Defence Research and Development Canada
Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

All Hazards Risk Assessment: Documenting the Calibration Process and Methodology

Ian Bayne
Jim Duncan
Maxsys

Scientific Authority:
Shaye K. Friesen
DRDC Centre for Security Science

Defence R&D Canada – Centre for Security Science

Contract Report
DRDC CSS CR 2013-016
September 2013

Canada

All Hazards Risk Assessment: Documenting the Calibration Process and Methodology

Ian Bayne
Jim Duncan
Maxsys

Scientific Authority:
Shaye K. Friesen
DRDC Centre for Security Science

Defence R&D Canada – Centre for Security Science

Contract Report
DRDC CSS CR 2013-016
September 2013

Scientific Authority

Original signed by Shaye K. Friesen

Shaye K. Friesen

Defence Scientist

Approved by

Original signed by Alain Goudreau

Alain Goudreau

Section Head - Risk Assessment & Capability Integration

Approved for release by

Original signed by Andrew Vallerand

Dr. Andrew Vallerand

DRDC Centre for Security Science, A/DRP Chair

Abstract

The All Hazards Risk Assessment (AHRA) Interdepartmental Risk Assessment Working Group (IRAWG) uses a scenario-based risk assessment to support federal emergency planning and policy development, and departmental efforts to prioritize risk within their respective mandates. In 2010, Public Safety Canada (PS) stated the requirement to create a common picture of all-hazards risks. They requested that results for the assessment of risks assessments derived from malicious (e.g., terrorist act; cyber-attack; organized crime, etc.) and non-malicious (e.g., natural disasters; industrial accidents, etc.) threats and hazards be presented on a common two-dimensional graph.

In support of this objective, the Centre for Security Science (CSS) is evaluating techniques to address these requirements including calibration of likelihood and impact estimates derived from different methods. CSS, working together with security, intelligence and risk domain experts, is evaluating approaches to calibrate malicious and non-malicious risk assessments.

This Contractor Report (CR) documents the likelihood calibration technique and the discusses the overall risk analysis process with a view to identifying implications for future work and to transition to a streamlined national risk assessment process. The main recommendation is that PS/CSS continue to collect data on the challenges associated with implementing the current process and methodology for one or two more cycles, as a basis upon which to assess the way forward, including the cost, benefit and risks.

Résumé

Le Groupe de travail interministériel (GTI) sur le cadre d'évaluation tous risques (ETR) évalue les risques au moyen de scénarios à l'appui de la planification des mesures d'urgence, de l'élaboration des politiques et des efforts ministériels pour établir l'ordre de priorité des risques conformément à leurs mandats respectifs. En 2010, Sécurité publique Canada (SP) a signalé qu'il fallait dresser un portrait commun de tous les risques. Il a été demandé que les résultats d'évaluation des risques découlant de menaces malveillantes (p. ex., acte terroriste, cyberattaque, crime organisé) et non malveillantes (p. ex., catastrophe naturelle, accident industriel) soient présentés avec un graphique bidimensionnel commun.

Pour atteindre cet objectif, le Centre des sciences pour la sécurité (CSS) considère les techniques pour répondre à ces exigences, y compris l'étalonnage des probabilités et les estimations de l'incidence obtenues de divers moyens. Le personnel du CSS collabore avec des experts dans les domaines de la sécurité, du renseignement et du risque afin d'examiner les approches permettant d'uniformiser les évaluations de risques malveillants et non malveillants.

Le présent rapport d'entrepreneur porte sur la technique d'étalonnage des probabilités et le processus global d'analyse des risques afin de déterminer l'incidence sur les travaux futurs et d'adopter un processus national d'évaluation des risques simplifié. Selon la recommandation principale, SP et le CSS devraient continuer de recueillir des données sur les difficultés liées à la mise en œuvre du processus et de la méthodologie actuels pour un ou deux cycles supplémentaires dans le but d'évaluer les prochaines étapes, y compris les coûts, les avantages et les risques.

Executive summary

All Hazards Risk Assessment: Documenting the Calibration Process and Methodology:

I. Bayne; J. Duncan, DRDC CSS CR 2013-016; Defence R&D Canada – CSS; September 2013.

Background: In 2010, Public Safety Canada (PS) stated the objective to create a common picture of risks derived from all hazards. To accomplish this objective, PS engaged federal departments in conducting risk assessments of risk event scenarios of federal interest, the results of which were to be placed on a common graph for presentation to stakeholders. A challenge is the availability of data for malicious threats compared to historical evidence for non-malicious hazards, and the different nature of risks in these two categories, necessitating different assessment methodologies. The calibration or mapping technique is intended to support comparative analysis and presentation of risk assessments of diverse threats and hazards in order to support emergency management planning.

Results: The Centre for Security Science (CSS), in collaboration with security, intelligence and other risk domain specialists, developed a prototype calibration process to support PS's objective. The calibration process was first implemented during the federal All Hazards Risk Assessment (AHRA) Cycle 2 (FY2012/13). PS plans to calibrate the risk scores of malicious scenarios for Cycle 3 (FY2013/14), and it is considering the value of the technique to support future cycles.

Significance: A preliminary investigation of approaches of other nations that have relatively mature national risk assessment programs suggests that the Canadian work on calibration is innovative and has potential. An analysis of the current approach and possible improvements would support a future decision on whether this technique is reliable, and/or there is a need to explore other tools and techniques.

Future plans: A number of factors affect decisions on future work including: the nature of malicious and non-malicious risks, which brings into question their presentation on a common graph; the source, timeliness and integrity of intelligence on malicious threats; and the experience from other nations' programs. From the perspective of transitioning the federal AHRA to a strategic national risk assessment, the suitability of currently employed scoring scales and techniques to regional and local levels needs to be assessed. Also, the choice and development of risk scenarios will depend on the level of assessment. In addition to the challenges associated with malicious versus non-malicious scenarios, the common presentation of risk assessment results will have to address issues related to scalability and presentation to diverse audiences with different planning horizons, risk exposure, and time, resource and other pressures.

Sommaire

All Hazards Risk Assessment: Documenting the Calibration Process and Methodology:

I. Bayne; J. Duncan, RDDC CSS CR 2013-016; R & D pour la défense Canada – CSS; septembre 2013.

Contexte : En 2010, Sécurité publique Canada (SP) a signalé qu'il fallait dresser un portrait commun de tous les risques. À cette fin, les ministères fédéraux ont été appelés à évaluer ces risques au moyen de scénarios d'intérêt fédéral et de présenter les résultats avec un graphique général à l'intention des intervenants. Un des problèmes rencontrés est la disponibilité des données sur les menaces malveillantes par rapport à la preuve historique quant aux dangers non malveillants, de même que la nature différente des risques dans ces deux catégories qui nécessite des méthodes d'évaluation distinctes. L'étalonnage ou la technique de schématisation vise à soutenir l'analyse comparative et la présentation d'évaluations de risques associés à diverses menaces à l'appui de la planification de la gestion des mesures d'urgence.

Résultats : Le Centre des sciences pour la sécurité (CSS) collabore avec des experts dans les domaines de la sécurité, du renseignement et du risque afin d'élaborer un processus d'étalonnage provisoire et d'atteindre l'objectif de SP. Ce processus a d'abord été mis en œuvre durant le deuxième cycle (AF 2012-2013) de l'évaluation tous risques (ETR). SP prévoit étalonner les cotes de risque des scénarios malveillants du troisième cycle (AF 2013-2014). Elle tient compte de la valeur de la technique à l'appui des prochains cycles.
Importance : Une enquête préliminaire sur les approches d'autres pays ayant des programmes nationaux plus ou moins établis d'évaluation des risques a révélé que les travaux canadiens concernant l'étalonnage sont novateurs et ont du potentiel. Une analyse de l'approche actuelle et des améliorations possibles serait utile pour éventuellement déterminer si cette technique est fiable ou si d'autres options sont à envisager.

Perspectives : Un certain nombre de facteurs influent sur les travaux futurs, y compris : la nature des risques malveillants ou non malveillants, ce qui remet en cause leur présentation avec un graphique commun; la source, la rapidité et l'intégrité du renseignement relatif aux menaces malveillantes; l'expérience liée aux programmes d'autres pays. Pour faire de l'ETR fédérale une évaluation stratégique du risque national, il faut examiner la pertinence des côtes et des techniques actuellement utilisées aux niveaux régional et local. Le choix et l'élaboration des scénarios de risques dépendront également du niveau d'évaluation. En plus des problèmes associés aux scénarios malveillants par opposition à ceux non malveillants, la présentation générale des résultats d'évaluation des risques devra aborder les questions liées à l'échelonnabilité et tenir compte des divers auditoires dont les horizons de planification, l'exposition aux risques et les contraintes de temps, de ressources et autres diffèrent.

