Adversarial Risk Analysis (at DHS)

David Banks

Department of Statistical Science

Duke University

Abstract

This talk reviews the history and some of the current practice for risk analyses that are done by the Department of Homeland Security. Besides the methods that have actually been used, we also discuss three alternatives proposed by people who have reviewed the DHS techniques.

This talk is based upon experience with the DHS while working at the Department of Transportation, the Department of Health and Human Services, and as a member of National Academy review panel on DHS risk methodology.

1. Just After 9/11

The Office of Homeland Security was established on October 8, 2001, under Tom Ridge. On November 25, 2002, it morphed into the Department of Homeland Security, subsuming the U.S. Immigration and Naturalization Service, the Transportation Security Administration, the U.S. Coast Guard, the Federal Emergency Management Agency, and other small agencies from various parts of the government.

The DHS was charged with consolidating "homeland security" into a single agency. However, the DHS did not have direct authority over the CIA or FBI, and there were inevitable turf battles with those groups as well as with other cabinet-level departments.

The DHS understood that it needed to do risk analysis. The early work was, understandably, improvisational and responsive.

Darrell Morgeson (from Los Alamos, then with DHS and subsequently IDA) led the early risk analysis efforts. The approach was to have have a small group of people sit together and think up possible threats. When they had a long list of threats, each threat would be rated from 1 to 5 in terms of ease and in terms of destructive consequences.

Then Morgeson would then brief the White House Homeland Security Council, and sometimes President Bush, on all threats that were classed as (5,5), (5,4) or (4,5).

This approach is crude but not foolish, and probably flagged many of the major vulnerabilities. The main drawbacks are that it did not explicity consider the costs of risk reduction, tradeoffs in investments in risk reduction, nor probabilities. But this may have been handled informally, or at higher levels than the internal DHS risk analysis process.

The method used by Morgeson's group is similar to "red teaming," which was developed by the DOD in the context of war games. The U.S. Army defines red teaming as a:

> **structured, iterative process executed by trained, educated and practiced team members that provides commanders an independent capability to continuously challenge plans, operations, concepts, organizations and capabilities in the context of the operational environment and from our partnersand adversariesperspectives.**

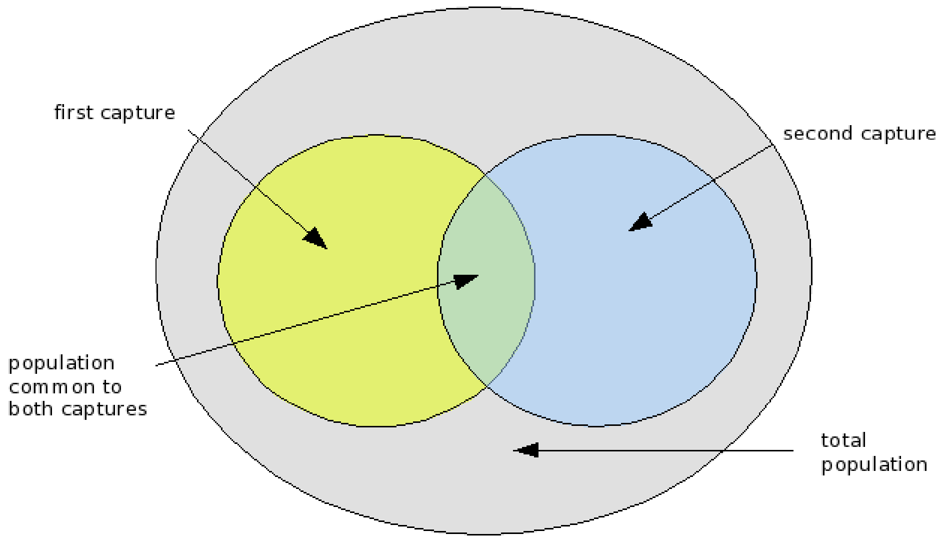
TRADOC News Service, July 13, 2005

Here one assigns a group of people to role-play terrorists under a specific scenario (i.e., specific goals and resources) and then see what kinds of attacks they develop.

Sometimes one extends this to "blue teaming", which operationalizes a game theory perspective.

