

Eliciting Uncertain Resilience Information for Risk Mitigation

Niles Egan
Stanford
negan@stanford.edu

Vladimir Menkov
Rutgers/DIMACS/CCICADA
vmenkov@gmail.com

Paul Kantor
Paul B. Kantor, Consultant
paulbkantor@gmail.com

Abstract

The literature of risk, mitigation, and resilience is rich in classifications and recommendations. The missing link is evaluation: ideally, data based; initially, based on expert judgment. We present a novel approach for eliciting probability distributions describing mitigation effectiveness. This approach can be used by subject matter experts (SMEs) who are not specialists in mathematics or engineering. A visual interface permits each expert to sketch a distribution by moving five colored dots on the user interface. The engine can weight and combine estimates from several SMEs into an aggregate density function suitable for presentation, and an aggregate cumulated distribution for use in Monte Carlo simulations. Additional supporting software adapts the tool for real-time support of virtual Delphi-type sessions involving multiple distributed experts. Use of the tool in a study aimed at controlling information and communication technology supply chain risks yields valuable information on those threats, and on the tool itself.

1. Introduction and Some Related Literature

Resilience and mitigation are widely used concepts, which overlap in many ways. Here we use “mitigation” to refer to policies and processes intended to reduce risk, and “resilience” to refer to the expected or stochastic reduction in risk, resulting from specific mitigations. There is an enormous literature on both of these concepts. For resilience in the context of supply chains, which is the focus of the work reported here, the review by Han *et al.* [1] found some 45 empirical studies and 70 that they characterized as conceptual. They identified some 11 distinct aspects of resilience for which metrics can and have been developed. These range from maintaining customer satisfaction to efficiency of responding to disruptions,

and effectiveness in reconstructing an impaired supply chain. In real world applications of the procedures described here, each Decision Maker (DM) will assign different weights for these quantifiable aspects, and each SME will have some particular subset in mind, when asked about the effectiveness of mitigations. In addition to the specific focus on supply chain resilience (SCRE), one may look more widely for metrics that can be applied to quantitatively assess the effectiveness of a mitigation, in terms of whatever measure of performance $P(t)$ is observed *over the course of time* t . A very thorough review of these, drawn from many sectors, including power and manufacturing is given by Chen *et al.* [2]. Those authors emphasize that the time integral of the deficit may not be an adequate measure of impact. This is particularly important when the disruption occurs early in the supply chain. In that case, a sudden shock may produce much greater disruption, as its effects hit all the downstream nodes or multiple agents rush to respond, than would an equally large total deficit that arises more slowly. This issue is discussed very recently by [3]. The world-wide disruptions attributable largely to COVID-19 (although the Suez canal blockage could have happened at any time) have made decision-makers at every level much more aware of the importance of supply chain resilience.

Decision-makers at various levels will have different performance measures. Auto-makers, for example, want to secure their own supplies of chips. A national level decision maker may be more concerned that *all* automakers maintain adequate supplies. Governors of individual states may (and did) compete for personal protective equipment; national decision makers want to ensure supplies go where they are needed. Because of these differences, eliciting estimates of mitigation effectiveness, which is the key goal of this research, must first decide whether to dig into the finest details, articulating specific threats and mitigations, or to seek a framework that will be useful and usable by any decision-making organization (DMO), with its own agreed-upon metrics and priorities. The research

reported here was done in the context of a broad study across many industry and government aspects of the Information and Communications Technology supply chain (ICTSC). Because of this broad scope it was not feasible to dig into fine details. Therefore the focus is on a broad framework, and supporting tools, that can be used by any DMO. This is in contrast to, for example, the detailed estimates developed for nuclear reactor risk assessment, where complex chains of events can have catastrophic consequences.

In addition to specifying the granularity, which we take to be broad, one should consider the degree of engagement expected from SMEs who are providing input. At one extreme, nuclear safety studies may ask for days of engagement from experts, who may do specific calculations to support their estimates. The present study is at the opposite extreme, limiting the engagement of SMEs to at most a few hours.

With these two specifications, the research needs an easy-to-use tool that is accessible to a broad range of SMEs without extensive training, and which recognizes that there will be many different measures of effectiveness. There is one more issue to consider: how shall we describe effectiveness? We have elected to describe effectiveness in terms of the *reduction in risk*, and, furthermore, to recognize that an expert cannot give a point value for this reduction, but may be able to describe a range and distribution across that range. Thus we are eliciting the probability distribution of a fraction. Technically, this appears the same as estimating an unknown probability, since both the probability and the fractional reduction in risk lie in the interval $[0, 1]$.

Review of classical methods, and the most-up-to-date extensions of these classical methods developed by [4], [5], [6] reflect their grounding in twentieth century pedagogy of statistics and probability. Those traditions strongly favor parametric distributions, such as the Gaussian or Beta distributions, with a preference for distributions having a conjugate Bayesian partner. That relation ensures that strict Bayesian multiplicative update to represent new information retains the parametric form of the prior distribution. While this makes analysis and graphical presentation of results easier, it may obscure the nature and degree of disagreement among the participating experts. E.g., with Gaussian models, two experts who agree that a parameter has variance 1 but disagree on the mean, favoring values $\mu = 5, 10$ will find their judgments combined to a uni-modal aggregate whose mode they both reject. Combining this with the knowledge that our findings will be used in computationally intensive simulation modeling, we prefer a simple but non-traditional model for the elicited

distributions.

