2017

Machine safety: Hands-on experimentation with risk estimation parameters and tools

François Gauthier
Université du Québec à Trois-Rivières

Yuvin Chinniah
Polytechnique Montréal

Damien Burlet-Vienney
IRSST, damien.burlet-vienney@irsst.qc.ca

Barthélemy Aucourt
Polytechnique Montréal

Stéphane Larouche
Université du Québec à Trois-Rivières

Institut de recherche
Robert-Sauvé en santé
et en sécurité du travail

Machine Safety

Hands-On Experimentation with Risk Estimation Parameters and Tools

François Gauthier
Yuvin Chinniah
Damien Burlet-Vienney
Barthélemy Aucourt
Stéphane Larouche

STUDIES AND RESEARCH PROJECTS

R-980

MECHANICAL AND PHYSICAL RISK PREVENTION

OUR RESEARCH
is working for you !

The Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST), established in Québec since 1980, is a scientific research organization well-known for the quality of its work and the expertise of its personnel.

Mission

To contribute, through research, to the prevention of industrial accidents and occupational diseases and to the rehabilitation of affected workers;

To disseminate knowledge and serve as a scientific reference centre and expert;

To provide the laboratory services and expertise required to support the public occupational health and safety network.

Funded by the Commission des normes, de l'équité, de la santé et de la sécurité du travail, the IRSST has a board of directors made up of an equal number of employer and worker representatives.

To find out more

Visit our Web site for complete up-to-date information about the IRSST. All our publications can be downloaded at no charge.
www.irsst.qc.ca

To obtain the latest information on the research carried out or funded by the IRSST, subscribe to our publications:

- *Prévention au travail,* the free magazine published jointly by the IRSST and the CNESST (preventionautravail.com)
- InfoIRSST, the Institute's electronic newsletter

Legal Deposit
Bibliothèque et Archives nationales du Québec
2017
ISBN : 978-2-89631-956-5
ISSN : 0820-8395

IRSST – Communications and Knowledge Transfer Division
505 De Maisonneuve Blvd. West
Montréal, Québec
H3A 3C2
Phone: 514 288-1551
publications@irsst.qc.ca
www.irsst.qc.ca

Machine Safety

Hands-On Experimentation with Risk Estimation Parameters and Tools

François Gauthier[1], Yuvin Chinniah[2], Damien Burlet-Vienney[3], Barthélemy Aucourt[2], Stéphane Larouche[1]

[1] Université du Québec à Trois-Rivières (UQTR)
[2] Polytechnique Montréal
[3] IRSST

STUDIES AND RESEARCH PROJECTS

R-980

This study was funded by the IRSST. The conclusions and recommendations are solely those of the authors. This publication is a translation of the French original; only the original version (R-940) is authoritative.

PEER REVIEW
In compliance with IRSST policy, the research results published in this document have been peer-reviewed.

ACKNOWLEDGMENTS

The authors would like to thank the following individuals and organizations for their support and valuable involvement in this study:

- The members of the follow-up committee: Johanne Doyon (Hydro-Québec), Michel Gingras (MultiPrévention), Wagui Geadah (Association sectorielle Fabrication d’équipement de transport et de machines [ASFETM]), Sèdoté Ghislain Hounkpe (Commission des normes, de l’équité, de la santé et de la sécurité du travail [CNESST]), Jacques Laroche (Associations de la santé et de la sécurité des pâtes et papiers et des industries de la forêt du Québec [ASSPPQ/ASSIFQ]), Marie Ménard (MultiPrévention) and Tony Venditi (Association sectorielle Fabrication d’équipement de transport et de machines [ASFETM]).

- The 25 subjects and their respective employers who generously gave their time for this study.

- Nicola Stacey, Nicola Healy and Colleen Butler of the Health and Safety Laboratory (Buxton, United Kingdom) for their expertise and active participation in preparing and carrying out this study.

- Joseph-Jean Paques for his sound advice.

ABSTRACT

In a risk assessment process as defined in international standard ISO 12100:2010, risk estimation is an essential step in which machinery designers and users determine risk levels and identify the most critical hazardous situations. Two previous studies funded by the Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST) revealed that the many risk estimation tools available take a wide variety of forms and that a number of their characteristics (e.g., parameters, architecture) can have a significant influence on the estimated level of risk. In this third instalment of the IRSST's research program on the assessment of machine risks, the impact of these characteristics was evaluated and a number of risk estimation tool construction rules were validated by means of an experimental study involving a variety of users, chiefly from industry.

Six tools were analysed: 4 that were selected from a list of 31 drawn from earlier studies, as well as tools provided by the IRSST and the Commission des normes, de l'équité, de la santé et de la sécurité du travail (CNESST), and by the Health and Safety Laboratory (HSL) in the United Kingdom. Some risk estimation parameters having flaws or potential biases were also selected. The risk estimation tools and parameters were applied to concrete scenarios of machinery-related hazardous situations by a representative sample of industry stakeholders, occupational health and safety officers, joint sector-based association advisors and trainers from various sectors. They were all asked to give their impressions and preferences with regard to the different risk estimation tools, along with their reasons.

With respect to the study of risk estimation parameters, five of the potential flaws established in the preceding instalment of this research program were analysed. The experimental results show that when subjects have trouble applying a parameter, they are generally able to associate it with a flaw in the parameter. The results also indicate fairly clearly that the impact of parameter construction flaws is not uniform. The type of flaw, its position on the parameter scale and the scenario in question influence its impact on the determination of the level of a parameter. The severity of harm parameters seem fairly robust and allow users to reach a solid consensus, despite the presence of flaws. On the other hand, the probability of occurrence of harm and the probability of occurrence of the hazardous event parameters are decidedly less robust. Besides the flaws in these parameters, the results obtained suggest that evaluating likelihood or probability is a difficult aspect of risk estimation that requires special attention.

For the study of the risk estimation tools, three criteria were used to determine potential application problems: (1) the ability to distinguish between scenarios with different risk levels, (2) the user's satisfaction with the result obtained and (3) the convergence of results (intersubject repeatability). By determining the origin of the problems found with the six risk estimation tools, the effect of the failure to follow certain construction rules for tool architecture could be confirmed. An architecture that gives more influence to a specific parameter (e.g., the first parameter on a graph) may amplify a divergence in results and reduce the ability to classify scenarios appropriately, especially if there are flaws in the most influential parameters. A matrix sensitive to the slightest change in the level of a parameter will have the same impact on one of its parameters when there is a flaw. An architecture that did not produce a uniform distribution of risk levels led to unsatisfactory results and to problems of scenario discrimination. A structure

that did not comply with ISO 12100:2010 did not reveal any impact on the risk estimation process.

It is hoped that these findings will help improve the robustness and reliability of existing tools and provide support for the risk assessment training currently provided by partners.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Background

Since 2005, the CNESST [Quebec's occupational health and safety board] has been implementing its Machine Safety Action Plan, which targets the hazards associated with moving parts on machines. The CNESST's various initiatives are based on a zero-tolerance policy when moving parts are in an area accessible to workers. Yet without a structured risk assessment method, it is hard to make the right decisions about the best ways to reduce machine-related risks (Lyon and Hollcroft, 2012; Hughes and Ferrett, 2005; Main, 2012; Pickering and Cowley, 2010). Risk assessment refers to a series of steps taken to examine the hazards associated with machines. It can be divided into two phases: (i) risk analysis and (ii) risk evaluation, as specified in international standard ISO 12100:2010. *Safety of Machinery – General Principles for Design – Risk Assessment and Risk Reduction*.

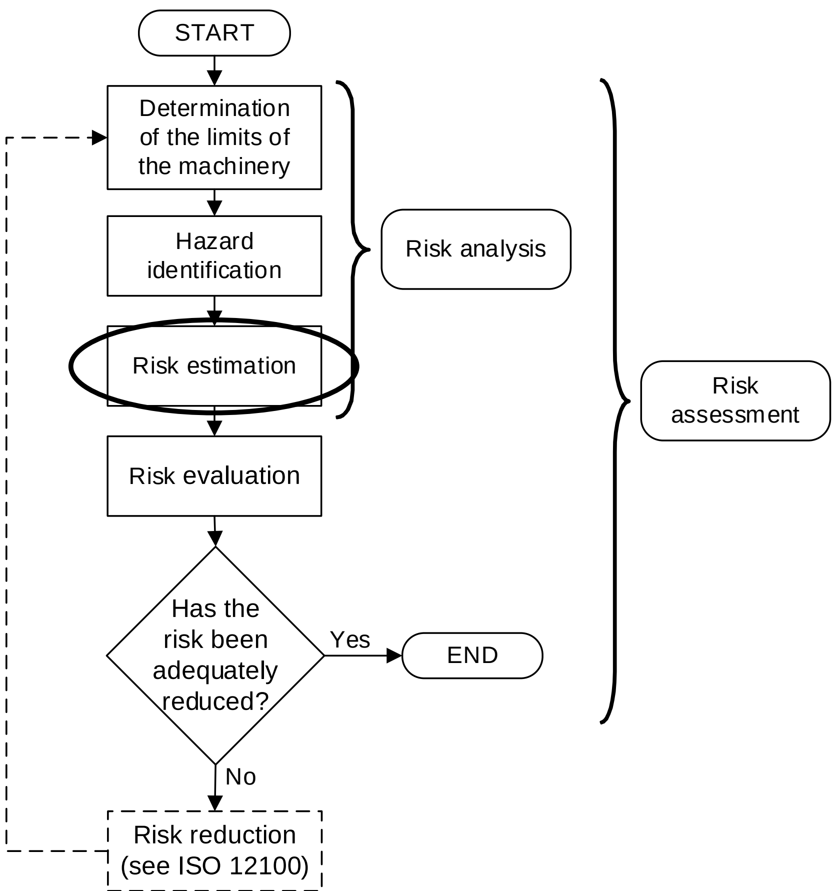

Figure 1. Risk assessment process based on standard ISO 12100:2010

Figure 1, taken from standard ISO 12100:2010, shows where risk estimation fits into the overall risk assessment process. This study focused on the risk estimation stage, which consists in estimating the level of risk inherent in each hazardous situation associated with using a machine. The figure shows that risk estimation is the last step in risk analysis, which leads to risk evaluation and risk reduction. Risk estimation is a critical step in prioritizing the risk reduction activities that follow a risk assessment process. Poor risk estimation may lead to the implementation of inappropriate or inadequate risk reduction measures on a machine.

1.2 Review of the Literature

According to standard ISO 12100 (2010), "the risk associated with a particular hazardous situation depends on the following elements: a) the severity of harm; b) the probability of occurrence of that harm, which is a function of 1) the exposure of person(s) to the hazard, 2) the occurrence of a hazardous event, and 3) the technical and human possibilities to avoid or limit the harm. Many tools are suggested in the literature for estimating risks based on these "elements" defined in the standard (Paques et al., 2005b). The tools are proposed both by accident prevention organizations and by companies. They are sometimes part of detailed training programs, while at other times they are presented alone in a separate publication describing how they are to be used. In these tools, the "elements," which are referred to as "risk estimation parameters" in this report, are combined or organized in a specific architecture so as to make the risk estimation process more systematic. An example of such a tool, with a two-parameter risk matrix architecture, is shown in the figure below.

Possibility of occurrence of harm	Severity of harm		
	Minor	**Moderate**	**Serious**
Unlikely	Low	Acceptable	Moderate
Possible	Acceptable	Moderate	Unacceptable
Likely	Moderate	Unacceptable	Unacceptable

Figure 2. Example of risk estimation tool

Etherthon (2007) confirmed that risk assessment results are obtained by collecting and analysing qualitative information on the severity of harm and health effects, and on the probability of occurrence of the events that could cause this harm. However, some experts in machine-related risk estimation have noted that the methods used in various European countries to assess machine risks, when such methods are available, may give different, if not contradictory, results. In some cases, they may even recommend different safety levels for the same machine (Charpentier, 2003). A certain variability in risk estimation results may be regarded as "natural," and therefore acceptable, but excessive dispersion may lead to inappropriate (inadequate or excessive) risk reduction measures being taken (Parry, 1999). Abrahamsson (2000) emphasizes the fact that some potential users of risk estimation tools perceive them as being unconvincing or frankly unusable. He has sought to validate various risk estimation tools, especially in the context of occupational exposure to chemicals (Abrahamsson, 2002). His research focused exclusively on the analysis of tool-related variables (e.g., model, parameters), without examining the other variables that might affect risk estimation (e.g., prior training, characteristics of people carrying out the risk analysis). He concluded that uncertainty is inherent in risk estimation, but that guidelines specific to different industries might help improve the estimating process.

Applying common risk estimation tools to machine safety requires interpreting often qualitative information, generally using an ordinal scale, as described by Stevens (1946). However, many risk estimation tools are not very precise or detailed (Chinniah et al., 2011). For instance, a qualitative verbal scale of the type *Highly unlikely*, *Unlikely* and *Likely* is used in some tools to determine the probability of occurrence of harm. Yet the degrees of the scale are not defined. It is therefore hard to determine the exact meaning of *Unlikely*, for instance. This type of design can create biases in the estimating process and significantly influence the final result (Duijm, 2015; Carey and Burgman, 2008; Christensen et al., 2003; Cox, 2008; Patt and Schrag, 2003). Beyth-Marom (1982) reported the results of an experiment conducted at a professional forecasting organization with subjects accustomed to giving verbal probability assessments ("probable," "possible," etc.). Her study highlighted the communication problems caused by verbal expressions of probability. It also revealed a high variability in the interpretation of such expressions, as Hubbard and Evans (2010) confirmed.

Nevertheless, and despite the problems inherent in qualitative ordinal scales described in the literature (Franceschini et al., 2004; Hubbard and Evans, 2010; Smith et al., 2009; Woodruff, 2005), the heavy use of these scales for risk estimation in areas where quantitative data are not easily available, as is the case in machine safety, cannot be ignored. There are also many advantages to using qualitative ordinal scales in risk estimation tools, such as providing a simple, effective approach and a clear framework for the systematic examination of hazardous situations (Ni et al., 2010). As a result, and as several authors have suggested, further research is required to "characterize in greater detail the conditions under which they are most likely to be helpful or harmful to decision making in risk management" (Cox, 2008; Lamy and Charpentier, 2008; Aven, 2012).

Given (a) the increasing use of qualitative risk estimation tools in machine safety, (b) the wide diversity of these tools and (c) the significant differences in the results they yield, the IRSST launched a thematic research program in 2004 with the goal of conducting an in-depth analysis of the tools proposed in the literature or used in industry (Paques and Gauthier, 2006). Two already completed studies in the program have shown that the many tools used in the risk estimation phase take a wide variety of forms and that many of their characteristics may significantly influence the risk level arrived at with them (Paques et al., 2005b; Paques and Gauthier, 2007; Chinniah et al., 2011). These studies have shown that the risk estimation results provided by the tools entail significant differences for the same hazardous situation. The area of application of the tool, its configuration and the details of its parameters seem to be the main factors contributing to this variability of results. Flaws or biases in tool construction likely to influence results under certain circumstances have been identified (Gauthier et al., 2010). These flaws or biases may be defined as characteristics specific to certain tools that, in previous studies, have shown their potential for generating incorrect estimates in certain cases.

In IRSST report R-684,[1] Chinniah et al. (2011) have suggested that because of these construction flaws, risk estimation in which different users apply different tools could lead to a wide dispersion of risk estimation results. A series of construction rules aimed at preventing flaws in risk estimation tools has also been proposed (Chinniah et al., 2011; Gauthier et al., 2012). The proposed construction rules are presented in Table 1.

[1] https://www.irsst.qc.ca/media/documents/PubIRSST/R-684.pdf

Table 1. Risk estimation tool construction rules and flaws (Gauthier et al., 2012)

<table>
<tr><th>Rule regarding</th><th>Application</th><th>Description of flaw</th><th>Construction rule</th></tr>
<tr><td rowspan="7">Flaws in risk estimation parameters</td><td rowspan="2">For parameters used</td><td>Poor definition of parameter</td><td>Define each parameter in explicit terms (e.g., probability of occurrence of hazardous event in relation to probability of occurrence of harm)</td></tr>
<tr><td>No definition of exposure interval</td><td>Define probability and exposure parameters in relation to exposure interval</td></tr>
<tr><td rowspan="5">For parameter levels</td><td>Poor definition of levels</td><td>Avoid using unique or vague terms to define parameter levels</td></tr>
<tr><td>Inconsistent definitions of different levels</td><td>Avoid using the same term or expression when describing two different levels of a parameter</td></tr>
<tr><td rowspan="2">Inadequate number of levels</td><td>Use from three to five levels for the severity-of-harm parameter</td></tr>
<tr><td>Use from three to five levels for the probability-of-occurrence-of-harm parameter</td></tr>
<tr><td>Gaps between levels</td><td>No discontinuity or gaps between parameter levels</td></tr>
<tr><td rowspan="7">Risk estimation tool architecture</td><td rowspan="3">For tool configuration</td><td>Inappropriate tool family</td><td>Risk graph or matrix</td></tr>
<tr><td>Non-standard configuration</td><td>Complies with ISO 12100:2010 (2 or 4 parameters)</td></tr>
<tr><td>Not calibrated for machine risks</td><td>Adapted for machine risk estimation</td></tr>
<tr><td rowspan="4">For the result provided by the tool (risk level or index obtained)</td><td>Insufficient number of risk levels</td><td>Define at least four risk levels</td></tr>
<tr><td>Non-uniform distribution</td><td>Uniform distribution of risk levels in matrix</td></tr>
<tr><td>Discontinuity in risk levels</td><td>No more than one difference in risk level between two adjacent cells</td></tr>
<tr><td>Excessive relative weight given to one parameter</td><td>Balance the relative influence of each parameter with the level of risk</td></tr>
</table>

These rules may potentially solve some of the significant variability problems noted in risk estimation results. They may also help users choose or improve existing risk estimation tools. However, further research was needed to confirm the impact of the tool flaws and to validate the proposed construction rules. This was the goal of the third instalment of the thematic research program on machine risk estimation, which took the form of an experimental study involving users from industry.

1.3 Research Objectives

Basically, risk estimation tools are used to identify qualitative distinctions between critical risks and those that are less critical. To provide satisfactory results, the tools must be designed so that different users will arrive at more or less the same risk level when applying the tool to a specific scenario. Likewise, the risk levels obtained by different tools for a given scenario should also have certain similarities. Last, when different tools are applied to different scenarios, they should arrive at results that classify the risks in the same order.

The assumption underlying this study was that excessive variability in risk estimations is due to construction flaws or biases in the tools. In other words, flaws or biases in the parameters and tools increase the variability of results, either the selected parameter level or the estimated level of risk, among users. The purpose of this study was therefore to gain a better understanding of the machine safety risk estimation process and determine the tool variables that can influence accurate estimation by the tools. Past experimentation suggests that well-designed tools and carefully defined risk estimation parameters can reduce the impact of the biases inherent in qualitative ordinal scales.

The objectives of this third IRSST thematic research program study on machine risk estimation were therefore to

1. Confirm the impacts of, and reasons for, the risk estimation parameter flaws established in the previous study (report R-684)

2. Confirm the impacts of, and reasons for, the failure to follow the construction rules established in the previous study (report R-684) regarding risk estimation tool architecture

Given that risk estimation tools consist of specific parameters and that the result of applying them (i.e., the risk level or index obtained) depends on how these parameters are defined and structured, this study focused on (a) risk estimation parameters examined separately and (b) complete risk estimation tools. It was expected that the flaws in the parameters and the failure to comply with construction rules regarding tool architecture would increase the dispersion of the experimental results, either the level of the selected parameter or the level of risk obtained.

This study was carried out by a team of researchers specialized in the machine risk assessment process from the Université du Québec à Trois-Rivières (UQTR), Polytechnique Montréal (PM), the IRSST and the Health and Safety Laboratory (HSL) in the United Kingdom. The two teams—the UQTR-PM-IRSST team in Quebec and the HSL team in the U.K.—decided to work together in order to increase their research potential by sharing expertise in the defining of detailed methodology, including the preparation of data-gathering instruments, and in the comparative analysis of the research results. However, the data gathering itself, which required the participation of users in industry, was conducted independently by the two teams.

The study's ultimate goal was to define clear criteria that can be used in industry to evaluate, select or develop machine safety risk estimation tools.

2. METHOD

To set up an experimental design that met the study objectives, the following methodological features were determined jointly by the UQTR-PM-IRSST and HSL teams:

- Scenarios on which the risk estimation parameters and tools would be tested
- Parameters and tools to be tested and the rationale for the choices
- Criteria used to select subjects
- Experimental protocol that determined how the parameters taken separately (first set of experiments) and the risk estimation tools (second set of experiments) would be applied by the subjects
- Questions concerning the estimation work in order to obtain subjects' views on the choices they made and the results obtained
- Results analysis methods, whether for parameters or tools

These features are dealt with below in three subsections: (i) the preparatory phase with the choice of scenarios, parameters, tools and subjects, (ii) the experimental protocol and (iii) the analysis of the experimental results.

2.1 Preparatory Phase

A number of methodological features used in the preparatory phase were based on IRSST report R-684, which was the Institute's earlier theoretical study on risk estimation tools (Chinniah et al., 2011).

2.1.1 Selection and Preparation of Hazardous Situation Scenarios

Four industrial-machine-related scenarios were selected from among the 20 developed and validated in the earlier theoretical study (Chinniah et al., 2011). In order to limit experimental bias, it was agreed that the selected scenarios had to meet the following conditions:

- Illustrate a mechanical hazard and be related to worker safety (and not worker health): accident processes leading to injury are generally speaking easier to interpret than those leading to occupational diseases, chiefly because of the problems involved in determining the exposure period required to cause disease
- Be fairly easy to understand and provide a sufficient level of detail: simple, non-specialized situations that can be evaluated by people who are familiar with machine safety in general

- Describe a work activity related to production rather than a less frequent activity such as maintenance: production activities involve more frequent operation of a machine, which makes estimating exposure easier

- Illustrate four distinct risk levels (i.e., low, mid-low, mid-high, high) and show a low standard deviation in their theoretical application based on the results of the earlier study: the risk levels of each situation are therefore more easily discernable

The selected scenarios, identified as A, G, M and S in the earlier study, are outlined in

Table 2. The level of risk associated with each scenario corresponded to the mean of the risk levels obtained with the 31 risk estimation tools of the earlier study (Chinniah et al., 2011).

Table 2. Description of the four scenarios selected for practical experimentation

Scenario	Title	Summary of work activity	Associated level of risk
A	Punching machine with mobile table	Functional demonstration of a punching machine at a trade fair	Low (47.7%)
G	Automated guided vehicle (AGV)	Movement of an automated guided vehicle (no operator) through a plant, following a yellow line painted on the ground	Mid–Low (61.9%)
M	Rewinder (papermaking machine)	Finishing tasks on a rewinder (removing the irregular parts of the roll) while the rewinder is switched on in manual mode	Mid–High (74.8%)
S	Robot	Worker changes a tool on a numerically controlled lathe. Presence of a robot that supplies the lathe with metal pieces	High (85%)

Appendix A sets out the scenarios as they were presented to the subjects for the experiments. Each scenario provided a picture of the machine and a description of the activity, the hazard, the hazardous situation, the hazardous event, its probability of occurrence, possible harm, exposure data and possibility of avoidance. This breakdown of the information describing a scenario was based on the elements that make up the accident process and risk, according to standard ISO 12100:2010.

