


September 1996

Risk Criteria for Approving or Rejecting Field Tests of High-Performance Weapons

Shaul Feller

Michael Maharik

Follow this and additional works at: <https://scholars.unh.edu/risk>

 Part of the [Engineering Commons](#), [Law Commons](#), and the [Social and Behavioral Sciences Commons](#)

Repository Citation

Shaul Feller & Michael Maharik, *Risk Criteria for Approving or Rejecting Field Tests of High-Performance Weapons*, 7 RISK 305 (1996).

This Article is brought to you for free and open access by the University of New Hampshire – School of Law at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in RISK: Health, Safety & Environment (1990-2002) by an authorized editor of University of New Hampshire Scholars' Repository. For more information, please contact ellen.phillips@law.unh.edu.

Risk Criteria for Approving or Rejecting Field Tests of High-Performance Weapons*

Shaul Feller & Michael Maharik**

Background

Testing any new technological system involves an implicit understanding: While aiming for success, failure is a legitimate outcome. Failure, or a whole spectrum of possible failures, should be taken into account when designing the test from the earliest stages. This paper discusses the implications for the risk management of modern, high-performance weapon systems testing. "High-performance" in this context is the combination of high maneuverability and potential long range; a guided missile is a typical example.

A primary concern in analyzing potential failures of such systems is the possible deviation from the pre-designed trajectory. The locus of all the points that may be reached by a specific weapon system, or its debris, is referred to as its "kinematic footprint;" a line enclosing this area is defined as the system's "kinematic envelope."¹ In general, the area of the target and its close surroundings, the expected impact zone for a successful flight, are relatively small compared to the kinematic envelope generated by a given combination of a tested system and its launch conditions.

If one insists on absolute safety, he or she must allocate a test area the size of the kinematic footprint and have it totally evacuated. Such a

* The authors are indebted to the late E. Ratzon (1905-1988) for his inspired ideas and pioneering work in the field of risk assessment in Israel.

** Dr. Feller is the Chief Safety Engineer at RAFAEL, Israel Armament Development Authority. He received his M.S. and Ph.D. (Physics) from the Hebrew University of Jerusalem.

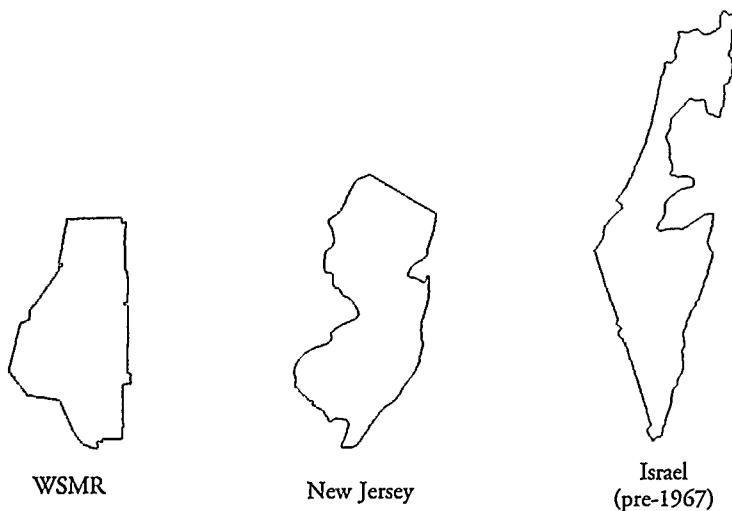
Dr. Maharik is the Safety Engineer, Ordnance Systems Division, RAFAEL. He received his B.S. (Aeronautical Engineering) from the Technion, Israel Institute of Technology, his M.B.A. from Tel-Aviv University, and his M.S. and Ph.D. (Engineering and Public Policy) from Carnegie Mellon University.

¹ This envelope is dictated by the capabilities of the tested system, which, in turn, depend on parameters of its propulsion system, aerodynamic configuration, control and guidance systems, and on the launch or release conditions (altitude, velocity, and angular orientation).

policy was tried at White Sands Missile Range (WSMR), New Mexico, in 1947, following an errant missile impact in a graveyard immediately outside Juarez, Mexico.² However, this conservative policy imposed extreme requirements that could not be met. Abandonment of the policy of absolute safety was indicated, which meant that imposing a certain level of risk on the public was unavoidable. As a result, engineering and probabilistic techniques for weapon-testing risk assessment, which reduced the evacuation requirement to areas smaller than those of the kinematic footprints, were later developed.

Figure 1

Size Comparison on Same Scale: White Sands Missile Range, New Jersey & Israel



The need to reserve large ground areas for weapon testing generates inconvenience even for large countries. For smaller ones, it is a serious problem. The U.S. has allocated an area of 6107 sq. miles for tests at WSMR.³ In contrast, the whole area of the pre-1967 State of Israel was 8031 and New Jersey is 7836 sq. miles. With such a size, one must adopt design approaches other than that of “absolute safety” to perform tests vital to national security.

² Winton G. Hammond & Roy E. Geisinger, *Reducing Safety Constraints Through Vehicle Design*, presented at the American Institute of Aeronautics & Astronautics Meeting on Launch Operations, Feb. 1970 (Paper 70-248).