This approach is hinted at, but not realized, in some other approaches that stem from ecology including, in particular, the MATCH project at Nottingham [7]. In that system a user may specify a distribution in terms of a histogram, which is completely not parametric. This has the advantage that an expert can express fairly complicated ideas, such as a trimodal distribution. On the other hand, it does lose the intuition associated with the notion of either the mode (most likely) or the median (50-50).

NOVELTY: We sought a simpler way of providing an adequate range of expression. We report on a new tool for eliciting and integrating “soft triangular” distributions and conditional probabilities. The new tools for group elicitation of quantitative findings about probability distributions [8] extend the classic methods for the combination of probabilities initiated by [9]. For reviews and other extensions on the combination of probabilities see also [10, 11, 12]. An important review of the whole field of probability estimates is given in [13].

CONTEXT: In summary, this work arose in estimating the **effectiveness of countermeasures or mitigations** used to protect or increase the resilience of supply chains. The particular sector of interest is the Information and Communication Technology sector, which includes both hardware and software. The particular cases discussed here concentrate on the hardware aspect. A notional model offered to SMEs is construction of a laptop by assembling items such as a motherboard, a keyboard and a screen, at a plant that forms the case to contain these parts and completes the assembly. Further details of this project are described in [14].

The remainder of this paper is organized as follows: in Section 2 we review the overall problem of eliciting distributions. In Section 3 we discuss the specific research context. We summarize the underlying mathematics in Section 4; in Section 5 we sketch the system, and in Sections 6,7 we sketch the visual and functional properties of the several interfaces. Section 8 provides a few representative findings, and in Section 9 we discuss possible conclusions and extensions, both methodological and project specific.

2. Eliciting Distributions

As noted, SMEs may express their uncertainty in a parametric (mean and variance) or non-parametric (sketch a curve) fashion. Some experts, with experience in the insurance industry, suggested using triangular distributions. Typically one elicits the most likely value

(mode) of the uncertain number, and constrains it with the highest and lowest reasonable values. This approach is sound if the expert's estimate is essentially symmetric about the mode, in which case the mode is essentially the median and the mean. If the range is far from symmetric, using a triangular distribution means that that one side is weighted much more than the other. Our conversations with experts suggest that the notion of a "50-50 value" (median) is more comfortable than asking for a "most probable value." Thus, if triangles are to be used, they should be allowed to "stretch" on one side, while remaining triangular on the other, so that the area, or total probability remains the same on both sides of the median.

To capture some of the virtues of both the histogram approach and the triangle approach, we developed a tool that permits the SME to specify median and range, and also increase the probability of extreme events in that range, as described below.

To apply this tool, and the underlying approach, one must narrow the threats and focus on countermeasures "suitable for quantitative assessment." Can we find a way to quantify the impact of a mitigation, in a language that does not depend on the specifics of the metric? We require only that the metric be defined on a ratio scale, so that the notions of 0% or 100% mitigation are meaningful. One expert might assess 90% effectiveness at preventing shipment delays. Another expert might characterize the loss as "a combination of lost orders, reputational damage, and overtime costs." and add "But I think that this mitigation will cut losses by about half." Of course the dollar value at a small company will be much less than the dollar value at a large company, for any given level of risk reduction.

Faced with this complex reality, and informants from both large (billion dollar) and small (less than \$50 million) firms, we sought a "precisely flexible" definition of effectiveness. We operationalize the meaning of "the amount of risk reduction" using two equations. The first is familiar throughout the field of risk management and sets the "unmitigated" risk, R_0 , associated to a threat as given by Equation 1.

$$R_0 = Consequences * Vulnerability * Threat \quad (1)$$

In engineering terms the three factors are understood as follows:

- Threat: the probability of the harmful event
- Vulnerability: the conditional probability that the event will "succeed" in causing harm(s)
- Consequences: the specific harm(s) that the event will produce, if it succeeds.

We find that subject matter experts are familiar with this "word equation" even if they lack engineering

background. Note that any specific mitigation might affect one, two, or all three of the factors. For example, if theft is the risk, a watchdog may apprehend a thief (reducing vulnerability) or deter a thief (reducing threat). SMEs may often disagree about which factor dominates and yet agree on the overall mitigation through the product of the factors.

Our elicitation therefore includes a working definition of effectiveness as the reduction in overall risk, *however the responding expert prefers to define it*. As an example we found that some experts considered a hurricane's damage to a production facility, while others considered that it may block truck transport. In sum, effectiveness E is defined implicitly by its role in Equation 2.

$$ReducedRisk = (1 - E) * R_0 \quad (2)$$

The heavy font used here calls attention to the fact that E is not a simple number, but is a random variable representing many kinds of uncertainty. That uncertainty may be due to specifics of the threat such as a hurricane, variation in the impact of the mitigation (e.g. attending to weather forecasts), or subjective uncertainty on the part of the expert about how to summarize his or her own experience in a numerical form. We cannot eliminate any of these uncertainties. However, formulating the effectiveness of mitigation in this way requires us to elicit the distribution of a **bounded** parameter.

3. Specific Research Context

This work is part of research examining the resiliency of the Information and Communication Technology Supply Chain against a variety of threats, both natural and man-made, discussed in [14]. The aim is to develop and pilot a methodology for assessing the mitigating impact, on resiliency, of specific countermeasures, while incorporating both stochastic and ontological uncertainties.