2.1.2 Selection and Preparation of Risk Estimation Parameters

The six types of risk estimation parameters tested in the first set of experiments were those recommended in machine safety by standard ISO 12100:2010, in particular:

- Severity of harm (S)
- Probability of occurrence of harm (Ph), a parameter that itself consists of three parameters:
  - Exposure frequency (Exf) and/or exposure duration (Exd)
  - Probability of occurrence of hazardous event (Pe)
  - Possibility of avoidance (A)

In the earlier theoretical study, six flaws associated with parameters were characterized (Table 1). The aim of the hands-on experiments conducted as part of this study was to test the impact of these flaws (except the flaw "poor definition of parameter," which was not evaluated) when they affect the different types of parameters. However, not all the test combinations were selected for the experiments (i) because of the difficulty of finding all the flaws on all types of parameters with the available bank of risk estimation tools (e.g., inadequate number of levels for Exd, Exf and A) and (ii) because of the fact that in the earlier study some flaws seemed to have had more impact on certain types of parameters (e.g., inadequate number of levels for S).

In the end, 20 parameters representing all types of risk estimation parameters were selected by the research teams for the experiments. The various parameters and their suspected associated flaws are described in Appendix B. To avoid possible bias stemming from subjects' ability to understand English, which would be hard to measure, the parameters were translated and presented in French.

2.1.3 Selection and Preparation of Risk Estimation Tools

Six tools were chosen for the second set of experiments. This number was a compromise to allow the time required for conducting the experiments. The following tool selection criteria were applied, in this order:

- Consider only risk estimation tools that (i) are currently used, (ii) are easy to use by following the written instructions available and (iii) were initially developed for machine safety
- Then consider tools that best meet the construction rules of the earlier study with respect to their architecture (Table 1). While the objective was to choose the best tools possible, they could still have certain construction flaws
- Choose tools with diverse estimating tendencies, according to the results of the earlier study

Based on these criteria, four risk estimation tools were first selected by research team consensus from the 31 tools identified in the earlier theoretical study (tools 19, 24, 69 and 89). Two further tools, numbered 91 and 114 in the earlier study, were also chosen because they were developed by the two main organizations taking part in the study, i.e., the IRSST (tool 91) and the HSL (tool 114). These tools do not follow all the construction rules for risk estimation tool architecture (Chinniah et al., 2011). The shortcomings in relation to the construction rules are as follows:

- For tool 91:
  - Non-uniform distribution (the possibility of arriving at risk levels 1 or 2 is greater)
  - Discontinuity in risk levels
  - Excessive relative weight given to one parameter (S)
- For tool 114:
  - Non-standard configuration (no parameter Pe)
  - Non-uniform distribution (the probability of arriving at each of the risk levels is not equal)

The characteristics of the selected tools are summarized in Table 3. To avoid possible bias stemming from the subjects’ ability to understand English, which would be hard to measure, the tools were translated and presented in French. Furthermore, the subjects did not have access to the tools in their original form (Appendix C) during the experiments. Only a standardized version of the tool parameters (i.e., without the risk level calculation process) was presented to them so that they would not be able to anticipate the resulting risk level (Appendix D).

Table 3. Risk estimation tools selected for experimentation

Tool	Number of levels							Risk level according to earlier study (%)						Reference
	Parameter						R	Tendency	Mean	Scenario				
	S	Ph	Exf	Exd	Pe	A				A	G	M	S	
19	3		2		3	2	4	Low	50	25	25	50	75	Ekelenburg et al. (1996)
24	4	4					4	High	85	75	100	100	100	ANSI B11.TR3 (2000)
69	4			2	3	2	11	Medium	64.1	27.3	45.5	63.6	81.8	Görnemann (2003)
89	3	4					6	Medium	63.1	50	66.7	66.7	66.7	MMMPIC (2002)
91	2		2		3	2	6	Low	50.8	33.3	66.7	33.3	33.3	ISO 14121-2:2007
114	4	4	4			4	4	High	78.4	66.7	66.7	33.3	100	HSL (2012)

R: risk level

2.1.4 Selection of Subjects for Experiments

To guarantee that homogeneous, reliable data would be obtained, the subjects taking part in the study had to have sufficient knowledge and experience about (i) the risk assessment process and (ii) the use of risk estimation tools. To be chosen to take part, candidates therefore had to meet the following criteria:

- Known to have a certain degree of experience in machine safety
- Assess themselves as having at least an "average" level of experience in risk analysis applied to machines (on a scale: low, average, good, very good)
- At a minimum over the last five years, they have carried out:
  - One risk analysis applied to machines if they have received official training in this area, or
  - Five risk analyses applied to machines if they have not received any specific training

This information on candidates was checked by means of a questionnaire during the recruitment process. In addition, to validate subjects' risk perception, they were asked to estimate, intuitively, the risk of each scenario using an analog scale (i.e., a 10 cm line). Note that they were not allowed to refer to this scale, presented on a separate sheet, later on. The results obtained were matched up with the associated level of risk presented in
Table **2**. The averages for each scenario were A: 36.4%; G: 35.6%; M: 66.4%; and S: 73.8%.

In accordance with the protocol, a sample of 25 participants was organized. This sample group of volunteers was fairly equally made up of (i) sector-based (joint or employer) OHS association advisers, (ii) maintenance staff or company safety officers and (iii) engineers specialized in machine safety. All the subjects signed a consent form, and the confidentiality of the information collected was guaranteed in accordance with the rules established by the Ethics Committee on Research Involving Human Subjects of the Université du Québec à Trois-Rivières (UQTR) (certificate number CER-12-180-06.10).

2.2 Experimental Protocol

To ensure consistency in the experiments, an experimental protocol was developed and then tested under real conditions by two members of the research team and two potential subjects.

The experiments were conducted at each subject's workplace. When their appointment was confirmed, subjects were asked to make available a quiet place along with a computer having an Internet connection for accessing the online questionnaires. The average duration of the experiments was 5 hr 30 min, not including breaks. The subjects were accompanied at all times by a researcher. The researcher took care of the computer input so that the subjects could concentrate on answering the questions. However, since many of the questions were fairly repetitive, the subjects were allowed to operate the computer during the experiment if they said they preferred to.

The researcher in charge of the experiments used two worksheets (one for each set of questions) developed in Excel, where a hyperlink was available for each experimental case (i.e., tool X/scenario Y; parameter X/scenario Y). Each hyperlink referred to the relevant questionnaire. All the questionnaires were developed using the online questionnaire software available on UQTR's secure intranet. This software was chosen because it offered all the required functionality, including data backup, security and extraction.

The two sets of experiments were done one after the other. However, to limit bias related to cognitive construction of the sequence of estimates by the subjects, the questionnaires were presented in a random order from one subject to the next: the Excel worksheets were mixed before the start of each of the two sets. A post-experiment questionnaire was also used to determine whether the subjects were familiar with one of the scenarios or tools, and to get their feedback on the experiments.

The detailed experimental protocol is presented in Appendix E. Regarding the tool experiments, it should be noted that the subjects did not have direct access during the process to the results associated with their choices. Only once all the tool parameter levels had been chosen could the researcher determine the resulting level of risk and present it to the subject. It was therefore impossible for subjects to "cook" their estimates.

All the questionnaires were validated by two members of the research team who had not been involved in developing them, with a view to ensuring the functionality of the system and the accuracy of the tool headings.

2.3 Analysis of Results

2.3.1 Analysis of Impact of Risk Estimation Parameter Flaws

The impact of the various flaws in risk estimation parameters was analysed on the basis of the results collected during the experimental phase, that is, for the application, by the subjects, of each parameter to each scenario:

- Level chosen by the subject
- Level of difficulty, on a scale of 1 to 5 (very easy to very hard), indicated by the subject
- Comments made by the subject

The levels chosen by the 25 subjects were broken down by percentage according to the different levels of each parameter. The mode was then determined for each case. The mode is defined here as the level of a given parameter chosen by the greatest number of subjects for a given scenario. The modal percentage is therefore the proportion of the 25 subjects who indicated the same level for a given parameter, applied to a given scenario. This percentage provided an indication of convergence (or divergence), i.e., the repeatability of the subjects' responses as a function of each flaw and each parameter (intersubject repeatability). Thus, the higher the modal percentage, the better the convergence of the results of applying a parameter. However, as the modal percentage is affected by the number of levels of each parameter, only values less than or equal to 60% were considered to be an indication of less good convergence of results.

Since the accuracy of the responses obtained could not be established with the modal percentage, the results were also compared with the established reference level of each parameter for each scenario. A reference level had been established in the earlier study (Chinniah et al., 2011) through a consensus of six experts (Delphi method), for the same scenarios and the same risk estimation parameters studied here. The comparison evaluated the performance of each parameter, as well as the potential impact of each flaw, by determining the percentage of subjects who chose the reference level.

Besides the convergence of the results and degree to which they matched the reference levels, an analysis of the ease of applying each parameter was also conducted. During the experiments, the subjects had to indicate, on a scale of 1 to 5, the level of difficulty of applying each parameter to each scenario according to the following statement: "The descriptions, the definitions or the number of levels of this parameter made my choice…" (Appendix E, Figure 18, question 3.2). The number of subjects who indicated a difficulty level of 4 (Fairly hard) or 5 (Very hard) was added up to arrive at a level-of-difficulty indicator.

The comments made by the subjects at each opportunity were counted up and classified in three categories:

1. General comments: Neutral or positive comments indicating, for example, the ease of applying a parameter in a given case or noting a specific characteristic of a parameter. For instance:

   – *No overlap, simple, clear description*

   – *Pretty clear. The breakdown made the choice easy*

   – *Plain and simple in this case: "On/Off"*

   – *The addition of the word "usually" makes the choice clearer*

   – *Enough levels. Definition is clear enough – reflects the case accurately*

2. Negative comments on the flaw in question: Comments made by subjects directly pointing to the flaw in the parameter. The number of negative comments from subjects about each flaw was used as an indicator of the perception and impact of the flaws studied. The nature of the comments also helped with understanding the impact of the various flaws on the risk estimation process. To associate the comments with the different flaws, linguistic tables were developed for each flaw. The tables provide keywords or key expressions that indicate links between comments and flaws. The research team compared each comment with the list of keywords. For instance, for the flaw "Gaps between levels," the following comments were noted, and the underlined words or expressions allowed the comment to be linked with the flaw:

   – *Contradictory choices? There's a <u>gap</u> between less than once a day and more than once an hour*

   – *<u>Interval missing</u>: twice per shift is not considered; it falls <u>between the two</u> parameters*

   – *There aren't enough levels to describe my choice: (lies <u>between the two</u>)*

   – *So what about less than once a day? Where does it fit? <u>Intermediate level missing</u>*

   – *Doesn't work. There's a <u>gap</u> between E1 and E2*

   – *There's a gap that's not covered, <u>between the two levels</u>*

3. Negative comments not related to a specific flaw: Comments indicating a problem or noting a negative aspect of the application of a parameter, with no connection to the flaw in the parameter. These comments were analysed in order to identify other issues (with parameter design) that the subjects may have observed.

On the basis of the collected data, the impact of the various flaws in the risk estimation parameters was analysed globally, and separately, for each type of parameter. The data were also cross-analysed.

2.3.2 Analysis of Impact of Risk Estimation Tool Flaws

Three criteria were used to analyse the data for each tool tested:

- **Ability to distinguish between scenarios** with different risk levels
- **User satisfaction** with result obtained
- **Convergence** of results (intersubject repeatability)

These criteria deal with three different aspects of the results. A tool that has good convergence, but cannot distinguish between scenarios with different risk levels will be useless. Similarly, a tool may have good convergence and may be able to discriminate effectively between scenarios, but if the users are not happy with it, they won't use it. A tool's ability to distinguish between scenarios is a reflection of its overall performance.

2.3.2.1 Ability to Distinguish Between Scenarios with Different Risk Levels

The comparison of the risk levels obtained by subject for the four scenarios served to determine whether each tool was able to distinguish between scenarios having different reference risk levels. In addition, to make sure that the scenarios were distinguished "correctly," their classification according to their calculated risk level had to be compared with the reference classification determined in the earlier study (Chinniah et al., 2011). As a reminder, according to that study, the scenarios were classified by increasing order of risk level as follows: scenario A, scenario G, scenario M and scenario S. For analysis purposes, the number of recurrences of the different cases described in Table 4 was recorded.

Table 4. Criteria used for the analysis of scenario classifications

Point of comparison	Condition
Number of classifications comparable with reference classification	– No reversal in relation to reference classification and, – No more than two consecutive scenarios with same risk level
Number of cases where three or four scenarios are estimated at same risk level	– Three or four scenarios with same risk level
Number of cases of reversal of scenario risk levels in relation to reference order	– Two non-consecutive scenarios with same risk level or, – Classification obtained strictly different from reference classification

As there was no specific critical threshold for the criterion of being able to distinguish one scenario from another, a qualitative analysis was performed to evaluate tool performance.

2.3.2.2 User Satisfaction

The answers to questions 2.2 and 2.3 on the questionnaires for the risk estimation tool testing (Appendix D) were used to assess user satisfaction with the results obtained. Two indexes were defined from these questions:

- Number of subjects who disagreed with the risk level obtained (subjects who answered “Somewhat disagree” or “Totally disagree”)
- Number of negative comments (total then broken down by nature of comment: risk level too high or risk level too low)

The number of subjects who disagreed with the risk level obtained was a reflection of the general level of user dissatisfaction. However, some users indicated that they agreed with the risk level arrived at, given the choices they made for each parameter. Several users were able to take a step back and commented that, regardless of the choices they had to make, the risk level obtained did not correspond to their perception. The number of negative comments therefore had to be considered.

The negative comments were broken down by type (i.e., risk level obtained too low or too high) in order to determine why the users disagreed.

The thresholds from which points were regarded as being a problem for a tool were:

- 7 subjects or more disagreed with the risk level obtained for a scenario
- 10 subjects or more made a negative comment about a scenario

2.3.2.3 Convergence of Results

As for the analysis of the impact of flaws in risk estimation parameters, the modal percentage was used to provide an indication of the convergence of the results obtained with the tools (see subsection 2.3.1). It is therefore the proportion of the 25 subjects who obtained the same risk level for a given tool. To compare the modal percentages among the tools, an equivalency scale had to be established, to reduce all the tools to four risk levels. The greater the number of risk levels, the more diluted the modal percentage will be. That would have the effect of decreasing the modal percentage for tools that have more risk levels.

To do so, the equivalency scale was defined using the percentage of the maximum as the point of comparison, as was done in the earlier study (Chinniah et al., 2011). Levels 1 to 4 were therefore defined by the respective value ranges of 0% to 25%, 25% to 50%, 50% to 75%, and 75% to 100%. To get the levels of tools with more than four risk levels to correspond to these four ranges, each level was defined by the upper limit of the value range. For instance, level 4 of tool 91 having six levels (with 1 being the lowest) corresponded to 66.7%, so it was considered to be equivalent to level 3 of a four-level tool. The level equivalencies described in Table 5 were then obtained. The modal percentage used to quantify the convergence is that of the tools reduced to four risk levels. Modal percentage values less than or equal to 60% were regarded as an indication of low convergence.

Table 5. Equivalency of risk levels between 4-level tools and 6-level and 11-level tools

Tools 19, 24, 114 4 risk levels	Tool 89 6 risk levels	Tool 91 6 risk levels	Tool 69 11 risk levels
1 (25%)	16.7% 6	16.7% 1	9.1% 0 18.2% 1
2 (50%)	33.3% 5 50% 4	33.3% 2 50% 3	27.3% 2 36.4% 3 45.5% 4
3 (75%)	66.7% 3	66.7% 4	54.5% 5 63.6% 6 72.7% 7
4 (100%)	83.3% 2 100% 1	83.3% 5 100% 6	81.8% 8 90.9% 9 100% 10

3. RESULTS – IMPACT OF FLAWS ASSOCIATED WITH RISK ESTIMATION PARAMETERS

3.1 Overall Convergence of Results by Parameter Type

The convergence of the results obtained for all subjects was analysed on the basis of the modal percentage of the subjects' responses, as established in subsection 2.3.1.

Table 6 presents the modal percentage obtained for each of the six parameter types studied as a function of the different flaws they contain and for all the scenarios (overall mean for the four scenarios). The table shows that the modal percentage varied significantly by parameter. The results for the parameter "severity of harm" were fairly consistent from one subject to the next, with a mean modal percentage of 83%. This means that on average, 21 of the 25 subjects obtained the same result in applying "severity-of-harm" parameters to all the scenarios.

Table 6. Modal percentage obtained for each parameter type, by flaw type

	Mean for all scenarios (%)						
Flaws	Severity of harm (S)	Probability of occurrence of harm (Ph)	Exposure frequency (Exf)	Exposure duration (Exd)	Possibility of avoidance (A)	Probability of occurrence of hazardous event (Pe)	**Mean by flaw**
Poor definition of levels	86	47		67	53		**63**
Inconsistent definitions of different levels	70	51				43	**55**
Gaps between levels	86	52	86	72			**74**
Inadequate number of levels	90						**90**
No definition of exposure interval		47					**47**
Mean by parameter	**83**	**49**	**86**	**69**	**53**	**43**	**65**
No flaws	89	46	79	75	60	72	**70**

That suggests that the parameter "severity of harm" was relatively robust despite the flaws, and that it led to a strong consensus among users on the level of severity of the potential harm that a situation might represent.

The exposure-related parameters (frequency and duration) were also fairly robust. In contrast, the parameters "probability of occurrence of harm," "probability of occurrence of the hazardous event" and "possibility of avoidance" did not perform to the same level with respect to the convergence of results, with modal percentages ranging from 43% to 53%. For these parameters, it seems that the subjects had more trouble reaching a consensus, especially for the probability parameters (probability of occurrence of harm and probability of occurrence of the hazardous event), for which the modal percentage did not exceed 52%. This outcome also suggests that the evaluation of probabilities is a difficult part of risk estimation.

The last column of Table 6 gives the mean modal percentage as a function of each flaw studied. It can be seen that some flaws seem to have a greater impact on the various parameters than others. For example, the flaw "inconsistent definitions of different levels," studied with respect to three types of parameters, seems to have a noticeable effect, with a mean modal percentage of 55%. When it is compared with the flaw "gaps between levels," studied with respect to four types of parameters, it can be seen that the latter flaw seems to have less of an impact on the parameters, with a mean modal percentage of 74%.

Table 6 also indicates the modal percentage obtained for parameters assumed to have "no flaws." It can be seen that, on average, the modal percentage for these parameters was slightly higher than that for the parameters with flaws. These "no flaw" parameters seem to have performed better for four of the six parameter types studied.

3.2 Overall Analysis of Level of Difficulty of Application by Parameter Type

The tables below present the mean results for all the scenarios regarding the difficulty of applying the parameters having the flaws studied. Table 7 presents the mean number of subjects who indicated that it was fairly hard or very hard (answer 4 or 5) to make their choice of level for each type of parameter, as a function of the flaws studied. It can be seen that the flaw "poor definition of levels" seems to have given subjects the most trouble, with a mean of 7.4 subjects having indicated a level of difficulty of 4 or 5 at each opportunity. The flaw "inconsistent definitions of different levels" was next, with a mean of 6.2 subjects. Table 8 presents the number of negative comments from subjects about the flaw in each case.

Figure 3 shows the relationship between the data in Table 7 and the data in Table 8, i.e., the number of subjects who indicated a level of difficulty of 4 or 5 and the number of negative comments about the flaws. It can be seen that the correlation is fairly good between the two results, with $R^2 = 0.87$. This correlation between the level of difficulty and the number of associated negative comments suggests that for all the flaws, when the subjects had trouble making their choices, they were generally able to attribute their problems to the presence of flaws.

Table 7. Mean number of subjects who indicated it was fairly hard or very hard to make their choices of level for each type of parameter, by flaw

	Mean for all scenarios						
Flaws	Severity of harm (S)	Probability of occurrence of harm (Ph)	Exposure frequency (Exf)	Exposure duration (Exd)	Possibility of avoidance (A)	Probability of occurrence of hazardous event (Pe)	**Mean by flaw**
Poor definition of levels	5.8	7.3		8.3	8.3		**7.4**
Inconsistent definitions of different levels	7.5	5.0				6.0	**6.2**
Gaps between levels	2.8	6.5	4.3	10.3			**5.9**
Inadequate number of levels	2.3						**2.3**
No definition of exposure interval		4.3					**4.3**
Mean by parameter	**4.6**	**5.8**	**4.3**	**9.3**	**8.3**	**6.0**	**5.2**
No flaws	2.3	6.8	7.5	4.3	5.3	6.3	**5.4**

Table 8. Mean number of negative comments related to each flaw, for each type of parameter

	Mean for all scenarios						
Flaws	Severity of harm (S)	Probability of occurrence of harm (Ph)	Exposure frequency (Exf)	Exposure duration (Exd)	Possibility of avoidance (A)	Probability of occurrence of hazardous event (Pe)	**Mean by flaw**
Poor definition of levels	10.6	13.0		10.8	11.8		**11.5**
Inconsistent definitions of different levels	11.3	9.3				10.0	**10.2**
Gaps between levels	5.0	6.0	10.0	3.0			**6.0**
Inadequate number of levels	5.8						**5.8**
No definition of exposure interval		1.0					**1.0**
Mean by parameter	**8.2**	**7.3**	**10.0**	**6.9**	**11.8**	**10.0**	**6.9**

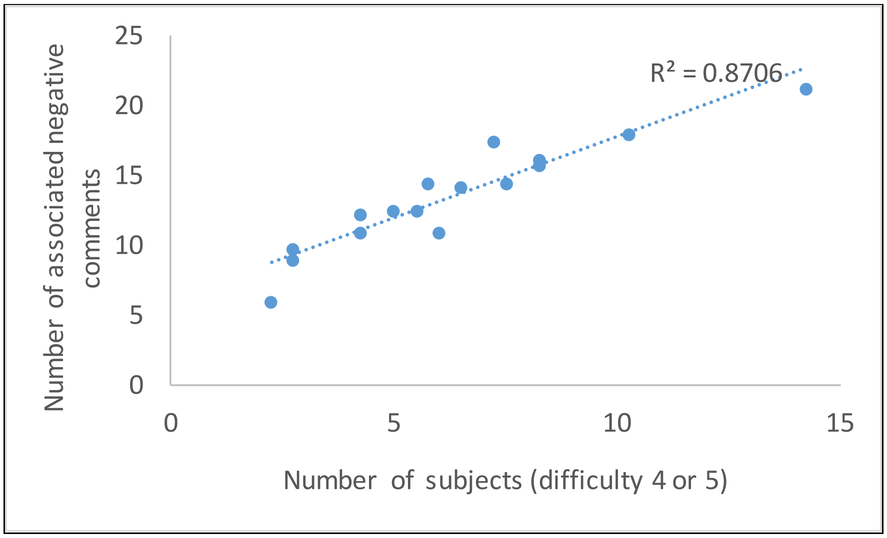

Figure 3. Relationship between number of subjects who indicated a level of difficulty of 4 or 5 and number of negative comments in connection with flaws

In contrast, as Figure 4 shows, when the data from Table 7 (number of subjects who indicated a level of difficulty of 4 or 5) are combined with the modal percentages presented in Table 6, the relationship is less clear. No correlation can be seen between the two results. So, the perception of the difficulty of applying the different parameters did not influence the convergence of the results among the subjects.