³ Department of Defense, Under Secretary of Defense for Research and Engineering, *Major Range and Test Facility Base Summary of Capabilities* (DoD 3200.11-D, 1983).

The Israeli Weapon Test and Evaluation Center (WTEC) "Shdema" is located in the south, and comprises 174 sq. miles of practically uninhabited desert. Military training areas adjacent to the test range provide some possibility of enlarging its ground area. However, the kinematic footprint for missile tests may sometimes be considerably larger than even that and include military bases, industrial facilities, and even general population communities.

Thus tests performed at the range may impose risk on these populations. The risk is involuntary and uncontrollable from the point of view of the exposed majority. Moreover, they are unaware of the risk or, at least, details that must be kept classified. In general, people are known to perceive involuntary, uncontrollable and unknown risks as higher than risks that have the same exposure levels, but are voluntary, controllable and known.⁴ These features add even more dimensions to the problem addressed by this paper.

The Problem

The management (or military chain-of-command) of a test range is responsible for its safety. The goal is to prevent personal injury and property damage as a by-product of range activities in and out of the range. In light of this goal, and remembering that absolute safety is an unrealistic requirement, decisions should be made on whether or not proposed test scenarios can be accepted.⁵ The core of the problem is thus the following question: Given the decision-maker's responsibility toward the safety of the exposed population, how should decisions be made? To be more specific, what risk acceptability criteria, if any, should be applied in support of such a decision-making process?

The term "risk acceptability" merits elaboration. One does not accept risk in isolation. One accepts options that entail some level of risk. Whenever the decision-making process takes into account benefits or other (non-risk) costs, the most acceptable option need not be the one with the least risk. "Acceptable risk" can thus be referred to the risk associated with the most acceptable option, or course of actions, in a

⁴ Paul Slovic, Baruch Fischhoff & Sarah Lichtenstein, *Fact and Fears: Understanding Perceived Risk*, in *Societal Risk Assessment: How Safe is Safe Enough?* (R. C. Schwing & W. A. Albers, Jr., eds. 1980).

⁵ A test scenario is the combination of the investigated weapon system, the test design, and the geographical and demographical setting.

particular decision problem.⁶ In our context, we assume that performing the test will benefit the security of the relevant society to such an extent, that some risk related to this option is acceptable.

Table 1⁷
Conditions for Standard-Setting

#	Condition ^a	Status ^b	Comment ^c
1	No choice among options is possible	+	Given geographical constraints, no other operative option is available
2	No selection among options is required	+	Derived from #1
3	Predictability is important	+	Tests are complicated and expensive; decision has to be made much earlier than date of operation
4	Regulators hope to shape future options	+	Decision policy should take into account future changes in technologies and population distribution
5	Competing technologies fall in the same jurisdiction	+	Almost all weapon tests are subject to the Ministry of Defense decision criteria
6	Category members are homogeneous	+	Most tests relate to the family of aero-explosive technologies
7	An explicit policy statement is attractive	+	Explicit policy is ethically required given security classification constraints
8	Value issues are sensitive	+	Dread, unfamiliarity, and lack of knowledge are concerned
9	Political resources are limited	+	The Defense Community faces numerous other hard decisions
10	Process is unimportant	+	Decision-making process is essentially similar for all cases
11	Awkward applications can be avoided	-	Decision policy sometimes needs further interpretation to meet this condition

a	Condition for favoring standard-setting over a case-by-case decision process.		
b	+ Condition is met; - condition is not met.		
c	Justification for suggested status.		

Once the notion of “acceptable risk” is clear, one may question when it makes sense to set and apply a structured set of criteria, as opposed to a case-by-case decision process (using techniques such as cost-benefit analysis or decision analysis). Fischhoff suggested a set of eleven conditions that favor standard setting.⁸ By closely inspecting

⁶ Baruch Fischhoff et al. *Acceptable Risk* (1989); Baruch Fischhoff, *Acceptable Risk: A Conceptual Proposal*, 5 *Risk* 1 (1994).

⁷ Adapted from Baruch Fischhoff, *Setting Standards: A Systematic Approach to Managing Public Health and Safety Risks*, 30 *Management Sci.* 823 (1984).

⁸ *Id.*

the case of weapon-system testing risks, and applying the principles of his standard-setting discussion, it can be shown that the first ten conditions are met (Table 1). In fact, by following a procedure to be discussed later, the eleventh condition can be met too.⁹

One may add a twelfth condition: Standard setting is preferable in the case of frequent rotation of decision-makers, when consistency is required. Bearing in mind the high turnover rate of military personnel, many of whom are intimately involved in the test approval process, this condition is also clearly met.

The need for structured, quantitative risk acceptability criteria is not unanimously agreed upon by the international weapon testing community. For example, in its time the U.S. Range Commanders' Council (RCC) promulgated the following risk management policy:¹⁰

Acceptable risk levels are not in general established. Risks are minimized to the extent feasible and then, based upon considerations of test objectives and national interest, the tests are performed or rejected.