Building on the distributions elicited by the methods presented here, the simulation team provides Monte Carlo distributions of the benefit to the overall network of each specific countermeasure, and of selected combinations of mitigations. The simulation results are further used in optimization calculations, using Mixed Integer Linear Programming to provide the optimal *reduction in risk* for each budget level under consideration.

We summarize here the key steps of the study.

- Review literature and government reports; focus on specific threat scenarios

- Talk with experts, to further focus and confirm selected threats
- Extensive (30+ minutes; N=34) one-on-one conversations with experts, particularly those whose experience supports quantitative estimates of countermeasure effectiveness. Zoom facilitates recording and transcription for qualitative review such as grounded [15] content analysis [16].
- Focus group discussions of four to six SMEs. Zoom technology makes it feasible to assemble outstanding experts but the mediated technology is tiring. The target of 90 minutes, with some garrulous experts has reduced the time available for data collection. We find that experts can assess some 7 to 12 (threat, countermeasure) pairs in 15 to 20 minutes. Brief training videos (3 to 4 minutes long) acclimate experts to the interface.
- Integration of the expert results for each specific focus group (examples below) in graphical form, for reporting and qualitative analysis
- Monte Carlo simulation and optimization studies, as described above.

4. Mathematical Models

4.1. The user interface algorithm

Each expert is invited to provide his or her estimate of the risk reduction probability distribution function in one of the two forms, depending on how the survey is configured: either a “distorted triangle” or a “bucket distribution”.

The bucket distribution is simply a piecewise-constant function with N (e.g. $N = 10$) “buckets”. In surveys using this distribution, the expert would indicate, by moving tokens into buckets, the probability that the risk reduction is (for $N = 10$) between 0 and 10%, between 10% and 20%, etc. This was developed replicating the least parametric model used by [7], as a “back-up” alternative, in case the “distorted triangle” (also called “five-point”) distribution should prove unacceptable to participating experts.

In surveys with distorted-triangle probability distributions, the shape of the distribution is controlled by five data elements: x_1, y_2, x_3, y_4, x_5 , where $0 \leq x_1 < x_3 < x_5 \leq 1$, $y_2 \geq 0$, and $y_4 \geq 0$. These are read off, by the algorithm, from the locations (see Figure 1 of the blue point at (x_3, y_{\max}) , the green points at $(x_1, 0)$ and $(x_5, 0)$, and the yellow points at $(x_1, y_2), (x_5, y_4)$.

The soft-triangle distribution is defined so that the distribution function $f(x)$ has the following properties:

- The support of the distribution is $[x_1, x_5]$.
- The median of the distribution is at x_3 .
- If $y_2 = y_4 = 0$, then $f(x)$ is unimodal with the mode at x_3 ; the distribution function $f(x)$ is continuous, increasing from $f(x_1) = 0$ to $f(x_3) = y_{\max}$, and then decreasing to $f(x_5) = 0$.
- If $y_2 = 0.5/(x_3 - x_1)$, then $f(x) = y_2$ on the entire segment (x_1, x_3) . Similarly, if $y_4 = 0.5/(x_5 - x_3)$, then $f(x) = y_4$ on the entire segment (x_3, x_5) .

In particular, with $y_2 = y_4 = 0$, we define $f(x)$ so that it is linear on the shorter wing of the distribution, and a monomial power function on the longer wing, the exponent being chosen so that x_3 remains the median. Thus, if $x_3 - x_1 < x_5 - x_3$, the distribution function is defined as

$$f(x) = (x - x_1)/(x_3 - x_1)^2 \quad (3)$$

for $x \in [x_1, x_3]$, and as

$$f(x) = \frac{1}{(x_3 - x_1)} \left(\frac{x_5 - x}{x_5 - x_3} \right)^k \quad (4)$$

with

$$k = 2 \frac{x_5 - x_3}{x_3 - x_1} - 1 \quad (5)$$

for $x \in [x_3, x_5]$.

The additional distribution parameters, y_2 and y_4 , are controlled in the user interface by the vertically movable yellow points, and allow the expert to individually modify the left and right wings of the distribution. By raising the left yellow control point (x_1, y_2) the user can create a discontinuity in the distribution, raising the left end of the left wing (and proportionally lowering the right end of that wing, so that the the area under the wing stays 0.5, keeping the median of the distribution at x_3). This is implemented by the following distribution (assuming $x_3 - x_1 < x_5 - x_3$):

$$f(x) = y_2 + R_2(x - x_1)/(x_3 - x_1)^2 \quad (6)$$

for $x \in [x_1, x_3]$, and

$$f(x) = y_4 + R_4 \frac{1}{(x_3 - x_1)} \left(\frac{x_5 - x}{x_5 - x_3} \right)^k \quad (7)$$

