Managing Risks: Statistically Principled Approaches to Combat Helmet Testing

Janice Hester, Thomas Johnson, and Laura Freeman

PROBLEM

Combat helmets protect troops against artillery rounds, mines, and small caliber bullets. Helmet designers strive to achieve high ballistic protection with lightweight helmets. Modern combat helmets are made from dynamic materials such as aramid and ultra-high molecular weight polyethylene fibers, which show more variability in performance than simpler armors. The Services conduct acceptance tests to evaluate the ballistic performance of each helmet design and production lot. The challenge to testers is to construct efficient tests to determine whether these helmets meet performance criteria.

BACKGROUND

Combat helmet designs are driven by the balance between increasing ballistic protection and decreasing weight. Starting in World War I, troops wore steel helmets to protect against artillery rounds. In 1985, the Personnel Armor System for Ground Troops (PASGT) helmet was fielded. The PASGT helmet was made from a laminate of ballistic material with aramid fibers, and it improved protection against fragments. In 2002, the U.S. Army replaced the PASGT helmet with the lighter weight Advanced Combat Helmet (ACH). The ACH and similar helmets are currently the most common helmets worn by U.S. troops. Recently, the U.S. Marine Corps developed the Enhanced Combat Helmet (ECH), which has a ballistic laminate of ultra-high molecular weight polyethylene fibers and provides some limited protection against small caliber bullets. Helmet designs continue to evolve, and the U.S. Army is pursuing two new helmet types – one that provides the protection of the ACH but is lighter weight and another that provides the protection of the ECH but is lighter weight. Figure 1 shows the evolution of combat helmets through the years.

> **IDA's analyses were central to the development of the most recent version of the improved, statistically principled acceptance test protocols for combat helmets.**

Beginning in 2007, congressional concern about the accuracy and consistency of body armor testing led to increased involvement in personal protective equipment by the Director, Operational Test and Evaluation (DOT&E). To address the concerns of Congress, DOT&E worked with the Services to develop test protocols for the ballistic components of First Article Testing (FAT) and Lot Acceptance Testing (LAT) for both body armor and combat helmets. In 2009, DOT&E asked IDA to expand its support for live fire

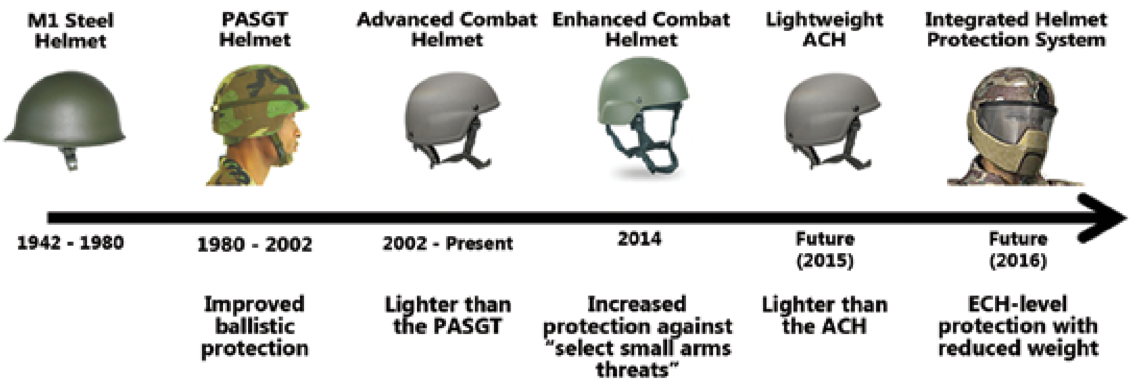


Figure 1. Evolution of DoD Combat Helmets

test and evaluation to include personal protective equipment. IDA's analyses were central to the development of the most recent version of the improved, statistically principled acceptance test protocols for combat helmets.

Helmets must protect against multiple ballistic threats; this article focuses on IDA's work on testing for resistance to penetration and ballistic limit estimation. Our work on evaluating resistance to penetration using statistically principled testing has led to an improved FAT protocol for aramid-based helmets. Our related research comparing newer design methods for fragment testing suggests that additional improvements to the protocols are possible for the estimation of ballistic limits. The statistical work discussed in this article is supported by frequent observations of helmet testing and continual analysis of helmet test data, which together ensure that the statistical studies are relevant to helmet testing.