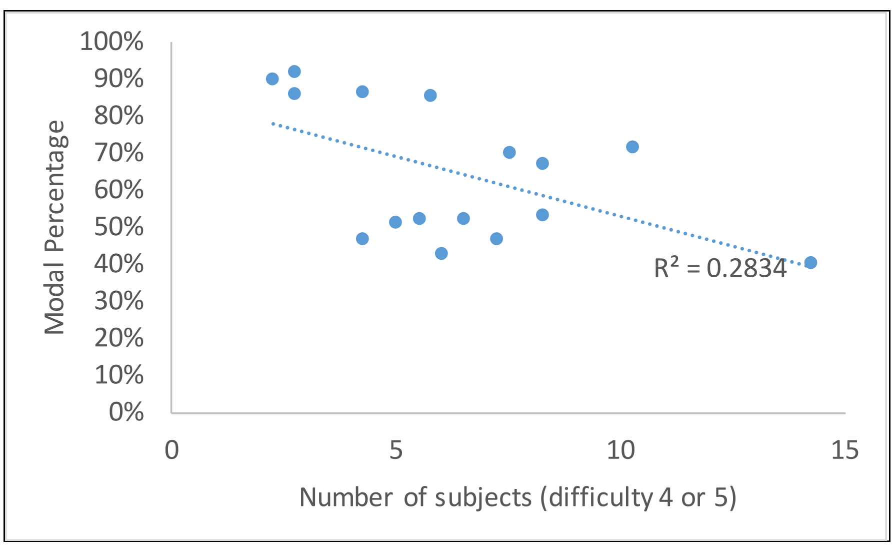

Figure 4. Relationship between number of subjects who indicated a level of difficulty of 4 or 5 and modal percentage

In-depth analyses of these general trends, for each type of parameter, are presented in the following sections.

3.3 Analysis of Flaws Associated with Parameter "Severity of Harm (S)"

Table 9 presents an overview of the analysis of the parameter "severity of harm."

Table 9. Analysis of the parameter "severity of harm" for the four scenarios

			Scenario A				Scenario G				Scenario M				Scenario S			
Severity of harm (S)	**Tool**	**No. of levels**	**Modal percentage**	**No. of subjects who had trouble responding***	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**
Poor definition of levels	33	3	100	2	4	13	68	9	14	18	88	9	13	15	100	2	3	5
Poor definition of levels	55	4	80	5	14	17	64	12	17	19	100	1	6	9	87	6	14	20
Inconsistent definitions of different levels	66	4	60	8	19	19	56	16	17	22	92	4	6	11	74	2	3	6
No flaws	69	4	92	1	0	6	68	8	0	14	96	0	0	4	100	0	0	2
Inadequate number of levels	91	2	100	0	3	3	60	9	11	12	100	0	6	6	100	0	3	3
Gaps between levels	102	3	56	4	4	12	88	6	7	13	100	1	6	8	100	0	3	3

*Number of subjects who said they found choosing "Fairly hard" or "Very hard" (subsection 2.3.1)

3.3.1 Description of Parameters and Flaws Studied

Six "severity of harm" parameters were used to assess the impact of the various flaws. Tables 10 to 15 show a compilation of the choices made by the subjects for each of these parameters and for each of the four scenarios. The percentage indicates the number of subjects out of 25 who chose a given level, the highest percentage being the modal percentage. For each scenario, the value in bold indicates the severity-of-harm reference level, as established on the basis of the equivalency scales from the earlier study (Chinniah et al., 2011). Note that the mode corresponded to the severity-of-harm reference level in 22 cases out of 24.

Two parameters (tools 33 and 55) concerned the flaw "poor definition of levels." Table 10 presents the results for the parameter "severity of harm" of tool 33. This parameter was selected because of the lack of information about choosing from among the tool's three severity-of-harm levels. It was deemed that the qualifiers "moderate," "serious" and "grievous" did not provide users with sufficient guidance.

Table 10. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "severity of harm" of tool 33

Scenario	Parameter level		
	Moderate injury or illness	Serious injury or illness	Grievous injury or illness, or death
A	**100**	0	0
G	**68**	32	0
M	0	13	**88**
S	0	0	**100**

Table 11 presents the results for the parameter "severity of harm" of tool 55. This parameter was selected because of the lack of information about choosing the severity-of-harm level. It was deemed that the qualifiers "minor" and "severe" did not provide sufficient guidelines to users.

Table 11. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "severity of harm" of tool 55

Scenario	Parameter level			
	Negligible: less than minor injury or occupational illness	Marginal: minor injury or occupational illness	Critical: severe injury or occupational illness	Catastrophic: death
A	**20**	80	0	0
G	0	**64**	36	0
M	0	0	**100**	0
S	0	0	13	**87**

Table 12 presents the results for the parameter "severity of harm" of tool 66. This parameter was selected to assess the flaw "inconsistent definitions of different levels" because the expression "minor injury" is used twice (first and second levels of the scale).

Table 12. Proportion of subjects (%) who selected a given level for each scenario, for the parameter “severity of harm” of tool 66

Scenario	Parameter level			
	Insignificant: possible minor injury	Marginal: minor injury and/or significant threat to the environment	Critical: single fatality and/or severe injury and/or significant damage to the environment	Catastrophic: fatalities and/or multiple severe injuries and/or major damage to the environment
A	**60**	40	0	0
G	12	**56**	32	0
M	0	0	**92**	8
S	0	0	**74**	26

Table 13 presents the results for the parameter “severity of harm” of tool 69. This parameter was selected as a parameter assumed to have “no flaws,” in other words, it did not have any of the flaws specifically analysed within the framework of this study.

Table 13. Proportion of subjects (%) who selected a given level for each scenario, for the parameter “severity of harm” of tool 69

Scenario	Parameter level			
	No harm	Low: trivial harm with no permanent results	Medium: serious harm with no permanent results	High: serious harm with permanent results, death
A	4	**92**	4	0
G	0	28	**68**	4
M	0	0	4	**96**
S	0	0	0	**100**

Table 14 presents the results for the parameter “severity of harm” of tool 91. This parameter, which has only two severity-of-harm levels, was selected to assess the flaw “inadequate number of levels.”

Table 14. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "severity of harm" of tool 91

Scenario	Parameter level	
	Slight injury (usually reversible), for example, scratches, laceration, bruising, light wound requiring first aid	Serious injury (usually irreversible, including fatality), for example, broken or torn-out or crushed limbs, fractures, serious injuries requiring stitches, major musculoskeletal troubles (MST), fatalities
A	**100**	0
G	40	**60**
M	0	**100**
S	0	**100**

Table 15 presents the results for the parameter "severity of harm" of tool 102. This parameter was selected to assess the flaw "gaps between levels." For this parameter, a significant gap can be seen between the two upper levels of the scale, which jumps from first aid to an injury causing permanent harm or death.

Table 15. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "severity of harm" of tool 102

Scenario	Parameter level		
	Minor: consequences not very serious	Significant: work has to stop, first aid is really needed	Disastrous: very serious accident (someone has been scarred for life, blinded or even killed)
A	56	**44**	0
G	4	**88**	8
M	0	0	**100**
S	0	0	**100**

3.3.2 Impact of Flaws on Convergence of Results

As mentioned earlier, the convergence of results for this parameter was good, with a mean modal percentage of 83%. This may indicate that the parameter is fairly robust and that the different flaws studied had relatively little impact on the determination of the severity of harm. Four cases in Table 9 are worthy of attention, however, as they had modal percentages lower than or equal to 60%.

The first two cases concern the flaw “inconsistent definitions of different levels” in the “severity-of-harm” parameter of tool 66, applied to scenarios A and G. Note that these two scenarios have the lowest severity reference levels. Table 12 presents the results obtained by the subjects in these two cases. For this parameter, the expression “minor injury” is used in the first two levels. This inconsistency in the definitions of different levels seems to have resulted in a greater dispersion of responses in the case of scenarios where these levels are conceivable options, i.e., scenarios with less serious risk.

The third case has to do with the flaw “inadequate number of levels” of tool 91 applied to scenario G. The possible harm “bruising, simple fracture” of scenario G straddles the two levels of this parameter (Table 14). Thus, 40% of the subjects opted for the first level (bruising) while 60% chose the second (fracture).

The fourth case concerns the flaw “gaps between levels” in the parameter “severity of harm” of tool 102, applied to scenario A. It can be seen that 56% of the subjects chose the first severity level and 44% the second level (Table 15). Note that the reference severity level is the second level for scenario A (punching machine with mobile table), with “bruises, cuts” as an indicator of the possible harm. For this parameter, the first severity level is defined as “minor: consequences not very serious,” while the second level is defined as “Significant: work has to stop, first aid is really needed.” In this case, it seems that the lack of a gap between these two adjacent levels may have influenced the subjects’ responses; “bruises, cuts” could be considered non-serious harm, but they might also require first aid. There is a significant gap between the two top levels of the scale, which go from first aid to an injury that causes permanent harm. The presence of this flaw seems to have worked in favour of better convergence of results for the higher-severity scenarios (M and S).

3.3.3 Impact of Flaws on Level of Difficulty of Applying Parameter

In addition to the modal percentage for each case, Table 9 also shows the number of subjects who indicated that they found it fairly hard or very hard to choose between the different levels in each case, as well as the number of negative comments made by the subjects. It can be seen that in a few cases, close to a third of the subjects (8 out of 25) found it fairly hard or very hard to make a choice.

Two cases concerning the flaw “inconsistent definitions of different levels” in the “severity-of-harm” parameter of tool 66, applied to scenarios A and G, are worth noting. These cases produced a relatively low modal percentage, as discussed earlier (see 3.3.1). This suggests that the problems the subjects faced resulted in a greater dispersion of their responses. The subjects’ comments also seem to confirm this observation. As Table 9 shows, 19 subjects made negative comments about this flaw for scenario A and 17 for scenario G. Here are some typical comments:

- *The difference between insignificant and marginal is slight. The two terms overlap.*

- *Two choices for a minor injury because of "and/or."*

- *Hard to distinguish between insignificant and marginal. Adding "possible" to the definition doesn't necessarily help me choose.*

The subjects also noted that the parameter associates a concept of probability with the determination of severity by using the word "possible" in the first level (see Table 12), as well as the concept of damage to the environment—aspects that seem to compound the inconsistency of the definitions of the levels of this parameter. It can be seen that the presence of this flaw in the severity parameter of tool 66 had an influence on the convergence of subjects' responses, made the choice of level difficult for many subjects, and that the subjects attributed this difficulty to the presence of the flaw in the parameter.

The flaw "poor definition of levels" in the parameter "severity of harm" of tool 33 prompted a lot of negative comments. For instance, for scenario M, 9 subjects indicated they had had trouble choosing the level, and 13 made a comment about the flaw. Scenario M (rewinder) gives "partial or complete amputation of upper limbs" as an indication of the possible harm. For the parameter "severity of harm" of tool 33, the first level of severity is defined as "serious injury or illness," while the second level is "grievous injury or illness, or death" (see Table 10). Although the modal percentage is relatively good in this case (88%) and corresponds to the reference severity level, it can be assumed that the subjects may have had a certain amount of trouble deciding whether amputation is a serious injury or a grievous injury. This hypothesis would seem to be confirmed by the subjects' negative comments about this flaw:

- *Arbitrary choice between serious or grievous.*

- *Have to interpret serious and grievous. In my opinion, an amputation is grievous... We're forced to interpret.*

- *Arbitrary choice to distinguish between serious and grievous.*

- *Define serious illness and grievous illness, grievous amputation?*

- *Difference between "serious" and "grievous"?*

The flaw "poor definition of levels" in the parameter "severity of harm" of tool 55 also gave rise to a lot of negative comments. Despite a relatively good modal percentage for all four scenarios taken together, the subjects often noted the lack of detail in the definition of a "minor injury or occupational illness" in relation to a "severe injury or occupational illness." Here again, many subjects pointed out the vague, imprecise definitions of the levels of the severity-of-harm parameter. In applying this parameter to scenario A, only 20% of subjects chose the level corresponding to the reference severity level (see Table 11).

Table 9 also shows that scenario G was the one that gave subjects the most trouble in making their choices. These problems are also reflected in the mean modal percentage, which, at 65%, was the lowest of the four scenarios. Scenario G (automated guided vehicle) gives “bruising, simple fracture” as an indication of the possible harm. This finding, together with the subjects’ comments, suggests that the flaws in these severity parameters could have a greater impact on the establishment of the severity level for intermediate harm, such as a simple fracture (scenario G), than on that for minor harm (e.g., the bruising of scenario A) or grievous harm (e.g., the amputation of scenario M and the death of scenario S).

The flaw “inadequate number of levels” in tool 91 did not appear to have much of an effect on the convergence of results or the level of difficulty perceived by the subjects. This flaw also garnered the lowest total number of negative comments from subjects, with 24 compared with an average of 43 for the other flaws. Nevertheless, virtually all the comments (23 out of 24) specifically concerned this flaw, for instance:

- *It’s good that there are examples. But there should be more categories.*
- *I would have liked to have more levels (I liked the fact there were examples).*
- *Lack of rungs on the scale.*
- *Not good: simple fracture on the same level as fatality. Missing some levels (too Boolean).*
- *Should be another level between the two; fracture is not covered.*
- *Not enough levels, choices per flaw.*

In other words, the subjects had no trouble choosing their levels, but for many of them, the choice didn’t feel right or they were forced into it because the parameter only had two levels.

3.3.4 Findings Regarding Impact of Flaws Associated with Parameter “Severity of Harm”

The observations and analyses presented in the preceding sections provide the basis for the following findings regarding the impact of the construction flaws on the parameter “severity of harm”:

1. The parameter “severity of harm” seems fairly robust and the convergence of results is less affected by flaws, in most cases, than the other risk estimation parameters.

2. The number of negative comments from the subjects in connection with the different flaws represented 74% of all the negative comments made (173 out of 234). This would seem to indicate that despite good convergence of results, the subjects were generally able to identify the flaws and recognize how they may negatively affect their choice of severity-of-harm level.

3. The flaws did not have a uniform impact on the application of the "severity-of-harm" parameters. The type of flaw, its position on the parameter's severity scale and the possible harm seem to influence its impact on the determination of the severity-of-damage level. For example, some flaws may not have any impact on the lowest and highest levels of harm, but have a greater influence on the establishment of the severity levels for intermediate harm, and vice-versa.

4. The flaws "poor definition of levels" and "inconsistent definitions of different levels" seem to have had the greatest impact on the subjects' determination of the severity-of-harm level. These flaws accounted for 174 of the subjects' 298 negative comments, and 130 of the comments concerned them specifically.

5. The flaw "gaps between levels" had variable effects, depending on the case. On the one hand, too great a gap between the levels of a parameter had little impact on the convergence of results and on the degree of trouble responding, as suggested by the results of the application of tool 102. On the other hand, this flaw was noted by subjects who made a total of 20 comments directly about it. The comments suggest that the gap prompted the subjects to choose a specific level in each case. However, a majority of the subjects mentioned that a level was missing from the scale for this parameter, which created a certain degree of discomfort in the selection process. Furthermore, the lack of a gap between two levels can have a considerable effect on the convergence of results. Moreover, further with regard to tool 102, nine subjects made comments (negative or positive) in support of this interpretation in applying it to scenario A, for which a modal percentage of just 56% was obtained.

6. The use of only two levels for determining the severity of harm may make it easier for users to choose. On the other hand, this situation seems to create a certain degree of unease among users. This flaw therefore had an effect similar to that of too great a gap between levels.

3.4 Analysis of Flaws Associated with Parameter "Probability of Occurrence of Harm (Ph)"

Table 16 presents an overview of the analysis of the parameter "probability of occurrence of harm."

Table 16. Analysis of the parameter "probability of occurrence of harm"

			Scenario A				Scenario G				Scenario M				Scenario S			
Probability of occurrence of harm (Ph)	**Tool**	**No. of levels**	**Modal percentage**	**No. of subjects who had trouble responding***	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**
Inconsistent definitions of different levels	6	5	44	4	12	14	44	2	7	11	63	5	10	13	54	9	8	12
Poor definition of levels	7	5	36	6	16	17	40	4	11	17	63	10	13	20	50	9	12	16
Gaps between levels	34	3	44	7	5	14	52	5	5	12	63	8	9	19	50	6	5	12
No flaws	41	6	68	6	0	18	40	5	0	12	38	7	0	14	38	9	0	12
Lack of exposure interval	89	4	48	3	1	10	60	4	1	11	46	5	0	12	33	5	0	9

*Number of subjects who said they found choosing "Fairly hard" or "Very hard" (subsection 2.3.1)

3.4.1 Description of Parameters and Flaws Studied

Five "probability of occurrence of harm" parameters were used to assess the impact of the various flaws. The tables below detail the results of the subjects' choices for each of these parameters and for each of the four scenarios. They show that the mode corresponds to the probability-of-occurrence-of-harm reference level in only 11 out of 20 cases.

Table 17 presents the results for the parameter "probability of occurrence of harm" of tool 6. This parameter was selected to assess the flaw "inconsistent definitions of different levels" because of the cross-referenced definitions of its first three levels, which use the terms "improbable," "remote," "unlikely," "conceivable" and "could occur." These expressions were deemed to be too similar on a semantic level.

Table 17. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "probability of occurrence of harm" of tool 6

Scenario	Parameter level				
	Improbable – probability close to zero	Remote – unlikely, though conceivable	Possible – could occur sometime	Probable – not surprised, will occur several times	Likely/frequent – occurs repeatedly / event only to be expected
A	0	4	44	**32**	20
G	0	20	44	**36**	0
M	0	8	**63**	25	4
S	4	33	**54**	8	0

Table 18 presents the results for the parameter "severity of harm" of tool 7, which concerned the flaw "poor definition of levels." This parameter was selected because of the lack of information about choosing from among the tool's five probability-of-occurrence-of-harm levels, which use only qualifiers that do not provide users with sufficient guidance.

Table 18. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "probability of occurrence of harm" of tool 7

Scenario	Parameter level				
	Remote	Improbable	Possible	Probable	Likely
A	0	4	32	**36**	28
G	4	20	28	**40**	8
M	0	4	**63**	33	0
S	8	17	**50**	13	13

Table 19 presents the results for the parameter "probability of occurrence of harm" of tool 34. This parameter was selected to assess the flaw "gaps between levels." For the parameter "probability of occurrence of harm," it can be seen that the lowest and highest levels are too far removed from the intermediate level. It is therefore hard to classify a situation having a probability of occurrence of harm that would be considered "low" without being very seldom, or "high" without being near certain.

Table 19. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "probability of occurrence of harm" of tool 34

Scenario	Parameter level		
	Low – occurs very seldom or never	Medium – reasonably likely to occur	High – certain or near certain to occur
A	12	44	**44**
G	20	52	**28**
M	25	**63**	13
S	50	**42**	8

Table 20 presents the results for the parameter "probability of occurrence of harm" of tool 41. This parameter was selected as a parameter assumed to have "no flaws," in other words, it did not have any of the flaws specifically analysed within the framework of this study.

Table 20. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "probability of occurrence of harm" of tool 41

Scenario	Parameter level					
	F – Highly improbable – probability cannot be distinguished from zero	E – Improbable – very unlikely to occur in life cycle	D – Remote – unlikely, but may possibly occur in life cycle	C – Occasional – likely to occur at least once in life cycle	B – Probable – likely to occur several times in life cycle	A – Highly probable – likely to occur frequently in life cycle
A	0	0	12	**12**	68	8
G	0	4	20	**32**	40	4
M	0	0	33	**38**	29	0
S	4	17	38	**25**	17	0

Table 21 presents the results for the parameter "probability of occurrence of harm" of tool 89. This parameter was selected in order to assess the effect of the flaw "lack of exposure interval." For this parameter, no indication was provided regarding the period of time (exposure interval) for which the probability was to be evaluated. Note, however, that with the exception of tool 41 above (assumed to have "no flaws"), none of the other "probability-of-occurrence-of-harm" parameters presented here gave any indication of an exposure interval. The parameter of tool 89 did not have any of the other flaws studied here.

Table 21. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "probability of occurrence of harm" of tool 89

Scenario	Parameter level			
	Very unlikely – could happen, but probably never will	Unlikely – could happen, but rare	Likely – could happen occasionally	Very likely – could happen frequently
A	0	20	**48**	32
G	0	32	**60**	8
M	0	**46**	42	13
S	8	**33**	25	33

3.4.2 Impact of Flaws on Convergence of Results

As mentioned earlier, the convergence of results for this parameter was low, with a mean modal percentage of 50%. Even the parameter of tool 41, assumed to have "no flaws," did not produce a better convergence of results. These results were similar for the four scenarios under consideration. These findings, combined with the fact that the mode corresponded to the probability-of-occurrence-of-harm reference level in only 11 cases out of 20, indicate that this parameter gave the study subjects a great deal of trouble. It is therefore hard to determine the impact of the flaws on these "probability-of-occurrence-of-harm" parameters based on the modal percentage.

3.4.3 Impact of Flaws on Level of Difficulty of Applying Parameter

Regarding the difficulty of applying the parameter, Table 16 shows five cases where close to a third of the subjects (8 out of 25) found it fairly hard or very hard to make their choices. These five cases were observed for scenarios M and S. Two cases concern the flaw "poor definition of levels" in the parameter "probability of occurrence of harm" of tool 7. As Table 18 shows, this parameter only uses qualifiers, without any other indications, to define the various levels of its scale. Furthermore, 13 subjects made negative comments focusing directly on this flaw for scenario M and 12 for scenario S. Here are some typical comments:

- *Levels need to be defined better.*
- *It's vague, we have to interpret. Hesitated between possible and probable... it's the same thing.*
- *The probability parameter itself is not clear (probability of harm versus probability [risk] level): It's all based on words ... no explanation, no clarification, no scale.*
- *The description is not sufficient to allow levels to be selected (possible as opposed to probable) for a communication error.*

– *No definition, so we have to interpret.*

There are two other cases for scenario S, which had the highest risk level of the four. The first concerned the parameter of tool 41, considered to have "no flaws." An analysis of all the comments the subjects made about this parameter shed new light on its construction. The parameter uses the concept of "life cycle" to define the exposure interval to be used for establishing the probability of occurrence of harm. Many of the negative comments specifically mentioned this concept of "life cycle." Many subjects said that this aspect of the level definitions made it harder for them to make their choices, for instance:

– *I don't like life "cycle." It doesn't mean anything to me.*

– *The concept of cycle just confuses things.*

– *The concept of life cycle is extremely hard to judge in a scenario like this one.*

– *More specific, but define life cycle and worker training.*

– *What life cycle (machine or duration of exposure)?*

– *I'm not sure about the equipment life cycle.*

These comments strongly suggest that a certain aspect of this parameter's construction was indeed harmful, even though it was assumed to have "no flaws."

The second case concerns the flaw "inconsistent definitions of different levels" in the parameter "probability of occurrence of harm" of tool 6. As Table 17 shows, 54% of the subjects chose the level "Possible – could occur sometime," which is the probability-of-occurrence-of-harm reference level, while 33% chose the lower level "Remote – unlikely, though conceivable." An analysis of the 12 comments from the subjects clearly shows that many of them were influenced by this flaw:

– *Hesitated between remote and possible. The definitions are very similar.*

– *Level descriptions are detailed, but fuzzy—similar words but don't really mean anything ... so I based my decision on my own perception. The parameter description leaves room for a judgment call (choose between 2 terms with a slightly different meaning).*

– *The descriptions don't present a clear choice between "remote" and "possible." When in doubt, I opt for the riskier of the two.*

– *The definitions of remote and possible are very similar.*

A last case concerns the flaw "gaps between levels" in the parameter "probability of occurrence of harm" of tool 34, applied to scenario M. For this parameter, eight subjects said they had some trouble making their choices, while nine made negative comments about the flaw. Despite a relatively acceptable modal percentage (63%), many of the subjects' comments reveal their uncertainty about the flaw, as the following examples show:

– *More levels are needed. Hesitated between low and medium. The gap between never and reasonably likely is high. What does "reasonably likely" mean? Result influenced by my experience.*

– *More categories needed. The terms aren't clear; not enough explanation.*

– *Not enough levels. Not enough gradation of levels. Chose closest by default.*

– *I think there should be at least one more level between low and medium. The probability isn't zero, but it's not medium either.*

It is worth noting that many of the comments refer to the fact that there is a sometimes significant gap between levels, but that the gap is the result of the lack of levels on the parameter's scale, which has only three.

