., body mass) of wearer

SOURCE: Office of Technology Assessment, 1992.

The vest could be certified to provide acceptable (viz., "90-percent") protection from lethal blunt trauma to wearers having $W_{1/3}T = 9.448$ $kg_{1/3}$-cm or greater. A certification of compliance could state the restriction in this way, or it could portray the restriction in graphical or tabular form, for example:

> This armor complies with NIJ-Std.0101.xx and provides 90-percent protection from lethal trauma from a stopped bullet to wearers weighing at least
>
> 54 kg and having a body-wall thickness of at least 2.5 cm, or
> 61 kg and having a body-wall thickness of at least 2.4 cm, or
> 70 kg and having a body-wall thickness of at least 2.3 cm, or
> 80 kg and having a body-wall thickness of at least 2.2 cm, or
> 92 kg and having a body-wall thickness of at least 2.1 cm, or
> 106 kg and having a body-wall thickness of at least 2.0 em.

This is more complicated and cumbersome than the current procedure. On the other hand, it could provide a rationale for certification of protection against blunt trauma caused by other than Type I bullets hitting Kevlar armor. It also would allow qualified certification of armor that would fail if required to provide the smallest wearers with acceptable protection from lethal blunt trauma.[15]

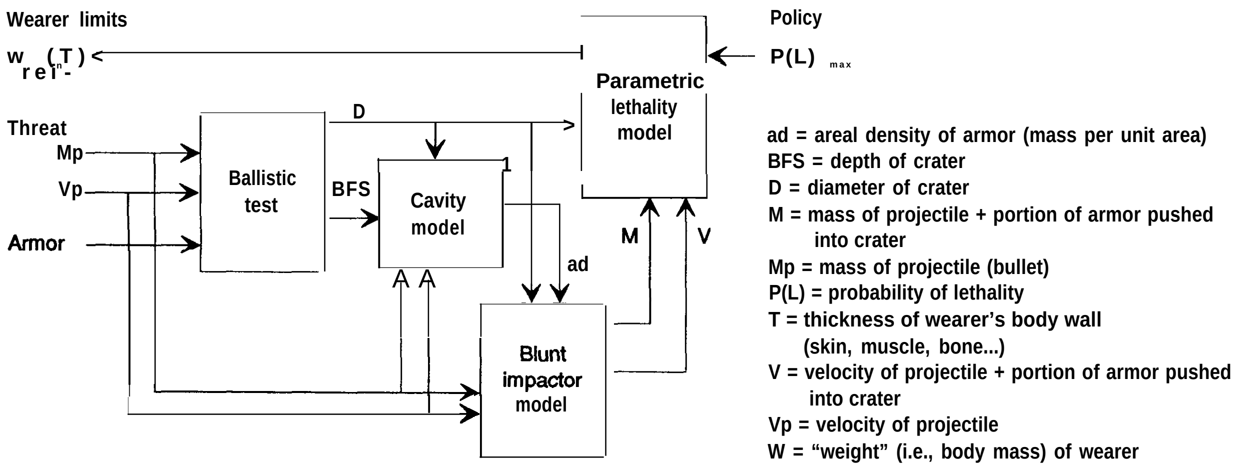
However, another drawback of the procedure must be addressed: it requires knowing the wearer's body-wall thickness, which is not readily measured. It might require a computed axial tomography (CAT) scan. This could be avoided, perhaps with some loss of reliability, by using a parametric lethality model that does not depend on T. For example, Clare et al. [35] developed a model of blunt-trauma lethality as a function of MV^2/WD. A related approach is to use, in a model that depends on T, an estimate of T in terms of other variables. For example, Sturdivan [132] has found that T is roughly proportional to $W_{1/3}$ in both goats and man. If a is the constant of proportionality, then one could use a $W1_{/3}$ in place of T in $MV^2/DW_{1/3}T$, resulting in a model that depends on $aM^2/DW^{2/3}$. OTA has determined that this procedure results in negligible reduction in goodness-of-fit to some data[16] but reduces goodness-of-fit to other data substantially.[17] Other such models could be developed; however, other things being equal, requiring a model not to depend on T may reduce the reliability with which it correctly predicts lethality.

15 For example, in a similar test also using a .38-cal., 10.2-g LRN bullet fired at a 7-ply, 1,000-denier, Kevlar-29 panel, the impact velocity was 787 fps (V_P= 240 rids), and theBFS was a crater 4.6 cm deep, with a circular base 6.0 cm in diameter. [114] This result would have failed the armor under NIJ-Std.- 0101.03, but the procedure discussed here would allow the armor to be certified for wearers havin$W^{1/3}T = 8.786$ $kg^{1/3}$-cm or greater. For example, the armor could be certified for wearers weighing 75 kg with body walls at least 2.1 cm thick.

16 For example, the data on lethality of blunt impacts to goat abdomen over the liver in table E-1.

17 For example, the data on lethality of blunt impacts to goat thoraxin table 1 of [35].

Figure E-9-Alternative Procedure for Estimating Probability of Blunt-Trauma Lethality From Backface Signature and Parametric Lethality Model



SOURCE: Office of Technology Assessment, 1992.

Revise BFS Limit(s) Based on Field Experience

The Army's initial medical assessment of body armor and the parametric lethality models described above are based on animal experiments performed before data were available on shootings of humans wearing such armor.

Now more than 20 assaults (but only 2 that resulted in death or critical injury) have been reenacted, several times each. OTA's analysis of the results (see app. D) concludes that the 44-mm BFS limit in NIJ Standard 0101.03 is smaller than necessary to limit the risk of death or life-threatening injury from a bullet that impacts at the maximum velocity for which protection is certified and is stopped by the armor to 10-percent, a goal specified by the NILECJ in 1976. However, the analysis does not show that the test reliably discriminates unsafe armor from safe armor; if it does, more reenactments will be needed to prove it.