Table of contents

List of figures and tables

Acknowledgements

CSS acknowledges the support of the Public Safety Canada (Julie Cranton and others), CSS defence scientists (Shaye Friesen and Dr. Simona Vega), security and intelligence specialists, the AHRA Interdepartmental Risk Assessment Working Group and Risk Sciences International (RSI) for their contributions to the development of the calibration technique.

CSS also acknowledges the contributions of risk management consultants (Ian Bayne and Jim Duncan) in the production of this Contractor Report.

1 Overview

1.1 Introduction

At the time this report was written, Public Safety Canada (PS) was in the third annual cycle of the All Hazard Risk Assessment (AHRA) implementation (FY 2013/14). The calibration process and methodology is being documented at this time as part of AHRA transition planning and to support future efforts in implementing the federal AHRA. Calibration is an innovative technique that is intended to facilitate the validation and communication of risk assessments across risk domains, and to better inform stakeholders with regard to resource allocation for risk treatment strategies and capability development.

In 2011, PS asked the Centre for Security Science (CSS) to investigate methods that would enable the display of malicious and non-malicious risks on a single, two-dimensional graph for presentation to senior management. In FY 2011/12, CSS started looking at the calibration technique starting with comparative analysis of malicious likelihood estimates. PS and CSS formed a working group that included security, intelligence and other risk domain specialists. The working group developed nine reference scenarios that could be used to facilitate mapping the likelihood estimates of malicious threats on to the same graph as the non-malicious scenarios that were being plotted using frequency and impact.

The calibration process and methodology are an extension of the impact/consequence analysis part of Step 3: Risk Analysis,[1] which is described in the AHRA Methodology Guidelines, hereafter referred to as the Guidelines (depicted in Figure 1). CSS is evaluating whether calibration should be integrated into the risk assessment scoring tool or it should remain a parallel process involving a separate expert elicitation process that exploits pairwise comparison and other techniques. PS and CSS are also reviewing options to calibrate impact assessments. The likelihood calibration technique was first implemented during Cycle 2, when results from a qualitative method were mapped on to the frequency-based likelihood scale used for the non-malicious scenarios. CSS and PS weighed the benefits of incorporating the pairwise comparison method and the existing reference scenarios into the rating tool, thus replacing the original qualitative method and eliminating the need for post-workshop calibration of results. However, as of May 2013, the full process has not been implemented and a decision has not been made on the way forward for impact calibration.

The Guidelines describe the risk analysis process including likelihood and impact assessment approaches (Chapter 3). Chapter 4 discusses risk evaluation but it does not describe the calibration methodology and process, which will presumably be added later when the process has stabilized and gained credibility. Risk analysis and risk evaluation are two steps in the federal AHRA process as indicated in Figure 1. A separate CSS report is being developed to explain the detailed the emerging calibration process. PS is reviewing options for calibration of impact analyses, which were developed by a contractor (i.e., RSI).

This contractor report (CR) should be read in conjunction with the AHRA Body of Knowledge (BoK), which considers options to transition the AHRA framework and methodology to a

[1] AHRA Methodology Guidelines (2012), Impact/Consequence Analysis: 22-40.

national risk assessment (NRA). One conclusion is that a strategic benefit of a NRA approach could be to bring the diverse regional perspectives into focus to support development a true national risk picture, which should support balance of investment decisions on multiple levels, and calibration could support consolidation of diverse risk assessments. The NRA could provide a valuable feedback loop to validate and refine the federal risk assessment process. This report examines the emerging calibration process and methodology with transition to a strategic NRA process in the 2-3 year timeframe in mind.

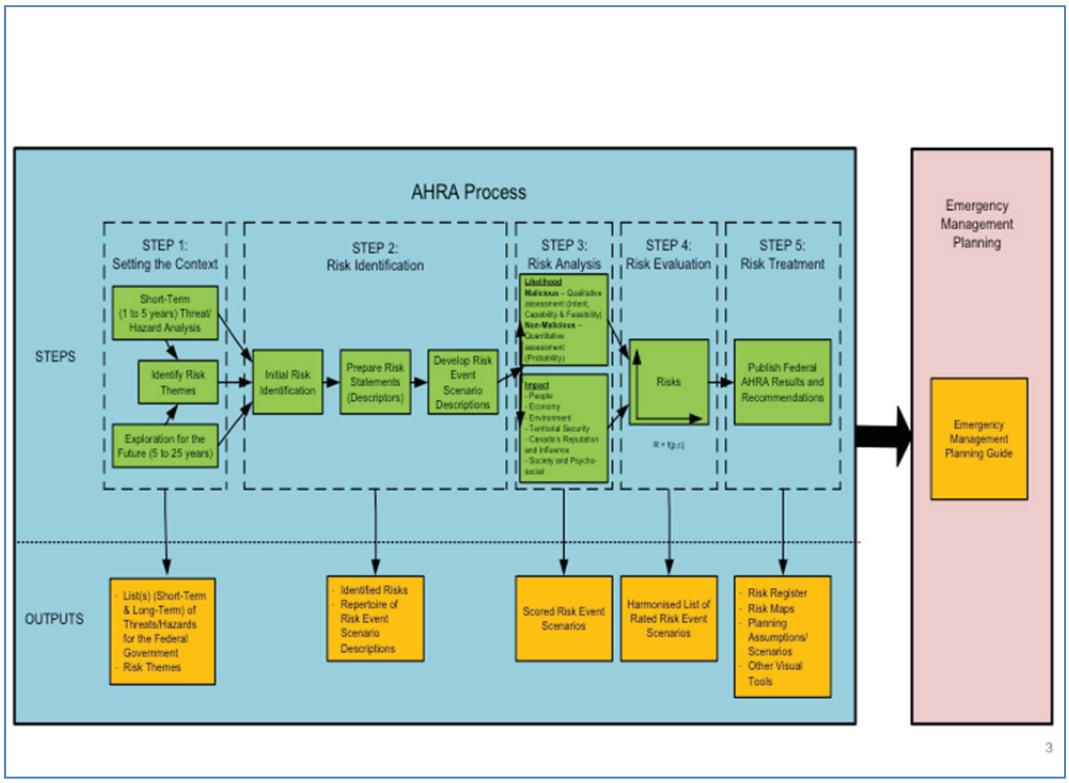


Figure 1: Federal AHRA Process

1.2 Background and Context

Documenting the calibration process and methodology is one activity within the CSS AHRA Transition Project (2012-2014), which consists of four tasks: All Hazards Risk Assessment (AHRA) Body of Knowledge (BoK) that summarizes the information baseline for the federal AHRA work during the 2006 to 2012 timeframe; documentation of the capability assessment methodology; automation support that investigates collaboration, data mining and information management technology; and strategic asset management that investigates automation support for a capability assessment methodology.

The AHRA Transition Project and the Interdepartmental Risk Assessment Working Group (IRAWG) activities are intended to support the advancement of the federal AHRA methodology, and to advance the risk assessment and capability integration components of the Canadian Safety and Security Program (CSSP).

From a federal AHRA methodology perspective, it is important to note that the original AHRA scope focused on the federal "all-hazard risk domain", which does not include the operational risk domain that is defined as "day-to-day issues confronting an institution", and that the Guidelines state that "these aspects may be considered in "Setting the Context", prior to identifying risks and assigning impact ratings…".[2] Furthermore, the AHRA considers a broad range of threats and hazards, and the methodology focuses on events that are above a certain threshold in magnitude. The technique of establishing magnitude thresholds is relevant to a NRA and the thresholds may vary for different regions depending on a number of factors including inherent resiliency, and time and space.

Two types of scenarios are considered in the calibration process:

- **'Actual' scenarios** are chosen and developed by primary departments. These include three types of scenarios – nominal, elevated and reduced for malicious threats and non-malicious hazards. For the purposes of this report, the term **anticipated** scenarios is preferred. They form the basis of the annual AHRA risk scoring workshop process; and

- **Reference scenarios** are the set of fictitious malicious scenarios developed by subject matter experts (SMEs) starting in 2011/12. Nine reference scenarios were developed using Expert Elicitation techniques and workshops with breakout groups to focus on specific scenarios. This set of scenarios formed the basis of the malicious threat likelihood calibration technique.

Figure 2 is an example of the two-dimensional graph that PS envisages can support summary briefings to senior management. The intent is to be able to display diverse risk scenarios on a common graph. PS uses the term "heat map". The graph is also called a "scatter plot". In the example below, there are no delineated cells and the size and shape of the circles are not used to differentiate risks.

[2] AHRA Methodology Guidelines (2010): page 6.