RESISTANCE TO PENETRATION

Combat helmets must demonstrate a high probability of stopping perforation from a 9mm handgun round, and some designs must also prevent perforation from another specified small arms round. Each helmet design comes in at least four sizes, and during FAT they are shot at five locations on the helmet and subjected to four separate environmental conditioning treatments. The FAT must provide confidence that all helmet sizes have acceptable performance under all test conditions. The primary statistical challenge for this component of testing is to design an efficient test that provides this confidence while still achieving a low risk of rejecting helmets with good performance.

The response of a combat helmet to a threat impact is stochastic, so resistance to penetration is characterized as the probability of a projectile completely penetrating through the helmet. This probability should be very low. The probability of penetration can be measured with increasing precision as the number of test shots increases, but helmet testing is expensive and destroys the tested helmets. Accordingly, tests should be efficient in the number of test articles they require.

Acceptance test designs should balance the risks of wrongly accepting a product that performs poorly and of rejecting a product that performs well. One important statistical tool for comparing acceptance test design is an operating characteristic (OC) curve, which shows the probability of accepting a helmet (passing the test) as a function of the true probability of penetration. Figure 2 shows OC curves for three notional tests that range in size from 75 to 450 test shots. The numbers of shots and allowable penetrations determine the shape of the curve, including the government's risk of accepting helmets with low performance and the manufacturer's risk that helmets with high performance will be rejected. The curves in Figure 2 all have the same manufacturer's risk; increasing the test size results in a steeper OC curve and decreases the government risk.

A helmet design's resistance to penetration can vary among the helmet sizes or across test conditions. A test with a single acceptance criterion on the helmet design's performance across all sizes and test conditions is therefore not sufficient. Instead, helmets must demonstrate performance across all sizes and conditions, which tends to increase the probability of incorrectly concluding that a helmet does not meet performance criteria.