If NIJ decides that a 10-percent risk is still acceptable (this is a policy choice implying a value judgment), the BFS limit could be increased. This might increase the risk to wearers of armor (perhaps only slightly) but might increase the frequency with which officers wear their armor. It would decrease the risk, to manufacturers, that armor that actually limits risk as required would fail the test.

To increase the confidence with which conclusions may be inferred (as in app. D), more reenactments of more assaults-especially assaults in which officers were killed or critically injured by stopped bullets-are needed. This will require monitoring assaults and collecting detailed data on those suitable and most important for reenactment.

If and when such reenactments have been performed, the Walker-Duncan procedure [164] could be used to revise any of the logistic models described in app. Din light of the new data. A new model with more parameters would have to be fitted, using separate-sample logistic regression, [9] to the cumulated data in order to estimate BFS limits for different cases—i.e., threat-, armor-, and wearer-dependent limits.

Specify Tests Other Than BFS

Someday, certification of acceptable protection from blunt trauma could be based in whole or in part on tests other than BFS measurements. Proposals include measuring pressure in the backing during impact, or measuring velocity and deformation simultaneously, to use in predicting lethal trauma according to a "viscous criterion. " These tests would require more sophisticated, expensive instrumentation than that currently used—primarily, a ballistic chronograph, a thermometer, and special rulers-and it is not yet known whether such tests would be more accurate than the current one or the other tests, discussed above, based on BFS measurements.[18]

Pressure Criteria

Some experts expect that the peak pressure measured in backing would be a better predictor of specific types of blunt trauma than would any test based on BFS. One such type is the laceration or rupture of arteries or other organs compressed suddenly by the intense pressure wave generated by the impact of a nonpenetrating bullet on armor. Such trauma has caused the death of one police officer, whose armor, in stopping a rifle bullet, penetrated his chest.[19]

Research has also demonstrated that a brief, intense pressure pulse, similar to the early portion of the pressure pulse generated by a nonpenetrating ballistic impact, may block conduction by cardiac nerves. [122, 123] Some deaths caused by automobile accidents and baseball impacts might be attributable to this mechanism or to apnea (cessation of breathing) or other effects. [154, 155] It might also be responsible for deaths caused by single blows of other types-e. g., the widely publicized classroom death caused by a blow delivered to the chest in a hitting game called a' 'cuss game.' Although deaths attributable to these mechanisms are apparently rare, tests based on BFS may not be a good predictor of them, because research has demonstrated that BFS is more strongly correlated with the later, longer, less intense portion of the pressure pulse than with the early, brief, intense portion. [84] However, correlation of peak pressure in backing with lethality in humans has not yet been established.

Viscous Criteria

Empirical research suggests that blunt trauma caused by automobile accidents, baseball impacts, and other causes may be classified as lethal or nonlethal based on the maximum value of the velocity of deformation times the factional compression of the body.[20] A blow is predicted to be lethal if the velocity of deformation times the fractional compression ever exceeds a certain threshold; this is called the "viscous criterion. " [156] [21] Using it in armor certification would require

1. using a backing that simulates the deformation-versus-time history of the human torso or can be calibrated with it, and
2. measuring velocity and deformation simultaneously.

Another hypothesis holds that lethality of blunt trauma may be predicted on the basis of maximum velocity and maximum deformation (or compression), [38, 84] which occur at different times. Such a model would be easy to use for certification, because the maximum velocity can be approximated as the impact velocity, which is already estimated in NIJ certification testing, and the maximum deformation of impacted tissue could be calibrated to crater depth in the inelastic backing, which is already recorded. Variants of the general hypothesis maybe tested for consistency with animal blunt-trauma data already collected.

For example, OTA fit the logistic model

$$P(L) = 1/(1 + \exp(-a - b\ \mathrm{In}(V) - c\ \mathrm{In}(\text{compression})))$$

to data on survival of 29 goats shot over the liver by blunt, nonpenetrating projectiles (see table E-l). V is projectile velocity at impact in m/s, and "compression' is maximum depth of abdominal indentation, in cm, divided by the cube root of the animal's body mass Win kg. The cube root of W was used as a proxy, or substitute, for the thickness of the body in the direction of indentation, which was not recorded. The best fit (maximum likelihood) was obtained with a = 13.0, b = -4.15, and c = 2.58, so the fitted model is

$$P(L) = 1/(1 + e^{-13} V^{4.15} / \text{compression}^{2.58})$$

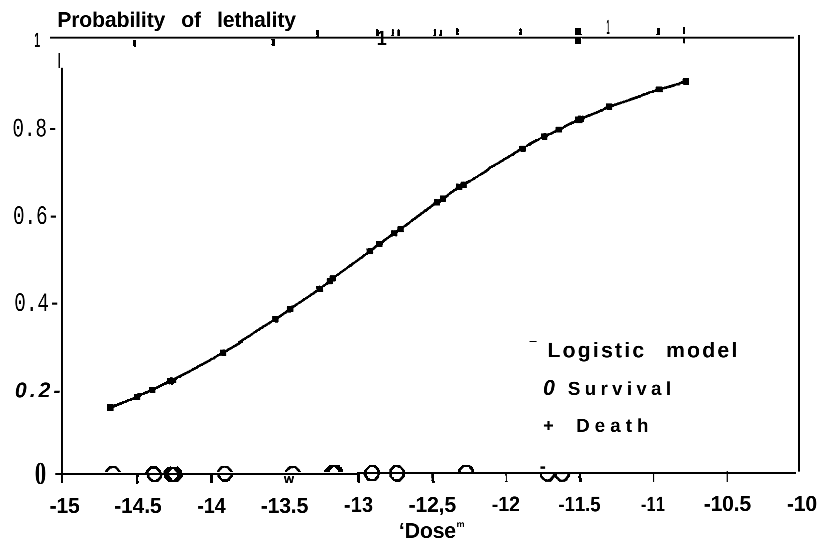
18 Iremonger and Bell [84] cited experiments that found viscous criterion values for impacts to be correlated with BFS depth, which in turn was found to be "not a sensitive measure of injury severity.' (However, their definition of the viscous criterion differed from that of Viano and Lau [156], whom they cited.) They speculated that pressure measurement, perhaps in combination with other measurements, might be abetter predictor of injury, but noted that 'further work is required in order to quantify the damaging effect of stress wave transmission."