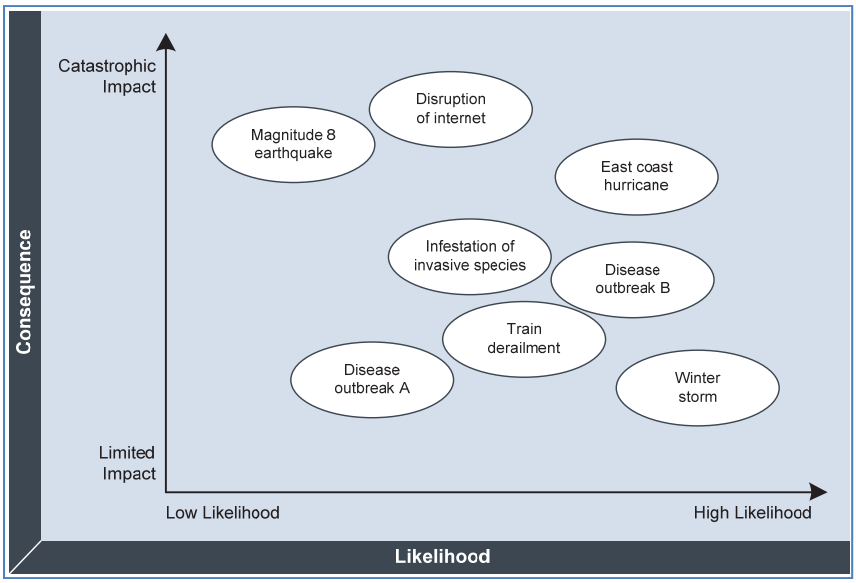


Figure 2: Sample of a Simplified Graphic (Based on a UK Model)

The Guidelines state that 'the intention of the [AHRA] process is … to produce a whole-of-government risk picture to support EM planning across federal government institutions and to ensure that interdependencies are recorded and managed" (2012: 2). The graph illustrates the ranking of risks assessed at the time, relative to the other risks being assessed. The technique uses a mathematical formula to display the output of the qualitative risk assessments. A challenge is that formulas vary across domains.

1.2.1 Scenario Development Process

The Guidelines describe the annual business cycle for the scenario-based risk assessment process and the process for developing risk event scenarios.[3] A CSS depiction of the four-steps within the broader AHRA context is given in Figure 3 below. The PS approach is that risk assessments "should be based on present day risk events" that are "under federal jurisdiction" in order to define the problem space and to place some limits on the number and type of risks being considered – to focus resources attention on risks that are relevant to federal mandates and EM planning horizon.[4]

[3] AHRA Methodology Guidelines (2010): page 16.
[4] Ibid., pp. 14, 11.

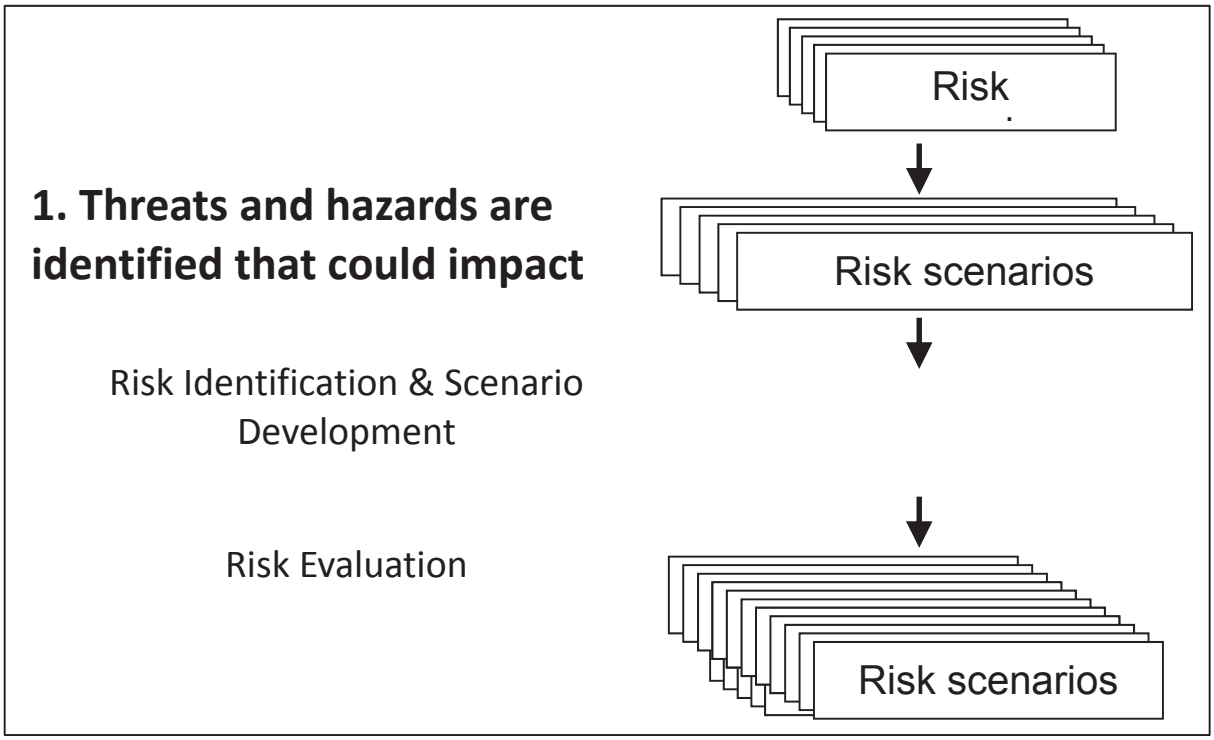


Figure 3: Scenario Development Process

The scenario development process identifies risk events in two timeframes – 5 years and 5-25 years. Scenarios describe risk events that are plausible in the next five years and that could have consequences that are possible within a 25-year timeframe.

1.2.2 Taxonomy

Departments may use the AHRA taxonomy to identify risk event scenarios. The Assistant Deputy Minister Emergency Management Committee (ADM EMC) provides guidance on priorities and scenarios selected for the annual cycles. The taxonomy tool that is being used to differentiate malicious and non-malicious threats and hazards is presented in Figure 4.

The terms malicious and non-malicious are not defined in the AHRA glossary; however, malicious and non-malicious threats / hazards can be described as follows:

- Malicious threats are intentional. The risks originate from threat actors like terrorists, organized crime and extremists; and
- Non-malicious threats include unintentional human-caused, health, accidents, technological failures and natural disasters.

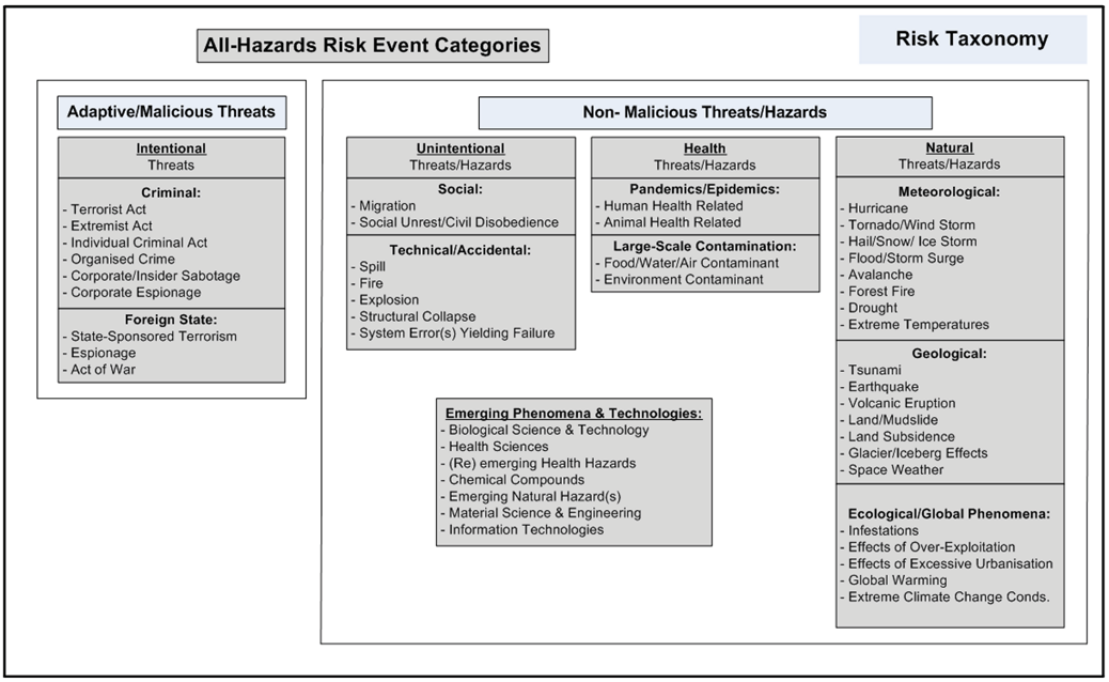


Figure 4: AHRA Taxonomy (2010)

1.3 Scope and Use Cases

This study focuses on the broader AHRA context including whether the technique should be embedded in the risk scoring tool or retained as a separate process. It also considers the existing impact assessment categories to support future work on calibration of impact assessments. Furthermore, this study also looks at other national risk assessment program approaches that are relevant to the calibration process and the graphical display diverse risk scenarios. These aspects are reviewed with the transition to a NRA in mind.