IDA developed an analytical framework for resistance to

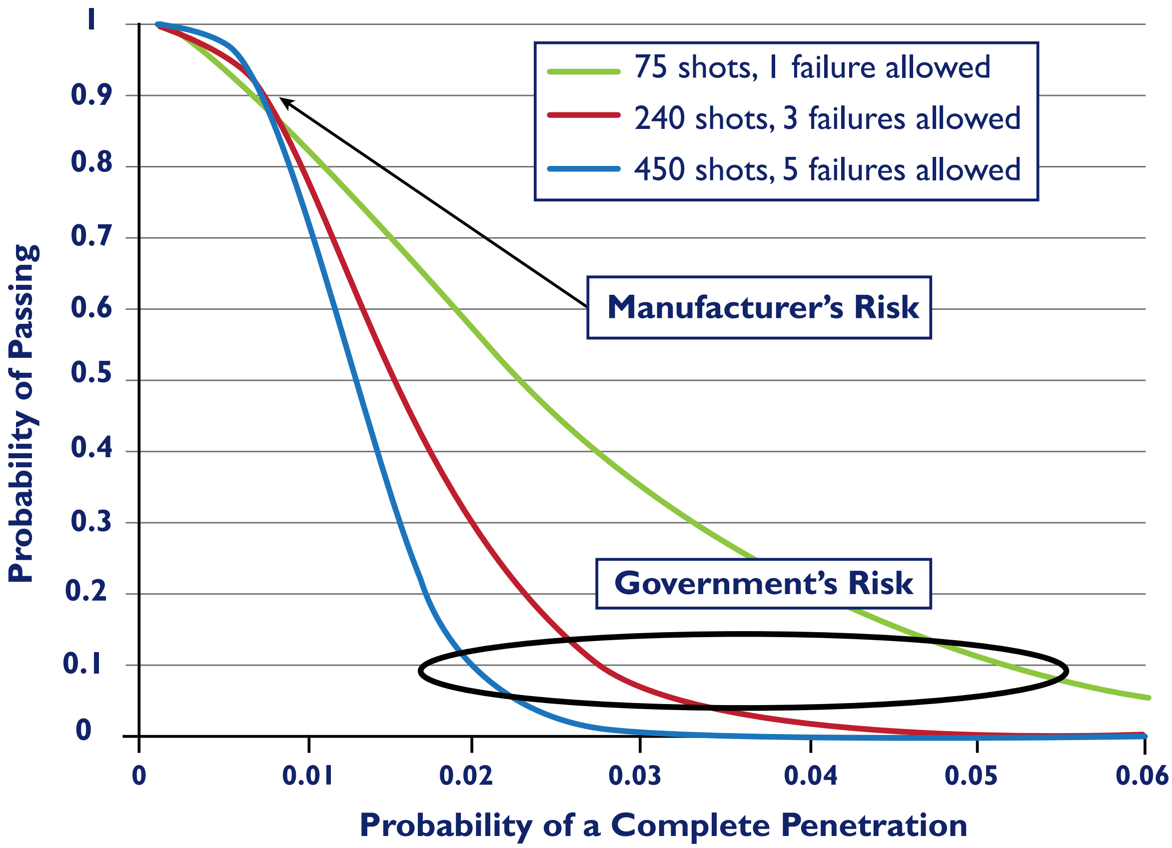


Figure 2. Operating Characteristic Curve for Test Sizes Ranging from 75 to 450 Shots

penetration testing during FAT that captures the tradeoff between the risks of accepting a helmet that performs poorly in one condition and of rejecting a helmet that performs uniformly well across all conditions. This is in contrast to simpler tests, like those shown in Figure 2, for which government and manufacturer risks are set for two different performance levels of the same characteristic (e.g., a single, aggregated probability of a complete penetration).

Instead of selecting a single pass/fail criterion, we select a set of pass/fail criteria that specify a maximum acceptable number of complete penetrations across all shots taken and a maximum acceptable number of complete penetrations within the shots taken on each individual test condition. For example, in the new protocol for aramid-based helmets, no more than three penetrations for the 9mm round are allowed across all sizes, environments, and locations (240 shots total). Of those three penetrations, no more than two can be in any one size. Similar criteria exist for environment and shot location.

Figure 3 shows the operating characteristic curves for the protocol for aramid-based helmets. The dotted blue curve shows the probability of passing the aggregate criterion (three allowed penetrations across all 240 shots) as a function of the aggregate probability of a complete penetration; the solid green curve shows the probability for each helmet size of passing the criterion on the individual size (two allowed on any single size) as a function of the probability of complete penetration for that helmet size; and the solid red curve shows the OC curve for passing all of the multiple test criteria simultaneously for the simple case in which the probability of a complete penetration does not vary among the helmet sizes or test conditions. Figure 3 illustrates how the statistical methodology IDA developed provides acceptable risk points both when all helmets have uniformly high performance and when one helmet size is different.

The key element of a hierarchical test is that if a helmet has uniform performance across the conditions, then the risk points for the full hierarchical test closely match the risk points for the aggregate criterion alone. The criteria on the individual conditions are selected such that, for a helmet with uniform performance, simultaneously passing the aggregate criteria and failing an individual criterion through random chance are unlikely. The benefit of this approach is that the FAT results are diagnostic and easy to interpret. If a helmet design's aggregate performance is low but uniform across the test conditions, then failing for the aggregate criterion is more likely than failing for one of the criteria on the individual conditions. On the other hand, if a helmet has high aggregate performance but a single low performing condition, then failing the pass/fail criterion on that condition is the most likely result. One drawback to this approach is that the aggregate and individual criteria cannot be specified independently. Finer control over these risk points is possible with more complex test