with the same k as above, and the values R_2 and R_4 chosen so that

$$\int_{x_1}^{x_3} f(x)dx = \int_{x_3}^{x_5} f(x)dx = 0.5. \quad (8)$$

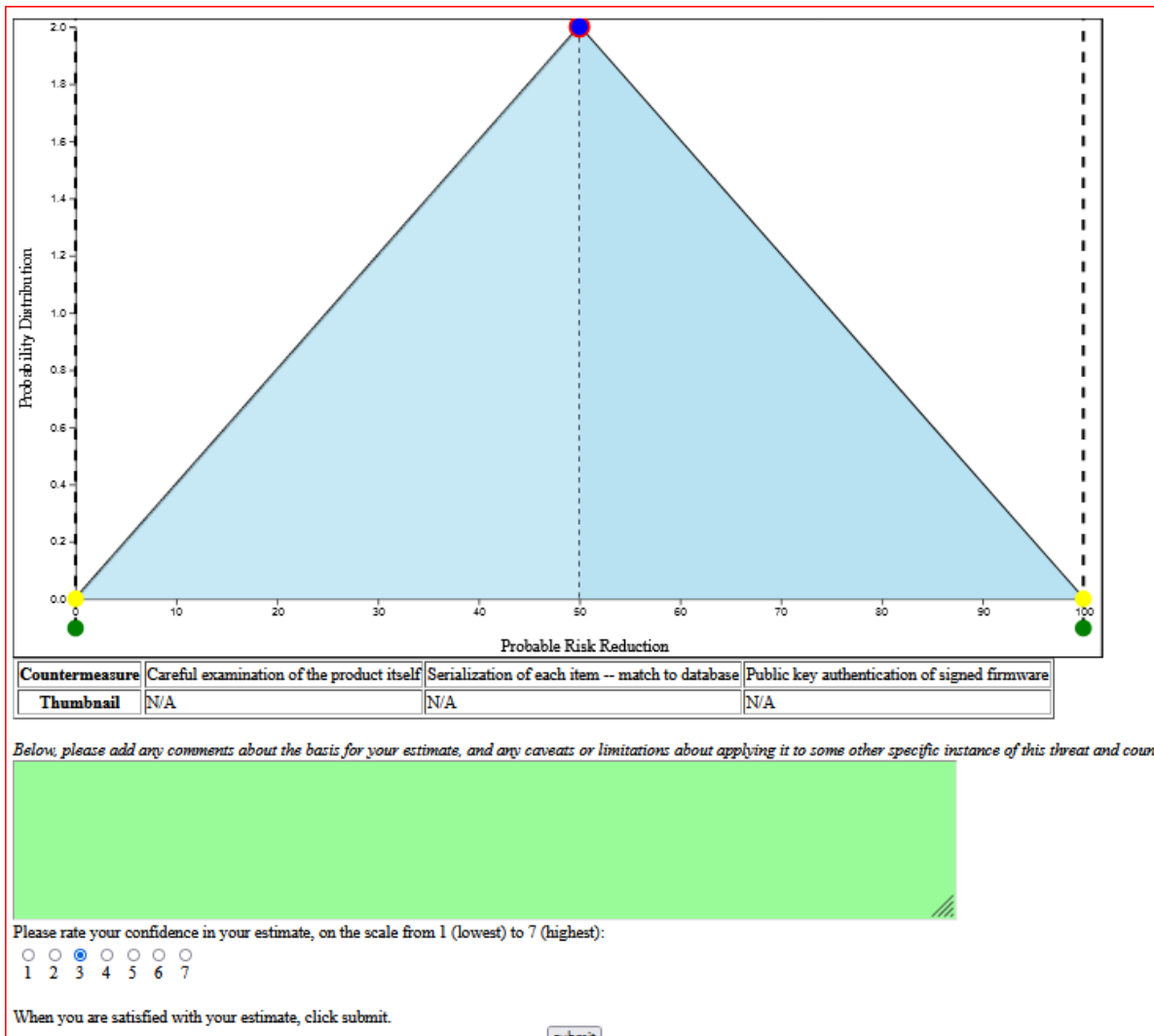


Figure 1. The subject matter expert's graphical interface.

4.2. Calculation of weighted means and medians

We provide several different measures of aggregation based upon: equally weighted, reported confidence, and narrowness of the range. For each weighting scheme we compute both the mean and the median of the cumulated distributions.

Weights. For both the averaged and median-cumulative distributions, a variety of weighting schemes are available, as follows.

1. Equal weights: $w_j = 1$ for each expert j who has provided an estimate for the countermeasure in question.
2. Explicit confidence-based weights: each expert is asked to explicitly rate his or her confidence in each estimate, as a positive integer number within

a certain range (1 through 7). This number (or a certain default, if none is provided) is directly used as the weight w_j of the estimate.

3. Width-based weights. This presumes that a high-precision estimate (e.g. "21-22% reduction") represents greater confidence or knowledge than a less-defined one ("between 0 and 50%"). Accordingly, the inverse of the width of the support of the probability distribution provided by the expert is used as the weight of his or her estimate, $w_j = 1/(x_5 - x_1)$.

Averaged density function. Assuming that n experts provided their probability estimates $f_j(x)$ ($j = 1, \dots, n$) for a given countermeasure, and a particular weighting scheme with weights w_1, \dots, w_n has been chosen, we compute (as the green curve, see Figure

6) the averaged probability density function given by Equation 9.

$$f_{\text{avg}}(x) = \frac{\sum_j f_j(x)}{\sum_j w_j} \quad (9)$$

Median cumulative distribution function. With an odd number of estimates with equal weights, we can compute the cumulative distribution function ($F_j(x) = \int_0^x f_j(t)dt$) for each one, and ask, for any given x , which of those cumulative distribution functions represents the median of the experts' opinions – that is, who is the expert k such that half of the remaining experts provided a lower value for the cumulative distribution function than k did (i.e., their $F_j(x) < F_k(x)$), and the other half provided a higher value than k did? More generally, we want to compute $F_{\text{med}}(x)$ such that for any x , it represents the weighted median of all experts' cumulative distribution functions. We define it as follows:

- It may happen that for this x there is an expert k such that only a “weighted minority” of experts provides a lower value of $F(x)$, and only another “weighted minority” provides a higher value. That is, $\sum_{j:F_j(x) < F_k(x)} w_j < 0.5$ and $\sum_{j:F_j(x) > F_k(x)} w_j < 0.5$. If such an expert exists, we view k as the “median expert” for x , and define, for this x , $F_{\text{med}}(x) = F_k(x)$.
- It may happen that no such “median” k exists, because the set of experts can be split into two non-overlapping subsets ($[1, \dots, n] = S_1 \cup S_2$, with $S_1 \cap S_2 = \emptyset$) with equal total weights ($\sum_{j \in S_1} w_j = \sum_{j \in S_2} w_j = 0.5 \sum_j w_j$) such that everybody in the first group gives a lower cumulative distribution function value than everybody in the second:

$$\max_{j \in S_1} F_j(x) = m < M = \min_{j \in S_2} F_j(x).$$

In this case, we use the midpoint of the gap between the two groups as the median value, i.e. define $F_{\text{med}}(x) = (m + M)/2$.

The function $F_{\text{med}}(x)$ constructed in this way behaves as a cumulative distribution function on $[0, 1]$, i.e. it is continuous and monotonically non-decreasing, with $F_{\text{med}}(0) = 0$ and $F_{\text{med}}(1) = 1$. Its derivative (which may not be continuous, even if those for each individual expert are) has the properties of a density function. Its numerical derivative is plotted (in gold, see Figure 6) to serve as the median of the several experts' probability density functions.

5. System Overview

For focus group processes input from the subject matter experts is captured and processed in real time. The system supports other management/research functions. A moderator can query the data in real time, in synchronous or focus group settings. Precise threats and countermeasures are refined by the participants, and can be captured in real time. An annotator interface replaces traditional flip charts, and summarizes specific threats, and selected countermeasures or mitigations. The moderator can check to see when any SMEs are not providing input, and when all estimates are in. The moderator can show the aggregated density function of the estimates, with each SME's median and width-based confidence signalled by a red dot. This can support Delphi iterations, if such convergence is desired. Finally, the system produces composite graphs for review by management decision-makers, or inclusion in reports and scientific communications, such as Figure 6.

6. SME Interface

Usability of the SME interface is crucial and informal anecdotal reports indicate that it is sufficiently usable. Formal usability study will require a separate project, exploring other domains requiring estimates of a fraction or probability.

To use the system, each SME chooses an anonymous working ID and proceeds to a selection dashboard, shown in Figure 4. This shows a thumbnail sketch for any countermeasures already assessed, and the SME may revise a judgment or add a new one. Either choice brings the expert to the central novelty of this elicitation tool, the “5-point graphical elicitation interface.” Instructions above the graphic interface shown in Figure 1. recall the threat and countermeasure with these guides.

- Assessing countermeasure: “Analyzing physical appearance against threat Counterfeit Parts”
- Please estimate the amount by which this Countermeasure reduces the risk of this threat.
- Please move the blue dot to the 50-50 or median point.
- Then move the green dots to sharpen the low and high ends of your estimate.
- Finally, you may lift the yellow dots to make the tails fatter if you desire.

Figure 5 shows how the distribution may change as the points are moved. This wide range of odd shapes

appears to be satisfactory to users, pending any more detailed research on the user experience.

7. Group/Delphi Interfaces

The tool has three additional interfaces, for an administrator, a moderator, and an annotator.

7.1. Administrator

| Group name | Distr. type | State | Survey closes automatically after 4 hr | Experts can enter rankings | M |
|-----------------------------------|-------------|----------|----------------------------------------|----------------------------|------------------|
| PaulNOV23 | T5 | FINISHED | true | true | Tue Dec 29 16:19 |
| Niles3 | T5 | FINISHED | true | true | Sat Jan 02 23:17 |
| 2_focusGroup | T5 | FINISHED | true | true | Mon Mar 15 19:2 |
| Massive-Input | T5 | OPEN | false | true | Mon Mar 15 22:0 |
| Annotator-Example | T5 | OPEN | false | true | Thu Nov 08 10:0 |

Figure 2. Administrator’s Dashboard displaying the groups (selection only).

An administrator may select any existing survey or focus group, or create a new one. By default a group closes after four hours, but can be set to remain open for surveys involving many respondents. The interface may be either the 5-point model described here or a 10-bin interface as presented by [7]. A small excerpt of the administrator interface is shown in Figure 2.

7.2. Moderator Interface

The moderator works with the group in real time, controlling things from an interface not shown here, and monitoring participant progress. In practice an open Zoom window is used to share a spreadsheet in which threats and countermeasures are recorded as they are refined by each group.

7.3. Annotator Interface

Threats and countermeasures details are added by the annotator during the discussion. Experts generally reach an *ad hoc* decision that a countermeasure may mitigate more than one threat, or may be subsumed by another. The annotator works in real time, with the interface shown in Figure 3. Threats are entered into the pink area, while each countermeasure is recorded separately in a green area. Each threat and countermeasure has both a short name, and a longer text description. These are replicated into the moderator’s and expert’s dashboards.