The RCC has not referred to the question "who defines the national interest." It is therefore assumed that this is up to the local Range Commander. Thus, one could find in the Safety Manual of a specific test range the following application of the general policy stated above, with regard to granting waivers:¹¹

A waiver [of safety policy] is granted [by the Commander, PMTC]... if, in the judgment of the Commander, PMTC, the risk involved is reasonable.

Recently, the policy has become more quantitative,¹² but the new guidelines need further workout to serve as decision criteria.

The United Kingdom has adopted a qualitative criterion, demanding "no significant foreseeable risk." One of the arguments for adopting this policy is that, given a numerical safety standard to achieve, the range user will select his or her assumptions and statistics to meet this standard rather than aim for the safest trial possible.¹³

⁹ Although the roots of our methodology go back to the 1970's, the issue of setting structured criteria is best demonstrated by using Fischhoff's conditions from 1984.

¹⁰ Range Commanders' Council, Range Safety Group, White Sands Missile Range, *Risk Analysis Techniques*, at 2.2. (Doc. 315-79, 1979).

¹¹ D. M. Altwegg, *PMTC Range Safety Handbook*, 4 (1976).

¹² J. R. Morrell, *Eastern Range Regulation 127-1 — Range Safety* (1993).

Several other issues should be contemplated when considering whether the case justifies specific decision criteria.

As stated above, weapon tests within small ranges impose risk on unaware civilian population. Recent years' risk communication policies in Western-type democratic societies call for public involvement in decision-making, especially with regard to issues that relate to public risk. Unfortunately, due to security reasons, such involvement is not feasible in our case.¹⁴ Structured acceptability criteria serve as a means of compensation for the missing interaction between government and the public, provided that they undergo a review process that is both professional and conscientious, even though not open for public debate.

Also, the specific conditions of continuous military threat create, in the case of the State of Israel, a balance of risk and benefit different than that of other, more peaceful states and may justify some increase of its risk-acceptability threshold. A Swiss colleague of the authors, some years back said: "You're 'lucky'; you are in a state of permanent war. Your standards may legitimately be more relaxed than ours." Otway and Winterfeldt also suggest that a risk that might be unacceptable in peacetime may be found acceptable in a wartime defense industry.¹⁵ In such circumstances, the enhanced imposed risk emphasizes the need for a compensatory-structured decision tool.

Besides ethical considerations, public trust is an issue. Trust is known to be fragile and "asymmetric": it is created rather slowly, but can be destroyed by a single mishap or mistake. This feature is sometimes referred to as a "tilted playing field".¹⁶ A severe conflict may arise if, following an accident, the public is confronted with an indefensible decision process on issues involving imposed risk. Formulating structured risk acceptability criteria contributes to the trust preservation by serving as a "one-sided contract" that the Defense Community has with the public. The "contract" serves to ensure that no arbitrary decisions are made by individual range commanders.

¹³ T. G. Wills, *Range Safety*, in *Design for Safety* (Royal Aero. Soc'y, 1991).

¹⁴ In fact, though seemingly foreign to democratic process, tight security by itself may eventually save life.

¹⁵ Harry J. Otway & Detlof von Winterfeldt, *Beyond Acceptable Risk: On the Social Acceptability of Technologies*, 14 *Policy Sci.* 247 (1982).

¹⁶ Paul Slovic, *Perceived Risk, Trust, and Democracy*, 13 *Risk Anal.* 675 (1993).

Finally, the population types exposed to risk also need consideration. Different types (such as soldiers or civilians) are exposed to different levels of risk during their day-to-day life. This state of affairs is acceptable to society, at least implicitly. If a criterion based on the background risk level of each exposed person is chosen, then these differences should be taken into account.

The Tool

Given pros and cons, above, we have adopted a policy of applying quantitative measures for go/no-go decisions related to proposed tests. A tool that reflects two commonly-agreed risk components (personal and group) has been developed. The first component, i.e., the risk to a given person, is a function of the investigated system, the size of the area that it can affect, and the presence of that person inside or outside the affected area, regardless of the presence of other persons there. The second component, i.e., the risk to society, depends (in addition to the above) on the number and distribution of people in the relevant areas. Thus, the decision tool is a set of benchmark number pairs. Each pair consists of two figures — the maximum allowed Individual Risk (IR) and the maximum allowed Societal Risk (SR):

- The IR is the probability that a given member of the exposed population becomes a fatality, due to weapon testing in the range, over a given time period.
- The SR is the statistical expected number of fatalities due to weapon testing, within the same population and over the same time-period.

Each pair corresponds to a specific type of the exposed population:

- Nonparticipating, uninformed general population (GP);
- Nonparticipating, uninformed workers in industrial facilities (IW);
- Defense-community nonparticipating and uninformed personnel (DN); and
- Defense-community personnel who are taking part in the test and are informed about the risks (DI).

Determination of the maximum allowed IR is based on a “ripple principle,” stating the following:

The existing background risk level of any population is not a constant. Rather, it is modulated by some ripple. Such a ripple reflects statistical variations of the annual risk levels

and also local fluctuations caused by governmental or municipal decisions that are justified by the general public's benefit. It is required that the integrated test-generated risk increment, contributed to the existing background risk level of a population by all the tests conducted over a given time-period, will not raise the average risk level of the most exposed members of that population beyond the ripple that modulates its background risk anyway — by “most exposed,” we literally mean a very small group of people.