3.4.4 Findings Regarding Impact of Flaws Associated with Parameter "Probability of Occurrence of Harm"

The observations and analyses presented in the preceding sections provide the basis for the following findings regarding the impact of the construction flaws on the parameter "probability of occurrence of harm":

1. The subjects seemed to have trouble determining the probability of occurrence of harm. The low convergence of results and the low rate of correspondence with the reference levels suggest that this parameter was a problem in most cases.

2. While it is hard to determine the specific impact of the various flaws in this parameter, the subjects were able to select these flaws specifically on a number of occasions.

3. The flaws "inconsistent definitions of different levels" and "poor definition of levels" were identified the most often by the subjects (on the basis of the number of associated negative comments). These flaws generated 93 negative comments concerning them specifically. This suggests that, even though the subjects did not always indicate that these flaws may have made the selection process harder, they were able to recognize these flaws in the parameter "probability of occurrence of harm" and their potential impact.

4. The flaws "gaps between levels" and "lack of exposure interval" were less frequently recognized by subjects in the parameters studied. That is especially true for the flaw "lack of exposure interval," which was noted in only two cases.

5. Several subjects objected to the concept of "life cycle," used to establish the exposure interval in the parameter of tool 41. This suggests that (i) this concept is not sufficiently precise to be useful, or that (ii) the subjects prefer to define the exposure interval

qualitatively themselves in estimating the probability of occurrence of harm. This second hypothesis seems supported by the fact that "the lack of exposure interval" was not identified as a flaw by most subjects.

6. The flaw "gaps between levels" seems to be perceived in the same way as "inadequate number of levels" when the number of levels of the probability-of-occurrence-of-harm scale is low. Regarding this flaw, a majority of the subjects mentioned that a level was missing from the scale for this parameter, which made them feel uncomfortable with the selection process.

3.5 Analysis of Flaws Associated with Parameter "Exposure Frequency (Exf)"

Table 22 provides an overview of the analysis of the parameter "exposure frequency."

Table 22. Analysis of parameter "exposure frequency"

			Scenario A				Scenario G				Scenario M				Scenario S			
Exposure frequency (Exf)	**Tool**	**No. of levels**	**Modal percentage**	**No. of subjects who had trouble responding***	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**
Gaps between levels	49	2	84	4	6	10	88	1	7	10	100	1	8	9	74	11	19	20
No flaws	67	4	64	7	0	12	64	9	0	11	100	3	0	6	87	11	0	12

*Number of subjects who said they found choosing "Fairly hard" or "Very hard" (subsection 2.3.1)

3.5.1 Description of Parameters and Flaws Studied

Only one flaw was examined for the parameter "exposure frequency": gaps between levels. The tables below show the results of the subjects' choices for the two parameters used and for each of the four scenarios. Table 23 presents the results for the parameter "exposure frequency" of tool 49. This parameter was selected to assess the flaw "gaps between levels." For this parameter, a significant gap between two levels can be seen for an exposure frequency that is greater than once a day, but less than once an hour.

Table 23. Proportion of subjects (%) who selected a given level for each scenario, for the parameter “exposure frequency” of tool 49

Scenario	Parameter level	
	E1 – Infrequent exposure (typically, exposure to hazard less than once per day or shift)	E2 – Frequent exposure (typically, exposure to hazard more than once per hour)
A	16	**84**
S	12	**88**
M	**100**	0
S	26	**74**

Table 24 presents the results for the parameter “exposure frequency” of tool 67, selected as a parameter assumed to have “no flaws,” in other words, it did not have any of the flaws specifically analysed within the framework of this study.

Table 24. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "exposure frequency" of tool 67

Scenario	Parameter level				
	1. Interval between exposures is more than a year	2. Interval between exposures is more than two weeks, but less than or equal to a year	3. Interval between exposures is more than a day, but less than or equal to two weeks	4. Interval between exposures is more than an hour, but less than or equal to a day. Where the duration is shorter than 10 min., the value may be decreased to the next level	5. Interval between exposures is less than or equal to an hour. This value is not to be decreased at any time
A	0	0	8	**28**	64
S	0	0	8	**28**	64
M	0	0	**100**	0	0
S	0	0	9	**87**	4

3.5.2 Impact of Flaws on Convergence of Results

The convergence of results for the parameter "exposure frequency" of tool 49 was good, with a mean modal percentage of 86%. Table 23 also shows that the correspondence between the subjects' responses and the exposure frequency reference levels was also very good. These results need to be assessed with care, however, as the modal percentage could not be less than 50% because the parameter only had two levels. Nevertheless, it would appear that in most cases, the subjects' choices converged toward the reference level despite the flaw "gaps between levels." Only scenario S seems have caused more hesitation, with a modal percentage of 74%.

For the parameter "exposure frequency" of tool 67 (assumed to have "no flaws"), the convergence was not as good for scenarios A and G.

3.5.3 Impact of Flaws on Level of Difficulty of Applying Parameter

As Table 22 shows, 11 subjects said they had had some trouble selecting the exposure frequency level for scenario S with the parameter of tool 49. In addition, 19 of the 20 negative comments about this parameter clearly and directly concerned the flaw in question (gaps between levels), for instance:

- *Contradictory choices? There's a gap between less than once a day and more than once an hour?*
- *Interval missing: twice per shift is not given as a possibility; it falls between the two parameters.*
- *There aren't enough levels to categorize my choice (it falls between the two).*

– *So what about less than once a day? Where does it fit? Intermediate level missing.*

– *Doesn't work. There's a gap between E1 and E2.*

– *There's a gap that's not covered, between the two levels.*

It is interesting to note that for the other three scenarios, only a few subjects said they had had trouble making their choices, and only a few negative comments were made about this flaw. Table 25 helps to understand the situation. For scenarios A, G and M, the flaw "gaps between levels" had no impact, as the information on exposure provided to the subjects was relatively consistent with the level definitions of the parameter for tool 49. For scenario S, however, the information provided specifically concerned the gap between the two levels: the exposure frequency is more than once a day, but less than once an hour. This means that subjects must choose a level more or less arbitrarily, which explains the trouble they had.

Table 25. Analysis of impact of flaw "gaps between levels" of parameter "exposure frequency" of tool 49

<table>
<tr><th>Scenario</th><th>Information on exposure</th><th>E1 – Infrequent exposure (typically, exposure to hazard less than once per day or shift)</th><th>E2 – Frequent exposure (typically, exposure to hazard more than once per hour)</th></tr>
<tr><td>A</td><td>A visitor remains in the work area for 5 minutes on average. The mobile table of the punching machine is moving 50% of the time. A visitor is present in the work area 20% of the time, on average, over a 10-hour day of exposure.</td><td></td><td>20% of the time over 10 hours for 5 minutes is equivalent to 2.4 exposures per hour</td></tr>
<tr><td>S</td><td>On average, the travel lane of the automated guided vehicle (AGV) is crossed 25 times by workers over the course of each 8-hour shift. It takes a worker 3 seconds to cross the AGV's travel lane completely. The AGV operates continuously during working hours.</td><td></td><td>25 times per 8-hour shift is equivalent to 3.1 exposures per hour</td></tr>
<tr><td>M</td><td>Each time a new roll is installed, i.e., once every two days. Each operation takes approximately 15 minutes.</td><td>Once every two days is less than once a day</td><td></td></tr>
<tr><td>S</td><td>Ten-minute job, twice every 8-hour shift.</td><td colspan="2">Exposure frequency greater than once a day, but less than once an hour</td></tr>
</table>

With respect to the parameter "exposure frequency" of tool 67 (assumed to have "no flaws"), a relatively high number of subjects had trouble with scenarios A, G and S. The subjects' comments suggest that the complexity of the definition of the fourth level of this parameter made it hard for them to choose, such as:

– *Long sentences that are hard to understand.*

– *Need to do calculations before I can respond.*

– *There are enough levels. The use of the interval is interesting, but makes it more complex. Level 4 is harder to understand, but I imagine you would get used to it with experience.*

– *Just ridiculous. Fuzzy. Really hard to respond.*

– *Complicated definitions (understanding them and applying them to the scenario).*

– *Hard to interpret the second part of the description of level 4.*

Note that the exposure frequency reference level for the scenarios in question (A, G and S) is this parameter's level 4, while for scenario M, the reference level is level 3. Level 4 is defined as follows (Table 24): "The interval between exposures is more than an hour, but less than or equal to a day. Where the duration is shorter than 10 min., the value may be decreased to the next level." The subjects clearly indicated that this type of description requiring an adjustment calculation is too complex to use. That seems also to have had an impact on the correspondence of the subjects' responses with the exposure frequency reference level, which was just 28% for scenarios A and G.

3.5.4 Findings Regarding Impact of Flaws Associated with Parameter "Exposure Frequency"

The observations and analyses presented in the preceding sections provide the basis for the following findings regarding the parameter "exposure frequency":

1. The flaw "gaps between levels" does not have a uniform impact on the application of the exposure frequency parameters. The position of the flaw on the parameter's scale, combined with the information on exposure, seems to influence the impact of the flaw on the determination of the exposure frequency level. As the results for scenario S showed, when a scenario involves a situation that concerns this gap, the convergence and level of difficulty reveal the impact of the flaw. The flaw was then highlighted in the comments of most subjects.

2. In determining the exposure frequency, a parameter level definition that is too complex has a significant impact on the convergence of results and on the degree of difficulty of the selection process. As the results for the "exposure frequency" parameter of tool 67 showed, when a calculation was required to choose a certain level, many subjects clearly identified this factor as detrimental to the selection process. It would therefore seem that this characteristic is an additional flaw that can affect the "exposure frequency" parameter and potentially other risk estimation parameters.

3.6 Analysis of Flaws Associated with Parameter "Exposure Duration (Exd)"

Table 26 provides an overview of the analysis of the parameter "exposure duration."

Table 26. Analysis of parameter "exposure duration"

			Scenario A				Scenario G				Scenario M				Scenario S			
Exposure duration (Exd)	**Tool**	**No. of levels**	**Modal percentage**	**No. of subjects who had trouble responding***	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**
Poor definition of levels	19	2	64	8	11	14	64	10	10	17	79	8	11	15	61	7	11	17
Gaps between levels	62	5	32	13	3	20	80	11	1	20	96	10	5	17	78	7	3	15
No flaws	91	2	56	5	0	13	72	2	0	12	88	5	0	15	83	5	0	14

*Number of subjects who said they found choosing "Fairly hard" or "Very hard" (subsection 2.3.1)

3.6.1 Description of Parameters and Flaws Studied

Three exposure duration parameters were used to assess the impact of two flaws on this parameter. The tables below detail the results of the subjects' choices for each of these parameters and for each of the four scenarios. They show that the mode corresponds to the exposure-duration reference level in only 11 cases out of 20.

Table 27 presents the results for the parameter "exposure duration" of tool 19, which concerned the flaw "poor definition of levels." This parameter was selected because of the lack of information about choosing from among the tool's two levels, which use only qualifiers that do not provide users with sufficient guidance.

Table 27. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "exposure duration" of tool 19

Scenario	Parameter level	
	1. Seldom to quite often	2. Frequent to continuous
A	**36**	64
S	**36**	64
M	**79**	21
S	**61**	39

Table 28 presents the results for the parameter "exposure duration" of tool 62. This parameter was selected to assess the flaw "gaps between levels." For this parameter, gaps can be seen between adjacent levels. For instance, exposure of two days per month would fall between the first two levels of the scale, and exposure of 12 hours per week would be between levels 3 and 4.

Table 28. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "exposure duration" of tool 62

Scenario	Parameter level				
	1. 2 hr/week; 1 day/month	2. 4 hr/week; ½ day/week	3. 8 hr/week; 1 day/week	4. 20 hr/week; half the time	5. 40 hr/week; all the time
A	32	20	**32**	8	8
S	**80**	0	0	4	16
M	**96**	0	0	4	0
S	**78**	4	9	9	0

Table 29 presents the results for the parameter "exposure duration" of tool 91, selected as a parameter assumed to have "no flaws," in other words, it did not have any of the flaws specifically analysed within the framework of the study.

Table 29. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "exposure duration" of tool 91

Scenario	Parameter level	
	F1 – Seldom to quite often and/or short duration of exposure	F2 – Frequent to continuous and/or long duration of exposure
A	56	**44**
G	**72**	28
M	**88**	13
S	**83**	17

3.6.2 Impact of Flaws on Convergence of Results

The convergence of results for these "exposure duration" parameters is quite variable, with a modal percentage ranging from 32% to 96%. The parameter of tool 19, with the flaw "poor definition of levels," presents the lowest mean modal percentage, although it is slightly above 60%. As

Table 27 shows, most of the subjects did not choose the exposure duration reference level for scenarios A and G with this parameter. For scenario M, the convergence of results was better. As Table 25 shows, the information on exposure provided for this scenario indicates quite clearly that the exposure duration was relatively short (15 minutes every two days). In contrast, for the other three scenarios, the information on exposure is less clear-cut, and the lack of precision in the descriptions of the two levels of the parameter of tool 19 ("Seldom to quite often" and "Frequent to continuous";

Table 27) may have had the effect of making the choice more random.

Regarding the flaw "gaps between levels" affecting the parameter "exposure duration" of tool 62, it can be seen that the modal percentage is good, except for scenario A, for which it is only 32% (Table 28). This parameter has five levels, and gaps can be noted between adjacent levels. However, the very broad dispersion of the subjects' responses, as well as the low modal percentage for this scenario for all the parameters studied, seems to indicate that what influenced the choices was not the gaps, but the interpretation of the information about exposure duration provided for scenario A. For the parameter "exposure duration" of tool 91 (assumed to have "no flaws"), the convergence was not as good for scenario A.

3.6.3 Impact of Flaws on Level of Difficulty of Applying Parameter

For the two tools (19 and 62) having a flaw in the "exposure duration" parameter, at least 8 subjects said they had had trouble selecting the exposure duration level for scenarios A, G and M. For the parameter "exposure duration" of tool 19, the presence of the flaw "poor definition of levels" was noted 43 times in the 63 negative comments. Here are a few examples:

- *Descriptions are not detailed enough.*

- *Not enough categories. Poorly defined.*

- *Easy for this case, but no clear explanations.*

- *Provide better definitions.*

- *Seldom, frequent and continuous should be described.*

- *Frequency is hard to determine without greater detail (limits between the two?). The two levels overlap.*

For the parameter "exposure duration" of tool 62, despite the flaw "gaps between levels," very few subjects blamed this flaw in their assessments, even though a lot of negative comments were made. As Table 28 shows, the subjects' responses converged relatively well toward the reference level, except for scenario A. However, a reading of the subjects' comments suggests that many of them found it hard to take the information on exposure frequency and duration (provided for each scenario) and transform it mathematically into data that could be used to determine the exposure duration. Here are some examples:

- *Hard to calculate frequency and duration at the same time.*

- *Duration should be distinguished from frequency.*

- *Mixing exposure time and frequency makes it harder to choose.*

- *By calculation. Surprising result; it's the lowest exposure level, but the fact that the workers cross over the lane 25 times a day seems pretty high to me.*

- *Duration and frequency should be isolated and described so they can be evaluated better. We have to make an "editorial" choice between them.*

- *The scale is not suited to all situations. For instance, it talks about hours, but the situation talks about seconds and number of times. The definition of the parameter is reversed in relation to the level with respect to duration and frequency ... not clear.*

Regarding the above, it is interesting to note that "exposure duration" is the parameter that generated the most negative comments from subjects, with an average of 17 negative comments per application, as against 14 for all other parameters. In comparison, the parameter "exposure frequency" generated only 12 negative comments per application. In addition, where flaws came into play, the mean modal percentage of the parameter "exposure duration" was 69.3%, whereas it was 86.5% for the parameter "exposure frequency." This suggests that in risk estimation, the concept of exposure duration could be harder to grasp and less robust than the concept of exposure frequency.

3.6.4 Findings Regarding Impact of Flaws Associated with Parameter "Exposure Duration"

The observations and analyses presented in the preceding sections provide the basis for the following findings regarding the parameter "exposure duration":

1. As for the other parameters, the flaw "poor definition of levels" had a significant impact on this parameter, with a mean modal percentage of 67% and 43 negative comments implicating it directly.

2. The impact of the flaw "gaps between levels" could not be confirmed for this parameter.

3. As for the parameter "exposure frequency," a parameter level definition that is too complex has a significant impact on the degree of difficulty of the selection process. As the results for the "exposure duration" parameter of tool 62 showed, when a calculation is required to choose a certain level, such as to convert frequency data into exposure duration, many subjects identified this factor as being detrimental to the selection process.

4. The perceived level of difficulty, the number of negative comments made by the subjects and the mean modal percentage of the "exposure duration" parameters may be an indication that this type of parameter is potentially less robust and harder to use than the parameter "exposure frequency" in estimating risk. The small number of tests for this parameter suggests that it should be interpreted with caution.

3.7 Analysis of Flaws Associated with Parameter "Possibility of Avoidance (A)"

Table 30 presents an overview of the analysis of the parameter "possibility of avoidance."

Table 30. Analysis of the parameter "possibility of avoidance"

			Scenario A				Scenario G				Scenario M				Scenario S			
Possibility of avoidance (A)	**Tool**	**No. of levels**	**Modal percentage**	**No. of subjects who had trouble responding***	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**
Poor definition of levels	57	5	48	9	13	19	44	5	12	18	58	13	12	14	63	6	10	14
No flaws	114	4	64	4	0	11	52	3	0	8	67	10	0	11	58	4	0	5

*Number of subjects who said they found choosing "Fairly hard" or "Very hard" (subsection 2.3.1)

3.7.1 Description of Parameters and Flaws Studied

Only the flaw "poor definition of levels" was evaluated for the parameter "possibility of avoidance." Table 31 presents the results for the parameter "possibility of avoidance" of tool 57, which concerned this flaw. This parameter was selected because of the lack of information about choosing from among the five possibility-of-avoidance levels, which use only qualifiers that do not provide users with sufficient guidance.

Table 31. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "possibility of avoidance" of tool 57

Scenario	Parameter level				
	1. Obvious	2. Likely	3. Possible	4. Rarely	5. Impossible
A	0	12	36	**48**	4
G	20	**44**	32	4	0
M	0	**13**	58	29	0
S	0	0	4	**63**	33

Table 32 presents the results for the parameter "possibility of avoidance" of tool 114, selected as a parameter assumed to have "no flaws," in other words, it did not have any of the flaws specifically analysed within the framework of the study.

Table 32. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "possibility of avoidance" of tool 114

Scenario	Parameter level			
	Possible: for all exposed people	Possible if trained: possible for people trained to recognize warning and how best to react, and warning allows sufficient time	Difficult: possible, but warning may not be obvious or time is limited	Impossible: no warning and/or not enough time to react
A	12	16	**64**	8
G	32	**52**	16	0
M	13	**8**	67	13
S	0	4	**58**	38

3.7.2 Impact of Flaws on Convergence of Results

The convergence of results for the parameter “possibility of avoidance” of tool 57 (with the flaw “poor definition of levels”) is fairly low, with a mean modal percentage of 53%. That is also reflected in the correspondence between the subjects’ responses and the possibility-of-avoidance reference levels presented in Table 31. For scenario M in particular, it can be seen that only 13% of the subjects chose the reference level. It would therefore seem that the subjects had trouble distinguishing among the intermediate levels for this parameter, i.e., “2. Likely,” “3. Possible” and “4. Rarely.”

For the parameter “possibility of avoidance” of tool 114 (assumed to have “no flaws”), a noticeably better convergence of results can be seen, with a mean modal percentage of 60%. However, the same problem of correspondence with the reference level for scenario M can be noted.

3.7.3 Impact of Flaws on Level of Difficulty of Applying Parameter

As Table 30 shows, the flaw “poor definition of levels” in the parameter of tool 57 seems to have had the effect of making it harder for subjects to choose the levels. All in all, the subjects said they had trouble making their choice 33 times with this parameter, compared with 21 times for the corresponding parameter of tool 114, which was assumed to have no flaws. Similarly, subjects made a total of 47 comments about this flaw. Scenario M seems to have given subjects the most trouble with the two parameters.

3.7.4 Findings Regarding Impact of Flaws Associated with Parameter “Possibility of Avoidance”

The observations and analyses presented in the preceding sections provide the basis for the following finding regarding the parameter “exposure duration”:

> As for the other parameters, the flaw “poor definition of levels” had a significant impact on this parameter, with a mean modal percentage of 53% and 47 negative comments implicating it directly.

3.8 Analysis of Flaws Associated with Parameter “Probability of Occurrence of Hazardous Event (Pe)”

Table 33 presents an overview of the analysis of the parameter “probability of occurrence of the hazardous event.”

Table 33. Analysis of parameter “probability of occurrence of hazardous event”

			Scenario A				Scenario G				Scenario M				Scenario S			
Probability of occurrence of hazardous event (Pe)	**Tool**	**No. of levels**	**Modal percentage**	**No. of subjects who had trouble responding***	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**	**Modal percentage**	**No. of subjects who had trouble responding**	**No. of negative comments about flaw**	**No. of negative comments in total**
No flaws	19	3	56	7	0	12	68	7	0	14	79	7	0	14	83	4	0	10
Inconsistent definitions of different levels	62	5	40	4	11	12	36	6	12	14	58	6	10	10	38	8	7	8

*Number of subjects who said they found choosing “Fairly hard” or “Very hard” (subsection 2.3.1)

3.8.1 Description of Parameters and Flaws Studied

The tables below show the results of the subjects’ choices for the two parameters studied and for each of the four scenarios. They show that the mode corresponds to the reference level for the probability of occurrence of the hazardous event in only four out of eight cases.

Table 34 presents the results for the parameter “probability of occurrence of hazardous event” of tool 19, selected as a parameter assumed to have “no flaws,” in other words, it did not have any of the flaws specifically analysed within the framework of the study.

Table 34. Proportion of subjects (%) who selected a given level for each scenario, for the parameter “probability of occurrence of hazardous event” of tool 19

Scenario	Parameter level		
	1. Low – so unlikely that it can be assumed occurrence may not be experienced	2. Medium – likely to occur sometime in the life of an item	3. High – likely to occur frequently
A	0	**44**	56
G	4	**68**	28
M	0	**79**	21
S	**8**	83	8

Only the flaw “inconsistent definitions of different levels” could be evaluated for the parameter “probability of occurrence of hazardous event.” Table 35 presents the results for the parameter “probability of occurrence of hazardous event” of tool 62, which concerned this flaw. This parameter was selected owing to the lack of consistency in the explanations about the “measures taken”; the four first levels indicate, in a more or less similar way, that measures are taken to reduce the probability of occurrence of the hazardous event. It can be seen that the third level refers to "measures partially taken," while the fourth level says "measures have started to be taken."