Red teaming is labor-intensive, but there is much to recommend it. It allows humans to use contextual knowledge and innovation without structuring the problem in artificial ways.

One can apply statistics to red teaming data. For example, it is important to know how many attacks were overlooked by Morgeson's group. To address this, one could have two red teams work independently, and then use Petersen's formula to estimate the number of attacks that neither group discovered.



The statistics of red teaming can be extended in various ways:

- ANOVA designs so that the scenario space is more thoroughly explored with less cost/effort/time.
- Rainbow games, in which participants do more sophisticated role-playing than is commonly done.
- Replicated games, to determine the variation in strategies and outcomes.
- More sophisticated models than Petersen's, which assumes equal catchability and a closed population (this leads to the Turing-Good estimate of the number of unobserved species, covariates describing catchability, and other interesting extensions).

This methodology could apply beyond national security (e.g., to corporate competition or political campaigns). It might be thought of as *Dungeons & Dragons* for grown-ups.

A separate approach to risk analysis was being developed at the Transportation Security Administration (over and above Tom Ridge's color-coded threat levels).

People wanted to find methods for profiling terrorists. The Department of Transportation developed secret passenger screening systems, and the underlying statistics was (in my opinion) dodgy. The process was driven as much by lawyers as by data, and the few and non-independent cases in the training sample were obviously problematic.

Somewhat after that, Admiral Poindexter convened a panel to study strategies for profiling terrorists. The more famous members included Alan Dershowitz, Steve Fienberg, Colonel Rafi Ron, Al Blumstein, Scott Atran, and others. The panel did not write a report, so it is not clear whether our discussion did more than inform Poindexter's thinking.

2. The National Academies Panel

In 2006 the National Academies was asked to review the DHS approach to risk analysis, specifically in the context of bioterrorism. That report is finished, and awaits only the signature of DHS before it is releaseed. It has been awaiting that signature for more than five months. I am not allowed to describe the panel recommendations until it is released, but I can discuss **my own** views.

The center of the evaluation concerned BTRA (Bioterrorist Threat Risk Assessment), a computer program developed by Battelle under contract for DHS. Battelle had won the contract in a competition that included Los Alamos and Sandia.

DHS has paid Battelle a lot of money. **In my opinion**, they did not get good value.

The panel was chaired by Greg Parnell, an OR expert from West Point. The members included people with expertise in counterterrorism, biological warfare, risk analysis, statistics, and related areas.

The BTRA model is large and takes several days to run. Using input derived from laboratory experiments, expert judgment (largely from CREATE or in-house scientists/managers), and DHS supplied parameters, BTRA produces a normalized ranked list of pathogens, ordered in terms of threat.

The BTRA algorithm is described in the report and has been laid out in open briefings and at the 2007 meeting of the Society for Risk Analysis. It involves a complex event/decision tree with depth 18. The nodes often involve complex calculation and simulation. For example, to estimate the number of people who would die in a bioterrorist attack, they use a complex SEIR model with about 40 different compartments.

I believe that BTRA fails in several ways:

- The model assigns probabilities to terrorist decisions that are not informed by a game-theoretic perspective (this partly results from a conflation of the event tree and the decision tree perspectives).
- The model assigns these probabilities by independent draws from a distribution, thereby describing the uncertainty about those probabilities; however, since the draws are independent, the sum of the probabilities at a decision branch does not add up to one.
- The algorithm normalizes the risks by dividing by the largest risk; this means that users do not know the absolute risk, which might be useful in decision making.
- The modeling is over-complex; e.g., to estimate the number of people who would die from a terrorist attack, I would express my belief with a simple Poisson distribution rather than churning through a complex model on each iteration.
- It takes too long to run, cannot handily explore "what-if" scenarios, and does not permit transparent sanity checks.
- The current version does not take account of economic consequences, only mortality and morbidity.

Several members of the panel have proposed and published alternative approaches to the general DHS problem:

Greg Parnell: This strategy simplifies the BTRA tree, converts it to a decision tree, and assigns probabilities in ways that conform to commonly accepted principles of mathematics. A key goal is to use commercial software and to speed up the calculations.