Given the ripple principle, the procedure by which the maximum allowed IR was determined was:

- The level of risk to which the population of Israel is in general exposed due to traffic accidents, industrial accidents and accidents at home, was investigated. The average was about 3×10^{-4} fatalities per person per year, while the amplitude of its annual fluctuations throughout the recent three decades was about 10^{-4} fatalities per person per year.
- The maximum allowed IR for the general public was set to be one tenth of the ripple, i.e., 10^{-5} per person per year; that of industrial workers was set as the ripple itself, i.e., 10^{-4} per person per year. As for the two other population types, the risk to which they are voluntarily exposed is, by definition, greater than that of the general public. Thus, on ethical grounds, maximum allowed IRs of 2×10^{-4} and 10^{-3} were allocated for nonparticipating, uninformed defense-community personnel and for defense-community personnel participating in the test and informed about the risks, respectively.

The value of 10^{-5} , taken as a criterion for acceptable general public IR, is generally consistent with published risk levels and criteria. For example, Otway and Erdmann argued that while accidents having the probability of the order of 10^{-6} per person per year are not of a great concern to the average person, people do recognize risks at level of 10^{-5} per person per year, warn their children about them and may accept a certain amount of inconvenience to avoid them. At an accident risk level of 10^{-4} per person per year people are willing to spend money to control the hazard.¹⁷ Kletz stated that we seem to accept risks from manmade events in the order of 10^{-7} per person per year for large number of people. He suggests that the maximum risk of death to

¹⁷ Harry J. Otway & Robert C. Erdmann, *Reactor Siting and Design from a Risk Viewpoint*, 13 Nuclear Eng. & Des. 365 (1970).

which any individual is exposed should be fixed at 10^{-5} to 10^{-6} per person per year.¹⁸ Okrent and Whipple also proposed the risk level of 10^{-5} per person per year for the most exposed individual as an acceptability criterion for activities that are “beneficial” to society, going up to 10^{-4} for “essential” activities.¹⁹ Specifically, Schneider established a value of 10^{-5} per person per year as a tolerable death probability for nonparticipating third parties in relation to explosive operations.²⁰ Bowen and Kaiser suggested similar values.²¹

Maximum allowed SR values were anchored on 5×10^{-5} annual fatalities allotted for the general public. This was initially chosen as twice the risk from meteorites (6×10^{-11} per person per year worldwide, i.e., 2.4×10^{-5} fatalities per year for the relevant 4×10^5 members of the general public exposed to test risks in Israel).²² Yet, a structured rationale for this number tries to ensure that when a large population is exposed, the vast majority is exposed to only a very tiny fraction of allowed IR levels. Thus, once the maximum allowed IR has been dictated, the maximum allowed SR serves as a control to limit the number of people exposed to high (though still allowed) levels of IR. The lower the actual IR, the larger the allowed number of people exposed.

This “exposed population limiting rationale” was considered so useful by itself, that it acquired a semi-independent role in establishing the allowed SR value. It was also compatible with our self-imposed obligations in upholding the “one-sided contract” with the public on the trust issue. Thus, the value indicated as a criterion was approximately one order of magnitude lower than numerous SR exposures in developed societies. It was ascertained that this strict goal could indeed be met in practice, given relevant features of our problem.

¹⁸ Trevor A. Kletz, *The Application of Hazard Analysis to Risks to the Public at Large*, in *Chemical Engineering in a Changing World* (W. T. Koetsier, ed. 1976).

¹⁹ David Okrent & Chris G. Whipple, *An Approach to Societal Risk Acceptance Criteria and Risk Management* (1977).

²⁰ T. Schneider, *Some Principles for a Quantitative Approach to Safety Problems in Explosive Storage and Manufacturing in Switzerland*, in *Minutes of the 17th DoD Explosives Safety Seminar*, Sept. 14–16, 1976, at 1445-1472.