The following calibration use cases inform this report:

- Calibration of the malicious threat and non-malicious hazard likelihood assessments;
- (Future) Calibration of impact assessments and mapping of impacts across categories;
- Feedback to the scenario development and prioritization process;
- Feedback to departmental risk assessment and scenario identification processes;
- Feedback to understanding of all hazards risk assessment as a framework and not a one-size-fits-all multi-criteria risk assessment technique;
- Feedback to evaluation of techniques and tools for a national risk assessment process; and
- Feedback to the overall risk analysis process (Step 3 of AHRA process).

1.4 Objectives

The main objective of the report is to summarize the existing malicious likelihood calibration process and methodology, and to present observations and ideas for transition to a NRA program. Therefore objectives include:

- Describe the context and rationale for the calibration process;
- Provide an overview of the malicious likelihood calibration methodology;
- Document concerns expressed by SMEs;
- Present observations on the six impact categories and their relevance for a NRA;
- Present deductions from a preliminary investigation of other national approaches to displaying the output of risk assessments on a single risk matrix; and
- Provide suggestions for future work.

1.5 Approach

The methodology consisted of a CSS document review and discussions with CSS SMEs who are participating in the development of the malicious likelihood calibration technique. The investigation included a preliminary investigation of selected national programs.

2 Findings

This section describes the scenario development process. It summarizes the likelihood calibration process work to date and the emerging options for a similar impact calibration technique. Some concerns that were expressed during the initial development of the malicious risk assessments are documented for future consideration. Finally, some first impressions related to calibration and the overall processes are included here and in the recommendations section for future consideration.

2.1 Scenarios

The basis of likelihood and impact assessment is the risk event scenario that describes the threat / hazard in terms of:

- **Who?** – Sometimes referred to as actor (e.g., individual, cell, organized group, lone wolf, syndicate, state-sponsored, etc.);
- **What?** - Description of the target or objective of the malicious attack (e.g., Individual/VIP, people/mass crowd, symbol/landmark, institution, critical infrastructure, trade gateway, etc);
- **How?** – Description of the method of attack sometimes referred to as a vector (e.g., CBRNE, weapons, cyber, disturbance/action, assault, etc); and
- **Where, Why, When?** – Description of the context, time dimension, motivation for the attack, location and other specific information that could inform the risk assessment (Optional).

Given the scenario, the risk assessment workshop participants are expected to rank the risk event scenarios by completing likelihood and impact estimates using the AHRA risk scoring tool, which does the mathematical calculations behind the scene, and ranks the risks based on the aggregated ratings (Excel spreadsheet).

2.2 Likelihood Calibration

Different types of risks use different approaches depending on the availability and quality of information. AHRA uses two basic approaches for malicious likelihood estimates, which are summarized below. The "old" process refers to the multi-criteria risk assessment without using likelihood calibration. The "new" process is intended to include likelihood and eventually impact calibration.

2.2.1 Malicious Likelihood Estimates – *The Old Process*

The IRAWG determined that is was not appropriate to estimate the frequency distribution for adaptive malicious threats. Therefore, the assessment was performed using a combination of threat, vulnerability and consequence analysis based on expert elicitation and intelligence reports (classified and/or open source). CSS developed the following generic "word ladders" to support estimation of the likelihood for malicious threats, which are included in the help page of the Excel risk scoring tool (see tables below).

Table 1: Technical Feasibility (Malicious Threat)

Technical Feasibility Table (Generic)	
Extremely Feasibility	Event **Extremely Feasible** due to the readily available and accessible Material or Equipment, and either the simplicity of execution or the ease of access to the necessary levels of knowledge and expertise.
Very Feasible	Event is **Very Feasible** due to the readily available and accessible Material or Equipment, and either the low complexity of execution or the ease of access to the necessary levels of knowledge and expertise.
Moderately Feasible	Event is **Moderately Feasible** due to the availability and accessibility to Material or Equipment, and either the moderate complexity of execution or the moderate ease of access to the necessary levels of knowledge and expertise.
Low Feasibility	Event has **Low Feasibility**, but is not impossible, due to the special Material or Equipment availability and controls and/or more difficult access, and either the very high complexity of execution or the very high effort to access the necessary advanced knowledge and skills levels.
Very Low Feasibility	Event has **Very Low Feasibility** due to the special Material or Equipment availability and controls and/or very difficult access, and either the extremely high relative complexity of execution or the extreme effort to access the necessary advanced knowledge and skills levels.

Table 1: Capability Assessment (Malicious Threat)

Capability Table (Generic)	
Extremely Capable	There are **ample documented cases** of evidence of an individual's or a group's capability to successfully execute these types of actions.
Very Capable	There are **several documented cases** of evidence of an individual's or a group's capability to successfully execute these types of actions.
Moderately Capable	There is **at least one documented case** of evidence of an individual's or a group's capability to successfully execute these types of actions.
Low Capability	There is **at least some evidence** of an individual's or a group's capability to successfully execute these types of actions.
Very Low Capability	There is **little-to-no evidence** of an individual's or a groups' capability to successfully execute these types of actions.

Table 2: Intent Assessment (Malicious Threat)

Intent	
Extreme Intent	There are **ample documented cases** of evidence of an individual's or a group's intention to execute these types of actions.
High Intent	There are **several documented cases** of evidence of an individual's or a group's intention to execute these types of actions.
Moderate Intent	There is **at least one documented case** of evidence of an individual's or a group's demonstrated intention to execute these types of actions.
Low Intent	There is **at least some evidence** of an individual's or a group's demonstrated intention to execute these types of actions.
Very Low Intent	There is **little-to-no evidence** of an individual's or a groups' intention to execute these types of actions.

The IRAWG assessed impacts using the Excel scoring tool, which included confidence assessments throughout the process. The scores were aggregated and plotted on a two

dimensional graph (likelihood and impact/consequences). The output was reviewed to confirm the relative ranking of the scenario compared to other scenarios. This work led to the development of the "new process", which is based on the calibration process and methodology.

2.2.2 Likelihood Calibration Process and Methodology – *The New Process*

The risk assessments are based on risk event scenarios that could disrupt critical federal services, overwhelm federal emergency management resources and/or prevent federal emergency plans from being executed. In the case of malicious threats, this can involve relying on classified information on specific threats within specific timeframes. It was determined that the techniques for ranking malicious and non-malicious threats and hazards should be different due to the nature of the risks, the sources and type of data, and the ability to predict future events or conditions that translate into unacceptable risk for federal institutions or the government as a whole. This reality of the problem space is the reason that a qualitative method (i.e., "old technique") was used. CSS and PS are experimenting with different calibration techniques to be able to compare assessments of diverse scenarios and to enable the outputs to be plotted on a common risk picture.

Non-Malicious Likelihood Assessment
The likelihood assessment for non-malicious threats and hazards consists of a risk frequency scale to avoid using probabilities. The likelihood scale is logarithmic with half scores for ratio increments of the square root of 10 and impact scale uses a logarithmic scale. This enables the scoring process to differentiate between high-likelihood but low impact events and low likelihood but high impact events.

For natural hazards the assessment is based on historical information and/or the best available science predictions, and modifier can be applied for anticipated changes that could change the frequency. For accidental and technological scenarios, the assessments are based on statistical evidence and expert opinion. The rating table from natural hazards can be used with modifiers applied to technology where emerging trends can be reasonable predicted to affect the rate of occurrence.

Malicious Likelihood Assessment
Under the "old' rating scale, the malicious likelihood assessment considers three criteria: Technical Feasibility; Capability; and Intent, which are combined into an aggregate likelihood rating. Rating for malicious scenarios relies mainly on expert judgment from the security and intelligence community augmented by S&T specialists including risk assessment, operations research and capability analysis. The method employs a subjective qualitative ranking of likelihood that considers the determined and adaptive nature of malicious threats.

The last part of the assessment is a judgment about the *intent* of the adversary. This assessment is based on intelligence analysis of the individual/organization, its sponsors and known objectives. The second part of the likelihood assessment assesses the ability of the adversary to carry out the attack (i.e., *technical feasibility)*. This assessment considers several factors: material; equipment; access to target or system; technical expertise; and access to critical information. The lowest combined score across the variables is used to define the lower limit of feasibility for the given scenario. The third element is *capability* or enabling capability, which is the assessment of the threat actor and the support capabilities to successfully carry out an attack. This assessment considers two factors: organizational capability including command, control, communications and

intelligence; and support & logistics. The aggregated score for the two variables uses the lowest rating.