Jerry Brown: With Carlyle and Wood, this approach solves for the Nash equilibrium of the game theory payoff matrix associated with each pair of attacker-defender options. The payoff is the expected value (or 90th percentile, or some other quantity) of the cost distribution. The method also does an exterior optimization to allocate defense investments across multiple defense options.

David Banks: With Anderson, this approach looks at the distribution of Nash equilibria rather than the Nash equilibrium of the mean cost in the game theory payoff matrix.

I believe that none of these techniques is a full solution, but that all of them are better than what BTRA does.

3. An Application to Smallpox

In 2002/2003, the U.S. government was visibly concerned about the possibility that terrorists would use smallpox.

On December 13, 2002, President Bush announced that smallpox was a major threat. In January, he declared a plan to vaccinate 11 million Americans against smallpox, starting with the military and first responders.

Public concern was high.

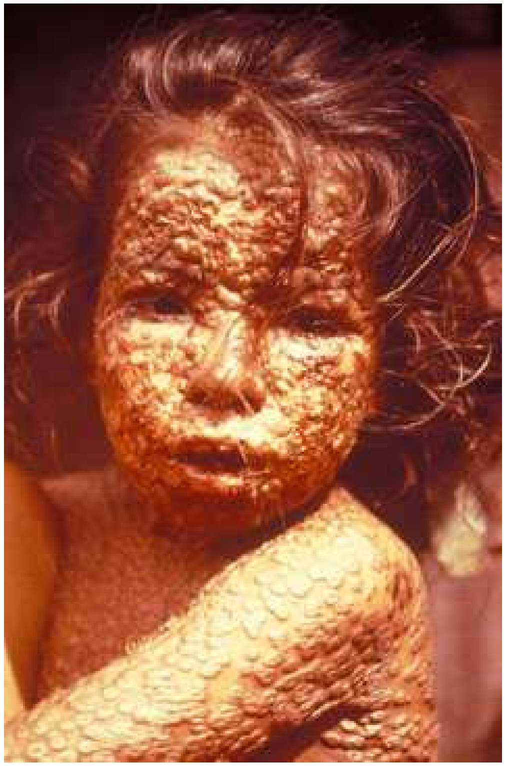
- The anthrax letter attacks in 2001 had shown that high-tech bioterrorism could happen.
- In 2001, senior government officials had conducted a high-profile exercise called 'Dark Winter' in which a smallpox attack on Oklahoma City quickly (and unrealistically) spread to the entire U.S. Ex-Senator Sam Nunn played the president, and the widely disseminated conclusion from the exercise was that the U.S. was not prepared to respond to a massive bioterrorist attack.
- Smallpox is a fearful disease, and the historical aura of horror lingers.
- Routine smallpox vaccination had been discontinued in the U.S. in 1972, and the reserve supplies of vaccine were limited.

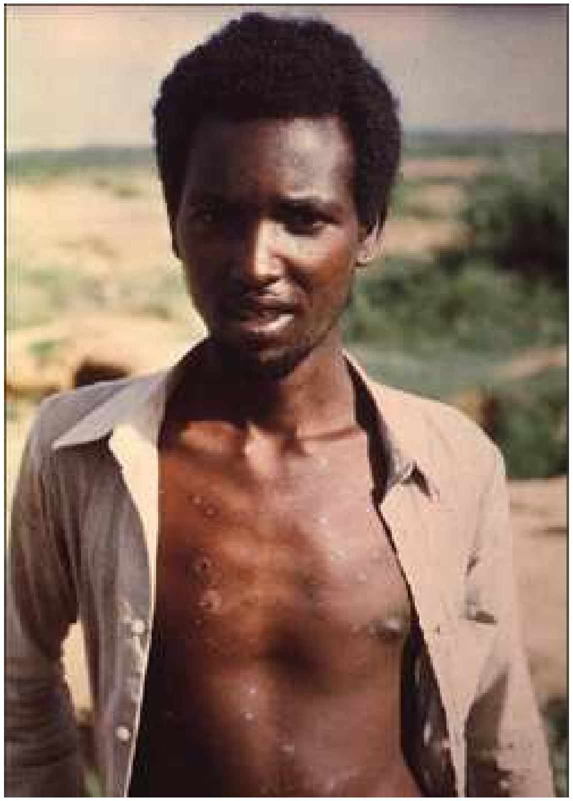
However, there was little clear evidence of an immediate smallpox threat, and the cost of proactive protection was high. Scientific advisors had very different opinions.