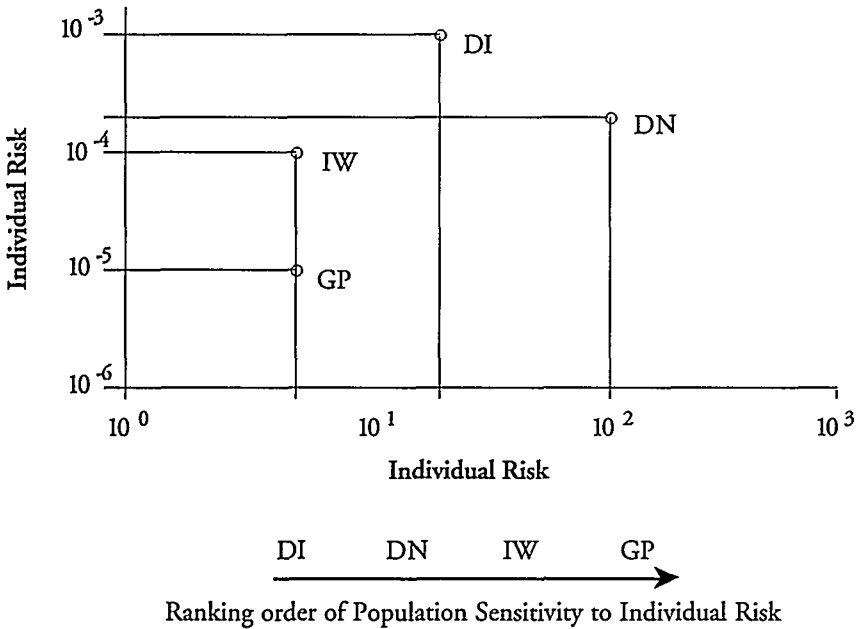
²¹ John Bowen, *The Choice of Criteria for Individual Risk, for Statistical Risks and for Public Risk*, in *Risk-Benefit Methodology* (D. Okrent, ed. 1975); G. D. Kaiser, *Overall Assessment*, in *High Risk Safety Technology* 105 (A. G. Green, ed. 1982).

²² See *supra* note 18.

A similar rationale, applied to industrial workers, set the maximum allowed SR for this population at the 5×10^{-4} level.

Figure 2²³

Location of Population Type on the Number Exposed — Individual Risk Plane



The two population types consisting of defense-community personnel present a special feature. As mentioned, it is assumed that a greater level of IR is tolerated by society for risks imposed on defense-community members than for risks imposed on civilian populations. Between the two defense-community groups, a higher risk is tolerated for participating members, who are intimately and voluntarily involved with the systems under test (DI), than for nonparticipating members (DN). However, as a kind of compensation for the relatively high levels of risks imposed on members of the test teams (DI), the allowed number of exposed people is reduced compared to the nonparticipating defense-community members. Figure 2 emphasizes this “tradeoff” between IR and allowed “Number Exposed” levels for the defense-

²³ GP: Nonparticipating, uninformed general population; IW: Nonparticipating, uninformed workers in industrial facilities; DN: Defense-community nonparticipating and uninformed personnel; and DI: Defense-community personnel who are taking part in the test and are informed about the risks.

community populations. As explained above, the IR-SR method serves as a tool for accomplishing this tradeoff.

The resulting values are in Table 2.²⁴ These values of risk acceptability criteria were approved by the Israeli MoD and are used in the range. Decisions on the acceptability of proposed tests are made on a per-test basis. The maximum risk levels permitted for a single test depend on the annual number of tests to which a specific type of population is exposed.

Table 2
Test Range Risk Acceptability Criteria

#	Population Type	Maximum Allowed Annual Increment of	
		Individual Risk	Societal Risk
1	GP ^a	10^{-5}	5×10^{-5}
2	IW ^b	10^{-4}	5×10^{-4}
3	DN ^c	2×10^{-4}	2×10^{-2}
4	DI ^d	10^{-3}	2×10^{-2}

a Nonparticipating, uninformed general population.

b Nonparticipating, uninformed workers in industrial facilities.

c Defense-community nonparticipating and uninformed personnel.

d Defense-community personnel who are taking part in the test and are informed about the risks.

Table 3 presents single-test acceptable risk levels, based on a set of assumptions regarding the annual number of exposures to risk from testing activities. Experience has shown that testing frequency decreases with increasing system's sophistication and energy. Also, more sophisticated and energetic systems result in larger kinematic envelopes. Thus, the population that exists in the danger areas of such tests is exposed to risk just few times yearly. The policy calls for rejecting risk levels that correspond to less than ten tests per year, even in cases of very infrequent operations.

²⁴ Shaul Feller, *Risks from Test-Ranges in Highly-Populated Environment* (Rafael, Haifa, Israel 1978) (in Hebrew); Shaul Feller, *On Acceptable Risks and Field-Testing* (Rafael, Haifa, Israel 1985) (in Hebrew).

A simplified procedure of decision-making, using those criteria, will consist of the following three (or more) stages:

- Estimating the actual risk number pairs for the proposed test scenario.
- Comparing this estimate with the benchmark values.
- Granting permission to go ahead if the actual pairs are lower than (or equal to) the benchmark values, or, otherwise, rejecting the proposed scenario and considering modifications that may produce lower actual values.

Table 3
Single-Test Acceptable Risk Levels

#	<i>Population Type</i>	<i>Annual Number of Exposures</i>	<i>Maximum Allowed per Test of</i>	
			<i>Individual Risk^a</i>	<i>Societal Risk</i>
1	GP ^b	50	2×10^{-7}	10^{-6}
2	IW ^c	50	2×10^{-6}	10^{-5}
3	DN ^d	200	10^{-6}	10^{-4}
4	DI ^e	200	5×10^{-6}	10^{-4}

^a Assuming a steady population in which the same individuals are exposed throughout the year, i.e., individuals are “accumulating” risk.

^b Nonparticipating, uninformed general population.

^c Nonparticipating, uninformed workers in industrial facilities.

^d Defense-community nonparticipating and uninformed personnel.