19 **The medical examiner attributed the cause of death not t. the penetration, per se, but to ''The shockwave created by the missile, ' which "lacerated the aorta, the pulmonary artery, and the vena cava immediately adjacent to the heart, resulting in death by insanguination into the thoracic cavities." [133] Although the speed of sound maybe so low in lung tissue that the pressure wave may have been supersonic (hence a shock wave) there, [38, 166] the pressure wave was probably subsonic (not a shock wave) in the aorta, the pulmonary artery, and the vena cava. However, even a subsonic pressure wave, if sufficiently strong, could cause the damage noted.**

20 The fractional compression is defined as the depth of deformation divided by the thickness of the body in the direction of deformation.

21 See also [94, 153, 158, 159]; cf. [84].

Figure E-10-A Logistic Model for Blunt-Trauma Lethality in Terms of Compression Times Velocity of Deformation (a "VIscous Criterion")



"dose" = 2.58 In(compression) -4.15 In(v) where
V = impact velocity in m/s
compression = (maximum depth of abdominal indentation in cm)/ $W^{1/3}$
W = body mass in kg

SOURCE: Office of Technology Assessment, 1992.

Figure E-10 shows the predicted lethality as a function of a ballistic "dose" defined by

"dose" = 2.58 In(compression) -4.15 in(V).

It may seem paradoxical that the model predicts that, of animals suffering comparable compression, those hit by higher velocity projectiles would be less likely to die.[22] Nevertheless, predictions of the model may be sensible if based on real data, because one would expect that, of similar animals, those hit by higher velocity projectiles would be more likely to suffer greater compression. What is surprising in this case is that those animals hit by higher velocity projectiles suffered less compression, on the average. Thus, although the apparently paradoxical form of the model is not surprising, the reason for it is. For whatever reason (perhaps mere chance), the data in table E-1 are peculiar, and one should doubt the validity of the OTA model based on them unless the peculiarity is explained or the model validated by other data.

Nevertheless, the model predicts the deaths and survivals in table E-1 with a likelihood ($6.2x10^8$) more than three times that ($1.9x10^8$) with which Sturdivan's model

$$P(L) = 1/(1 + e^{-29.0}(MV^2/W^{1/3}TD)^{4.34}),$$

predicts them. This shows that the simple viscous criterion considered here predicts lethality better than a logistic model in terms of $MV^2/DW^{1/3}T$. More complicated viscous criteria considered by OTA fit slightly better, but not as well as a nonviscous logistic model,

$$P(L) = 1/(1 + e^{14.2} M^{-32.1} V^{10.9} D^{42.6} \text{Depth}^{-5.06} W^{-11.0} T^{-0.249}),$$

which predicts the outcomes with a likelihood of $2.3x10^{-6}$, which is 37 times the likelihood with which OTA's viscous model predicts the outcomes and more than a hundred times the likelihood with which Sturdivan's model predicts the outcomes. It is possible that a logistic model predicting lethality or injury in terms of the viscous criterion proposed by Viano, Lau, and colleagues could predict outcomes of other experiments (in which the required measurements are recorded) better than OTA's viscous model, or other logistic models, would. However,

[22] Similar results are common in logistic models that depend on correlated variables, such as velocity and compression in this case. The predictions of such an apparently paradoxical model are usually reasonable if the variables do not have values outside the range of values of the data to which the model was fitted.

more research would be needed to find out whether this is true.

To summarize, it is plausible that pressure criteria could predict blunt-trauma lethality from some, possibly rare, causes better than other criteria discussed here. However, there is as yet no basis for expecting that criteria based on pressure measurements in backing would significantly improve predictions; future research may, or may not, provide such a basis. Measurement of backing pressure for certification or acceptance tests based on pressure criteria would require instrumentation costing hundreds or thousands of dollars. Viscous criteria may predict lethality of ballistic blunt trauma as well as or better than parametric models developed by the Army for the NILECJ. However, it is reasonable to expect that more general parametric models including but not restricted to viscous criteria maybe better predictors of blunt-trauma lethality. Some, but not all, viscous criteria would require expensive instruments for measuring and recording backing indentation and velocity histories.

ASSURING QUALITY AT POINT-OF-SALE AND IN SERVICE

Revise NIJ Std. 0101.03 to Apply to Lot-Acceptance Testing

Some of the issues of enforcement and quality control discussed in appendix C would be solved if NIJ revised its armor certification process to be a lot-certification process rather than a model-certification process, with a separate style-certification process.

To execute this option, NIJ would have to

1. Revise the current standard to apply to lot testing, as NILECJ-0101.00 [141] did.[23]
2. Define "lot" precisely. (Must a lot be homogeneous? Why?)
3. Specify the number of samples from each lot to be tested, or a way to calculate the number from statistical criteria such as maximum probability of accepting a lot more than 1 percent of which is defective.
4. Ensure that the samples to be tested are selected randomly from each lot.

Definition of Lot

The definition of lots is usually guided by the following principles [60, 107]:

- *Lots should be natural units in commerce.*
- *Lots should be homogeneous—all* units in a lot should be made in the same time period by the same workers using the same equipment and materials, which in turn should be from the same lot, etc.