3.1 Smallpox Facts

- In 1967, WHO estimated that there were 15 million cases, with 2 million deaths.
- In 1979, WHO certified the eradication of smallpox (the effort cost $300 million).
- In developed nations, the most recent natural outbreak was in Yugoslavia in 1972. In the U.S., the last outbreak had been in NYC in 1947; 6.3 million people were inoculated within three weeks.
- There are two forms: *Variola major* (hemorrhagic) and *Variola minor*. After infection, the incubation period lasts about 12 days, and the disease becomes obvious and infectious on the 12-15th day. By the 28th day, the patient is dead or recovering.
- Smallpox is less infectious that measles or influenza. In developed nations, no recent epidemic has progressed beyond the second generation.
- The fatality rate from *Variola major* is between 3% and 35%. For *Variola minor* is about 1% with modern treatment.

The last natural case of hemorrhagic smallpox killed Rahima Banu, a two-year-old girl in Bangladesh in 1975. The last natural case of Variola minor was a Somali cook named Ali Maow Maalin, who survived.





In 1978 there was an accidental release from a Birmingham laboratory. A medical photographer died. This led to the elimination all stocks except for archival samples maintained in the U.S. and Russia.

3.2 The Smallpox Decision

In 2002, U.S. policy-makers felt that terrorists might make three kinds of smallpox attack:

- A major attack, involving multiple cities and weaponized virus.
- A minor attack, similar to the anthrax letters.
- No smallpox attack.

And the kinds of defenses under consideration were:

- Stockpiling vaccine
- Stockpiling vaccine and increasing biosurveillance
- Stockpile, increase biosurveillance, inoculate all first responders
- Inoculate essentially everyone.

To explore these possible defenses, the U.S. held a series of public meetings and sought scientific advice. D. A. Henderson went around the country giving lectures, analysts weighed intelligence, political support was calculated, op-ed pieces got written, and so forth.

The CDC recommended against widespread inoculation. The vaccine kills about 1-3 people per million, and about 3 people per 100,000 undergo extensive hospitalization. About 30% of recipients miss the next day of work.

But President Bush and other political leaders wanted to buy and use massive amounts of vaccine.

In this setting we describe a tabletop exploration of the game theory and risk analysis approach. The analysis was performed at the FDA, but not disclosed at the time.

In game-theoretic terms, the payoff matrix for this problem is:

	No Attack	**Minor Attack**	**Major Attack**
Stockpile	C_{11}	C_{12}	C_{13}
Biosurveillance	C_{21}	C_{22}	C_{23}
First Responders	C_{31}	C_{32}	C_{33}
Mass Inoculation	C_{41}	C_{42}	C_{43}

Note: Ideally, the option of not even stockpiling vaccine could have been part of this table. However, FDA management ruled against that exploration.

A classical game theory person would use the minimax theorem to find the optimal play for U.S. policy-makers. But this overlooks many problems.

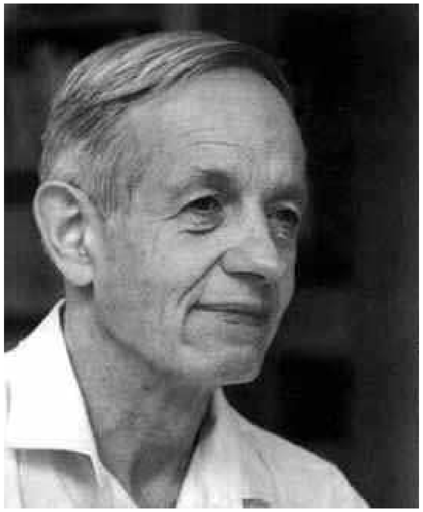
Some of the problematic assumptions of classic game theory include:

- **Perfect knowledge of the payoff matrix entries.** But this might be improved by risk analysis.
- **A zero-sum game.** For this application, the zero-sum formulation is probably reasonably accurate.
- **Rational minimaxing opponents.** Kahnemann and Tversky (and any thoughtful adult) recognize that decision-making is not rational in a strong sense. The Kadane-Larkey Bayesian version of game theory offers a potential solution.
- **Equivalent utility functions.** Terrorists may prize targets differently the U.S. does. The Kadane-Larkey formulation can also address this.
- **One-time games with simultaneous moves.** Our matrix represents a normal-form version of game theory; the extensive form applies to sequential alternating moves. Dynamic programming is probably a better tool.
- **Sufficient resources for all attacks.** Some of the options are much more expensive than others. Additionally, terrorists must have regular visible successes to ensure a continual supply of committed volunteers. These circumstances suggest that portfolio analysis should be an important tool.

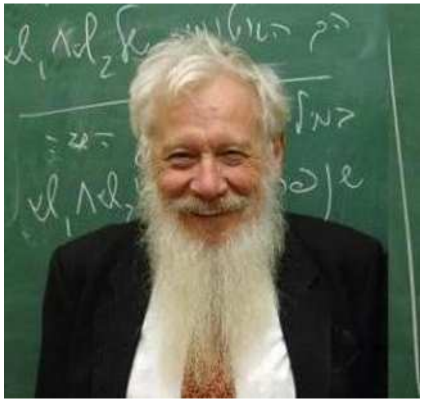
If we believed the assumptions, the von Neumann (1928) showed that the minimax solution is optimal. The U.S. picks the defense with the smallest row-wise maximum cost, and the terrorist picks the attack with the largest column-wise minimum cost.

If the common cell is not the one that attains the U.S. minimum and the terrorist maximum, then randomization is used. This gives a stable solution.

von Neumann



Nash



Aumann

Game theory has weakened some of the more unrealistic assumptions. There are ways to handle imperfect knowledge of the payoff matrix, asymmetric information, repeated play and, to a limited extent, asynchronous moves.

Nonetheless, there is an important alternative. In 1982, Jay Kadane and Pat Larkey developed a Bayesian version of game theory in which one ends up putting prior distributions over the actions of one's opponent and then choosing the action that minimizes the expected loss.

In many ways, the minimum expected loss criterion seems more robust and reasonable than the minimax criterion. And it seems well-suited to applications in which there is partial information from intelligence about the interests and activities of terrorists.

This tabletop exercise uses both criteria and compares the result.

Statistical risk analysis also plays a central role. As previously indicated, the costs in the payoff table are unknown. Risk analysis can find reasonable distributions for these random costs.

Consider the cost C_{11} in the top-left cell. It represents the cost to the U.S. of stockpiling vaccine when, in fact, no terrorist attack using smallpox is made.

$$\begin{aligned} C_{11} \quad = \quad & \textbf{cost to test diluted Dryvax + cost to test Aventis vaccine} \\ & \textbf{+ cost to make 209 x 106 doses + cost to produce VIG} \\ & \textbf{+ logistic/storage/device costs.} \end{aligned}$$

The Dryvax and Aventis vaccines had been kept in storage for years. Their residual potency was unclear, and it was hoped that these could be diluted to stretch the amount of vaccine available. Such testing involves laboratory trials and in vitro challenges.

Two experts pooled their opinions and decided that the cost for such testing might be uniformly distributed between \$2 million and \$5 million.

The FDA said that new vaccine production cost was not random. The contract specified \$512 million. (Actually, when last I checked the cost overrun had gone to almost \$1 billion.)

VIG cost might be N(\$100 million, \$20 million^2).

Logistics costs might be N(\$940 million, \$100 million^2).

The other costs in the matrix can be developed in similarly Delphic ways. (And although one should be very skeptical of such exercises, in fact this is the way that CREATE, DHS, and many other counterterrorism groups are proceeding in trying to build expert judgment into their analyses.)

Note that the different costs in the matrix are correlated. If the stockpiling costs turn out to be higher than expected, those higher costs should also appear as a summand in every cell in that same row. Similarly, if the unknown number of cities that are attacked in the third column is unusually high, that same large value should drive every cell in that column. The payoff table is a random matrix with a complex correlation structure.