Malicious likelihood assessment combines the three scores to produce the overall likelihood ranking. The overall capability score is determined by summing the technical feasibility and enabling capability scores using a predetermined table, which produces a single score for overall capability. The final likelihood ranking is the result of combining the overall capability score and the Intent score using the notion of the *weakest link* (i.e., least capability that could still carry out a successful attack).

Elicitation of expert judgments is the critical aspect of the assessment malicious threats. Intelligence experts do not use probability estimates to evaluate threats. At the end of the AHRA Cycle 2, CSS described the calibration process as follows.

> A calibration process was developed following the 2010-2011 AHRA cycle that sought to establish a correspondence between the likelihood of malicious events and the non-malicious frequency scale. The process sought to make the likelihood assessment for malicious events compatible with frequency-based estimates such as: "once every 10 years", or equivalently, "10-1 times per year.
>
> This was accomplished via a set of hypothetical "reference" scenarios that, by design, were intended to describe events ranging from very frequent (such as once per year) to very rare (once per 100,000 years). Nine reference scenarios were developed, to fit into three broad categories: high, medium and low likelihood. Each likelihood category was represented by a "triplet" of three reference scenarios, which were ranked relative to each other by doing a pair-wise comparison within the triplet. The likelihood estimate, on the non-malicious scale, of at least one of the hypothetical reference scenarios in the triplet was required, which was accomplished through consensus among a group of intelligence experts, given assumptions about the feasibility and capability components, and the intent.
>
> Once one scenario was "pinned" to the frequency scale, the other two followed by virtue of their relative position within the triplet. The process was repeated for each of the three "triplets" of reference scenarios covering the same likelihood category; at the end, all nine reference scenarios were positioned on the frequency scale, in other words, their likelihood was described in terms that are compatible with frequency-based estimates.
>
> The set of nine hypothetical scenarios served as reference points by which the likelihood of actual malicious scenarios was then estimated, by assessing their likelihood relative to the likelihood of one or more of the scenarios in the reference set. The process was piloted during the 2011-2012 AHRA cycle and proved successful."[5]

The pairwise comparison technique was applied for each set of three scenarios for high, medium and low thresholds. Two scenarios were presented at a time. The working group of experts answered two questions for each pair of scenarios: which scenario is more likely and how much

[5] CSS, Verga e-mail, 7 Feb 13 – FY 2011/12 (Cycle 2)

more likely is it? The output of this process was the relative likelihood for each scenario. The process was repeated to compare high, medium and low likelihood scenarios across the six impact categories. The output of this two-step process was the relative ranking for all nine reference scenarios, which could then be used to make an expert judgment on the accuracy of the predictions for the "anticipated" scenarios. Figure 5 summarizes the emerging malicious likelihood calibration process.

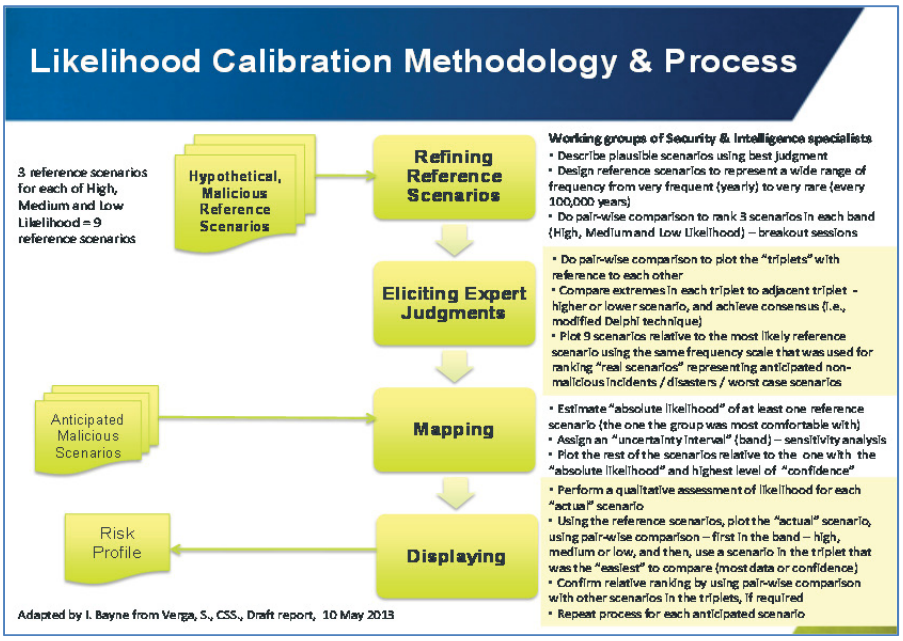


Figure 5: Emerging Malicious Likelihood Calibration Process

2.3 Non-Malicious and Malicious Impact Assessment

Impact assessments for malicious and non-malicious risk use a common impact assessment framework based on six impact categories that are described in the AHRA Methodology Guidelines: people, economy, environment, Canada's reputation and influence, territorial security, and societal and psycho-social effects. Each category is assessed using an order of magnitude approach. The technique varies by category. For some categories, a logarithmic scale is used to aid in differentiating impact severity, while for others, the technique is more qualitative. The impact assessments can be completed at the same time or separately from the likelihood assessment. The Guidelines describe the impact categories in detail. The calibration process for comparative analysis of impact assessments has not been developed.

2.4 Concerns

A review of communications during Cycle 2 indicates that SMEs expressed concerns with the likelihood assessment approach and/or calibration. These comments are summarized here for future consideration.

- **Intel community**: The S&I SMEs were "very clear" about not using probabilities for likelihood estimates for malicious, adaptive threats due to the lack of historical data and the availability of context- and time-sensitive intelligence reports (various discussions); and "…intelligence community is adverse towards percentages, probabilities, numbers etc. A word ladder would be their preference. Instead of putting a log scale would it be better initially to indicate low/high annual likelihood on the scale". (PS, 29 November 2011).

- **Transport Canada:** "TC continues to disagree with the process of evaluating the scenarios. We feel the process lacks rigour, and scoring is based more on gut feelings than science. We just don't want the scoring to give senior managers false confidence." (TC, 24 February 2012); and "In talking to the risk SMEs within TC, their main concern relates to possibly giving senior managers a false sense of confidence with the results, when several process gaps (or assumptions) exist and aren't awarded the proper caveats (where they can't be fixed for lack of benchmark." (TC, March 2012); and,

- **CSIS, TC and HC**: The attendees at a working group of experts did not agree with the results of the calibration exercise. Therefore, PS proceeded with using the results that were generated at the risk scoring workshops. PS decided to present the results on one graph, and to differentiate between the malicious and non-malicious scenarios (different colors). PS stated that, "in the next cycle of the AHRA, we will explore a more thorough exercise for calibrating the likelihood scales (PS, 26 September 11).

2.5 Limitations

The Guidelines discuss using deterministic methods such as models and simulations. Some departments have such tools. For example, Environment Canada uses plume modeling for chemical release scenarios. NRCan uses Hazards US (HAZUS) to provide input for the impact of an earthquake scenario. The outputs from these more detailed models provide input to the assessment process. The availability of such tools and the resources to apply them in other departments are systemic constraints, which would be relevant at the regional level for a NRA.

2.6 Impact Calibration Process and Methodology - Options

A contractor presented options to PS for calibrating impact scoring scales in March 2013. Options consist of:

- **Calibration by Design**: The AHRA scoring tool would include calibration with no SME intervention. This would enable a one-step assessment process, but it may not provide assessors with visibility of the process and it may not improve confidence in the assessment or calibration techniques; and
- **Post-Hoc Calibration**: Keeps the scoring and calibration tools separate, which means that the calibration could be done by the same group or a third party after the risk scoring workshop.

The main assumption expressed by the contractor is that, "a calibration exercise is required to determine the equivalency between scores in each impact category"[6]. A number of comparison methods are presented. A problem with any of the methods is that the participants may not have experience with the techniques described, which would have learning curve, and therefore, cost, time and support implications. The contractor did not present cost estimates of the work to embed the calibration in the risk scoring tool. The paper also presents alternatives for elicitation that are potentially labour-intensive, and there would be dependencies on people who may not have the relevant experience and access to the relevant documentation.

Other observations on the contractor's presentation include that, with the exception of decision support tools, there is no mention of electronic voting, crowdsourcing or other information technology alternatives to the AHRA scoring tool. There is also no discussion of weighting of categories. This would require a "political" or value judgment to be placed on each criterion; however, this could be more powerful than the current method of differentiating levels of impact through multiple pre-assigned "modifiers". There is no cost / benefit / risk assessment for the status quo and the options. Therefore, the decision requested appears to be to pick one or more options to invest more money in the risk scoring tool.