Unlike a “flip chart,” the annotator’s interface is not exposed to the SMEs in real time. Instead, we have shared a (Google sheet) table of threats and

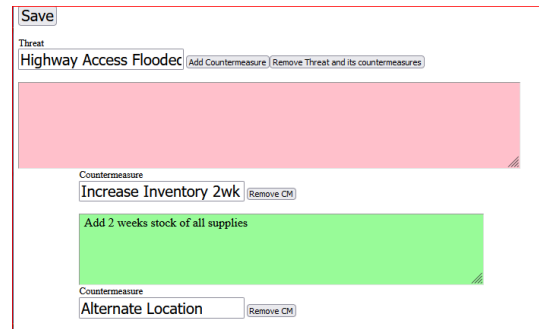


Figure 3. Annotator’s interface (example) .

countermeasures, with the moderator adding notes. Screen space prevents replicating the physical focus group model in which charts on the wall record what has been agreed upon. A technology that shows both the shared document and the faces of all participants would benefit this part of the work.

The moderator and annotator can both access the moderator dashboard and, to open the system for judgments, send the SMEs a hyperlink to the login page. We have not imposed password protection, as the link is a long random string, unlikely to be hacked during a brief 90 to 120 minute focus group or Delphi session.

8. Representative Findings

The findings of any particular focus group are best summarized in two different ways. For human inspection and examination a visual array of aggregated density curves, is most effective. A typical example is given in Figure 6. In this particular instance, equal weighting has been used. Columns each represent a threat; rows represent the countermeasures to that threat. Elements of the same row in different columns generally represent different countermeasures. In this matrix the columns represent: counterfeit parts; extreme weather; and onboarding a new vendor. The mitigations are (top down): For THREAT=Counterfeits: careful examination of each part, serialization and database review, and public key authentication of firmware; for THREAT=Extreme weather: monitor weather forecasts, diversify suppliers and transport, and regularly exercise plans (a fourth, periodic audits of supply chain is not visible in this extract); and for THREAT=Onboarding a new vendor: contractually require background and security checks, agree to security measures for transmission and storage of intellectual property, and quality control inspection of the facility.

Usually a wide-ranging discussion of threats and

mitigations precedes the quantitative evaluation. For example one group noted that “thoroughly vet the management of a potential supplier” can be agreed to include “monitor the financial stability of a potential supplier” although, strictly speaking, one of them was attached to the onboarding threat, and the other attached to preventing counterfeits. We find that participants quickly move to the essential elements of each proposed countermeasure, and agree, for that session, to represent a particular mitigation by an agreed-upon phrase. All discussions are recorded and transcribed, to support careful review. Two groups may settle on significantly different interpretations of the same key phrase. For this reason we do not propose to further aggregate the results from distinct groups, even when the “names” of

the countermeasures may have been the same.

In each aggregated plot, Figure 6, we show both the mean and the median of the experts’ estimates, with the mean in green and the median in gold. There are competing theoretical justifications for considering either alternative. Generally the mean behaves in the most intuitive way when examining the probability densities visually. On the other hand, the median makes better sense when the aggregation is used in simulation, since it is robust to the outliers. There is not room here to show those cumulated curves. As explained in Section 4.1, the median is defined by reference to the cumulated distributions.

In each aggregated plot there is a red dot for each of the individual experts. The x-coordinate reflects that

| Focus Group HP-FG: Expert's Dashboard | | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| Focus group stage (as of Mon Jun 14 00:34:13 UTC 2021): Open . You can enter your first estimates. If/when instructed by the moderator, you can also enter a second estimate. | | |
| Threat and countermeasures | Your estimates (please only enter one estimate for each (threat,countermeasure) pair, until instructed by the moderator to enter another one) | Thumbnail |
| Threat: Counterfeit Parts Unlawful or unauthorized reproduction, substitution, or alteration that has been knowingly mismarked, misidentified, or otherwise misrepresented to be authentic. | | |
| Countermeasure: Supply chain visibility Consulting an industry database of counterfeit parts | You have not made an estimate yet. Make it now! | N/A |
| Countermeasure: Analyzing physical appearance Image processing to assess the visual appearance | You have not made an estimate yet. Make it now! | N/A |
| Ranking link will appear here after you have rated at least 2 countermeasures. | | |
| Threat: Hurricane Extreme weather event | | |
| Countermeasure: Weather forecast Detection of threat (hurricane) using weather forecast | You have not made an estimate yet. Make it now! | N/A |
| Countermeasure: Geographical diversity Diversifying suppliers to ensure continuous availability of raw materials in case of extreme weather events in one geographic location. | You have not made an estimate yet. Make it now! | N/A |
| Ranking link will appear here after you have rated at least 2 countermeasures. | | |
| Threat: Vendor Fraud Threat while onboarding a new vendor | | |
| Countermeasure: Certification or a verification database A group of companies comes together to verify a list of vendors or using a third-party certification to countermeasure vendor fraud. | You have not made an estimate yet. Make it now! | N/A |
| Ranking link will appear here after you have rated at least 2 countermeasures. | | |
| [Back to the dashboard] © 2020-2021 Rutgers University - State University of New Jersey Focus Group: HP-FG Expert's pen name: pkJUN13a | | |

Figure 4. Experts Dashboard.