Table 35. Proportion of subjects (%) who selected a given level for each scenario, for the parameter "probability of occurrence of hazardous event" of tool 62

Scenario	Parameter level				
	p0=1: event hard to imagine (measures consistent with state of the art)	p0=2: event imaginable, but unusual (measures taken)	p0=3: event is possible (measures partially taken, clear deficiencies)	p0=4: the event may be expected to occur (measures have started to be taken)	p0=5: the event should be expected to occur (no existing measures)
A	0	4	40	**20**	36
G	0	36	**36**	28	0
M	0	13	**58**	21	8
S	0	**29**	38	21	13

3.8.2 Impact of Flaws on Convergence of Results

As Table 33 shows, the convergence of results for the parameter "probability of occurrence of hazardous event" of tool 62, presenting the flaw "inconsistent definitions of different levels," is significantly lower than that for the corresponding parameter of tool 19, assumed to have no flaws. As Table 35 shows in detail, the subjects' responses were fairly widely dispersed for this parameter. With a mean modal percentage of 43%, it had the lowest rate of convergence of all the parameters submitted to the subjects as part of this study. This suggests that the inconsistency of the definitions involving the concept of "measures taken" to reduce the probability of occurrence of the hazardous event meant that the subjects had to deal with confusing concepts that resulted in broader dispersion of their responses. This finding would also seem to be supported by the subjects' comments.

3.8.3 Impact of Flaws on Level of Difficulty of Applying Parameter

Surprisingly, this low rate of convergence of the results of applying the parameter of tool 62 is not reflected much in the level of difficulty indicated by the subjects. On the other hand, the vast majority of the negative comments made (40 related comments out of 44 in total) do refer specifically to the flaw. Here are some examples:

- *I hesitated between 3 and 4. What do "measures have started to be taken" and "measures partially taken" mean? Seem like synonyms.*

- *Define "measures have started to be taken" and "measures partially taken," and at the same time define "measures."*

- *There is no clear distinction between level p03 and level p04 with respect to the start of incomplete measures.*
- *The () make me hesitate between 4 and 5. Went with the first part of the definition, which doesn't fit with the second part.*
- *Overlap between choices 3-4-5.*
- *Hard to distinguish between 2 and 3.*
- *I hesitated between 2 and 3; they're very similar.*
- *The concept of "measures taken" is not clear. The number of categories is good.*
- *Levels 2 to 5 are too similar.*

3.8.4 Findings Regarding Impact of Flaws Associated with Parameter "Probability of Occurrence of Hazardous Event"

The observations and analyses presented in the preceding sections provide the basis for the following finding regarding the parameter "probability of occurrence of hazardous event":

> As for the parameters "severity of harm" and "probability of occurrence of harm," the impact of the flaw "inconsistent definitions of different levels" seems significant. So, despite the fact that the subjects did not always indicate that this flaw may have made the selection process more difficult, they were still able to recognize the presence and potential impact of the flaw on the parameter "probability of occurrence of hazardous event."

4. RESULTS – ANALYSIS OF APPLICATION OF RISK ESTIMATION TOOLS

4.1 Tool 19

4.1.1 Performance of Tool in Classifying Scenarios

The scenario classification results achieved with tool 19 are shown in Table 36.

Table 36. Scenario classification obtained with tool 19 compared with reference classification

	Number of occurrences
Classification comparable to reference classification	15
Cases where 3 or 4 scenarios estimated to have same risk level	4
Cases of reversal of scenario risk levels in relation to reference order	6

Table 36 shows that four subjects arrived at classifications that were not determined. It also shows that six subjects reversed the order of one or more scenarios in relation to the reference order (e.g., estimating scenario S to be of lower risk than scenario M). The other 15 subjects obtained a scenario classification comparable to the reference classification. However, in terms of distinguishing between scenarios with different risk levels, tool 19 had trouble characterizing scenarios of low or mid-low risk. Of the 15 subjects who arrived at a classification comparable to the reference order, 13 were unable to distinguish between scenarios A and G (17 of the 25 subjects in total). On the other hand, mid-high and high risk scenarios were characterized properly. This was the case, in particular, for scenario S, for which 18 of the 25 subjects arrived at a distinctly higher risk level than for the other scenarios. In addition, 11 subjects clearly distinguished between scenarios M and S, in that order. In comparison, only two subjects arrived at the strictly lowest risk level for scenario A.

4.1.2 Convergence and User Satisfaction

Table 37 presents a summary of the data regarding the analysis of convergence and user satisfaction with tool 19.

Table 37. Summary of analytical data for tool 19

Scenario	**Satisfaction**				**Convergence**
	Number of subjects who disagreed with risk level obtained*	Number of negative comments			Modal percentage (tool reduced to 4 levels)
		Level too high	Level too low	Total	
A	6	3	4	7	52
G	6	3	6	9	64
M	10	0	11	11	92
S	7	5	5	10	44
Total	29	11	26	37	63

*Number of subjects who answered "Somewhat disagree" or "Totally disagree" (subsection 2.3.2.2)

The overall convergence for tool 19 is close to the critical threshold of 60%. It is below the threshold for the low- and high-risk scenarios (A and S, respectively). Intersubject repeatability for scenario M, in contrast, was high (92%).

Based on the number of subjects who disagreed with the risk level arrived at and the number of negative comments, the subjects were not satisfied with tool 19 for mid-high- and high-risk scenarios (M and S, respectively). For scenario M, 11 of the 25 subjects said they thought the risk level arrived at with this tool was too low. The 10 comments regarding the results for scenario S were divided. This result is related to the low modal percentage. It is worth noting that the scenario which had the best convergence of results was also the one that generated the most dissatisfaction.

4.1.3 Main Findings Regarding Tool 19

The observations and analyses presented in the preceding sections point to the following problems with tool 19:

- Difficulty identifying low- or mid-low risk scenarios
- Relatively low level of user satisfaction for mid-high-risk scenario (level obtained too low)
- Low modal percentage for high-risk scenario, related to divided comments about the risk level arrived at (too low or too high)

4.2 Tool 24

4.2.1 Performance of Tool in Classifying Scenarios

The scenario classification results achieved with tool 24 are shown in Table 38. The classification was undetermined for 5 subjects, while 10 reversed the order of one or more scenarios in relation to the reference order. Twelve subjects arrived at a scenario classification comparable to the reference classification. That represents less than half of the subjects.

Table 38. Scenario classification obtained with tool 24 compared with reference classification

	Number of occurrences
Classification comparable to reference classification	12
Cases where 3 or 4 scenarios estimated to have same risk level	5
Cases of reversal of scenario risk levels in relation to reference order	10

Regarding the tool's ability to distinguish between scenarios with different risk levels, 42 of the subject/scenario pairs were strictly distinguished. This indicates that in 58% of cases, the risk level could not be distinguished for two (22 times) or three (5 times) scenarios.

4.2.2 Convergence and User Satisfaction

Table 39 presents a summary of the data regarding the analysis of convergence and user satisfaction with tool 24.

Table 39. Summary of analytical data for tool 24

Scenario	Satisfaction				Convergence
	Number of subjects who disagreed with risk level obtained	Number of negative comments			Modal percentage (tool reduced to 4 levels)
		Level too high	Level too low	Total	
A	6	2	4	6	48
G	2	2	0	2	48
M	5	3	0	3	76
S	2	2	0	2	68
Total	15	9	4	13	60

Tool 24 did not exceed any critical threshold for the user satisfaction criterion. Nine of the 13 comments mentioned a calculated risk level that was too high (across the four scenarios), while four felt that the calculated risk level for scenario A was too low.

Unlike the satisfaction criterion, which was respected (fewer than seven subjects disagreed with the risk level obtained), the overall convergence of tool 24 only just managed to reach the critical threshold (set at 60%). In particular, slightly less than half the users arrived at the same risk level for scenarios A and G.

4.2.3 Main Findings Regarding Tool 24

The observations and analyses presented in the preceding sections support the following findings regarding tool 24:

- Difficulty classifying scenarios (four A/G reversals, four M/S reversals, in particular)
- Good user satisfaction with results obtained
- Low convergence for scenarios A and G

4.3 Tool 69

4.3.1 Performance of Tool in Classifying Scenarios

The scenario classification results achieved with tool 69 are shown in Table 40.

Table 40. Scenario classification obtained with tool 69 compared with reference classification

	Number of occurrences
Classification comparable to reference classification	20
Cases where 3 or 4 scenarios estimated to have same risk level	0
Cases of reversal of scenario risk levels in relation to reference order	5

Tool 69 stands out from the others in its ability to distinguish between scenarios with different risk levels and categorize them according to the reference classification. Twenty of the 25 subjects achieved a classification comparable to the reference classification. Twelve subjects arrived at the reference classification exactly. In addition, having accurately identified 76% of the subject/scenario pairs, this tool was the only one to differentiate among the scenarios in a majority of cases. Twelve times, two scenarios were given the same risk level. Each time, it was either the scenarios A and G, or the scenarios M and S, which were not distinguished properly. With this tool, the greatest number of subjects arrived at a classification identical or close to the reference order.

The good performance of tool 69 in terms of distinguishing between scenarios, regardless of the classification, may be related to the fact that it has 11 risk levels.

4.3.2 Convergence and User Satisfaction

Table 41 presents a summary of the data regarding the analysis of convergence and user satisfaction with tool 69. The convergence of results obtained with this tool is 72% for scenarios

A, M and S. On the other hand, for scenario G it is just 56%, which is below the critical threshold.

The scenarios with low and mid-low risk levels (A and G, respectively) gave rise to user dissatisfaction. These two cases reached or exceeded the critical thresholds for the number of subjects who disagreed with the risk level obtained and for the number of negative comments. In contrast, the scenarios with mid-high and high risk levels did not seem to cause the subjects any problems, as no thresholds were reached for scenarios M and S.

Table 41. Summary of analytical data for tool 69

Scenario	**Satisfaction**				**Convergence**
	Number of subjects who disagreed with risk level obtained	Number of negative comments			Modal percentage (tool reduced to 4 levels)
		Level too high	Level too low	Total	
A	15	0	21	21	72
G	10	2	8	10	56
M	5	1	2	3	72
S	5	4	2	6	72
Total	35	7	33	40	68

4.3.3 Main Findings Regarding Tool 69

The observations and analyses presented in the preceding sections support the following findings regarding tool 69:

- Very good performance in terms of scenario classification
- Very good at distinguishing between scenarios, probably due to the fact that it has 11 risk levels
- High level of user dissatisfaction for low or mid-low risk scenarios, with the comments indicating that the risk level obtained was too low. That is probably due to the fact that the lowest risk level is rated “zero” on the tool’s scale
- Convergence is relatively good, with the exception of scenario G

4.4 Tool 89

4.4.1 Performance of Tool in Classifying Scenarios

The scenario classification results achieved with tool 89 are shown in Table 42. Nearly half of the subjects arrived at a classification different from the reference order (9) or an undetermined order (4). The other 12 subjects obtained a classification in accordance with the expanded reference order. Tool 89 therefore did not perform very well when it came to classifying the scenarios.

Table 42. Scenario classification obtained with tool 89 compared with reference classification

	Number of occurrences
Classification comparable to reference classification	14
Cases where 3 or 4 scenarios estimated to have same risk level	9
Cases of reversal of scenario risk levels in relation to reference order	4

Regarding the tool's ability to distinguish between scenarios with different risk levels, 35 of the subject/scenario pairs were clearly distinguished. This indicates that in 65% of cases, the risk level could not be distinguished between two (26 times) or among three (3 times) or even four (1 time) scenarios.

4.4.2 Convergence and User Satisfaction

Table 43 presents a summary of the data regarding the analysis of convergence and user satisfaction with tool 89.

Table 43. Summary of analytical data for tool 89

	Satisfaction				**Convergence**
	Number of subjects who disagreed with risk level obtained	Number of negative comments			Modal percentage (tool reduced to 4 levels)
Scenario		Level too high	Level too low	Total	
A	7	5	0	5	64
G	6	5	2	7	48
M	2	3	0	3	64
S	6	2	6	8	52
Total	21	15	8	23	57

Tool 89 met all the user satisfaction criteria. Seven subjects did not agree with the risk level arrived at for scenario A, but only 5 negative comments were received. Users therefore seemed generally satisfied with the tool.

Although the critical threshold for the number of negative comments was never reached, there were still 23 such comments in all (out of a possible 100), which represents almost a quarter. There were enough to indicate that the nature of them was significant: 15 out of the 23 negative comments were to the effect that the risk level obtained with the tool was too high.

The convergence criterion was close to or below the critical threshold for each scenario. As a result, the mean convergence for the four scenarios did not meet the convergence criterion.

4.4.3 Main Findings Regarding Tool 89

The observations and analyses presented in the preceding sections support the following findings regarding tool 89:

- Middling performance in terms of classifying scenarios
- Good user satisfaction, but convergence of results was low
- Comments indicate that risk levels obtained were too high

4.5 Tool 91

4.5.1 Performance of Tool in Classifying Scenarios

The scenario classification results achieved with tool 91 are shown in Table 44.

Table 44. Scenario classification obtained with tool 91 compared with reference classification

	Number of occurrences
Classification comparable to reference classification	10
Cases where 3 or 4 scenarios estimated to have same risk level	3
Cases of reversal of scenario risk levels in relation to reference order	12

Tool 91 performed poorly in terms of scenario classification. Only 10 subjects came up with a classification comparable to the reference classification. In addition, it was the tool that produced the most risk level reversals (12) in relation to the reference order.

Note that scenario A (with the lowest reference risk level) was always classified at the low or mid-low risk level and that scenario S (with the highest reference risk level) was always classified at the mid-high or high level. However, the scenarios with an intermediate risk level were harder to situate using tool 91. Scenario G ended up in all four possible positions, while scenario M was in three different positions (but never at the lowest risk level of the four scenarios).

Regarding the tool's ability to distinguish between scenarios with different risk levels, 39 of the subject/scenario pairs were accurately distinguished. This indicates that in 61% of cases, the risk level could not be distinguished for two (26 times) or three (3 times) scenarios.

4.5.2 Convergence and User Satisfaction

Table 45 presents a summary of the data regarding the analysis of convergence and user satisfaction with tool 91.

Table 45. Summary of analytical data for tool 91

	Satisfaction				**Convergence**
	Number of subjects who disagreed with risk level obtained	Number of negative comments			Modal percentage (tool reduced to 4 levels)
Scenario		Level too high	Level too low	Total	
A	8	1	7	8	80
G	12	3	3	6	44
M	13	0	15	15	76
S	7	0	8	8	52
Total	40	4	33	37	63

Users were more dissatisfied with tool 91 than with any of the other tools, considering the number of subjects who disagreed with the risk level obtained (40). It prompted 37 negative comments in total. Note, in particular, that this tool seemed to have trouble with the mid-high risk scenario (scenario M), as it exceeded the critical thresholds of the two satisfaction indexes and was the reason for close to half of the negative comments. In addition, all the negative comments concerning this scenario made the same point: the calculated risk level was too low. This tendency to underestimate seems to have been independent of the scenario, according to 33 of the 37 comments in all that mention this.

Convergence was good for scenario A (low risk level) and, surprisingly, for scenario M (mid-high risk level). In contrast, the convergence criterion was not met for scenario G (mid-low risk level) or scenario S (high risk level).

4.5.3 Main Findings Regarding Tool 91

The observations and analyses presented in the preceding sections support the following findings regarding tool 91:

- Poor performance in terms of scenario classification
- Difficulty distinguishing between low and mid-low risk scenarios
- Good convergence for scenarios A and M, but poor for scenarios G and S
- Deep user dissatisfaction stemming, according to the comments, from a tendency to underestimate

4.6 Tool 114

4.6.1 Performance of Tool in Classifying Scenarios

The scenario classification results achieved with tool 114 are shown in Table 46.

Table 46. Scenario classification obtained with tool 114 compared with reference classification

	Number of occurrences
Classification comparable to reference classification	15
Cases where 3 or 4 scenarios estimated to have same risk level	3
Cases of reversal of scenario risk levels in relation to reference order	7

The risk level could not be distinguished in 77% of cases for two (34 times) or three (3 times) scenarios. Note that it seems to be hard, with this tool, to distinguish between low and mid-low risk scenarios, as well as between mid-high and high risk ones. However, despite the poor rates of scenario distinction, tool 114 did perform relatively well in classifying scenarios with different risk levels, compared with the other tools. Given that 15 of the 25 subjects arrived at a scenario classification comparable to the reference classification, it ranked joint second, tied with tool 19.

4.6.2 Convergence and User Satisfaction

Table 47 presents a summary of the data regarding the analysis of convergence and user satisfaction with tool 114.

Table 47. Summary of analytical data for tool 114

Scenario	Satisfaction				Convergence
	Number of subjects who disagreed with risk level obtained	Number of negative comments			Modal percentage (tool reduced to 4 levels)
		Level too high	Level too low	Total	
A	6	5	1	6	80
S	5	5	2	7	60
M	5	6	0	6	76
S	6	3	0	3	96
Total	22	19	3	22	78

Tool 114 met all the user satisfaction criteria. However, 19 of the 22 negative comments made by the subjects who disagreed with the risk level calculated by the tool indicated that, in their opinion, the level was too high.

Convergence was also good for tool 114, with a mean of 78%. The critical threshold was just achieved, with a modal percentage of 60% for scenario G (mid-low risk level).

4.6.3 Main Findings Regarding Tool 114

The observations and analyses presented in the preceding sections support the following findings regarding tool 114:

- Middling performance in terms of classifying scenarios
- Very low rate of distinguishing between low and mid-low risk scenarios, as well as between mid-high and high risk ones
- Good user satisfaction and good convergence (borderline for scenario G)

The comments indicated that the risk level calculated by the tool was too high, contrary to what was found in the earlier study (tendency to underestimate).

5. DISCUSSION

The main objective of this study was to conduct hands-on experiments in order to confirm the real and perceived impact of flaws and biases in risk estimation tools and the configuration of their parameters. The ultimate goal of this advance in knowledge is to be able to propose robust, reliable configurations for risk estimation tools and define enlightened criteria for the evaluation of existing tools and the development of specific new ones.

With respect to risk estimation parameters, five of the potential flaws established in the preceding instalment of this research program were analysed:

1. Poor definition of levels
2. Inconsistent definitions of different levels
3. Inadequate number of levels
4. Gaps between levels
5. Lack of exposure interval

With respect to risk estimation tool architecture, four of the potential flaws established in the preceding instalment of this research program were analysed:

1. Non-standard configuration
2. Non-uniform distribution
3. Discontinuity in risk levels
4. Excessive relative weight given to one parameter

The main findings arrived at in this study are assessed in the following sections.

5.1 Impact of Flaws on Risk Estimation Process

The first general finding concerns the subjects' ability to recognize the flaws in the construction of the risk estimation parameters. The experimental results show that when subjects have trouble applying a parameter, they are generally able to associate it with a flaw in the parameter. This finding is supported by the strong correlation between the number of negative comments and the level of difficulty indicated (see Figure 3).

On the other hand, subjects' perception of the difficulty of applying different parameters is not well correlated with the convergence of results (see Figure 4). This means that even when the level of difficulty perceived is low, convergence in relation to the levels chosen by the subjects may sometimes be poor. Then the subjects choose different levels depending on their individual understanding of the relationship between the construction of the parameter and its levels, and the hazardous situations. This finding may be an indication that the presence of some flaws, in certain parameters and certain cases, may influence, or even bias, the process whereby the subject selects a level, without that influence being perceived negatively by the subject. Some flaws may therefore sometimes influence the risk estimation process without users realizing it.

The results also indicate fairly clearly that the impact of parameter construction flaws is not uniform. Significant variations can be seen, not only by type of parameter, but also by whatever hazardous situation scenario is analysed.

For the types of parameters, those concerning the severity of harm seem relatively robust and yield a solid consensus among users regarding the level of severity of the potential harm involved in a situation, despite the presence of flaws. That does not, however, rule out the fact that the subjects may recognize the flaws, as 74% of all the negative comments from subjects concerned the various flaws. On the other hand, the probability parameters (probability of occurrence of harm and probability of occurrence of the hazardous event) are significantly less robust, which generally means poorer convergence. Besides the flaws in these parameters, the results obtained suggest that evaluating probability is a difficult aspect of risk estimation that requires special attention. Similarly, the experimental results suggest that the parameter “exposure duration” is more sensitive to flaws and harder to apply than the parameter “exposure frequency” in risk estimation.

The nature of the flaw, its position on the parameter scale and the scenario in question also influence its impact on the determination of the level of a parameter. This may reveal itself as much in the convergence of results as in the level of difficulty or the number of negative comments associated with it. For example, some flaws have no impact on determining the severity of harm when the potential harm is very high, but they can have a greater impact on it when the potential harm is low or medium. The highly variable impact of the flaw “gaps between levels” also provided evidence for this finding. That could explain the behaviour of some risk estimation tools analysed in the earlier study, which had very clearly overestimated or underestimated the risk of some of the 20 hazardous situation scenarios (Chinniah et al., 2011).

5.1.1 Impact of Flaw “Poor Definition of Levels”

The flaw “poor definition of levels” is the one that had the greatest impact on the choice of risk estimation parameter levels. This flaw was evaluated on the basis of four types of parameters:

- Severity of harm (S) (two statements evaluated)
- Probability of occurrence of harm (Ph)
- Exposure duration (Exd)
- Possibility of avoidance (A)

The results obtained for this flaw were:

- A mean modal percentage of 63% (compared with 65% for all flaws)
- A mean of 7.4 subjects who indicated a difficulty level of 4 or 5 on each application (compared with 5.2 for all flaws considered together)
- A mean of 11.5 negative comments in connection with the flaw, for each application (compared with 6.9 for all flaws taken together)
- A correspondence of the mode with the reference level 16 times out of 20

In light of the analysis of these experimental results, it can be confirmed that this flaw does have an impact on the risk estimation process. Both the quantitative results and the qualitative analysis of the subjects' comments confirm the significant negative effect of the flaw. Its influence is clear for the four types of parameters with which it was evaluated.

Thus, as advanced in the earlier study in this research program, the parameter levels must be defined appropriately in order to make the risk estimation process more robust. Some tools use only figurative terms or expressions to define the various levels of their parameters (e.g., "possible" or "probable" for the parameter Ph, or "seldom to quite often" for the parameter Exf). The drawback of figurative terms and expressions is that they leave a lot of room for user interpretation. Does "possible" have the same meaning for all users? What is meant exactly by "quite often"? Given the lack of precision in these terms, anyone who uses a tool of this kind may interpret each level differently from someone else. This interpretation problem is mitigated when detailed definitions are given. Employed jointly with figurative terms or expressions, detailed definitions can provide users with a better analytical structure, reducing the problems they face and promoting greater convergence of risk estimation results. These observations confirm what other authors have pointed out (Carey and Burgman, 2008; Christensen et al., 2003; Cox, 2008; Patt and Schrag, 2003).

5.1.2 Impact of Flaw "Inconsistent Definitions of Different Levels"

The flaw "inconsistent definition of different levels" also had a major impact on subjects' choices of risk estimation parameter levels. This flaw was evaluated with three types of parameters:

- Severity of harm (S)
- Probability of occurrence of harm (Ph)
- Probability of occurrence of hazardous event (Pe)

The results obtained for this flaw were:

- A mean modal percentage of 55% (compared with 65% for all flaws)
- A mean of 6.2 subjects who indicated a difficulty level of 4 or 5 on each application (compared with 5.2 for all flaws considered together)
- A mean of 10.2 negative comments in connection with the flaw, for each application (compared with 6.9 for all flaws taken together)
- A correspondence of the mode with the reference level 8 times out of 12

As for the flaw "poor definition of levels," the experimental results confirm that the flaw "inconsistent definitions of different levels" also had a major impact on subjects' choices of levels for the risk estimation parameters. The impact of an inconsistency in level definitions was observed on the three types of parameters for which it was evaluated. With respect to modal percentage (intersubject convergence), this flaw had the worst score (55%), having an effect even on the parameter "severity of harm," which the other flaws had relatively little influence on. It also had an effect in particular on the parameter "probability of occurrence of hazardous event (Pe)," with a modal percentage of 43% and 40 negative comments. Qualitative analysis of subjects' comments also confirmed the major negative impact of inconsistency in parameter level definitions on the risk estimation process.