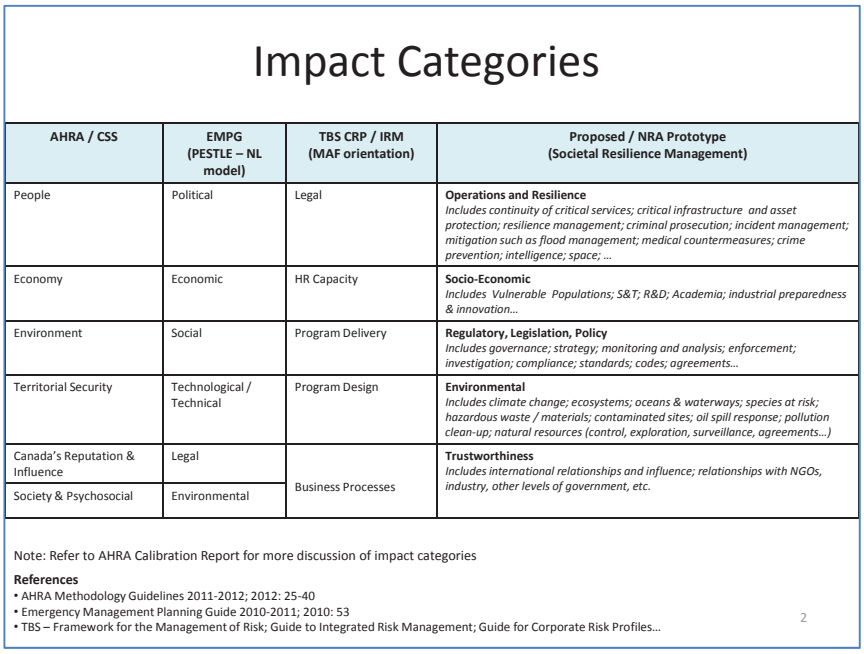
2.7 Observations on the Impact Assessment Categories

Study team observations on the existing six categories influenced by experience with federal risk management activities include:

- **Categories**: The impact categories have been reviewed beyond the AHRA community including presentations to international counterparts (e.g., DHS);
- **Consistency**: The categories are different than the ones in the EMPG and TBS Management of Risk, and they appear to be skewed to health and safety risks (See comparison Figure 6);
- **Clarity**: Although the consequences are different (health and safety vs. psycho-social), more than one category deals with the impact on people, which appears to contradict the Guidelines stated intention to make the categories orthogonal;
- **Complexity**: Cost impacts are consolidated in the economic impact. This puts the onus on the assessor who may have very limited knowledge of cost estimating;
- **Variance**: Categories do not cater for dynamic ranges of values;
- **Category Selection**: There is no category for operations, which would presumably be critical at the regional level (i.e., for NRA), and for convergence with capability assessment and balance of investment decision support (i.e., the risk assessment is a beginning, not an end in itself);
- **Frequency Scales**: While the frequency scales may help decision makers understand scaling of impact and the assessments be mathematically coherent from an academic / research perspective (e.g., scale up by orders of magnitude), it is not clear that the extreme values are relevant or that they improve the quality of the assessment (e.g., non-malicious hazard – 11 levels – with most infrequent being 100,000 years); and
- **Value**: The real value of the impact assessment may be in adding clarity to the risk event scenarios, not in the enabling the assessment itself. For the actual assessment, the scales are complicated and, at least at the regional level, probably not that relevant unless the

[6] Paoli, G., 28 March 2013.

analysis is focused on treatment options for a specific risk domain (e.g., public health; insurance).



Impact Categories

AHRA / CSS	EMPG (PESTLE – NL model)	TBS CRP / IRM (MAF orientation)	Proposed / NRA Prototype (Societal Resilience Management)
People	Political	Legal	**Operations and Resilience** *Includes continuity of critical services; critical infrastructure and asset protection; resilience management; criminal prosecution; incident management; mitigation such as flood management; medical countermeasures; crime prevention; intelligence; space; …*
Economy	Economic	HR Capacity	**Socio-Economic** *Includes Vulnerable Populations; S&T; R&D; Academia; industrial preparedness & innovation…*
Environment	Social	Program Delivery	**Regulatory, Legislation, Policy** *Includes governance; strategy; monitoring and analysis; enforcement; investigation; compliance; standards; codes; agreements…*
Territorial Security	Technological / Technical	Program Design	**Environmental** *Includes climate change; ecosystems; oceans & waterways; species at risk; hazardous waste / materials; contaminated sites; oil spill response; pollution clean-up; natural resources (control, exploration, surveillance, agreements…)*
Canada's Reputation & Influence	Legal	Business Processes	**Trustworthiness** *Includes international relationships and influence; relationships with NGOs, industry, other levels of government, etc.*
Society & Psychosocial	Environmental		

Note: Refer to AHRA Calibration Report for more discussion of impact categories

References
- AHRA Methodology Guidelines 2011-2012; 2012: 25-40
- Emergency Management Planning Guide 2010-2011; 2010: 53
- TBS – Framework for the Management of Risk; Guide to Integrated Risk Management; Guide for Corporate Risk Profiles…

2

Figure 6: Impact Categories (an alternative framework)

In essence, the federal AHRA scenario process is using a bottom-up approach (i.e., departments identify scenarios from an institution or mandate perspective) that does not appear to be focusing attention on several serious scenarios. A NRA program could help to validate federal scenarios and add value to the federal AHRA activity by focusing attention of priorities including those that affect multiple regions.

Table 4: First Impressions of Impact Categories

Category	Comments
People	• The tool has 17 judgments. Many of the categories deal with human health effects that may or may not be known for some time. This would probably not support a forward-focused, capability assessment of medical countermeasures, vaccine production and distribution capacity, and other large-scale requirements. • The value of the assessment of non-fatal heath impacts is not intuitive. • Society and Psycho-social also deals with People impacts, which violate the principle of having independent categories (Guidelines, page 23). • It may be more useful to consider socio-economic as a category, especially for a region-based NRA (proposed in the AHRA BoK).

Category	Comments
Economy	• The scale has 17 levels ($10K to $100B, and user defined). This would cover catastrophes on a scale with recent disasters – Deepwater Horizon; and the Japan quake-tsunami-Fukushima Daiichi meltdown[7]. A West Coast earthquake is a realistic scenario and the impact scale would be adequate. For another comparison, the Netherlands (NL) uses Euro 10M to Euro 50B. However, the scale could be simplified by providing fewer levels and providing a range of values. • An alternative could be to express the financial impact in percentage of operating budgets, which would be more scalable. For example the total operating budget of BC plus other factors including the response from across Canada could yield a financial model. BC must have such estimates readily available for reasonable worst-case scenarios. • This category could be used to explicitly connect AHRA to critical infrastructure (assets, services).
Environment	• Maximum duration of impact is 6 years, which is not a catastrophic environmental disaster • A worst-case scenario could be based on the Deepwater Horizon Disaster and the financial and environmental impact in the billions. • The extent modifiers are ambiguous and their use begs the question, so what? • A confidence level is required six times. This seems to be overkill and redundant. Why not use one confidence assessment for the category or overall assessment?
Territorial Security	• Modifiers are largely irrelevant • This category is ambiguous and not comprehensive (e.g., violation of sea limits) • An alternative approach would be to have a category for operations, which could be described for any organization or jurisdiction including defence, aid to the civil power, humanitarian operations, illegal fishing, human trafficking and smuggling, etc. • A broader interpretation could also encompass national sovereignty, and the continuity / resilience of government operations and the provision of services to Canadians, etc.
Reputation & Influence	• Reputation is a private sector term. Public trust is more relevant, but it is too narrow. Trustworthiness is a broad term that could be more useful. • The category is not well defined. Influence refers to public reaction, but this is not reflected in the tool in this category, although it should be addressed in the psychosocial category.
Society and Psychosocial	• This category is focused on emergency social services, and health and medical response and recovery. • It appears that it would take significant time to elicit expert judgement on most of the tool inputs. If the assessment were done individually, then it would presumably be difficult to aggregate and/or validate the assessment.

[7] The World Bank estimates that the costliest natural disaster of all time is US$235B, and the full cost will not be known for years (Japan earthquake / tsunami / Fukushima Daiichi meltdown, 2012). (After Catastrophe, Carlson, S., The Chronicle Review, 6 May 2013)

2.8 Observations on Risk Scoring Tool

While some members of the project team for this particular study have not used the tool in a workshop setting or witnessed a workshop, some first impressions are presented based on experience in risk assessments, critical infrastructure criticality analysis and business impact analysis in support of organizational resilience for multiple federal institutions and multiple levels of government including participation in a national level scenario-based, executive risk assessment workshop.