In addition, a lot should have at least enough units to provide the samples required for quality assurance (see item 3 above). For economy, the lot size should be many times the sample size, so that the cost of testing, including the cost of the samples, could be amortized over the units remaining after sampling for testing.[24]

Units of Commerce

The natural unit of commerce in armor varies widely; a large order may consist of tens of thousands of units,[25] while for custom armor it is often 1 unit. If the current test procedure is retained, shipping 1 unit of certified custom armor would require producing 7 units from which 6 could be sampled at random for testing. Even more samples would be required if high statistical confidence in high reliability[26] were demanded.

Lot Homogeneity

In some approaches to quality control, it is important that a lot be homogeneous, i.e., that all units in the lot be alike. In the approach to acceptance sampling described above in *Increase Total Shots and Allow Penetrations,*[27] lot homogeneity is important because it provides a rationale for *assuming* that all units in a lot have the same reliability, so that the reliabilities of the units not

---

23 NIJ Guide 100-87, *Selection and Application Guide to Police Body Armor [145],* might also need to be revisal.

24 An alternative for attaining hi@ confidence with small sample sires is to use Bayesian methods of risk assessment, which are explicitly subjective and hence controversial. However, they have been used to assess the safety of nuclear power plants and space launch vehicles. [11]

25 A large order may consist of tens of thousands of units of various sizes. We argue that size may affect ballistic resistance both in tests and in service, so otherwise similar armor of various sizes should not be considered a single lot, according to the usual definition of a lot.

26 Viz., probability of passing.

27 The approach is a form of 'acceptance sampling on the basis of parameters.''

tested may be inferred from the results of the tests of the units selected from the lot to be tested. This assumption may be wrong, and it may be unnecessary.

- It may be *wrong* because subtle, unnoticed variations in manufacturing processes could cause the reliabilities of apparently identical units to differ. Ballistic test results could be subjected to a statistical test to decide whether they are.[28] But,
- It may be *unnecessary*, depending on type of reliability one is interested in. Two distinctly different concepts of reliability that should be distinguished are (1) the reliability of an individual unit of armor, and (2) the (''average") reliability of a lot, which is, by definition, homogeneous in the lot. In either case, a lot could be any set of armor labeled as such by the manufacturer-not necessarily homogeneous in ballistic resistance nor in any other respect, such as size-provided it passes statistical tests, based on the results of ballistic tests, to limit the risk of accepting bad armor as well as the risk of rejecting good armor.

  Concept (l)-of the reliability of an individual unit-is problematical in the classical, frequentist interpretation of probability, which holds that reliability (i.e., probability of success) is a meaningful concept only if it is possible to conduct identical, repeated trials.[29] However, if the individual units of a lot may differ, perhaps invisibly, and especially if the purpose of testing is to determine whether they do differ, then tests of samples from the lot cannot be *assumed* to be *identical* repeated trials.[30] [31]

  Concept (2), the reliability of a lot (which an adherent to concept (1) could call the average reliability of a lot), is an admissible concept in the classical paradigm of statistical inference. Sampling and testing (e.g., as described above in *Increase Total Shots and Allow Penetrations)* provides information directly about the reliability of a lot, which may be all that some consumers care about. But, together with information about lot size and sample size, it also provides information about the distribution of the individual reliabilities in a lot.

Sample Size

In fact, if one is concerned about individual reliabilities in a lot, the minimum sample size will be determined by the lot size, themmaximum acceptable risk of accepting unreliable armor, and the maximum acceptable risk of rejecting reliable armor. If one is concerned only about the reliability of a lot, the minimum sample size will not depend on the lot size, but only on the maximurn acceptable consumers' and producers' risks.

It is simpler to illustrate this by focusing on the number of tests required (rather than the number of shots required), the number of test-failures allowed (rather than number of penetrations allowed), and the probability that a unit will pass the test (rather than the reliability of stopping each shot) .32 Also, for purposes of this discussion, we consider a' 'unit" of armor to be a set of however many identical garments are required for a test-e.g., 4 garments for a 2-caliber wet/dry NIJ test of standard-type ballistic resistance, or 1 garment for a l-caliber wet-only or dry-only test of special-type ballistic resistance. An 8.53-percent probability of passing a 48-shot test corresponds to a 95-percent geometric-mean single-shot probability of passing (the boundary between "bad" and "marginal" armor in the example above[33]), and a 95.3-percent probability of passing a 48-shot test corresponds to a 99.9-percent geometric-

28 For example, a 2-sided, 1-sample Kolmogorov-Smirnov test [45] could be used to test goodness of fit to a binomial distribution, which the number of passes would have if all units had the same probability of passing. It gives an upper bound on the statistical significance—i.e., a significance level—at which a discrete distribution, such as a binomial distribution, may be rejected.

29 If so, the reliability is the limit that the relative frequency (i.e., fraction) of successes is almost certain to approach as then-r of trials increases without bound.

30 One can nevertheless contrive scenarios in which the reliability of an individual unit of an inhomogeneous lot would make sense in the classical paradigm. For example, even though lot 1 may contain only 1 unit of size-38 model A armor, one could argue that it is meaningful to speak of its reliability, because one could, if one wanted, make additional units of size-38 model A armor and test them. This still assumes, however, that their properties-including the invisible ones being tested—would be identical.

31 The reliability of an individual unit is a meaningful concept in the *Bayesian* paradigm of statistical inference [11, 80, 81].

32 Otherwise, it would be necessary to introduce such arcane concepts as the arithmetic mean (i.e., the average) of the geometric-mean single-shot probabilities of passing.