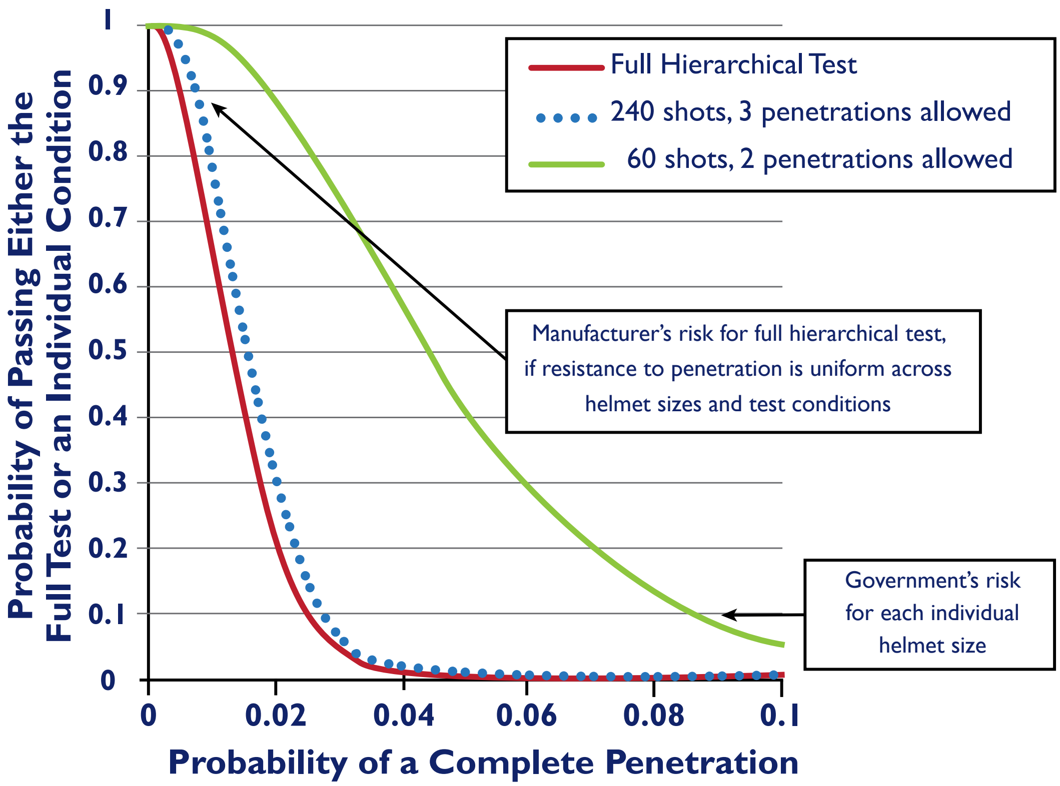


Figure 3. Operating Characteristic Curves for the DOT&E Aramid Helmet Protocol

designs that include the possibility for multiple rounds of testing.

BEHIND HELMET BLUNT TRAUMA

The rapid deformation of a combat helmet following a ballistic impact creates a potential for blunt trauma injury even when the projectile does not completely penetrate the helmet; the deforming helmet shell can impact the wearer's head. To mitigate this risk, the helmet's deformation following an impact with the 9mm test round is measured during testing and compared to established upper limits. Figure 4 shows the image of a head form filled with clay before a shot is taken (left) and after (right); the maximum deformation is measured from the deepest location in the clay indent.

The FAT and LAT protocols include a procedure for assessing the measured deformations against the established upper limits. Deformation requires a different

Figure 4. Clay Helmet Head Form Before the Shot (left) and After the Shot (right) Illustrating the Helmet Deformation into the Clay Channel

analysis method than resistance to penetration, because deformation is a continuous metric (a measured value) rather than a binomial (success/failure) metric. IDA used simulation studies to investigate the best approach to writing a protocol for deformation that accounts for the multiple test conditions. We showed that Analysis of Variance (ANOVA) can be applied within a FAT to test individual conditions while controlling overall risks.