Similar to the use of frequency in lieu of uncertainty or probability, the reliance on hypothetical statistics could skew the impact assessments and make it very difficult to calibrate across categories. For example, while fatalities may be a useful metric for considering public health hazards and emergency planning strategies, the values do not scale easily, and they are not relevant when it comes to operational countermeasures and capabilities. Differentiating between types of fatalities (e.g., adult, children) may be useful for some elements of a capability (e.g., technology; emergency social services; evacuation; medicine…), but this level of granularity would not appear to inform the development of a broad-based, system of systems, national risk treatment strategy[8].

Similarly, using orders of magnitude to scale up financial impacts is problematic. It may be more relevant to consider percentages of operating budgets or ranges. This technique could be used for organizations, sectors and different levels of government. Using fixed numbers is inflexible, and it can be misleading and reduce confidence. It is also not clear how the tool is used to differentiate risk event scenarios in the two timeframes – near-term – five years; and future 5-25 years, and the process for getting reliable input for consideration of "future" scenarios is unclear (2012:11).

2.9 Observations on Other Nations' Approaches

This section briefly summarizes other nations' approaches for likelihood and/or impact assessment, and for displaying and communicating the output of risk assessments to management. Nations considered included: United Kingdom (UK), United States (US), Australia / New Zealand (AS/NZ) and the Netherlands (NL). Table 5 highlights some elements of the evolving national concepts (not in any particular order).

[8] It is noted that departments do ask CSS for advice about this level of granularity.

Table 5: National Risk Assessments – Concepts / Assumptions

Nation	Concepts / Assumptions
NL	"We deliberately avoided the traditional 'risk is likelihood times consequence', because this tends to suggest a strictly quantitative interpretation and because reducing 'risk' to a single number conceals two fundamental dimensions. After the risk scenarios have been assessed for these two components [likelihood, impact] they are merged and an overall picture of the various types of risks is created (DNRA, p.23)."[9] NL considers a 5-year period in two ways – first, when an event is likely to occur, and second, when decision makers expect "significant allocation of additional capabilities for adequate safety management…"[10] "An apparent weakness of the Dutch NRA [DNRA] is that there is little external validation, control, or understanding of the expert opinion process for the various hazards and threats."[11] "The DNRA's two-dimensional risk diagram can only offer an approximate ordering of national risks, whereby the position of malicious threats seems far more uncertain than that of natural or technological hazards. This makes any overall risk ordering less reliable."[12]
AS	"The AS government's approach to Critical Infrastructure Resilience goes beyond risk management and business continuity planning (which to a large extent only addresses reasonably foreseeable risks) to also address hazards and risks that are unforeseen or unexpected… A resilience approach to managing risks…encourages organizations to develop a more organic capacity to deal with rapid-onset shock. This is in preference to the more traditional approach of developing plans to deal with a finite set of scenarios." "Perception bias can often permeate an organization's thinking about foreseeable risk. This bias tends to discount scenarios that have not occurred in the recent experience of the decision maker, and bypasses a serious attempt to prove or disprove their plausibility. The constantly changing nature (and accelerating rate of change) of the economy, technology and society mean that past events are not an adequate guide to determining plausible future hazards."[13]
UK	"The estimates of frequency and consequences for each of the events considered were compared where appropriate. No effort was made to create a simple "risk judgment" for any event type because it was deemed infeasible to aggregate all consequence types into a single metric."
US	A National Research Council report concluded, "A fully integrated analysis…is likely to be inaccurate or misleading…. The risks presented by terrorist attack and natural disasters cannot be combined in one meaningful indicator of risk, and so an all hazards risk assessment is not practical (pp. 8-9)."[14]

[9] Vlek, C. (2013), How Solid is the Dutch (and UK) NRA…? Risk Analysis; Society for Risk Analysis (SAR): pp. 14-15, section 5.1.
[10] Ibid, p.11
[11] Ibid, p. 18, section 7.2.
[12] Ibid, p. 19, section 9.4.
[13] Australian Government's CIR Strategy, 2010: 3
[14] Vlek, C. (2013), p. 19, section 8.4.

Deductions for a NRA based on a preliminary review of other national programs include:

- Canada does not have the same legislative structure or financial flexibility and therefore, Canada will have to develop a solution that is practical, relevant, credible and affordable;
- All nations realize the limitations of quantitative and qualitative risk analysis techniques, and all countries with NRA programs are experimenting with a variety of techniques to improve the quality and reliability and usefulness of risk information;
- A Canadian NRA framework should address the requirement to manage classified and sensitive but unclassified information, including government and industry information;
- A Canadian NRA should keep malicious and non-malicious risk assessments separate when it comes to displaying the outputs to decision makers;
- A Canadian NRA should consider a multi-level approach to scenarios with regional scenarios being more focused on emergency operations, and the capability and capacity to cope with the impact of malicious threats and non-malicious hazards, and manage the consequences over time.

3 Conclusions and Recommendations

The conclusions from the review of the emerging calibration methodology and process, and a preliminary benchmark analysis are presented below, followed by recommendations for future work for the calibration technique and/or transition to a NRA in the 2-3 year timeframe.

3.1 Conclusions

It is too early to tell if the calibration technique has tangible value. The "calibration" work should continue for at least one or two more cycles, and then PS/CSS should re-assess the way forward (e.g., cost vs. benefit). Two potential benefits are:

- Technique to perform independent validation of the output of federal scenario-based risk assessments, and to support refinement and sustainment of a Federal Risk Profile and library of scenarios for analysis of "incidents of national significance"[15];
- Part of a tool set that can be used to validate and compare Regional Risk Profiles and (critical infrastructure) Sector Risk Profiles; and
- Technique to support development of a National Risk Profile, preferably with separate components and display techniques for malicious and non-malicious threats and hazards.

PS/CSS should develop a National Risk Assessment Framework that considers:

- All hazards, and threat/hazard- and risk domain-specific scenario perspectives;
- An systematic approach to independent validation of risk assessments and risk treatment prioritization;
- Extension of the analysis to include cost, benefit and risk analysis of treatment options so that the cost of the preferred treatment strategy is presented to decision makers together with the risk ranking; and
- Convergence of risk and capability assessment methodologies and processes.

There is no common or best practice methodology to do risk assessments for malicious and/or non-malicious risks. Nations are trying to balance mathematical and expert judgment approaches. There are multiple approaches to defining threats and hazards (e.g., no best of breed taxonomy). CA can learn from others and contribute to the evolving body of knowledge, which is generally accepted to be increasing in importance. PS work on calibration has the potential to improve confidence in the output of risk assessment processes. However, the work is in its early stages and a long-term vision and commitment would help to exploit the work to date. It may make sense to focus on calibration of multi-criteria risk assessments of malicious threats and use a simplified two-dimensional assessment for non-malicious hazards (P*I), recognizing that some risk domains substitute frequency for probability. The AS two-step concept may be very suitable for Canada to screen risk scenarios and focus more resources on the more serious risk scenarios.

[15] This term should be defined.

3.2 Recommendations

From the perspective of transition to a strategic national risk assessment, it is recommended that a streamlined, operationally-oriented mapping approach that is supported by a parallel comparative analysis process would be appropriate. This process could use calibration and/or other techniques. Benefits of a parallel process would be to validate federal (and regional) assessments and to support development of a library of anticipated and reference scenarios that consider federal, regional and cross-border perspectives.

In the near-term (Cycle 3), next steps for the calibration technique include:

- Finish calibrating the malicious scenarios for this cycle;
- Present the raw results back to federal departments IRAWG;
- Consolidate results and prepare final report for the 2012-2013 cycle; and
- Present result to ADM-EMC in the fall (2013).

Recommendations for future PS and/or CSS consideration include:

Likelihood Assessment and Calibration

- Keep likelihood calibration as a separate technique until it is more mature. That is, departments should continue to use the qualitative multi-criteria assessment process and then, validate the assessment using expert elicitation and calibration;
- Distinguish among assessments that use frequency, probability and likelihood assessments when presenting results to decision makers;
- Document why probabilistic risk assessments (PRA) are not appropriate for adaptive malicious threats, and identify situations where PRA is relevant for the Canadian context (e.g., engineering and environment domains); and
- Conduct critical analysis of the use of frequency for natural disasters and other non-malicious threats (i.e., past history is not necessarily a reliable predictor of future occurrence or severity).

Impact Assessment and Calibration

- Perform critical analysis of AHRA risk scoring tool and associated mathematics and confidence assessment; and
- Differentiate impact and consequences, and develop taxonomies (AHRA BoK, 2013);

NRA Transition

- Review the process for defining the problem space including constraints on the scenario selection process given that regions have a different decision making and EM planning framework, and different pressures;.
- Review the magnitude threshold identification framework for applicability and portability to a NRA;
- Investigate options for weighting impact categories, preferably with different categories (prototype included in this report);
- Perform critical analysis of AHRA impact categories (refer to prototype above). The AHRA BoK recommends performing a benchmark analysis of existing regional approaches, and review of international experience, to determine the practicality of transitioning to a common tailorable framework; and

- Consider the strategic value of the NRA as a means to (externally):
    - Validate the federal AHRA scenario library, methodology, processes and techniques, and
    - Validate P/T/FNI and regional equivalent programs and risk management activities.