33 See *Increase Total Shots and Allow Penetrations,* above.

mean single-shot probability of passing (the boundary between "marginal" and ''good' armor in the example above) .34

Suppose now, for example, that a lot consists of 10 units, that 2 of the units are selected randomly and tested, and that both pass. Exact l-sided binomial confidence limits on the average passing probability are easily calculated for this case;[35] the average passing probability is at least 0.0853 with 99.3-percent statistical confidence. If the average passing probability were no greater than 0.0853, there would be no more than a 0.7-percent chance that the results would have been as good as those obtained. Thus the consumers' risk is only 0.7-percent.[36 37]

There is, however, a greater risk that one or more or the units in the lot has a passing probability lower than 0.0853. The probability of a pass (the reliability of the lot) is the sum (over all units) of the probability that the unit will be selected times the probability that it will pass if tested. Each unit has the same probability of being selected: the reciprocal of the lot size. Thus probability of a pass is the average of the individual probabilities of passing. In the present example, the 2 units tested could each have a passing probability of 0.4265 while the 8 units not tested could have a passing probability of O, and the average passing probability would be 0.0853. By such calculations one may deduce lower confidence limits on individual passing probabilities from the lower confidence limits on the average passing probability. In general, individual passing probabilities may be much lower than the average passing probability, at the same confidence level, especially if the lot size is much larger than the sample size. In contrast, confidence limits on the average passing probability are insensitive to lot size, but sensitive to sample size.

If a maximum acceptable consumers' risk and a maximum acceptable producer's risk are specified, one may prepare a control chart, such as the example shown in figure E-11, to indicate whether a lot must be rejected to limit the consumers' risk or accepted to limit the producers' risk. The chart is for l-percent maximum consumers' risk of accepting a lot with a passing probability worse than $0.95^{48} = 0.0853$ and l-percent maximum producers' risk of rejecting a lot with a passing probability better than 0.999* = 0.9531. These illustrative values are arbitrary; similar charts could be prepared for other choices. Figure E-12 shows how the control limits (the boundaries of the must-accept and must-reject regions) change as the maximum acceptable consumers' and producers' risks are increased to 5 or 10 percent.

What should be done if the test results lie in the discretionary region between the lower and upper control limits? In the interest of reproducibility, such a decision should not be made arbitrarily on a case-by-case basis; a policy (even if arbitrary) governing such cases should be established. One option would be to require testing to continue; this might well consume all the armor in a lot, but it would not violate either the maximum acceptable consumers' risk or the maximum acceptable producer's risk. Another option would be reject the lot; this would be consistent with a desire to minimize consumers' risk without exceeding the maximum acceptable producer's risk. The opposite extreme would be reject the lot; this would be consistent with a desire to minimize producer's risk without exceeding the maximum acceptable consumers' risk. Many other policies are conceivable; the choice would be a value judgment for NIJ.

To recapitulate, specification of sample sizes implies a judgment about the risk NIJ will accept of accepting a lot with more than a maximum allowable percentage of "defective" units. (See box E-l.) A clearer alternative would be to specify the maximum acceptable risks explicitly and a means of calculating the sample sizes they require in specific cases (e.g., for sequential testing).

34 A better definition of ''bad'' would include a trauma-survivability criterion, for example: For purposes of this standard, "bad armor'' is armor having a (geometric) mean stopping probability of no greater than 0.95 or a probability per shot of exceeding the backface signature limit of greater than 0.05.

35 The 1-sample Kolmogorov-Smirnov test [19, 45] also provides 1-sided confidence limits on the average passing probability, but they are conservative, not exact, for discrete distributions such as the binomial distribution.

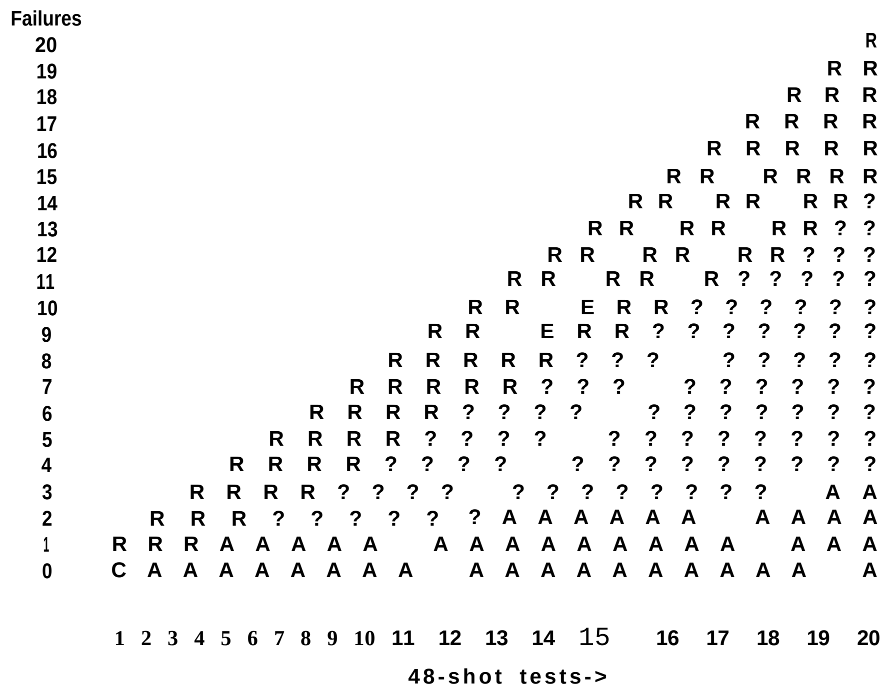
36 This is also the significance level—i.e., probability of error—at which one can reject the hypothesis that the lot is bad—i.e., has a probability of passing lower than 8.53 percent.

37 If the ballistic test were a V_{50} test (or some other test that results in a' 'score" rather thanapass or failure), a 1-sided, l-sample Kolmogorov-Smirnov test [19, 45] could be used to calculate a kind of consumers' risk or significance level-the probability that the actual distribution of V_{50}s in the lot exceeds the empirical distribution of measured V_{50}s (i.e., is worse) by some specified margin.