^e Defense-community personnel who are taking part in the test and are informed about the risks.

Applications

The following examples, all of which are taken from actual design and operational activities in the test range, exhibit some of the possible applications of the criteria described above.

Determination of Risk/Population Parameters

Suppose that the risk is homogeneously distributed within the kinematic envelope of the investigated system, i.e., the value of the actual IR is a constant over the area enclosed by this envelope. If this value is smaller than the benchmark IR (thus satisfying the first risk acceptability condition), than the ratio (Benchmark SR / Actual IR)

dictates the maximum permissible number of exposed people. For example, if the actual levels of individual risks are the maximum permitted, then the maximum permissible numbers of people exposed are 5, 5, 100 and 20 for the four types of population, respectively (see Table 2). As mentioned above, the lower the IR, the higher the permissible number of people exposed: for example, if the maximum IR in a single test is not larger than 10^{-9} , then, under the assumption of 50 exposures per year, one can permit a presence of up to 1000 people (general public) in the area.

Alternatively, if the number of people exposed to risk is given (again, under the assumption of evenly-distributed risk). In this case, the ratio (Benchmark SR / Number Exposed) dictates the maximum permissible IR, as long as it is smaller than the benchmark IR.

More Complex Risk Distribution

In general, the risk throughout the kinematic footprint is not uniform. Its distribution depends on the probabilities of possible failures of the tested system, and their resulting behavior and trajectories. Each specific failure (or type of failure) results in an impact area that represents assembly and launch tolerances. The probability of impact following such a failure is unevenly distributed within this impact area, with the higher probability densities in its center. However, for certain classes of failures (e.g., when a radar-seeking missile homes on a false radiation source), the impact area may shrink into a single point not necessarily at the center, where the whole probability of impact is condensed.²⁵

A complete probabilistic risk assessment of a proposed weapon field test thus comprises of the following:

- Collecting the relevant information in detail.
- Performing failure analysis of the investigated system (with failure modes and their associated probabilities as outcomes).
- Calculating trajectories following failures, i.e., trajectories deviating from the nominal test plan.
- Estimating impact areas resulting from failures.
- Projecting the proposed test geometry on the test arena, and eliciting the test's kinematic envelope.

²⁵ Although in this case the missile itself operates properly, such scenario should be regarded as a failure of the test system as a whole.

- Calculating IR and integrating SR levels throughout the kinematic footprint of the tested system. For example, if each zone corresponds to a single failure and vice versa, and if the probability of hitting each zone following the corresponding failure is evenly distributed throughout this zone, then the calculation is performed as follows (the equations refer to any single type of the exposed population):

$$IR_i = P_f \times P_{i/f} \times (A_{Ki} / A_i)$$

$$SR = \sum_i (IR_i \times N_i)$$

where:

- IR_i - Individual Risk at zone hit by failure i ;
 - P_f - Probability of the weapon system's failure;
 - $P_{i/f}$ - Probability of failure i , given failure;
 - A_{Ki} - Mean Area of Effectiveness (MAE) following failure i , given impact;
 - A_i - Area of zone i ;
 - SR - Societal Risk;
 - N_i - Size of population at area hit by failure i .
- Decision-making: test approval, test rejection, or further iterations until safety demands are met.

The Real World

In practice, the distribution of risk is even more complex than the above example. A failure (or combination of failures) may occur at any point along the designed trajectory; the impact probability distributions resulting from failures are uneven; and, in addition, overlapping of impact zones are common. To take all these features into account, one must divide the kinematic footprint into "cells," and sum up, for each cell, the contributions of all possible failures, occurring at any point along the trajectory, that may lead to impact inside this cell. A detailed description of this procedure is beyond the scope of this paper. Several specific examples of such a procedure are described elsewhere.²⁶

²⁶ Shaul Feller & Michael Maharik, *Probabilistic Risk Assessment of Weapon-Systems Field-Testing: Accounting for System's Complexity and Unfamiliarity*, in Proceedings of PSAM-II — An International Conference on Probabilistic Safety Assessment and Management (George E. Apostolakis & J. S. Wu, eds. 1994).

Flight Termination System Reliability

The objective of flight termination is to prevent the tested object from transgressing the boundaries of a pre-determined permitted flight sector of the kinematic footprint, by “imposing conditions of zero lift, yaw and thrust.”²⁷ There are several possible methods for terminating a missile’s free flight — cutting it into two or more aerodynamically-unstable parts, ripping its rocket-motor open or activating its warhead (provided one is installed). A Flight Termination System (FTS) greatly reduces the risks imposed by the test but will never provide absolute safety outside the pre-determined sector, due to its own possibility of failure. The schematic description of the probabilistic safety aspect of missile flight termination, shown in Figure 3, illustrates this point.²⁸

Suppose that in a proposed test scenario the actual SR is greater than the benchmark SR, i.e., the integration of IRs gives (Actual SR) = $K \times$ (Benchmark SR), $K > 1$.