3.3 Methodology Evaluation Framework

Future analysis of the calibration methodology and process should consider using a performance management framework that considers:

- When and how do organizations that directly or indirectly support the mandate become engaged in the scenario development and risk assessment processes?
- What is the optimum way to engage subject matter experts (e.g., productivity, consistency, repeatability)?
- How many low confidence input ratings[16] mean that there is unacceptably low confidence in either the scenario or the risk assessment (area for further targeted study)?
- Does the outcome of the calibration process justify the level of effort?
- Should the calibration, and preferably impact category weighting, be included in the risk scoring tool, or should calibration by "experts" be a parallel activity with a separate report that could constitute an independent validation process with "calibration" as one comparative analysis technique (i.e., identify the trade-offs)?
- What are reasonable approaches for calibration and/or independent validation for a regionally-based NRA methodology?

[16] The Guidelines state that assessing confidence in inputs to the risk scoring tool for different impact categories is **optional**. Using this feature consistently should be an opportunity to "calibrate" the risk assessment and scenario development processes. The confidence ratings do not change the results. They only change the size of the ellipses as an aid to focus management attention. There is no mandatory requirement to include a judgment on confidence as part of the assessment (2012: 24).

References

AHRA Body of Knowledge, draft, CSS, May 2013.

AHRA Methodology Guidelines, 2011-2012 (Chapter 3), PS, 2012

Australian Government, *Critical Infrastructure Resilience Strategy*, 2010.

Emond, Edward J., *Developments in the Analysis of Rankings of Operational Research*, DRDC CORA TR 2006-37; DRDC – Centre for Operational Research & Analysis.

Expert Elicitation Task Force; *Expert Elicitation Task Force White Paper*, Science and Technology Policy Council, US Environmental Protection Agency; August 2011.

Friesen, S.K., *Chemical, Biological, Radiological/Nuclear and Explosive Risk Assessment*, DRDC CSS N 2008-08; DRDC – Centre for Security Science.

Leung, K. and Verga, S., *Expert Judgement in Risk Assessment*, DRDC CORA TM2007-57; December 2007; DRDC – Centre for Operational Research & Analysis.

McFee, Dr. Chris, et al, *Blackett Review of High Impact Low Probability Risks*, UK Government Office for Science, URN 12/519; 2011.

National Emergency Risk Assessment Guidelines, National EM Committee (2010), developed by the Tasmanian State Emergency Service, Emergency Management Australia, 2010.

National Risk Register of Civil Emergencies, 2012 Edition, Cabinet Office, UK, 2010.

Paoli, G., *A Description and Comparison of Options to Calibrate Impact Scoring Scales for AHRA,* Risk Sciences International, presentation to PS, 28 March 2013.

Proposed Calibration of Malicious Likelihood and Non-Malicious Likelihood, Draft Summary, Risk Sciences International, 28 November 2011.

The AHRA Process: Scoring Impact and Likelihood, PS and CSS, 2 February 2012 (PowerPoint).

The Department of Homeland Security's Risk Assessment Methodology: Evolution, Issues and Options for Congress, CRA Report for Congress, Order Code RL33858, 2 February 2007.

Working with scenarios, risk assessment and capabilities in the National Safety and Security Strategy of the Netherlands, Working Group of Academics, October 2009.

Verga, S., *Summary of the "calibration" of malicious likelihood for the All Hazards Risk Assessment scenarios*, draft working paper, CSS, 10 May 2013.

Vlek, C., "*How solid is the Dutch (and the British) National Risk Assessment? Overview and Decision-Theoretic Evaluation"*, Risk Analysis journal, Society for Risk Analysis, DOI: 10.1111/risa.12052; 2013.

List of symbols/abbreviations/acronyms/initialisms

AHRA	All Hazard Risk Assessment (CA framework and methodology)
AS	Australia
BoK	Body of Knowledge (AHRA information baseline, draft, CSS, 2013)
CA	Canada
CSIS	Canadian Security Intelligence Service
DNRA	Dutch National Risk Assessment
GC	Government of Canada
GIS	Geographic Information System
HAZUS-MH	Hazards United States – Multi-Hazard (FEMA program; NRCan & CSS involved and adapting to Canadian environment)
HC	Health Canada
IRAWG	Interdepartmental Risk Assessment Working Group (federal AHRA community)
NL	Netherlands
NRA	National Risk Assessment (UK, CA)
NRCan	Natural Resources Canada
NRR	National Risk Register (UK)
NSRA	National Security Risk Assessment (UK)
NZ	New Zealand
PRA	Probabilistic Risk Analysis
PS	Public Safety Canada
P/T/FNI	Provinces/Territories/First Nations & Inuit (unofficial term used on AHRA BoK)
S&I	Security and Intelligence
S&T	Science and Technology
SME	Subject Matter Expert
SNRA	Strategic National Risk Assessment (US, DHS)
TC	Transport Canada
TM	Technical Memorandum
UK	United Kingdom
US	United States

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)

Ian Bayne
Maxsys INC.

2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.)

UNCLASSIFIED
(NON-CONTROLLED GOODS)
DMC A
REVIEW: GCEC APRIL 2011

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

All Hazards Risk Assessment: Documenting the Calibration Process and Methodology:

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Bayne, I.; Duncan, J.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)	6b. NO. OF REFS (Total cited in document.)
September 2013	36	16

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Contract Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Centre for Security Science
Defence R&D Canada
222 Nepean St. 11th Floor
Ottawa, ON Canada K1A 0K2

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
CSSP-2012-TI-1108	W7714-11-5085 TASK-7

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
	DRDC CSS CR 2013-016

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

The All Hazards Risk Assessment (AHRA) Interdepartmental Risk Assessment Working Group (IRAWG) is using a scenario-based risk assessment to support federal emergency planning and policy development, and departmental efforts to prioritize risk within their respective mandates. In 2010, Public Safety Canada (PS) stated the requirement to create a common picture of all-hazards risks. They requested that results for the assessment of risks assessments derived from malicious (e.g., terrorist act; cyber attack; organized crime) and non-malicious (e.g., natural disasters; industrial accidents) threats and hazards be presented on a common two-dimensional graph.
In support of this objective, the Centre for Security Science (CSS) is evaluating techniques to address these requirements including calibration of likelihood and impact estimates derived from different methods. CSS, working together with security, intelligence and risk domain experts, are evaluating approaches to calibrate malicious and non-malicious risk assessments. This Contractor Report (CR) documents the likelihood calibration technique and the discusses the overall risk analysis process with a view to identifying implications for future work and to transition to a streamlined national risk assessment process. The main recommendation is that PS/CSS continue to collect data on the challenges associated with implementing the current process and methodology for one or two more cycles, as a basis upon which to assess the way forward, including the cost, benefit and risks.

Le Groupe de travail interministériel (GTI) sur le cadre d'évaluation tous risques (ETR) évalue les risques au moyen de scénarios à l'appui de la planification des mesures d'urgence, de l'élaboration des politiques et des efforts ministériels pour établir l'ordre de priorité des risques conformément à leurs mandats respectifs. En 2010, Sécurité publique Canada (SP) a signalé qu'il fallait dresser un portrait commun de tous les risques. Il a été demandé que les résultats d'évaluation des risques découlant de menaces malveillantes (p. ex., acte terroriste, cyberattaque, crime organisé) et non malveillantes (p. ex., catastrophe naturelle, accident industriel) soient présentés avec un graphique bidimensionnel commun.
Pour atteindre cet objectif, le Centre des sciences pour la sécurité (CSS) considère les techniques pour répondre à ces exigences, y compris l'étalonnage des probabilités et les estimations de l'incidence obtenues de divers moyens. Le personnel du CSS collabore avec des experts dans les domaines de la sécurité, du renseignement et du risque afin d'examiner les approches permettant d'uniformiser les évaluations de risques malveillants et non malveillants. Le présent rapport d'entrepreneur porte sur la technique d'étalonnage des probabilités et le processus global d'analyse des risques afin de déterminer l'incidence sur les travaux futurs et d'adopter un processus national d'évaluation des risques simplifié. Selon la recommandation principale, SP et le CSS devraient continuer de recueillir des données sur les difficultés liées à la mise en œuvre du processus et de la méthodologie actuels pour un ou deux cycles supplémentaires dans le but d'évaluer les prochaines étapes, y compris les coûts, les avantages et les risques.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

All Hazards; Risk Assessment; Emergency Management