In the tabletop exercise, we tried to have two experts independently assess each unknown quantity, and then reconcile their judgments. That was not always possible, and the intent here is to illustrate the strategy.

It is unlikely that the overall decision is grossly sensitive to minor differences in the risk modeling at this level.

Other terms that experts assessed include:

- The value of a human life was treated as fixed, at \$750,000. (This follows the DOT human capital model; non-market methods tend to give higher values.)
- The number of cities attacked in the third column: they judged this to be a shifted Poisson with mean 5 and shift 2.
- The number of key personnel to be inoculated: this was guessed to be uniform between .5 million and .6 million.
- The number of smallpox cases per attack: this was guessed to be gamma with mean 10 and sd 100.
- The cost to treat one smallpox case: normal with mean \$200,000 and sd \$50,000.
- The economic costs of a single attack: gamma with mean \$5 billion and sd \$10 billion.

The latter cost dominates the values in the cost table, pushing the minimax decision rule towards significant investment.

Regarding the payoff table as a random matrix, then one realization of it looks like the following:

	No Attack	**Minor Attack**	**Major Attack**
Stockpile	**$1.5B**	**$16.3**	**$29.3B**
Biosurveillance	**$2.7B**	**$5.6B**	**$15.1B**
First Responders	**$2.7B**	**$6.2B**	**$11.6B**
Mass Inoculation	**$2.4B**	**$2.9B**	**$4.4B**

Many such realizations are generated, and for each the minimax decision and the minimum expected loss decision are calculated. The minimum expected loss rule requires estimates of the conditional probabilities of attack.

One finds the proportion of realizations for which each of the four investment decisions (Stockpile, Biosurveillance, Key Personnel Inoculation, Inoculate Everyone) is the superior decision. This is an estimate of the probability that a given strategy is the best.

The following table gives estimates of the probabilities of a smallpox attack given terrorist knowledge of which protective investment the U.S. made. No input was provided by an intelligence analyst working in this area, but the overall magnitude seems reasonable.

	No Attack	**Minor Attack**	**Major Attack**
Stockpile	.95	.040	.010
Biosurveillance	.96	.035	.005
First Responders	.96	.039	.001
Mass Inoculation	.99	.009	.001

If the U.S. were to only stockpile vaccine, then the probability of no smallpox attack is .95, the probability of a single attack is .04, and the probability of multiple attacks is .01. Similarly, one reads the conditional attack probabilities for other investments across the row.

The minimum expected loss criterion multiplies the probabilities in each row above by the corresponding costs in the corresponding row of the payoff table, and then sums across the columns. The best defense has the smallest sum.

We ran 100 simulations of the payoff table and found the best decisions under the minimax and minimum expected loss (MEL) criteria. (We also considered a scoring function that took account of the magnitude of the differences between the choices.)

	Mimimax	**MEL**
Stockpile	13	99
Biosurveillance	21	0
First Responders	27	0
Mass Inoculation	39	1

If we had been allowed to consider the option of not even stockpiling vaccine, that might have been preferred by MEL.

One can do sensitivity analysis by examining the influence of different assumptions about the probability of attack or different models for generating the random payoff table.

Exploration found a probability table that made the MEL criterion give results similar to the minimax criterion:

	No Attack	**Minor Attack**	**Major Attack**
Stockpile	.70	.20	.10
Biosurveillance	.80	.15	.05
First Responders	.85	.10	.05
Mass Inoculation	.90	.05	.05

	Mimimax	**MEL**
Stockpile	13	15
Biosurveillance	21	29
First Responders	27	40
Mass Inoculation	39	16

4. Closing Comments

Adversarial risk analysis is a key problem with applications far beyond DHS. It requires a combination of game theory, statistical risk analysis, and portfolio theory.

One issue in DHS applications is that federal agencies solve local risk problems rather than seek better global solutions. For example, investing in counterintelligence may protect against many kinds of attack, whereas the smallpox vaccine only protects against one.

Another issue is risk transfer. Hardening one target may simply shift the threat. Also, from a game theory perspective, there is a perverse incentive to be the second softest target.