Thus, as advanced in the earlier study in the research program, although they define each of their levels in a relatively detailed fashion, some risk estimation parameters use terms that are inappropriate, confusing or too similar semantically. The examples of the parameters drawn from tools 6, 62 and 66 are good illustrations of the various forms that this construction flaw can take.

To avoid confusing some users, full, precise definitions need to be provided, so that there is no ambiguity about which levels are which. The consistency of the terms used to classify the levels defined by a given parameter is also important (Curry and Burgman, 2008; Christensen et al., 2003; Theil, 2002; Willquist and Torner, 2003; Mosteller and Youtz, 1990). To take an example, using the expression "minor injury" in the first two levels of the scale of the parameter "severity of harm" of tool 66 should be avoided. Whatever the parameter, its levels should progress from the lowest to the highest, and the terms used should reflect this progression so that users can clearly distinguish between levels and select the one that corresponds to the risk situation they are estimating.

5.1.3 Impact of Flaw "Inadequate Number of Levels"

This flaw was evaluated with a single type of parameter, i.e., "severity of harm (S)." The results obtained were:

- A mean modal percentage of 90% (compared with 65% for all flaws)

- A mean of 2.3 subjects who indicated a difficulty level of 4 or 5 on each application (compared with 5.2 for all flaws considered together)

- A mean of 5.8 negative comments in connection with the flaw, for each application (compared with 6.9 for all flaws taken together).

The experimental results suggest that the use of just two levels to establish the severity of harm makes it easier for users to choose, but that users were sometimes uncomfortable making the choice and felt as though they were being forced into it. For tool 91 that the experiments were conducted with, the binary nature of the definition of potential harm, based on whether the harm is reversible or not, seemed to leave a number of subjects ill at ease, to judge by their comments. With this type of parameter, an irreversible injury (e.g., loss of a fingertip) is considered to be on the same level as a worker fatality. In some cases, a user might not feel comfortable choosing the right level. When the number of thresholds of a given parameter is inadequate, some thresholds tend to cover too many different, if not extreme, situations.

So, while the small number of experimental results is not sufficient to draw conclusions with any certainty, questions need to be raised about the impact of this type of parameter construction on the perception and functioning of the risk estimation process. As a result, the recommendation from the earlier study to the effect that risk estimation parameters should normally have from three to five levels is a cautious approach that is worth following. Standard ISO14121-2 (2007) indicates that parameters should have a minimum and a maximum number of levels, but does not specify what that number should be.

However, it is also possible that this characteristic does not constitute a flaw for all types of parameters. An earlier study showed that the parameter “possibility of avoidance” has only two levels in 73% of cases (Packets et al., 2005a).

5.1.4 Impact of Flaw “Gaps Between Levels”

This flaw was evaluated on the basis of four types of parameters:

- Severity of harm (S)

- Probability of occurrence of harm (Ph)

- Exposure frequency (Exf)

- Exposure duration (Exd)

The results obtained for this flaw were:

- A mean modal percentage of 74% (compared with 65% for all flaws)

- A mean of 5.9 subjects who indicated a difficulty level of 4 or 5 on each application (compared with 5.2 for all flaws considered together)
- A mean of 6.0 comments in connection with the flaw, for each application (compared with 6.9 for all flaws taken together)
- A correspondence of the mode with the reference level 13 times out of 16

The experimental results suggest that the impact of the flaw "gaps between levels" is variable. It can have virtually no influence on convergence (modal percentage) when the gap is very large; it leads or even forces subjects to choose a specific level in each case. It is then perceived in the same way as the flaw "inadequate number of levels" and makes users feel ill at ease when they are selecting their levels. The gaps between levels may be more significant when a parameter's scale only has a few levels. However, the flaw "gaps between levels" may also be found in a parameter having several levels, as is the case for the parameter "exposure duration" of tool 62, which has five.

The position of the flaw on the parameter's scale, combined with information about the scenario under consideration, can also affect its impact. When a scenario corresponds to a situation that concerns a gap in a parameter's scale, the perceived level of difficulty and the subjects' comments reflect its impact. If that is not the case, then the flaw may have no impact whatsoever.

It can therefore be concluded that this flaw can have an impact on the risk estimation process, but whether it will be big or small depends on the circumstances. Being aware of the flaw is important, however, as it can give rise to underestimations or overestimations for certain specific scenarios. In addition, while a significant gap between two levels may favour better convergence of results, the gap can make users feel quite uncomfortable, which can harm the perception of the tool and the smooth functioning of the risk estimation process.

5.1.5 Impact of Flaw "Lack of Exposure Interval"

This flaw was evaluated with only one type of parameter, i.e., "probability of occurrence of harm (Ph)."

The results were:

- A mean modal percentage of 47% (compared with 65% for all flaws)
- A mean of 4.3 subjects who indicated a difficulty level of 4 or 5 on each application (compared with 5.2 for all flaws considered together)
- A mean of 1.0 negative comments in connection with the flaw, for each application (compared with 6.9 for all flaws taken together)

For the reasons given earlier (see subsections 3.4.1 and 3.4.2), it is hard to draw any conclusions about the impact of the flaws affecting the "probability of occurrence of harm." First of all, of all the probability-of-occurrence-of-harm statements studied, only the one from tool 41 (assumed to have "no flaws") included an indication about the exposure interval. Second, the convergence of results is poor in all cases. In addition, several subjects said the concept of "life cycle," used to establish the exposure interval in the parameter statement of tool 41, "bothered" them.

These results may indicate:

- That the concept of "life cycle," used to establish the exposure interval, is not precise enough to be useful
- That users prefer to define the exposure interval themselves, qualitatively, when estimating the probability of occurrence of harm
- Or that the lack of an exposure interval is a flaw that may influence, or even bias, the process whereby subjects select a level, without that influence being perceived negatively by them

This last hypothesis could be one of the explanations for the very low convergence (modal percentage) of the results for the parameter "probability of occurrence of harm." In connection with this hypothesis, an exhaustive analysis of 412 comments from subjects (for all statements of the parameter "probability of occurrence of harm" applied to the four scenarios) turned up only three comments about the lack of indication of an exposure interval. It is clear that the subjects did not perceive this lack of information as a factor influencing the risk estimation process.

Further research will therefore be needed to gain a better understanding of the reasoning that leads to a qualitative estimation of the probability of occurrence of harm (Carey and Burgman, 2008; Christensen et al., 2003).

5.1.6 Others Flaws and Biases

For the two types of exposure parameters studied (frequency and duration), a parameter level definition that is too complex had a significant impact on the convergence of results and on the degree of difficulty of the selection process. As the results for the "exposure frequency" parameter of tool 67 and the "exposure duration" parameter of tool 62 showed, when a calculation was required to choose a certain level, many subjects clearly identified this factor as detrimental to the selection process. It would therefore seem that this characteristic is an additional flaw that can affect the choice of a level for the "exposure frequency" parameter and potentially other risk estimation parameters.

5.2 Impact of Risk Estimation Tool Architecture That Fails to Follow Construction Rules

Through identifying the origin of the problems observed in the use of the six risk estimation tools, the impact of the failure to follow certain tool architecture construction rules is discussed in this subsection.

5.2.1 Tool 19

The level of satisfaction was fairly low for the mid-high risk level scenario (scenario M). The subjects arrived at a medium risk level (level 2 out of 4), although their comments suggested they thought they should have obtained a high risk (level 3 out of 4). A close look at the parameter choices shows that the subjects were consistent: S2 (serious – irreversible injury), Ex1 (seldom to quite often), Pe2 (medium) and A1 (possible under specific conditions). The modal percentage for this scenario was high (92%). The dissatisfaction problem is therefore related to the breakdown of risk levels, as choosing a severity level of S2 (serious – irreversible injury) leads in two thirds of cases to a low or medium risk level. This problem is an instance of the flaw "non-uniform distribution of risk levels." For this tool, 21 parameter combinations out of 36 (58%) lead to the two lowest risk levels.

The problem identifying low and mid-low risk scenarios is another indication of the non-uniform distribution of risk levels. When the lowest severity level is chosen, as for scenarios A and G, the frequency of exposure and the possibility of avoidance do not influence the risk level obtained: 8 of the 12 possible combinations lead to low risk.

The low modal percentage for the high-risk scenario (S) stems from the divergence of the choices made with respect to severity of harm and exposure frequency. These parameters have flaws which became apparent with scenario S:

- Severity of harm: confusion between worst harm and most likely harm. As the difference is not specified in the definition, the subjects chose different severity levels (e.g., 2. irreversible or 3. death)

- Exposure frequency: inadequate number of levels (two levels) and poor definition of levels (seldom to quite often; frequent to continuous)

These flaws are found on the first two parameters of the risk graph, and these parameters have the greatest relative weight on the result. As Figure 5 shows, a one-level change in severity, while the other parameters are kept unchanged (e.g., S2/Ex1/Pe3/A1 compared with S3/Ex1/Pe3/A1), can cause the risk level to jump from 2 (medium) to 4 (extremely high). The combined effect of the flaws in these parameters and their influence explains the divergence in the risk level obtained (the subjects obtained all four risk levels).

Tool 19 thus illustrates the negative impact that the following flaws can have:

– An architecture that gives an excessive relative weight to one or more parameters (e.g., first parameter in a graph) and that features a discontinuity in risk levels.

  The impact (e.g., divergent results) occurs in particular when there are construction flaws in the parameter that has the most influence on the result (e.g., poor definition of levels, inadequate number of levels).

– An architecture that does not lead to a uniform distribution of risk levels.

  That leads to dissatisfaction with the results obtained and to problems distinguishing between scenarios.

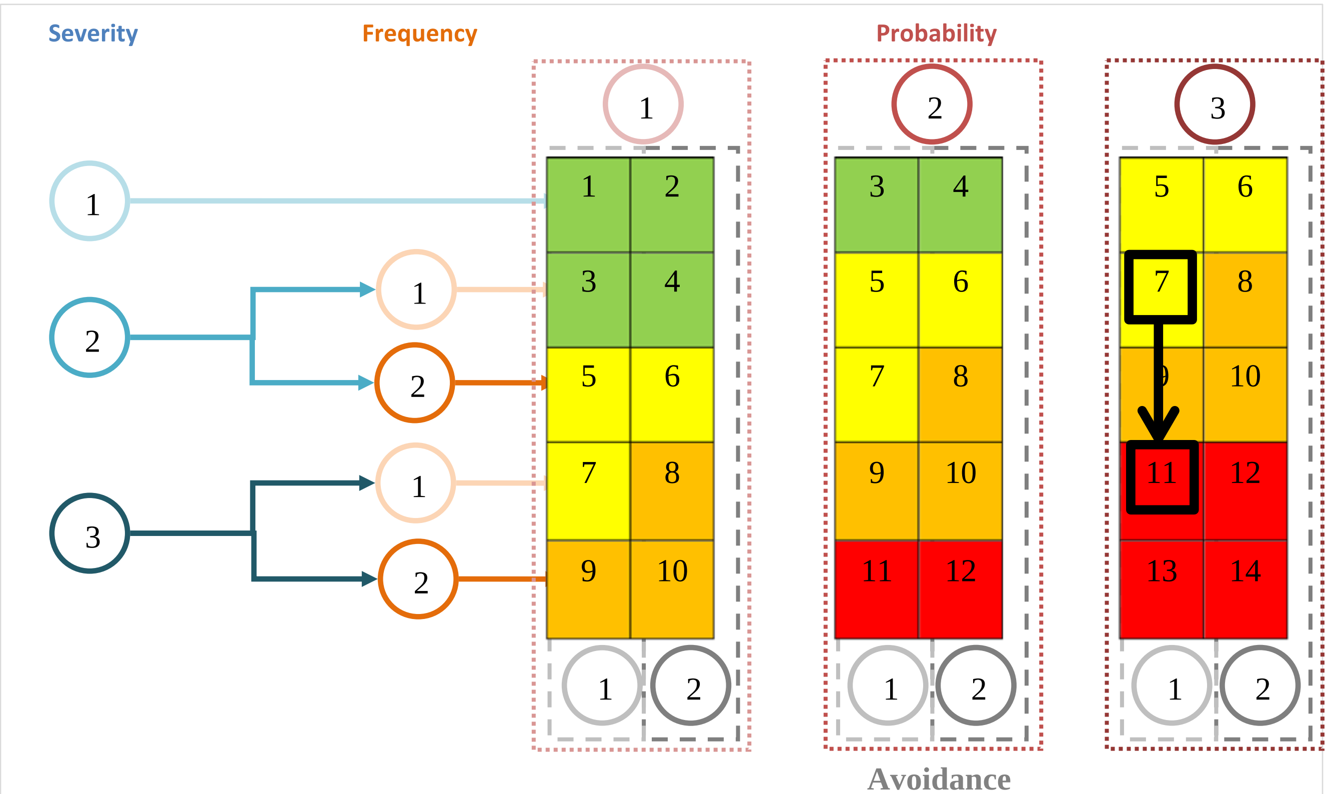

Figure 5. Illustration of the impact of a change in level for the parameter "severity of harm" on the risk level obtained with tool 19

5.2.2 Tool 24

The low modal percentages obtained with tool 24 for scenarios A and G (48%) and the difficulty the tool had in classifying them in the right order stem from the poor convergence of results for the probability of the occurrence of harm. The modal percentage for the probability of the occurrence of harm in these scenarios was approximately 50%. The divergence is attributable to two flaws in this parameter: “poor definition of levels” (e.g., word and associated definition not always appropriate) and “gaps between levels” (2. unlikely and 3. likely) The impact of these flaws can be seen in the risk levels obtained, as a two-parameter matrix with 16 possible combinations for 4 risk levels is sensitive to the choice of level of a parameter.

Tool 24 therefore illustrates the negative impact that a flawed parameter can have in a sensitive matrix, especially respecting divergent results for the risk level obtained and scenario classification.

5.2.3 Tool 69

Strong dissatisfaction was expressed with tool 69 for low and mid-low risk scenarios. Subjects’ comments suggested the risk level obtained with the tool was lower than their perception of the situation. This problem may be due to two things:

- The lowest risk level is designated “0.” The subjects didn’t like this designation, as for them zero risk does not exist.

- The non-uniform distribution of the matrix, with close to 50% of the parameter combinations (23/48) giving a risk of 0/11 or 1/11. In addition, when a low risk level is chosen, the choice of exposure parameter has no influence. The same thing was observed for tool 19.

The lower modal percentage for scenario G can be explained chiefly by the combined problems in the choice of severity of harm and frequency of exposure, in contrast with the other scenarios. These parameters have flaws that became apparent with scenario G:

- Severity of harm: inconsistent definitions of different levels, with groups of words that are repeated for several levels (e.g., trivial harm with no permanent results; serious harm with no permanent results).

- Exposure to harm: inadequate number of levels (i.e., only two)

The flaws concern the first two parameters of the risk graph. These parameters have a greater relative weight on the result. A change of one level in severity or exposure to harm, without changing the levels of the other parameters (e.g., S2/Ex2/A1/Pe2 compared with S3/Ex2/A1/Pe2), can cause the risk level to rise from 0 to 4. The combined effect of the parameter flaws and their influence is what explains the divergent risk level results obtained for scenario G.

Tool 69 thus illustrates the negative impact that the following flaws can have:

- An architecture that gives greater influence to one or more parameters (e.g., first parameter on a graph) and that is characterized by a discontinuity in risk levels.

The impact (e.g., divergent results) occurs in particular when there are construction flaws in connection with parameters that have a greater relative weight on the result (e.g., poor definition of levels, inadequate number of levels).

- An architecture with a non-uniform distribution of risk levels.

  That leads to dissatisfaction with the results obtained.

- Designating a risk level as “0” (i.e., subjects dissatisfied because zero risk does not exist).

5.2.4 Tool 89

The low modal percentages obtained with tool 89 (mean of 57%) and the difficulty the tool had in classifying the scenarios in the right order stem from the poor convergence of subjects’ choices for the probability of the occurrence of harm. The modal percentage for this probability was 48% for scenario A, 56% for scenario G, 64% for scenario M and 44% for scenario S. According to the subjects’ comments, the definitions of the levels of this parameter are not clear enough (see Table 21), which could explain the divergence. It is reasonable to surmise that this parameter, initially assumed to have “no flaws,” may in fact be affected by the flaw “poor definition of levels.” In addition, this divergence in the choice of the probability parameter is not mitigated by the matrix of tool 89. With only 2 parameters and 12 possible combinations for 6 risk levels, the risk level obtained is sensitive to any change in the level of a parameter.

Tool 89 therefore illustrates the negative impact that a flawed parameter can have in a matrix sensitive to the slightest change in a parameter level, especially concerning divergent results for the risk level obtained and scenario classification.

5.2.5 Tool 91

The subjects had trouble identifying low risk (A) and mid-low risk (G) scenarios using tool 91. This problem stems from the tool’s architecture, which features a non-uniform distribution of risk levels: 15 of the 24 possible combinations of parameters (62.5%) lead to a risk index of 1 or 2 (out of 6). In particular, when slight severity (S1) is chosen, regardless of what other parameters are selected, the risk level will be 1, except in one case (level 3 of the parameter “probability of occurrence of harm”).

Similarly, many negative comments concerned an estimated risk level lower than what the subjects perceived, especially for scenario M. The risk level choices made by the subjects converge and seem consistent for the scenario: serious injury (amputation), seldom exposed (once every two days), low probability of occurrence, and possibility of avoidance under certain conditions. However, the risk level obtained by the subjects with these choices using the risk graph seems low in relation to the potential severity of the situation and the frequency of exposure. The same observation can be made with regard to scenario S. Here again, the problem is related to the non-uniform distribution of the risk levels on the graph.

Last, tool 91 performed poorly in scenario classification, especially of scenarios G and M; only 10 subjects arrived at an order that matched the reference order. Poor convergence was also noted for scenarios G (44%) and S (52%). This problem stems from the construction flaws of the parameters:

- Severity of harm: "inadequate number of levels," "inconsistent definition of different levels," which can be seen with respect to scenario G (see Table 14).

- Exposure frequency and/or duration: "inadequate number of levels" and "poor definition of levels." For scenarios A, G and S, the choice of the exposure frequency or duration appeared to be random, as the subjects' choices were distributed almost evenly between the two levels.

The flaws are in the first two parameters of the risk graph. These parameters have a greater relative weight on the result. A change in severity level without any change in the levels of the other parameters (e.g., S1/Ex2/Pe2/A2 compared with S2/Ex2/Pe2/A2) can cause the risk level to jump from 1 to 5. The combined effect of the parameter flaws and their influence is what explains the divergent risk level results obtained for scenario G.

Tool 91 illustrates the negative impact that the following flaws can have:

- An architecture that gives greater influence to one or more parameters (e.g., first parameter on a graph) and that is characterized by a discontinuity in risk levels.

  The impact (e.g., divergent results, scenario classification) occurs in particular when there are construction flaws in connection with parameters that have more influence on the result (e.g., poor definition of levels, inadequate number of levels).

- An architecture that leads to a non-uniform distribution of risk levels.

  That in turn led to dissatisfaction with the results obtained and to problems distinguishing between scenarios.

5.2.6 Tool 114

Tool 114 performed well in classifying scenarios having different risk levels. However, when using it, subjects had trouble distinguishing between low and mid-low risk scenarios, on the one hand, and between mid-high and high risk scenarios, on the other. A reading of the parameter choices made by the subjects reveals that the risk levels obtained for each scenario were mostly in distinct squares of the risk matrix. The inability to distinguish between certain scenarios therefore stems from the way risk levels are grouped together in the matrix.

A slight problem of convergence was observed only in the case of the mid-low scenario (scenario G; modal percentage of 60%). However, when all the parameter choices made by the subjects are analysed, it can be seen that there are notable divergences for several scenarios:

scenario A, exposure frequency (modal percentage of 44%, hesitation between levels 3 and 4); scenario G, possibility of avoidance (56%, levels 1 and 2); scenario S, possibility of avoidance (48%, levels 3 and 4). A problem of convergence of the risk level obtained was observed only in the case of scenario G because of the way the matrix is constructed. However, this construction corrects the divergences observed for scenarios A and S. Hesitation between levels 3 and 4 for the frequency of exposure or the possibility of avoidance did not have any influence on the risk level obtained.

A slight tendency to overestimate risks was noted, as more than 85% (19/22) of the subjects' comments were to this effect. An analysis of the choices for the different parameters suggests that the overestimate was related to the parameter "exposure frequency," which was flawed by a poor definition of its levels. Frequent exposure (level 3 out of 4) is defined as exposure at least once a day. For the scenarios presented, this level accounted for close to 50% of the choices made, drawing the risk levels obtained upwards toward the two highest levels.

Tool 114 performed well overall, based on our analysis criteria. Its non-standard configuration (no parameter for the "probability of occurrence of the hazardous event") did not seem to have any particular impact on the result. On the other hand, this tool illustrates how parameter divergences can in some cases be mitigated by matrix architecture (breakdown of levels).

5.2.7 Summary of Impact of Failure to Follow Construction Rules for Risk Estimation Tool Architecture

The analysis of the problems found with the six risk estimation tools confirmed the impact of the failure to follow certain construction rules for tool architecture. Table 48 provides an overview assessment and recommendations regarding the failure to follow certain construction rules.

Table 48. Summary of impact of failure to follow construction rules for risk estimation tool architecture

Failure to follow a construction rule	Explanation/Source	Impact	Ex.	Notes regarding tool architecture
• Non-uniform distribution	Significantly higher probability of arriving at certain risk levels.	Dissatisfaction with risk level obtained and trouble distinguishing between certain scenarios.	19 69 91	Distribute uniformly the probability of arriving at each risk level. Some risk graphs may conceal this flaw.
• Excessive relative weight given to one parameter • Discontinuity in risk levels	Structure that gives greater weight to one parameter over the risk level obtained. This may result in jumps in the risk level obtained, related to the choice made for the influential parameter, e.g., first parameter chosen in a graph.	Amplifies divergence of risk level results when there are flaws in the most influential parameters.	19 69 91	Pay special attention to a flaw (e.g., inadequate number of levels, poor definition of levels) on the first parameter chosen on a risk graph. Risk graphs generally give greater weight to the first parameter chosen and may conceal a discontinuity in risk levels.
• Non-standard configuration	Parameter lacking to estimate the probability of occurrence of harm, e.g., Pe and A.	No significant impact noted.	114	N/A
• Flaws other than those listed in Table 1	Matrix sensitive to the least change in level of a parameter. High number of risk levels relative to number of possible parameter combinations.	Amplification of divergence of risk level results when there is a flaw in one of the parameters.	24 89	Pay special attention to a flaw (e.g., inadequate number of levels, poor definition of levels) on the two parameters of the matrix. Use appropriate grouping of risk levels to mitigate flaws in parameters (e.g., 114).
	Risk level designated “0.”	Strong dissatisfaction with risk level obtained.	69	Do not use a risk level designated “0.”
• Inappropriate tool family • Not calibrated for machine risks • Inadequate number of risk levels	Not tested as part of this study.			