As it is, the test is unacceptable and should be rejected. Some modification should thus be made in the test configuration. For example, an FTS can be incorporated into the system under test. Risk will still be imposed on people outside the permitted flight sector, but now, referring again to the previous simplified model, one gets, for every failure mode, Actual $IR_i = (1 - R_{FTS}) \times IR_i$, where R_{FTS} is the FTS reliability. Assuming that all the area within this sector is evacuated so that it does not contribute to the risk, Actual SR = $(1 - R_{FTS}) \times$ SR.

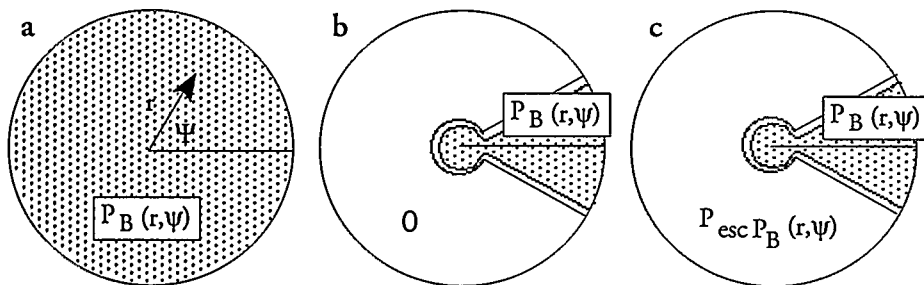
To reduce the actual SR to the permitted level, one must have $1 - R_{FTS} \leq (1 / K)$, i.e., $R_{FTS} \geq 1 - (1 / K)$. For example, if (Actual SR) / (Benchmark SR) = 100, then one must have $R_{FTS} \geq 0.99$. In this way, the set of risk acceptability criteria is used to define a quantitative performance requirement actually applied in designing an engineering safety device.

²⁷ Air Force Systems Command, Armament Division, *Range Safety* (Reg. 127-1, 1984).

²⁸ Adolf H. Knothe, *Range Safety — A Necessary Evil*, Aerospace Engineering, June 1961, at 20-21, 70-76.

Figure 3²⁹

Missile Flight Termination — A Schematic Probabilistic Description



a No range safety action; impact is possible at any point within the kinematic envelope. The impact probability density P_B , however, will vary with distance r from the launcher and azimuth deviation ψ from the intended flight line.

b If range safety action were absolutely reliable, the impact probability outside a given sector could be reduced to zero; inside, it would remain unchanged. However, there would be an accumulation of impact probabilities right along the boundary lines owing to the range safety action taken on missiles that otherwise would have achieved impact outside.

c Since range safety action cannot be absolutely reliable, there is a chance P_{esc} that flight termination command remains without result; a missile may thus escape the range safety action. Consequently, the impact probability density outside the given sector will be as shown.

“Grey” Decision Zones

Risk acceptability criteria such as those suggested in Table 2 are susceptible to criticism that they may lead to rejection of otherwise attractive alternatives, whose risk levels are just beyond the threshold of acceptability.³⁰ Such complaints arise especially when it is very expensive (in terms of money or test data, the latter sometimes of primary importance) to gain the small decrease in risk needed to achieve compliance. Situations like this may create a temptation to overstep the benchmark standard.

²⁹ From Knothe, *supra*.

³⁰ Stephen L. Derby & Ralph L. Keeny, *Risk Analysis: Understanding “How Safe is Safe Enough?”* 1 Risk Anal. 217 (1981).

In our case, a “grey decision zone” is defined if the actual risk levels are found exactly equal to the maximum permitted levels, or within half an order of magnitude above. In such cases, the first thing to do is to re-examine the assumptions at the base of the calculations (e.g., probability distributions of failures and impact areas, annual number of exposures). From our experience, analysts tend to inject large, and sometimes unnecessary, margins of safety, expecting that even with these margins the results will meet the criteria with “room to spare.” In border cases like this, these “generous” margins should not be used, and the assumptions should be defined as accurately as possible.

If, following this re-examination, one still finds himself in the “grey decision zone,” our policy is to re-evaluate the risk-benefit balance of the proposed test and approve or reject the test accordingly. This re-evaluation should only be performed by a higher authority that does not belong to the test establishment.

On the face of it, this policy may seem to serve as a mechanism for carrying out tests that should, by right, have been rejected. Nevertheless, the reader should bear in mind that the core of the matter is not risk per-se but a balance between risk and benefit. The reader should also remember that, in this context, the benefit is preservation of human life. Thus, one deals with a risk-risk balance.³¹ When these risks and “benefits” are very evenly balanced, the problem deserves a further evaluation. As mentioned above, the mandate for the evaluation and resulting decision is transferred to higher authorities.³²

Past Experience

Comparison of the risk acceptability thresholds discussed above with experience gained during two decades of weapon testing activity produces the following observations:

- IR levels in inhabited, or temporarily occupied, areas within the kinematic envelope were always found to be at least one order of magnitude below the maximum permitted level.

³¹ Lester B. Lave, *Health and Safety Risk Analyses: Information for Better Decisions*, 236 Science 291 (1987).

³² This procedure contributes toward the removal of awkwardness from cases located in the “grey decision zones,” and thus to meeting Fischhoff’s eleventh condition for favoring standard-setting (*see* Table 1).