Figure E-n-Example of Control Chart for Acceptance Testing

1% Consumer's Risk, $p_B = .950^{48} = 0.0853$

1%. Producers' Risk, $p_G = .999^{@} = 0.953$



Legend:
R= REJECT—Consumers' Risk too great if accepted
A= ACCEPT—Produoers' Risk too great if rejected
?=Could ACCEPT or REJECT
C= Conflict must ACCEPT and REJECT (so require more tests)
p^B=maximum probility that bad armor will pass(definfiionof bad armor).
p_G=minimum probability that good armor will pass (definition of good armor).
SOURCE: Office of Technology Assessment 1992.

Sample Selection

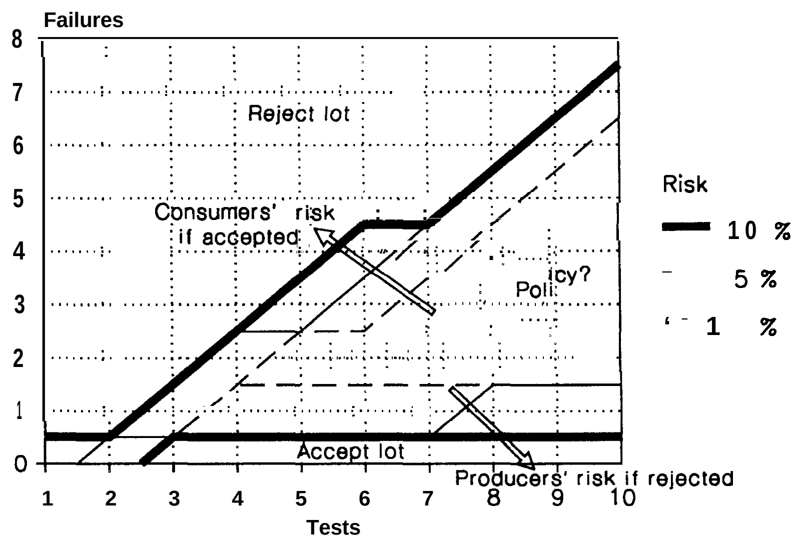
A lot-certification process could require a lot submitted for sampling and testing to be inventoried, tagged, and sampled by (or as prescribed by) NIJ, and the samples to pass a sequential test such as that described above. The armor need not all be shipped to NIJ; it could be inventoried and sampled on the manufacturer's premises by an agent of NIJ. The samples would be sealed and shipped for testing, while the balance of the armor would remain sealed on the manufacturer's premises until the samples are certified to have, or found not to have, the specified level of ballistic resistance,

All armor labeled as belonging to the lot would have to be inventoried. Marketing a unit of armor labeled as belonging to a lot that has been certified when in fact the unit was kept aside from, or produced after, the NIJ inventory and sampling would be false and deceptive labeling, an offense punishable under existing statutes enforced by the FTC. However, detecting such a practice would require a government surveillance program, which could be run by NIJ. It might require undercover purchases on the open market, which might require substantial funding, unless sellers agree to reimburse the costs of obtaining the samples randomly.

Quality-Control Options

Some manufacturers have extensive in-house quality-control programs; here we consider how purchasers and wearers could be assured of product

Figure E-12—Testing More Samples Can Reduce Both Consumers' and Producers' Risks



This figure shows the boundaries between the rejection, indeterminate, and acceptance regions of figure E-1 1, as well as boundaries for 5-percent consumers' and producers' risks and for 10-percent consumers' and producers' risks. (For all cases, $p_B = 0.0853$ and $p_G = 0.953$.)

SOURCE: Office of Technology Assessment, 1992.

quality by an independent third party, such as NIJ, with expertise and a vested interest in quality assurance, and none in armor sales.

In general, the testing and certification could be done by the government or by the private sector (e.g., UL or HPWLI), with or without government (NIJ or OSHA) supervision, and could be voluntary or compulsory. However, a compulsory program, such as would be authorized by enactment of H.R. 322, might be limited to inspection and ballistic testing of products (e.g., lot certification). The alternatives described in this section would require intimate access to the manufacturing process and the cooperation with the manufacturer; they are probably only feasible if voluntary.

An alternative to certifying lots is to certify models (as is now done) and also test samples of units of certified models produced after certification to decide whether they differ significantly from the samples tested for model certification. If they do, certification of the model would be suspended until the production process is corrected. If the decision is made by statistical inference, this is called statistical process control (SPC). Other options rely more on inspection-of samples of armor as well as the production process—and less on ballistic testing, to attain a desired level of confidence in product quality.

In one option for SPC, NIJ would require V_{50} measurements [38] as part of the certification test, to provide a baseline against which V_{50}s of future samples of the same model could be compared to check consistency of physical properties. However, certification of a model would not depend on the measured V_{50}s; it would continue to depend on a test of ballistic resistance, such as those specified by NIJ Standard 0101.03.

At least two V_{50}s would have to be measured in certification testing to establish upper and lower control limits-values within which V_{50}s of later samples must lie if they are to be considered consistent with the samples tested for certification. The upper and lower control limits would also depend on certain assumption-e.g., that V_{50}s of baseline samples are normally distributed-and on how many standard deviations from the mean the

38 As specified by MIL-STD-662D [138]; see also app. C.

Box E-l—Lot Sampling and Acceptance Testing in NILECJ-Std.-0101.00

NILECJ-Std.-0101.00, unlike later versions of the standard, contained a section (4.1) on quality assurance and an appendix (A) on sampling. [141] The apparent purpose of these sections was to provide guidance to manufacturers, retailers, and, especially, purchasers, who might want to specify quality-assurance provisions in a purchase agreement. The text of the standard specified ballistic tests, suggested procedures and sample sizes for lot testing, but did not describe the certification process, Apparently the NILECJ considered certification of lots, but left the definition of "lot" so vague that a rnanufacturer could call his entire production of a given model a "lot," The standard recommended that a sample of more than one unit should be tested if the lot size was larger than 8 units. However, the de facto certification process required a sample of only one unit from a lot of arbitrary size. This violated the only explicit quality-assurance requirement of NILECJ-Std.-0101 .00:

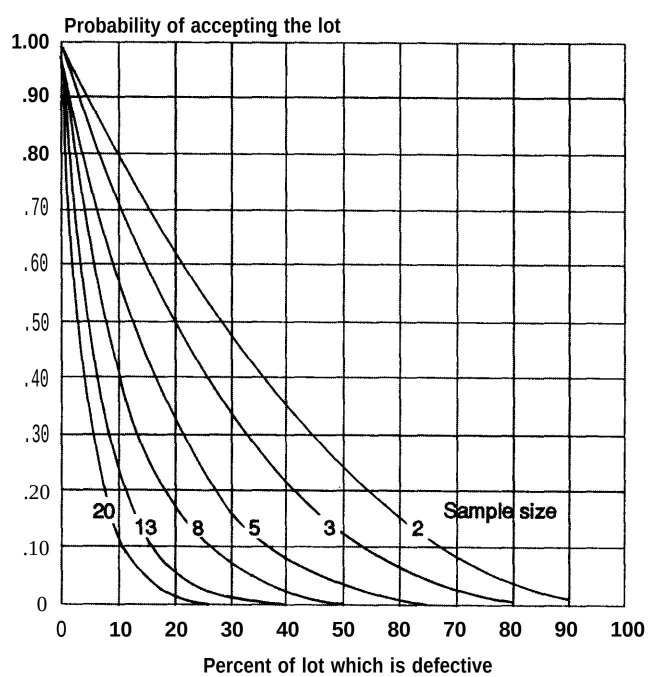
> A ssmple of each lot shall be taken for test at random, using a table of random numbers or an equivalent procedure.

If the entire production (including future production) of a given model is considered to be a lot, then one cannot, in the present, select a sample from it at random for testing. In effect, this "random sampling" requirement, the essence of which survives in the current standard, precludes considering the entire production of a model to be a lot Hence we consider certification of compliance with NILECJ-Std.-0101 .00 or its successors to be a design certification rather than any sort of lot certification--that is, it attests to the potential ballistic resistance of units of a certain design but provides no information on the actual ballistic resistance of production units. Section 4.1.1 of NILECL-Std.-0101.00 provided the following advice on sample size:

> The number of complete armors selected for test from each lot may be in accordance with the table below. This table is considered to be a reasonable compromise between an acceptable level of quality and the cost of testing. However, any desired sample size maybe selected by the purchaser, and should be specified in the purchase document. For a discussion of statistical considerations, see appendix A.

The standard recommended a sample size of 1 unit for a lot size of 1 to 8 units, and a sample size of 20 units for a lot size of 151 or more units. The recommendations imply judgments about the acceptability of risk as indicated in figure 4 of appendix A to the standard reproduced here.

Effect of Sample Size on the Probability of Accepting A Lot, As a Function of the Percent of the Lot That Is Defective



SOURCE: National Institute of Law Enforcement and Criminal Justice, 1972.

control knits should be, which can be deduced from the maximum probability of error allowed in deciding that the production process is ''out of control" when a sample's V_{50} falls outside the control limits. A typical but arbitrary choice is to choose upper and lower control limits 3 standard deviations above and below the mean; these are called '3-sigma' control limits. [31] Only 0.3 percent of the V_{50}s of samples produced by a process "in control" would lie outside 3-sigma control limits, if the V_{50}s of baseline samples were indeed normally distributed.

Once the control limits are established based on certification test results, samples of units of the model produced thereafter would be selected randomly (e.g., each unit produced having a l-percent chance of being selected) and their V_{50}s would be measured. If the V_{50} of any sample is outside the control limits, the production process would be judged to be out of control, and certification of the model would be suspended until the production process is corrected (so that sample V_{50}s again fall within the control limits).

Control limits based on certification test results could be used for other purposes, even if NIJ did not want to use them for SPC. For example, purchasers could use them as benchmarks for acceptance tests: A purchaser could make acceptance of a lot contingent on samples having V_{50}s within the control limits, or above the lower control limit. They could also be used to investigate the possibility of false or deceptive labeling: For example, if armor of a certified model failed to perform as rated in service, its V_{50} could be measured and compared to the control limits. If outside, it would indicate that the production process was out of control when the unit was produced, even if inspection revealed the failed armor to be identical in appearance to the units submitted for certification testing.

Advocates of V_{50} tests for quality testing propose that nondeformable fragment-simulating projectiles (FSPs) [139] be used, instead of bullets, for the V_{50} tests, because, being machined from steel instead of cast from lead, they are more uniform (and more penetrating) than any bullet,[39] and FSP V_{50}s of similar samples generally have less variance, than do ballistic V_{50}s of similar samples. However, they also cost more (a .22-caliber FSP costs about $1.50), and the 3-sigma control limits for ballistic V_{50}s are no more likely to be exceeded than are 3-sigma control limits for FSP V_{50}s of similar samples, although the former would be farther apart.

An advantage of using V_{50} tests, instead of pass/fail tests, for SPC is that many fewer tests (or shots) are required to establish control limits or thereafter discern an anomaly in quality at a specified level of statistical significance. One could, for example, calculate 3-sigma control limits for the number of passes (O or 1) of one .03 test, but this test statistic would not be normally distributed.[40 41] The number of passes in 30 or more .03 tests would be approximately normally distributed, but obtaining such a statistic would require submission of 180 samples of armor, and shooting at least 120 of them!