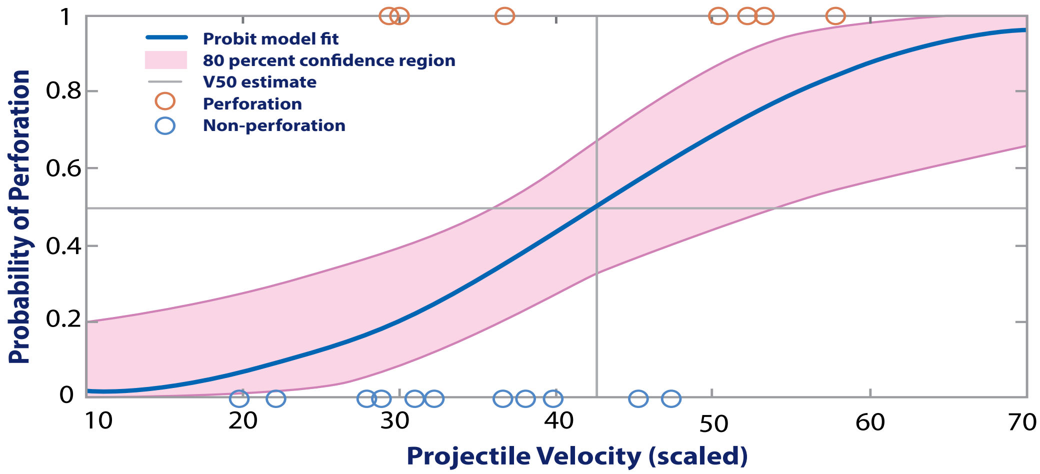
FRAGMENT THREATS AND BALLISTIC LIMIT ESTIMATION

The Services set helmet performance requirements for the minimum ballistic limits against several standard fragment simulants; these limits are incorporated into the DOT&E protocols. The ballistic limit is the velocity at which a projectile completely penetrates the helmet 50 percent of the time. While the velocity corresponding to a lower probability of penetration would be a better measurement of ballistic protection, the 50th percentile has been used historically because it is the percentile that is measured with the greatest precision.

The test and academic communities have developed several different procedures for determining the ballistic limit of armor through testing. IDA performed a simulation study to determine which of six published procedures would be the most efficient and accurate if used for helmet testing. Each procedure combines a set of rules for selecting shot velocities, terminating testing, and calculating the ballistic limit. To ensure that the simulation results were relevant, the simulation incorporated historical helmet performance data.

To estimate the ballistic limit, testers vary the velocity of the test fragment between shots in a prescribed manner with the goal of finding a velocity range in which there is a mix of penetrations (failures) and non-penetrations (successes). The orange and blue circles in Figure 5 are example data for a ballistic limit test; they illustrate the spread in helmet performance for velocities near the ballistic limit.

Under the current test procedures, which are known as the “up-down method,” testers select each shot velocity by increasing or decreasing the velocity based on the previous shot’s outcome. Once testers achieve a predetermined equal number of complete penetrations and helmet successes within a fixed velocity range, they stop the test and estimate the ballistic limit as the arithmetic mean of this set of shots. The up-down method is not statistically rigorous for multiple reasons, but in particular it frequently does not use all of the data to determine the ballistic limit. For example, if eight shots are required to get three successes and three failures in the required velocity range, the analysis throws away the other two data points. Newer test design and analysis methods use generalized empirical model fits based on all the data to both determine the next shot in the test sequence and characterize the probability of penetration as a function of the projectile’s velocity.



Orange circles indicate shots that perforated the helmet while blue circles indicate shots that were stopped by the helmet. Note that there is a velocity range in which the results vary.

Figure 5. Example Data from a Ballistic Limit Test

The most important result of IDA's simulation study was that, regardless of the test design method used to select shot velocities, using model fitting and maximum likelihood estimation along with the associated criteria for stopping the test resulted in a more efficient test than the up-down method. Figure 5 shows a probit model fit to the example test data; this is an example of an empirical model fit that uses maximum likelihood estimation to assess the probability of penetration as a function of velocity. The measured ballistic limit is an estimate of the true ballistic limit, but also includes error due to variability in the helmet's performance near the ballistic limit. By using maximum likelihood estimation, the ballistic limit can be estimated with fewer shots on average without increasing either the bias (the difference between the measured and the true value) or the variance in the estimate.