6. COMPARATIVE ANALYSIS OF RESULTS OBTAINED BY HSL

6.1.1 Background

As specified in subsection 1.3, this research project was conducted jointly with a team from the Health and Safety Laboratory (HSL) in the United Kingdom. A decision was made to work together so that expertise could be shared in the development of a detailed research method, including the preparation of data-gathering instruments, and in the comparative analysis of results.

However, the actual data gathering, which required the participation of users from industry, was carried out independently by the two teams: UQTR-PM-IRSST and HSL. Given the constraints specific to how HSL operates, its researchers conducted their experiments in a different way from the UQTR-PM-IRSST team, but using exactly the same basic components (tools, scenarios).

The HSL team therefore carried out experiments using tools 24 and 91 comparable to those presented in the preceding sections of this report with scenarios G and M, but without any part concerning individual parameters. Thus, from online experiments, the HSL obtained results from 59 subjects for these two scenarios.

Other data-gathering approaches were also followed by this team, and semistructured interviews were conducted with 13 subjects to get their comments and impressions. It is important to note that caution should be exercised in considering these comparative results, as the tools and scenarios used by the HSL team were presented in their English version, whereas those used by the UQTR-PM-IRSST team were in French. Differences in interpretation may have influenced some of the results obtained by each team.

6.1.2 Experimental Results

The results of the experiments (conducted online by the HSL team) for tools 24 and 91 with scenarios G and M were compared with the equivalent results obtained by the UQTR-PM-IRSST team. They are summarized in tables 49 and 50.

Table 49. Comparison of risk levels obtained by UQTR-PM-IRSST team and by HSL team, for tool 24

	Scenario G		Scenario M	
Risk levels of tool 24	IRSST n = 25 (% of subjects)	HSL (online) n = 59 (% of subjects)	IRSST n = 25 (% of subjects)	HSL (online) n = 59 (% of subjects)
Negligible	4	13	0	1
Low	40	31	0	1
Medium	48	34	24	31
High	8	22	76	68

Table 50. Comparison of risk levels obtained by UQTR-PM-IRSST team and by HSL team, for tool 91

	Scenario G		Scenario M	
Risk levels of tool 91	IRSST n = 25 (% of subjects)	HSL (online) n = 59 (% of subjects)	IRSST n = 25 (% of subjects)	HSL (online) n = 59 (% of subjects)
1 and 2	68	70	56	49
3 and 4	28	10	40	43
5 and 6	4	20	4	8

This analysis shows that the risk levels obtained by the various groups of subjects are on the whole similar with respect to their modal percentage. Some differences can be noted, however, in the distribution of results for scenario G with tool 24. Still, overall, the results suggest that the tendencies regarding the dispersion of risk estimation results are not influenced by where the subjects come from.

6.1.3 Results with Respect to Subjects' Perception

The data gathered by the HSL team regarding subjects' perception of the use of the risk estimation tools were compared with similar data collected by the UQTR-PM-IRSST team.

Comparable tendencies could be seen in the two sets of data. As the UQTR-PM-IRSST team observed, the number of subjects who disagreed with the results of tool 91 was high; the majority of the subjects interviewed by the HSL team said they felt the same tool underestimated the risk in most cases. For tool 24, the subjects studied by both teams thought that the results were generally satisfactory.

In the semistructured interviews conducted by the HSL team, the subjects identified and recognized the impact of certain flaws in the risk estimation parameters. As the analyses presented earlier showed, tools 24 and 91 (see subsections 5.2.2 and 5.2.5) do have these flaws on some of their parameters. The observations noted by the HSL team are summarized below:

- *Poor definition of levels* and *inconsistency in definitions used:* One recurring comment made by subjects was that the terminology used to define some parameters made the choices hard. In some cases, these definition problems prompted participants to point out overlapping between levels.

- *Inadequate number of levels* and *gaps between levels:* Subjects commented on the fact that for some parameters, the reduced number of levels made the choice simple, but they felt uncomfortable with it. They also noted that there were sometimes significant gaps between two levels of a parameter ("*big jump*"). The parameter "severity of harm" of tool 91, with just two levels, was cited as a typical example. The subjects clearly stated their view that all parameters should have more than two levels.

The subjects interviewed also raised other concerns regarding tool design:

- The subjects were surprised to see that tool 24, a two-parameter matrix, produced results comparable to those of tool 91, which uses four parameters. They did, however, question the perceived reliability of two-parameter tools. Some subjects said that a two-parameter tool may appear to be too simple and have less credibility than a tool with several parameters. On the other hand, the subjects also said they preferred tools that were visually simpler.

- The parameter "possibility of avoidance (A)" was identified by some subjects as not required in risk estimation.

- The parameters concerning probability were those that gave subjects the most trouble when estimating risks.

- The subjects indicated a preference for parameters whose level definitions included examples, such as for the severity of harm with tool 91 (G1 – Slight injury, usually reversible; for example, scratches, laceration, bruising, light wound requiring first aid).

- The subjects indicated that, when a level referred to historical data, the tool was harder to use, as was the case for the probability of occurrence of the hazardous event with tool 91 (O2 – Low: event related to a technical failure of probability higher than or equal to 10E-5 breakdowns/hour [1 breakdown/100,000 hours]).

- As with the results obtained by the UQTR-PM-IRSST team, the subjects interviewed by the HSL did not note the flaw "lack of exposure interval," even though it was present in the two tools tested.

6.1.4 Conclusion of Comparative Analysis

Although the studies conducted by the two teams differed in terms of the extent of their results and some methodological aspects, certain common tendencies can be seen in the results. The experiments showed similar tendencies with respect to the dispersion of results under the effect of flaws and the subjects’ level of agreement with the risk level obtained. In addition, in both cases, the subjects were able to identify the flaws in the parameters and report on the impact of the flaws on the trouble they had choosing the risk level corresponding to a given situation and on the performance of the risk estimation tools.

7. CONCLUSION

The nature of this study meant that it could not be based on a totally watertight methodology. A number of hard-to-control variables may have influenced some of the results. The goal of the study was to observe the behaviour of flawed tools and parameters when used to estimate risks, without necessarily seeking to assess the users' cognitive processes. It is important to recognize, however, that cognitive factors do have an influence on the estimating process and that this study was not able to shed light on them or control them totally. This is one of the study's limitations.

Nevertheless, within the boundaries of these methodological considerations, this study helps to understand the impact that flaws in the parameters and architecture of risk estimation tools used in machine safety can have on results of risk level estimations. The results show that these flaws can lead to low convergence of the risk levels arrived at by different subjects, for a given hazardous situation, and to subject dissatisfaction with tool performance and accuracy. In most cases, the subjects were able to recognize a flaw when it made it harder from them to choose the level corresponding to a given situation. These observations were also confirmed by the study conducted by the HSL team.

The results also indicate that the impact of parameter construction flaws is not uniform. The type of flaw, its position on the parameter scale and the scenario in question influence its impact on the determination of the level of a parameter. In addition, the parameter "severity of harm" is relatively robust despite various flaws, but the parameters "probability of occurrence of harm" and "probability of occurrence of the hazardous event" are far more affected by them. Besides the flaws in these last two parameters, the results suggest that evaluating probability is a difficult aspect of risk estimation that requires special attention.

With regard to tool architecture, the results show that an architecture that gives greater influence to one parameter can amplify divergent results and reduce the ability of the tool to classify scenarios appropriately. This effect is also amplified when the parameter in question is itself affected by a flaw. A matrix sensitive to the slightest change in the level of a parameter will have the same impact when there is a flaw in one of its parameters. An architecture that does not feature a uniform distribution of risk levels will lead to problems distinguishing between scenarios and to subject dissatisfaction with the results obtained (Chinniah et al., 2015).

These experimental results bolster the validity of many of the construction rules set out in the earlier study (Chinniah et al., 2011). The rules will help considerably reduce subjectivity in the risk estimation process and address certain problems related to the significant variability in estimated risk levels. The results will therefore help improve the robustness and reliability of existing tools and provide support for the risk assessment training currently provided by partners.

To ensure that the users of machine safety risk estimation tools will benefit from the findings of this study, the next step should be to write up a self-diagnostic guide providing a clear, concrete presentation of the flaws in parameters and tool architecture, the construction rules to follow and examples of good and not-so-good wording. This guide would be useful to those tasked with choosing or improving risk estimation tools.

BIBLIOGRAPHY

Abrahamsson, M. (2000). *Treatment of uncertainty in risk based regulations and standards for risk analysis*, Report 3116, Lund University, Sweden, 82 pages.

Abrahamsson, M. (2002). *Uncertainty in quantitative risk analysis- characterisation and methods of treatment*. Department of Fire Safety Engineering, Lund University, Sweden.

ANSI B11.TR3 (2000). *ANSI Technical Report - Risk assessment and risk reduction - A guide to estimate, evaluate and reduce risk associated with machine tools,* American national Standard.

ANSI/RIA R15.06 (1999). *American National Standard for Industrial Robots and Robots Systems - Safety Requirements*.

Aven, T. (2012). "Foundational Issues in Risk Assessment and Risk Management." *Risk Analysis*, 32, 1647-1656.

Beyth-Marom, R. (1982). "How Probable is Probable? A Numerical Translation of Verbal Probability Expressions," *Journal of Forecasting*, Vol. 1, No. 3, pp. 257 - 269.

Carey, J.M. et Burgman, M.A. (2008). "Linguistic Uncertainty in Qualitative Risk Analysis and How to Minimize It." *Annals of the New York Academy of Sciences*, 1128, (1) 13-17.

Charpentier, P. (2003). *Projet européen RAMSEM- Développement et validation d'une méthode d'appréciation du risque machine basée sur les principes de la norme EN 1050*. Projet A.5/1,058 de l'INRS.

Chinniah, Y., Gauthier, F., Lambert, S. et Moulet, F. (2011). *Experimental Analysis of Tools Used for Estimating Risk Associated with Industrial Machines*, Rapport de recherche R-684, IRSST, 77 p.

Chinniah, Y., Gauthier, F., Burlet-Vienney, D. et Aucourt, B. (2015) *Analysis of two risk estimation tools applied to safety of machinery*, International conference – Safety of Automated Industrial Systems, Bonn, Germany, November 2015.

Christensen, F.M., Anderson, O., Duijm, N.S. et Harremoes, P. (2003). "Risk Teminology - a platform for common understanding and better communication." *Journal of Hazardous Materials*, 103, (3) 181-203.

Company A (2002). *Identification des dangers et risques en Santé/Sécurité*, internal document.

Company P (2003). *Risk assessment and risk reduction*, internal document.

Company R (2004). *Évaluation des risque - Partie 2: Évaluation des mesures de réduction des risques*, internal document.

Company X (1997). *Tableau d'analyse de risque* (sans titre), internal document.

Cox, L.A. (2008) "What's wrong with Risk Matrices?," *Risk Analysis*, Vol. 28, No. 2, pp. 497-512.

CSA-Q634-91 (1991). *Risk Analysis Requirements and Guidelines*, Canadian Standard Association.

CSST (2002). *Sécurité des machines, Phénomènes dangereux, situations dangereuses, évènements dangereux, dommages*, Commission de la santé et de la sécurité du travail du Québec, DC 900-337 (07-02), 15 p.

Duijm, N.J. (2015). “Recommendations on the use and design of risk matrices,” *Safety Science*, Vol. 76, No. 7, pp. 21-31. DOI:10.1016/j.ssci.2015.02.014

Ekelenburg, V.H.P., Hoogerkamp, P., Hopmans, L.J. (1996) *A practical Guide to the Machinery Directive* (Traduit du hollandais par D. Brown), Mechanical Engineering Publications Ltd.

Etherton, J. (2007). “Industrial Machine Systems Risk Assessment: A Critical Review of Concepts and Methods,” *Risk Analysis*, Vol. 27, No. 1, pp. 71-82.

Etherton, J., Main, B., Cloutier, D., Christensen, W. (2008). “Reducing Risk on machinery: A Field Evaluation Pilot Study of Risk Assessment,” *Risk Analysis*, Vol. 28, No. 3, pp. 711-721.

Franceschini, F., Galetto, M., Varetto, M. (2004) “Qualitative Ordinal Scales: The Concept of Ordinal Range,” *Quality Engineering*, Vol. 16, No. 4 pp. 515-524.

Gauthier, F., Chinniah, Y., Lambert, S. (2012) “Analysis and Classification of the Tools for Assessing the Risks Associated with Industrial Machines,” *Journal of Occupational Safety and Ergonomics*, Vol. 18, No. 2, 245–265.

Gauthier, F., Moulet, F., Chinniah, Y., Stacey, N., Healy, N. (2010) *A Comparative Analysis of Risk Estimation Tools for Industrial Machines*, International conference – Safety of Automated Industrial Systems, Tampere Finland, June 2010.

Gondar Design (2000) *Risk assessments*, http://www.purchon.co.uk/safety/risk.html, 5 p.

Görnemann, O. (2003). *SICK AG Scalable Risk Analysis & Estimation Method* (SCRAM), ISO/TC199 WG 5 N 0049, 12 p.

HSL (2012) *How to complete a methodical risk estimation*, Health and Safety Laboratory.

Hubbard, D., Evans, D. (2010) “Problems with scoring methods and ordinal scales in risk assessment,” *IBM Journal of Research and Development*, Vol. 54, No. 3.

Hughes, P. et Ferrett, E. (2005). *Introduction to Health and Safety at Work*, 2nd Edition, 2nd ed. Oxford, Elsevier Butterworth-Heinemann.

IEC 62278 (2001). *Railway applications - The specification and demonstration of Reliability, availability, Maintainability and safety (RAMS)*, International Electrotechnical Committee.

ISO 12100 (2010). *Safety of machinery -- General principles for design -- Risk assessment and risk reduction*, International Standard.

ISO 14121-2 (2007). *Risk Assessment - Part 2: Practical guidance and examples of methods*, International Standard.

ISO/TS 14798 (2006). *Lifts (elevators), escalators and moving walks -- Risk assessment and reduction methodology (annex C)*, International Standard.

Kazer, B.M. (1993). “Risk Assessment: A Practical Guide,” *The Health and Safety Practionner*, Institution of Occupational Safety and Health.

Lamy, P., Charpentier, P. (2009). “Estimation des risques – Recensement des méthodes et subjectivité des paramètres de l’estimation,” INRS, *Hygiène et sécurité du travail*, ND 2305-214-09, pp37-44.

Lyon, B. et Hollcroft, B. (2012). “Risk Assessments top 10 pitfalls and tips for improvement.” *Professional Safety*, 57, (12) 28-34.

Main, B.W. (2012). *Risk Assessment Challenges and Opportunities*, Ann Arbor, USA, Design Safety Engineering inc.

Main, B.W. (2004). *Risk Assessment: Basics and benchmarks*, Design Safety Engineering Inc., 485 p.

MIL-STD-882D (2000). *Standard Practice for System Safety (Appendix A)*, US Department of Defense.

Mosteller, F. et Youtz, C. (1990). “Quantifying probability expressions.” *Stat Sci* (5) 2-12.

Ni, H., Chen, A., Chen, N. (2010). “Some Extensions on Risk Matrix Approach,” *Safety Science*, Vol. 48, pp. 1269-1278.

Paques, J.-J. (2005). *Results of exploratory tests on tools for assessing the risks associated with industrial machines*, 4th International Conference Safety of Industrial Automated Systems, September 26-28, Chicago, Illinois, USA

Paques, J.-J. et Gauthier, F. (2006). “Thematic program: Integrated projects on risk assessment tools for industrial machinery,” *HST-CDN (Hygiène et sécurité du travail)*, ND 2259-205-06, p. 33-40.

Paques, J.-J. et Gauthier, F. (2007). “Analysis and Classification of the Tools for Assessing the Risks Associated with Industrial Machines,” *Journal of Occupational Safety and Ergonomics*, 13(2), p. 173-187.

Paques, J.-J., Bourbonnière, R., Daigle, R., Doucet, P., Masson, P., Micheau, P., Lane, J. et Tardif, J. (2005a). *Transfert de compétences en formation sur la gestion de la sécurité des machines et les moyens de protection*. IRSST, Rapport de recherche R-394, 105 p.

Paques, J-J., Gauthier, F., Perez, A., Charpentier, P., Lamy, P. et David, R. (2005b). *Bilan raisonné des outils d’appréciation des risques associés aux machines industrielles*, IRSST, Rapport R-459, 64 p.

Paques, J-J., Perez, A., Lamy, P., Gauthier, F., Charpentier, P. et David, R. (2005c). *Reasoned review of the tools for assessing the risks associated with industrial machines: Preliminary results*, 4th International Conference Safety of Industrial Automated Systems, September 26-28, Chicago, Illinois, USA.

Parry, G.W. (1999). *Uncertainty in PRA and its implications for use in Risk-informed decision making.* Proceedings of the 4th International conference on probabilistic safety assessment and management, PSAM 4, Edited by Mosleh, A. & Bari, R.A., New York.

Patt, G.A. et Schrag, D.P. (2003). “Using specific language to describe risk and probability.” *Climatic Change*, 61, (1-2) 17-30.

Pickering, A. et Cowley, S.P. (2010). “Risk Matrices: Implied accuracy and false assumptions.” *Journal of health and safety research and practice*, 2, (1) en ligne.

Raafat, H. (1995). *Machinery Safety: The Risk Based Approach*, Technical Communication Publishing Ltd., 70 p.

Ruge, B. (2004). *BASF Risk Matrix as Tool for Risk Assessment in the Chemical Process Industries*, BASF.

Smith, E.D., Siefert, W.T., Drain, D. (2009). “Risk Matrix Input Data Biases,” *Systems Engineering*, Vol. 12, No. 4, pp. 344-359.

Stevens, S.S. (1946). “On the theory of scales of measurement.” *Science*, Vol. 103, No. 2684, pp. 677-680.

SUVA (2002). *Méthode SUVA d’appréciation des risques liés aux installations et appareils techniques*, Caisse nationale Suisse d’assurance en cas d’accidents.

The Metal Manufacturing and Minerals Processing Industry Committee (2002). *A Guide to Practical Machine Guarding*, Queensland Government - Workplace Healt and Safety.

Theil, M. (2002). “The role of translations of verbal into numerical probability expressions in risk management: a meta-analysis.” *Journal of Risk Research*, 5, 177-185

Willquist, P. et Torner, M. (2003). “Identifying and analysing hazards in manufacturing industry – a review of selected methods and development of a framework for method applicability.” *International Journal of Industrial Ergonomics*, 32, 165-180.

Woodruff, J.M. (2005). “Consequence and likelihood in risk estimation: A matter of balance in UK health and safety risk assessment practice.” *Science Direct* 345-353.

APPENDIX A – SCENARIOS USED FOR EXPERIMENTS

Table 51. Scenario A

Punching machine with mobile table	
Activity	Functional demonstration of a punching machine at a trade fair. The machine operates in automatic mode, punching holes in sheet metal, which is placed on a mobile table and moves unpredictably in the directions indicated by the arrows in the picture.
Hazard	Movement of the mobile table.
Hazardous situation	A visitor to the trade fair is standing next to the mobile table (possible movements along two axes), as the photo shows.
Hazardous event	A visitor is struck by the table when he steps into its path of movement.
Probability of hazardous event	Pressure-sensitive mats on the ground do not protect the hazard zone completely (e.g., if a visitor stands like the person in the picture).
Possible harm	Bruises, cuts.
Exposure	Visitors stay at the stand for 5 minutes on average. The machine's mobile table is moving 50% of the time. On average, a visitor is at the stand 20% of the time over the course of a 10-hour day.
Possibility of avoidance	Visitors are not warned about the hazards of the machine and do not have any prior knowledge about it. The mobile table moves at a speed of 1 m/s. Movements that are under way or about to happen are indicated by a warning light.

Table 52. Scenario G

Automated guided vehicle (AGV)	
Activity	An automated guided vehicle (no operator) moves through a plant following a yellow line (traffic lane) painted on the ground (predetermined path).
Hazard	Movement of AGV in direction of arrow.
Hazardous situation	AGV operates in same area as workers.
Hazardous event	A worker is struck by AGV.
Probability of hazardous event	An accident of this kind has already occurred and, since then, workers have been informed of the AGV's predetermined path (traffic plan) in their work area. No barriers prevent workers from entering the vehicle's path.
Possible harm	Bruising, simple fracture.
Exposure	On average, the vehicle's traffic lane is crossed by workers 25 times over the course of each 8-hour shift. It takes a worker 3 seconds to cross the vehicle's travel lane completely. The AGV operates continuously during working hours.
Possibility of avoidance	The AGV moves at a speed of 10 km/h and has a pressure-sensitive bar (bumper bar, circled in photo) that stops the vehicle if it hits something. The AGV is also equipped with a beeper and a rotating orange warning light. The yellow line painted on the ground indicates the vehicle's predetermined path (traffic lane).

Table 53. Scenario M

Rewinder (papermaking machine)	
Activity	Remove irregular parts of the roll that could cause technical problems when it is used (e.g., in a printing press). The rewinder is on in manual mode.
Hazard	Nip points (getting drawn in by the roll).
Hazardous situation	The hands of the two workers are close to the nip points.
Hazardous event	Accidental start-up of the roll as a result of poor communication between the workers near the roll and the person controlling the movement of the roll.
Probability of hazardous event	The workers communicate orally. They are 5 m apart. The shop is fairly quiet.
Possible harm	Partial or complete amputation of upper limbs.
Exposure	Each time a new roll is installed, i.e., once every two days. Each operation takes approximately 15 minutes.
Possibility of avoidance	The speed of the roll is reduced and its movement is controlled by a hold-to-run control. Stoppage of the movement of the roll is instantaneous.

Table 54. Scenario S

Robot	
Activity	Worker changes a tool on a numerically controlled lathe used to machine metal parts. A robot supplies the lathe with metal pieces to be machined and then removes them when the machining has been done.
Hazard	Movement of robot toward worker.
Hazardous situation	Worker is standing in robot's path. Robot is powered up and remains in stand-by.
Hazardous event	The worker is struck by the robot. The robot receives a start-up command as a result of a failure of the programmable logic controller (PLC) that controls it.
Probability of hazardous event	The robot is controlled by a standard PLC and not by a safety PLC.
Possible harm	Multiple fractures, concussion, death.
Exposure	Ten-minute job, twice every 8-hour shift.
Possibility of avoidance	The worker has his back turned to the robot and is wearing hearing protection. The end of the robot moves very quickly, at around 2 m/s. The robot has visual and sound warning devices. No record of past failure, recent or earlier.