- SR levels for each of the defense-community population categories have never been anywhere near the maximum permitted levels. In fact, given the calculated IRs, one could accept the presence of a much higher number of people — up to several orders of magnitude — than actually present within the kinematic envelope. This margin reflects the relatively high benchmark SR allocated to those populations.
- Yet, when considering risks imposed on civilian populations, the actual SR levels were sometimes found to be inconveniently close to the benchmark numbers. For instance, SR levels of 10^{-6} , 1.5×10^{-6} , 8×10^{-6} , and even 1.5×10^{-5} fatalities per test were obtained in some specific cases, as compared to the permissible 5×10^{-6} level per test (assuming a frequency of ten tests per year).

In some border cases, the test design was changed to meet the requirements. In others, the “grey decision zone” procedure was applied, and the decision was referred to a higher management authority. However, even though the discussed tests were approved and successfully performed, the inconvenience associated with small difference between the actual SR and the benchmark SR has remained on the agenda. This inherent grain of inconvenience is unavoidable. In fact, it serves as an ethical safeguard, by preventing the decision-makers from becoming too complacent.

As for the uncertainties associated with the actual risk values: The estimated basic probability of the weapon system’s failure (P_f , see above) was based on many years’ statistics of missile testing in the U.S. We adopted this probability value, although our experience produced a significantly lower number. In the next step, the uncertainties associated with the models and conditional probabilities were estimated, and the sensitivity of the resulting risks to these uncertainties was assessed. It was found that once the choice of a P_f value was made, the outcome was very insensitive to both model and conditional probabilities uncertainties: less than half an order of magnitude altogether. Choosing, as we did, a conservative SR value, and given our demonstrated weapon systems’ failure frequency, we concluded that our criteria do indeed address the issue of possible uncertainties.

Enlarging the Picture

When dealing with a probabilistic risk assessment one must bear in mind what may be called the “Engineering-Probability Principle” (EPP):

It is legitimate to rely on probabilistic arguments, estimates, and criteria, only when all that is “engineeringwise” feasible and reasonable has indeed been introduced into the investigated system and test design to prevent failure and/or mitigate its outcome.

A strict adherence to this principle is an essential condition of the “one-sided” contract that the Defense Community has with the public, given the existing security classification constraints. In a way, it also serves as a safeguard against a possible tendency, mentioned above, to manipulate statistics to meet numerical standards.³³ The following are several applications of the EPP:

- Typically, the danger areas of most proposed tests can be divided into two zones: a relatively small central zone and a large peripheral one. Probability of impact is concentrated in the former, while the rest of it is sparsely distributed in the latter. The EPP forbids the presence of nonparticipating personnel in the central zone even if probabilistically acceptable. Thus, the central impact zone is always totally evacuated from nonparticipating personnel.
- Prior to a large-scale test, one must always verify that there is no singular concentration of people within the potential impact area, e.g., a central sporting or entertainment event. This contributes toward preventing multi-fatality accidents even when the statistical expected value of fatalities is formally acceptable. In this way, even though the existing set of criteria does not explicitly relate to catastrophic events, the EPP serves as an informal tool for reduction of the probabilities of such events.
- Even when the weapon system kinematic envelope exceeds the test range boundaries, both IR and SR probabilistic criteria are sometimes met with “room to spare,” e.g., in the case of air-to-air missiles with inert warheads tested over a sparsely populated desert. However, following the EPP, a mandatory FTS is required whenever the kinematic envelope is not fully contained within the test range boundaries.

³³ See *supra* note 13.

- Separation of hazard variables is highly recommended. Dynamic weapon tests that must include a live warhead (e.g., when the goal of the test is investigating terminal effects) are to be designed with as short a range as possible. Weapon tests that are planned for maximum range (e.g., when system aerodynamic performance is investigated) are to be designed with an inert warhead.

Other engineering tools that back the use of probabilistic considerations are detailed elsewhere.³⁴

Closure

Measuring and valuing risk is multifaceted; research in this area is dynamic, and concepts are introduced continuously. Against this background, a workable tool that compresses the essential topics of risk analysis and management into a set of test range risk acceptability criteria has been developed and implemented. This tool is the basis of a structured decision-making process, is useful even in extreme, complex situations, and is adaptable to risk management developments. In spite of its compressed nature, this tool also addresses the variety of social complexities and perceptions in reaction to the use of technology, as well as to the profile of the specific society which it is meant to serve.³⁵ This ethically-based approach contributes to preserving the credibility of involved institutions, and thus to public trust in them — a measure of primary importance in public acceptance of technology.³⁶



³⁴ Michael Maharik, *Safety Design of High-Performance Weapon Testing* (Rafael, Haifa, Israel 1986) (in Hebrew).

³⁵ Brian Wynne, *Risk and Social Learning: Reification to Engagement*, in *Social Theories of Risk* (Sheldon Krinsky & Dominic Golding, eds. 1992).

³⁶ Harry J. Orway, *Public Wisdom, Expert Fallibility: Toward a Contextual Theory of Risk*, in *Social Theories of Risk*, *supra*.