Misestimating helmet performance prior to testing can lead to poor choices of fragment velocity during testing, which can increase both the dispersion of and the bias in the ballistic limit estimate. Our simulation study demonstrated that test designs that use generalized linear modeling (i.e., three-phase optimal design (3POD) and Neyer's Method) to select the shot velocities are less sensitive to initial misestimates of helmet ballistic limit than the up-down method. Figure 6 shows the median bias and interquartile range (25th and 75th percentiles shown as the lines extending from each marker) of the ballistic limit estimates for each method for a range of initial misestimates in the variance; the most desirable result is a bias of zero with a narrow interquartile range. Note that the starting assumptions about helmet performance were intentionally misestimated to show the robustness of each method to having limited knowledge of the actual performance variability around the ballistic limit for the helmet design under test.

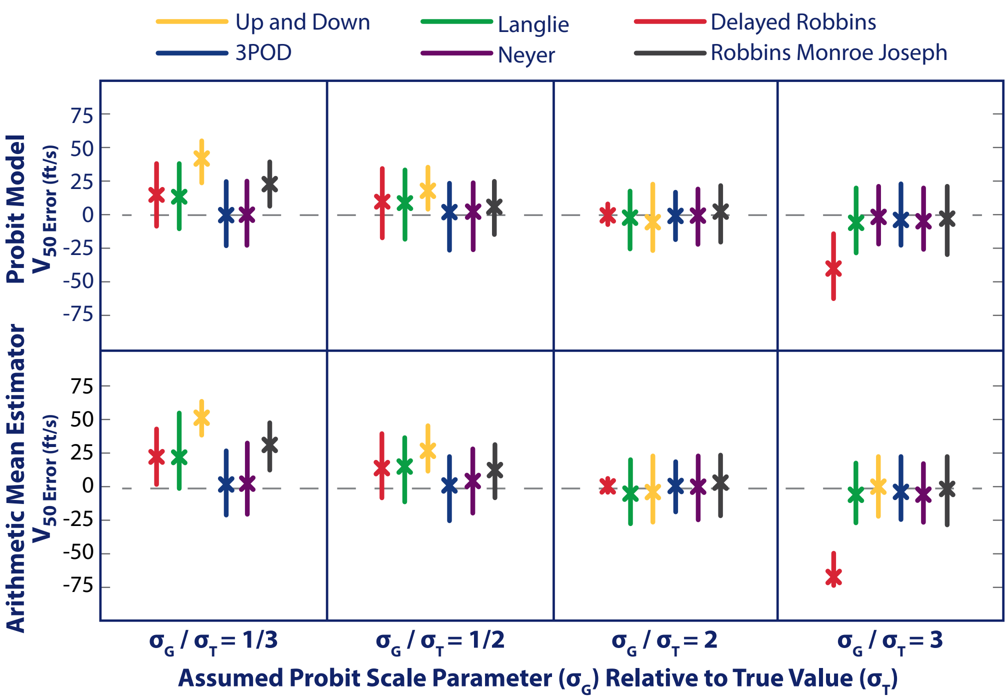


Figure 6. Simulation Results Comparing the Maximum Likelihood Estimate from the Probit Model and the Arithmetic Mean Estimator for Various Test Strategies

CONCLUSION

The Services will continue to pursue lighter helmets and improved ballistic performance. The application of statistically principled test designs will help ensure that new combat helmets have acceptable ballistic performance. IDA has developed innovative design methods that have improved helmet testing protocols for resistance to penetration, while balancing risks to both government and manufacturer across multiple conditions. IDA's research on ballistic limit design and analysis methods shows that further improvements can be made to existing protocols. Making these improvements will ultimately provide a better understanding of helmet ballistic performance, resulting in better equipment for our soldiers.

Dr. Hester is an Adjunct Research Staff member in IDA's Operational Evaluation Division. She holds a Doctor of Philosophy in physics from Princeton University.

Dr. Johnson is a Research Staff member in IDA's Operational Evaluation Division. He holds a Doctor of Philosophy in aerospace engineering from Old Dominion University.

Dr. Freeman is an Assistant Director in IDA's Operational Evaluation Division. She holds a Doctor of Philosophy in statistics from the Virginia Polytechnic Institute and State University.