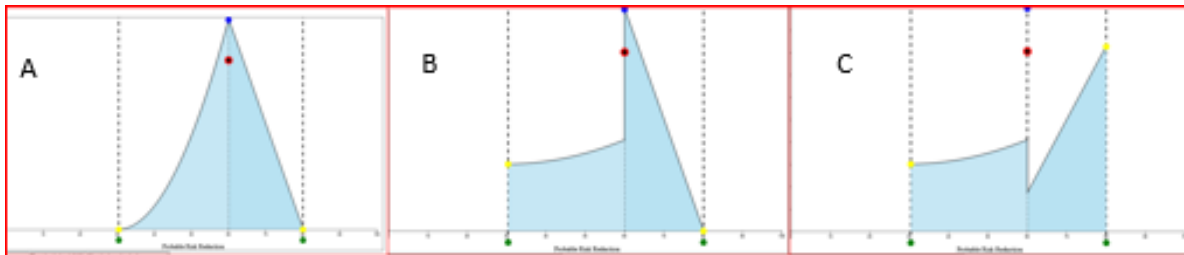


Figure 5. Three adjustments of the triangle. A: median moved right, towards 100% and range defined. B: Longer tail increased. C: Shorter tail increased.

expert's estimate of the median, while the height of the dot represents that expert's confidence. In these examples that height is computed from the tightness of the range. This scheme makes it immediately clear when there are substantial disagreements among the experts. If the goal were to develop an appearance of consensus it might make sense to conduct one or two Delphi rounds to seek a more coherent picture. That might, however, conceal significant uncertainties about the true effectiveness of the given countermeasure. In the example curves shown here the focus group comprised four industry experts with a total of about 60 years' experience, and they were fairly confident in their assessments.

To illustrate how one may reason from this information, it appears clear that the third countermeasure for Counterfeit parts is judged strongly more effective than the other two. The peak is farther to

the right, and sharper. Similarly, for Extreme weather, the second and third countermeasures are clearly judged more effective than the first. Even if one were to remove the strongly held dissenting view at the low end, the remainder is clearly broad, representing great uncertainty as to just how effective monitoring weather forecasts can be. (Parenthetically we note that in discussion one SME emphasized that the weather information is only as valuable as the plan that is implemented when bad weather is predicted.)

9. Conclusions and Discussion

The tool presented here may support further research in two ways. First, there are six methods of aggregation: three weighting schemes and two ways of computing the representative distribution (see Section 4.1). Second, although only each expert's most recent distribution is

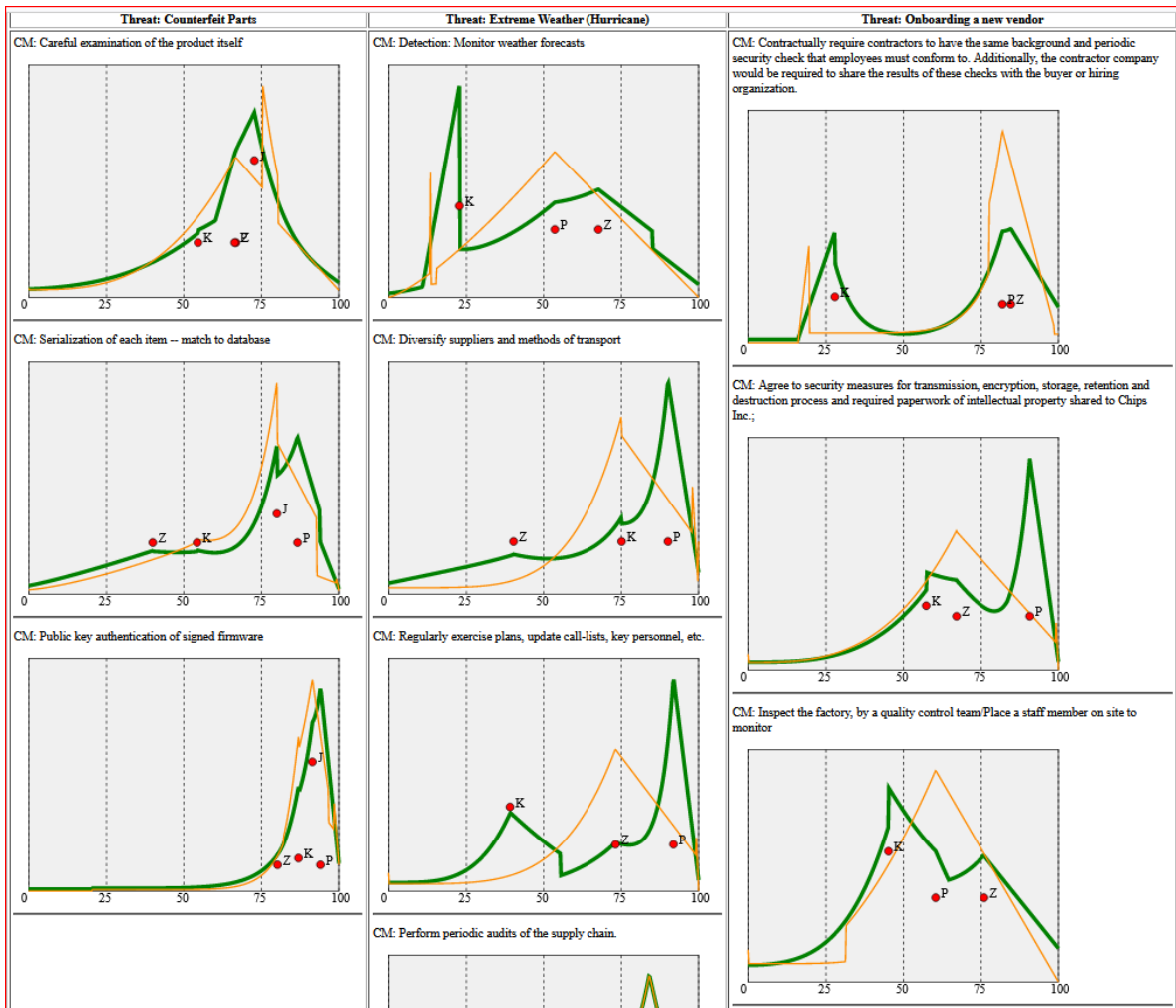


Figure 6. An array of aggregated probability density functions. (See text)