Thus FSP V_{50} tests are an economical means of detecting a *statistically significant change in* armor and are used for this purpose by the military and by some manufacturers of police armor. However, a statistically significant change in FSP V_{50} may or may not denote an unacceptable change in the type of ballistic resistance in which confidence is sought. A statistically significant change in FSP V_{50} would be grounds for subjecting additional samples to inspection and ballistic-resistance testing, but not necessarily for concluding that ballistic resistance has become unacceptable. The converse should also be considered: an unacceptable change in the type of ballistic resistance in which confidence is sought may not be reflected in a statistically significant change in FSP V_{50}. Experts believe that it would, but it would be difficult to prove that it would, for all types of bullets and armors.

FSP V_{50} tests may be more acceptable to some purchasers and wearers for SQC than certification-type tests or ballistic V_{50} tests, for psychological reasons:

1. Because the tests are different from the certification test, manufacturers might approach periodic retesting without the trepidation some

39 See, e.g., T.A. Abbott, "The Variation of the Geometry of Fragment Simulators,'' pp. 205-218 in [134].

40 Binomial confidence limits could be used in this case, if the probability of passing were assumed to be constant when the process is in control, Or a Kohnogorov-Smirnov test in any case.

41 Another issue is that, for the process to be "in control," the probability of passing would have to be 99.7 percent—much higher than is necessary for armor to have better than even odds of being certified.

feel when contemplating repeated testing with the NIJ .03 test.
2. Purchasers and wearers who might be wary of armor certified to have been penetrated by bullets (as in a ballistic V_{50} test) might accept armor certified to have been penetrated by FSPs, which are laboratory instruments (not bullets like those used by criminals).

Other options rely more on inspection and lesson ballistic testing to attain a desired level of confidence in product quality. Some options rely on inspection of the production process as well as inspection of samples of armor. A voluntary program resembling the Classification program of Underwriters Laboratories (UL) would be based on the following principles:[42]

1. Testing to a nationally recognized standard.
2. Publication of the test results in a report that includes a comprehensive description including photos and drawings of the products.
3. Publication of a list of manufacturers and specific products that have demonstrated by tests compliance with the requirements.
4. Factory follow-up inspections at least four times a year using the report described in item 2 to assure that production units are identical to the unit which was submitted for and passed the testing.
5. Annual sample retest—this involves selection of a representative sample during one of the inspection visits and returning it to the test laboratory for retest to assure continued compliance.
6. Products produced under such a program would carry the mark of the third-party certification laboratory. This would facilitate user identification of those products that have been deemed to be in compliance with the standard.
7. The test laboratory shall maintain tight control of its mark. Compliance failure at either the factory follow-up inspection, item number 4, or annual retest, item number 5, would require corrective action, removal of the certification mark, or holding of shipment of the affected units. Additionally, certification marks could easily include lot traceability identifiers which could facilitate a recall as a last resort.

A manufacturer seeking to have a product Listed or Classified by UL pays UL to inspect and test initial samples of the product to determine whether the product meets UL standards for safety from fire and electrical shock (e.g., in the case of Listing) or some other standard (in the case of Classification). If so, and if the manufacturer agrees to allow (and pay) UL to conduct a limited number of surprise inspections of the manufacturer's production and quality-control processes (including some tests of randomly-selected production items), then UL Lists or Classifies the product, and permits the manufacturer to affix a seal ("mark") indicating that the product is Listed or Classified by UL[43]

The cost of UL or UL-like procedures for assuring the quality of body armor would depend on the standard to which they should comply, which in turn might specify how samples are to be selected, inspected, and tested, and the confidence (if any) with which the tests are to assure that the samples are identical to the original test articles or, in any case, provide the ballistic resistance required.

One option would be to test intitial samples for model certification in accordance with NIJ Standard 0101.03 or a similar standard, and thereafter to base certification of product quality (viz., similarity to the initial samples) on audits of the manufacturer's production and quality-control processes and on selection, inspection, and ballistic testing of production samples.

The feasibility of intitial testing by UL was demonstrated in June 1988, when UL conducted a series of tests of body armor for TAPIC in accordance with NIJ Standard 0101.03. The testing was overseen by a staff member of the NIST Law Enforcement Standards Laboratory to verify that the work was in conformance to the .03 standard and consistent with its interpretation at LESL. UL now estimates that such initial testing of a model could be performed for about $3,000 and about $1,500 for each additional model from the same manufacturer) tested at the same time.

An ongoing followup inspection program typically involves a basic annual charge of $435 plus an inspection fee of $72 per hour spent by the UL inspector at the manufacturing facility. UL estimates

42 Isaac I. Papier, Managing Engineer, Burglary Detection and Signalling Department, Underwriters Laboratories, Inc., personal communication, Aug. 5, 1991.

43 Today, UL Lists no armor garments but does test and certify a broad range of products that provide ballistic protection.

that a basic followup service for NIJ-like armor Classification would require 4 annual visits, each about 1 or 2 hours long, if the manufacturer's quality-control program is in good order. On one of the visits, the UL inspector would select random samples (not necessarily including samples of all models) for testing, the cost of which would be extra but much less than that for initial testing, because not all models would be tested and no report would be generated. [112] Hence the recurring annual cost to a manufacturer could be little more than about $700 to $1,000.

This option would provide neither quantitative estimates of the confidence in the program nor (the other side of the coin) of the probability of failure-i.e., the probability that a unit of production armor Classified by UL as complying with the standard of ballistic resistance actually does not (or fails a ballistic test, which is not quite the same thing). Some manufacturers might hesitate to participate in it, because they would perceive the unannounced factory inspections as intolerably intrusive.

Although this option for UL Classification would not provide purchasers of UL-Classified armor with quantitative estimates of risks, other options could. For example, lot-acceptance testing and certification, as described above, could be done in the context of UL Classification if the NIJ standard were revised to apply to lots instead of models.

If NIJ reconsiders UL Classification or an analogous option and solicits bids for such a program, several independent test laboratories might respond by proposing programs.