sketched in the visual summary, the five parameters of every submission are retained, and can support later analysis and Delphi iterations. Some prior studies suggest “best mix” guidelines for a focus group, such as [9]. Our early findings suggest that some experts have specific quantitative knowledge, based on internal engineering studies. Other experts have a broader industry-wide view, often in association with a specific plan or tool for increasing resilience. At this point it is not known how best to mix such experts to produce the most effective group aggregations.

This tool has the potential to rationalize discussion of the quantitative benefits of mitigations and countermeasures, complementing the cost considerations that currently control adoption. While the visual summaries of Figure 6 are helpful in conversations with top managers, the concrete aggregated distributions will support computations of both the expected benefit of a mitigation and the probability that risk exceeds some maximum acceptable level. What appears most novel here is the combination of a “top-down” formulation of the Effectiveness concept, coupled with a practical tool for “bottom up” elicitation of an uncertain parameter, from the most knowledgeable SMEs.

The long term goal of this research is to influence and transition to practice in government and industry. Thus we plan not only to demonstrate the method, but also to share tools that can be used to replicate this analysis for specific instances, where the necessary information is often proprietary or classified. Insofar as such studies yield generalizable findings about mitigation effectiveness it is hoped that use of this methodology will also encourage and facilitate information sharing, so that every organization can apply its limited resources more effectively.

10. Acknowledgements

Acknowledgement: This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 17STQAC00001-05-00. **Disclaimer:** The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security. The authors thank Fred Roberts, Elsayed Elsayed, W. Grace Guo, and Eric Rosenberg for helpful conversations, and three reviewers for helpful comments and careful review, which have improved the presentation.

References

- [1] Y. Han, W. K. Chong, and D. Li, “A systematic literature review of the capabilities and performance metrics of supply chain resilience,” *International Journal of Production Research*, vol. 58, no. 15, pp. 4541–4566, 2020.
- [2] Y. Chen, E. A. Elsayed, and Z. Huang, “Systems resilience assessments: A review, framework and metrics,” *To appear in the International Journal of Production Research*, 2021.
- [3] I. Giannoccaro and A. Iftikhar, “Mitigating ripple effect in supply networks: the effect of trust and topology on resilience,” *International Journal of Production Research*, pp. 1–18, 2020.
- [4] W. G. Stillwell, D. Von Winterfeldt, and R. S. John, “Comparing hierarchical and nonhierarchical weighting methods for eliciting multiattribute value models,” *Management Science*, vol. 33, no. 4, pp. 442–450, 1987.
- [5] E. J. Bonano, S. Hora, R. Keeney, and D. Von Winterfeldt, “Elicitation and use of expert judgment in performance assessment for high-level radioactive waste repositories,” tech. rep., Nuclear Regulatory Commission, Washington, DC (USA). Div. of High-Level Waste Management; Sandia National Lab. (SNL-NM), Albuquerque, NM (United States), 1990.
- [6] R. L. Keeney and D. Von Winterfeldt, “Eliciting probabilities from experts in complex technical problems,” *IEEE Transactions on engineering management*, vol. 38, no. 3, pp. 191–201, 1991.
- [7] S. Mitchell and I. Dunning, “MATCH Elicitation tool,” <http://optics.eee.nottingham.ac.uk/match/uncertainty.php>, 2021.
- [8] P. B. Kantor, “Soft triangles for expert aggregation,” *arXiv preprint arXiv:1909.01801*, 2019.
- [9] R. T. Clemen and R. L. Winkler, “Combining probability distributions from experts in risk analysis,” *Risk analysis*, vol. 19, no. 2, pp. 187–203, 1999.
- [10] M. A. Burgman, *Trusting judgements: how to get the best out of experts*. Cambridge University Press, 2016.
- [11] C. Werner, T. Bedford, R. M. Cooke, A. M. Hanea, and O. Morales-Nápoles, “Expert judgement for dependence in probabilistic modelling: A systematic literature review and future research directions,” *European Journal of Operational Research*, vol. 258, no. 3, pp. 801–819, 2017.
- [12] A. O’Hagan, C. E. Buck, A. Daneshkhan, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow, *Uncertain judgements: eliciting experts’ probabilities*. John Wiley & Sons, 2006.
- [13] A. O’Hagan, C. E. Buck, A. Daneshkhan, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow, “Uncertain judgements: eliciting experts’ probabilities,” 2006.
- [14] W. Guo, P. Kantor, E. Elsayed, E. Rosenberg, R. Lei, S. Patel, B. Ruskey, and F. Roberts, “Supply chain threats and countermeasures: From elicitation through optimization,” *HICSS*, 2022 (to appear).
- [15] A. Strauss and J. M. Corbin, *Grounded theory in practice*. Sage, 1997.
- [16] M. Phillips and J. Lu, “A quick look at nvivo,” *Journal of Electronic Resources Librarianship*, vol. 30, no. 2, pp. 104–106, 2018.