APPENDIX B – PARAMETERS TESTED IN EXPERIMENTS

Table 55. Parameters tested in experiments and types of associated flaws

Tool	Definition of parameter	Flaw	Reference
Severity of harm (S)			
33	– Moderate injury or illness – Serious injury or illness – Death, grievous injury or illness	Poor definition of levels	Main (2004) p. 155–157
55	– 4) Negligible: Less than minor injury or occupational illness – 3) Marginal: Minor injury or occupational illness – 2) Critical: Severe injury or occupational illness – 1) Catastrophic: Death	Poor definition of levels	Company X (1997)
66	– Insignificant: Possible minor injury – Marginal: Minor injury and/or significant threat to the environment – Critical: Single fatality and/or severe injury and/or significant damage to the environment – Catastrophic: Multiple fatalities and/or multiple severe injuries and/or major damage to the environment	Inconsistent definitions of different levels	IEC 62278 (2001)
69	– No harm – Low: Trivial harm with no permanent results – Middle: Serious harm with no permanent results – High: Serious harm with permanent results, death	No flaws	Görnemann (2003)
91	– S1 Slight injury (usually reversible), e.g., scratches, laceration, bruising or light wound requiring first aid. – S2 Serious injury (usually irreversible, including fatality), e.g., broken or torn-out or crushed limbs; serious injuries requiring stitches, fatalities	Inadequate number of levels	ISO 14121-2:2007
102	– 1. Minor: Non-serious consequences – 2. Significant: Work has to stop, first aid is required – 3. Disastrous: Very serious accident (someone has been scarred for life, blinded or even killed)	Gaps between levels	Gondar (2000)
Probability of occurrence of harm (Ph)			
6	– Improbable – Probability close to zero – Remote – Unlikely, though conceivable – Possible – Could occur sometime – Probable – Not surprised, will occur several times – Likely/Frequent – Will occur repeatedly / event only to be expected	Inconsistent definitions of different levels	Kazer, BM. (1993)
7	– Remote – Improbable – Possible – Probable – Likely	Poor definition of levels	Raafat, H. (1995)
34	– Low – Very seldom or never – Medium – Reasonably likely to occur – High – Certain or near certain to occur	Gaps between levels	Main (2004) p. 164–165

Tool	Definition of parameter	Flaw	Reference
41	– F – Highly improbable – Probability practically zero – E – Improbable – Very unlikely to occur in life cycle – D – Remote – Unlikely, but may possibly occur in life cycle – C – Occasional – Likely to occur at least once in life cycle – B – Probable – Likely to occur several times in life cycle – A – Highly probable – Likely to occur frequently in life cycle	No flaws	ISO/TS 14798 (2006)
89	– Very unlikely – Could happen, but probably never will – Unlikely – Could happen, but rare – Likely – Could happen occasionally – Very likely – Could happen frequently	No definition of exposure interval	The Metal Manufacturing and Minerals Processing Industry Committee (2002)
Exposure frequency (Exf)			
49	– E1: Infrequent exposure (typically, exposure to hazard less than once per day or shift) – E2: Frequent exposure (typically, exposure to hazard more than once per hour)	Gaps between levels	ANSI/RIA R15.06 (1999)
67	– 1. Interval between exposures is more than a year – 2. Interval between exposures is more than two weeks, but less than or equal to a year – 3. Interval between exposures is more than a day, but less than or equal to two weeks – 4. Interval between exposures is more than an hour, but less than or equal to a day. Where the duration is shorter than 10 min., the value may be decreased to the next level – 5. Interval between exposures less than or equal to an hour. This value is not to be decreased at any time.	No flaws	ISO 14121-2:2007
Exposure duration (Exd)			
19	– Seldom to quite often – Frequent to continuous	Poor definition of levels	Ekelenburg et al. (1996)
62	– 1. 2 hr/week; 1 day/month – 2. 4 hr/week; ½ day/week – 3. 8 hr/week; 1 day/week – 4. 20 hr/week; half the time – 5. 40 hr/week; all the time	Gaps between levels	SUVA (2002)
91	– F1 Seldom to quite often and/or short duration of exposure – F2 Frequent to continuous and/or long duration of exposure	No flaws	ISO 14121-2:2007
Probability of occurrence of hazardous event (Pe)			
19	– 1. Low – so unlikely that it can be assumed occurrence may not be experienced – 2. Medium – likely to occur sometime in the life of an item – 3. High – likely to occur frequently	No flaws	Ekelenburg et al. (1996)

Tool	Definition of parameter	Flaw	Reference
62	– P0 = 1: Event hard to imagine (measures consistent with state of the art) – P0 = 2: Event imaginable, but unusual (measures taken) – P0 = 3: Event is possible (measures partially taken, clear deficiencies) – P0 = 4: Event may be expected to occur (measures have started to be taken) – P0 = 5: Event should be expected to occur (no existing measures)	Inconsistent definitions of different levels	SUVA (2002)
Possibility of avoidance (A)			
57	– 1. Obvious – 2. Likely – 3. Possible – 4. Rarely – 5. Impossible	Poor definition of levels	Company P (2003)
114	– Possible: For all exposed people – Possible if trained: Possible for people trained to recognize warnings and how best to react, and warning allows sufficient time – Difficult: Possible, but warning may not be obvious or time is limited – Impossible: No warning and/or not enough time to react	No flaws	HSL (2012)

APPENDIX C – RISK ESTIMATION TOOLS STUDIED

This appendix presents the architectures of the six tools tested, in their original form. The wording of each tool, as it was presented to the subjects, is given in Appendix D.

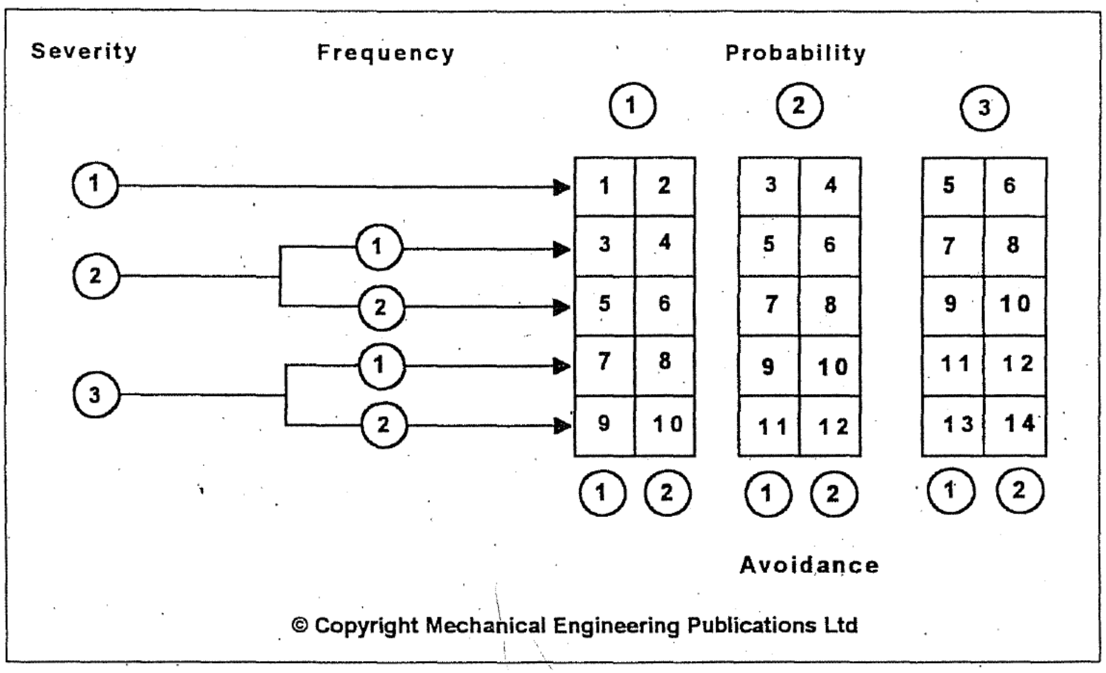

Figure 6. Risk graph of tool 19 (Source: Ekelenburg et al., 1996)

Probability of Occurrence of Harm	Severity of Harm: Catastrophic	Serious	Moderate	Minor
Very Likely	High	High	High	Medium
Likely	High	High	Medium	Low
Unlikely	Medium	Medium	Low	Negligible
Remote	Low	Low	Negligible	Negligible

Figure 7. Risk matrix of tool 24 (Source: ANSI B11.TR3, 2000)

	Severity of Harm *	Exposure to Harm *	Harm avoidance *	Probability of Ocurrence *		
				Low	Middle	High
START	No Harm **	-	-	0	0	0
	Low	-	Avoidable	0	0	1
			Not avoidable	0	1	2
	Middle	Low	Avoidable	1	2	3
			Not avoidable	2	3	4
		High	Avoidable	3	4	5
			Not avoidable	4	5	6
	High	Low	Avoidable	5	6	7
			Not avoidable	6	7	8
		High	Avoidable	7	8	9
			Not avoidable	8	9	10
				Resulting Risk Level		

Figure 8. Risk graph of tool 69 (Source: Görnemann, 2003)

PROBABILITY **How likely COULD it happen**	**CONSEQUENCE: How severely could it hurt someone?** CATASTROPHIC kills, disables permanent injury	 MAJOR significant injury, NOT permanent	 MINOR first aid only, no lost time
VERY LIKELY could happen	1	2	3
LIKELY could happen occasionally	2	3	4
UNLIKELY could happen but rare	3	4	5
VERY UNLIKELY could happen, probably never will	4	5	6

Figure 9. Risk matrix of tool 89 (Source: The Metal Manufacturing and Minerals Processing Industry Committee, 2002)

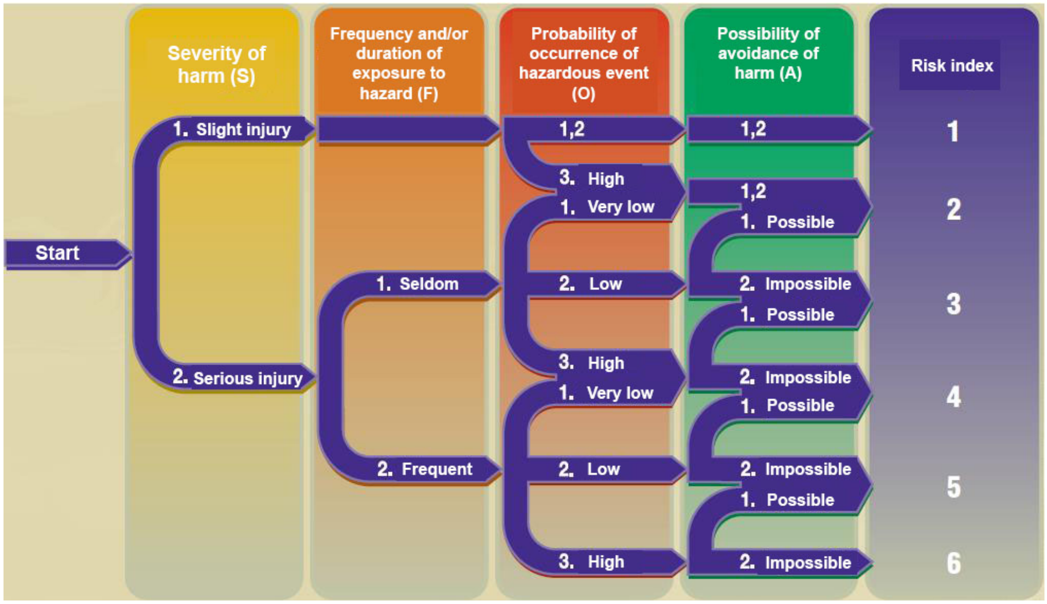

Figure 10. Risk graph of tool 91 (Source: CSST, 2002)

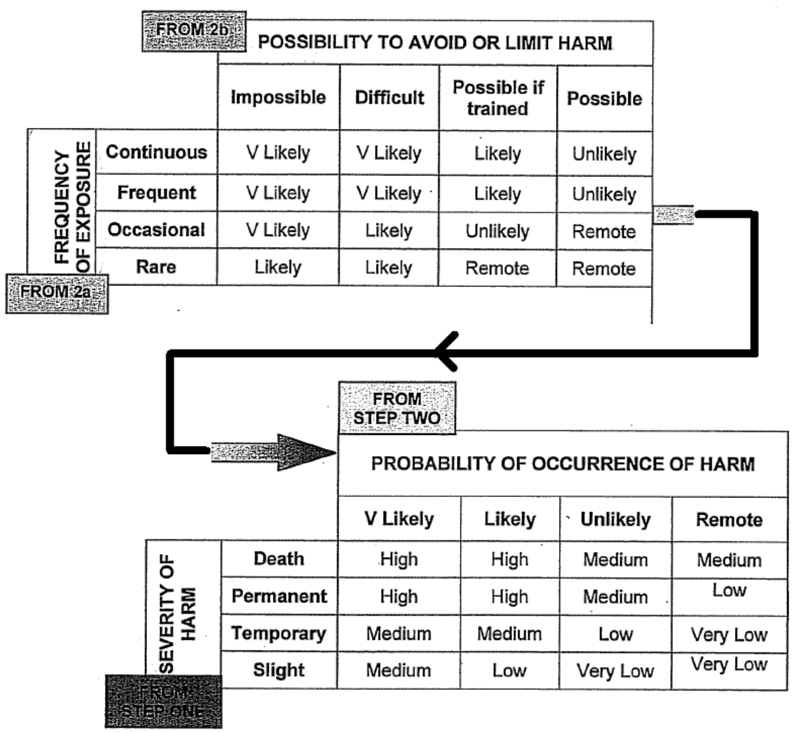

Figure 11. Risk matrix of tool 114 (Source: HSL, 2012)

APPENDIX D – QUESTIONNAIRE USED FOR TESTING OF RISK ESTIMATION TOOLS

Section 1/2: Risk estimation

This risk estimation tool has 4 parameters. Please estimate the risk level for the scenario indicated above by selecting the appropriate levels for the different parameters listed below.

1.1 Tool 19: Severity (of the possible harm)

☐ 1. Slight (normally reversible) injury or damage to health
☐ 2. Serious (normally irreversible) injury or damage to health
☐ 3. Death

1.2 Tool 19: Frequency and duration of exposure of persons to the hazard

☐ 1. Seldom to quite often
☐ 2. Frequent to continuous

1.3 Tool 19: Probability of occurrence of an event which can cause harm

☐ 1. Low – so unlikely that it can be assumed occurrence may not be experienced
☐ 2. Medium – likely to occur sometime in the life of an item
☐ 3. High – likely to occur frequently

1.4 Tool 19: Avoidance – the technical or human possibilities to avoid or limit the harm

☐ 1. Possible under specific conditions
☐ 2. Scarcely possible

Section 2/2: Opinion on level of risk estimated by tool

2.1 Risk level obtained with tool 19

☐ 1–4 Low risk
☐ 5–7 Medium risk
☐ 8–10 High risk
☐ 11–14 Extremely high risk

2.2 The risk level obtained with this tool is an accurate reflection of the information presented in the scenario

☐ Totally agree
☐ Somewhat agree
☐ Neither agree nor disagree
☐ Somewhat disagree
☐ Totally disagree

2.3 Explain your answer

____
____
____

Figure 12. Questionnaire for tool 19

Section 1/2: Risk estimation

This risk estimation tool has 2 parameters. Please estimate the risk level for the scenario indicated above by selecting the appropriate levels for the different parameters listed below.

1.1 Tool 24: Severity of harm

- ☐ Minor – no injury or slight injury requiring no more than first aid (little or no lost work time)
- ☐ Moderate – significant injury or illness requiring more than first aid (able to return to same job)
- ☐ Serious – severe debilitating injury or illness (able to return to work at some point)
- ☐ Catastrophic – death or permanently disabling injury or illness (unable to return to work)

1.2 Tool 24: Probability of occurrence of harm

- ☐ Remote – so unlikely as to be near zero
- ☐ Unlikely – not likely to occur
- ☐ Likely – may occur
- ☐ Very likely – near certain to occur

Section 2/2: Opinion on level of risk estimated by tool

2.1 Risk level obtained with tool 24

- ☐ Negligible
- ☐ Low
- ☐ Medium
- ☐ High

2.2 The risk level obtained with this tool is an accurate reflection of the information presented in the scenario

- ☐ Totally agree
- ☐ Somewhat agree
- ☐ Neither agree nor disagree
- ☐ Somewhat disagree
- ☐ Totally disagree

2.3 Explain your answer

____

____

____

Figure 13. Questionnaire for tool 24

Section 1/2: Risk estimation

This risk estimation tool has 4 parameters. Please estimate the risk level for the scenario indicated above by selecting the appropriate levels for the different parameters listed below.

1.1 Tool 69: Severity of harm

- ☐ No harm
- ☐ Low: trivial harm with no permanent results
- ☐ Middle: serious harm with no permanent results
- ☐ High: serious harm with permanent results, death

1.2 Tool 69: Exposure to harm

- ☐ Low: seldom or very short exposure to harm
- ☐ High: often or short to longer exposure to harm

1.3 Tool 69: Harm avoidance

- ☐ Avoidable: harm can be normally avoided
- ☐ Not avoidable: harm avoidance is seldom or not possible

1.4 Tool 69: Probability/likelihood of occurrence

- ☐ Low: harm will occur very seldom
- ☐ Middle: harm is possible, but not necessary
- ☐ High: harm is mostly consequence of exposure

Section 2/2: Opinion on level of risk estimated by tool

2.1 Risk level obtained with tool 69

☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

2.2 The risk level obtained with this tool is an accurate reflection of the information presented in the scenario

- ☐ Totally agree
- ☐ Somewhat agree
- ☐ Neither agree nor disagree
- ☐ Somewhat disagree
- ☐ Totally disagree

2.3 Explain your answer

____
____
____

Figure 14. Questionnaire for tool 69

Section 1/2: Risk estimation

This risk estimation tool has 2 parameters. Please estimate the risk level for the scenario indicated above by selecting the appropriate levels for the different parameters listed below.

1.1 Tool 89: How severe the injury could be (consequences)

- ☐ Minor: first-aid only, no lost time
- ☐ Major: maiming, significant injury, not permanent
- ☐ Catastrophic: kills, disables, permanent injury

1.2 Tool 89: Likelihood of the hazard causing an injury (probability)

- ☐ Very unlikely: could happen, but probably never will
- ☐ Unlikely: could happen, but rare
- ☐ Likely: could happen occasionally
- ☐ Very likely: could happen frequently

Section 2/2: Opinion on level of risk estimated by tool

2.1 Risk level obtained with tool 89

- ☐ 1. Measures must be taken immediately to control this risk
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6. Take care of other priorities first

2.2 The risk level obtained with this tool is an accurate reflection of the information presented in the scenario

- ☐ Totally agree
- ☐ Somewhat agree
- ☐ Neither agree nor disagree
- ☐ Somewhat disagree
- ☐ Totally disagree

2.3 Explain your answer

____

____

____

Figure 15. Questionnaire for tool 89

Section 1/2: Risk estimation

This risk estimation tool has 4 parameters. Please estimate the risk level for the scenario indicated above by selecting the appropriate levels for the different parameters listed below.

1.1 Tool 91 (CSST): Severity of harm: S

- ☐ S1 Slight injury (usually reversible), for example, scratches, laceration, bruising, light wound requiring first aid
- ☐ S2 Serious injury (usually irreversible, including fatality), for example, broken or torn-out or crushed limbs, fractures, serious injuries requiring stitches, fatalities

1.2 Tool 91 (CSST): Frequency and/or duration of exposure to hazard: F

- ☐ F1 – Seldom to quite often and/or short duration of exposure
- ☐ F2 – Frequent to continuous and/or long duration of exposure

1.3 Tool 91 (CSST): Probability of occurrence of hazardous event: O

- ☐ O1 Very low: mature technology, proven and recognized in safety applications
- ☐ O2 Low: event related to a technical failure of probability higher than or equal to 10E-5 breakdowns/hour (1 breakdown/100.000 hours); or event caused by the actions of a qualified, experienced, trained person carrying out a single task, etc.
- ☐ O3 High: event related to a technical failure of probability higher than or equal to 10E-3 breakdowns/hour (1 breakdown/1,000 hours); or event caused by the actions of a person having no experience or no specific training.

1.4 Tool 91 (CSST): Possibility of avoidance or reduction of the harm: A

- ☐ A1 Possible under some conditions
- ☐ A2 Impossible or rarely possible

Section 2/2: Opinion on level of risk estimated by tool

2.1 Risk level obtained with tool 91 (CSST)

- ☐ 1 = Lowest
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6 = Highest

2.2 The risk level obtained with this tool is an accurate reflection of the information presented in the scenario

- ☐ Totally agree
- ☐ Somewhat agree
- ☐ Neither agree nor disagree
- ☐ Somewhat disagree
- ☐ Totally disagree
- ☐

2.3 Explain your answer

____

____

____

Figure 16. Questionnaire for tool 91

Section 1/2: Risk estimation

This risk estimation tool has 3 parameters. Please estimate the risk level for the scenario indicated above by selecting the appropriate levels for the different parameters listed below.

1.1 Tool 114 (HSL): Severity of harm

- ☐ Slight: first aid needed, but no time of work or change of duties required Temporary: injury or ill-health requiring time off work from which essentially a full recovery normally expected (i.e., no loss of quality of life)
- ☐ Permanent: disability or health impairment which is normally irreversible, having impact on quality of life
- ☐ Death: injury or damage to health resulting, within a short period, in the death of operator and/or any other person in vicinity

1.2 Tool 114 (HSL): Frequency of exposure

- ☐ Rare: exposure not anticipated during normal use
- ☐ Occasional: exposure possible during normal use
- ☐ Frequent: exposure at least once a day
- ☐ Continuous: exposure every use or all the time during use

1.3 Tool 114 (HSL): Possibility to avoid or limit harm

- ☐ Possible: for all exposed people
- ☐ Possible if trained: possible for people trained to recognize warning and how best to react, and warning allows sufficient time
- ☐ Difficult: possible, but warning may not be obvious or time is limited
- ☐ Impossible: no warning and/or not enough time to react

Section 2/2: Opinion on level of risk estimated by tool

2.1 Risk level obtained with tool 114 (HSL)

- ☐ Very low
- ☐ Low
- ☐ Medium
- ☐ High

2.2 The risk level obtained with this tool is an accurate reflection of the information presented in the scenario

- ☐ Totally agree
- ☐ Somewhat agree
- ☐ Neither agree nor disagree
- ☐ Somewhat disagree
- ☐ Totally disagree

2.3 Explain your answer

____

____

____

Figure 17. Questionnaire for tool 114

APPENDIX E – EXPERIMENTAL PROTOCOL

All the experiments are to be conducted according to the following general procedure:

1. Read, explain and sign the consent and ethics form.
2. Have subject fill in the identification questionnaire.
3. Provide subject with a printed copy of the four standardized scenarios (Appendix A), specifying that:
   - Each scenario concerns a specific hazard, even if other hazards may be visible in the photograph. The photograph serves to support the description, but not to provide additional information.
   - All available information required for the experimentation is contained in the scenario description. The researcher present may not provide any additional information, for fear of creating bias in relation to the other subjects. All subjects have access to the same information.
4. Mix the two Excel worksheets to ensure that the tests are done in a random order, from one subject to the next, in order to limit possible bias.
5. Apply the six tools to the four scenarios by filling in the 24 online questionnaires (see section below).
6. Apply the 20 parameters, divided into six categories, to the four scenarios by filling in the 24 online questionnaires (see section below).
7. Fill in the post-experiment questionnaire to check whether the subject is familiar with one of the scenarios or tools, and to collect his or her views on the experiments.

Applying the six tools (in random order) to the four scenarios should take an average of 2 hr, 15 min. This part of the experimentation proceeds as follows:

1. Read through the scenario in the first column of the Excel worksheet:
   a. Ask the subject to explain the scenario to verify his or her understanding of it.
   b. If necessary, ask the subject to reread some passages of the description.
   c. The subject must then intuitively put a cross on a 10-cm line (analog scale) to indicate his or her estimation of the risk related to the scenario. This scale, presented on a separate sheet of paper, may not be consulted by the subject later on.
2. Fill in the questionnaire for the tool on the first row of the Excel worksheet (Appendix D):
   a. The subject chooses the appropriate level for each parameter in the tool. The researcher may not supply any information. Subjects may change their choices, provided they haven't said they have finished.
   b. Once the subject has finished, the researcher indicates on the questionnaire the risk level obtained with the choices made by the subject.

c. Subjects are then asked to indicate their level of satisfaction with the result obtained (i.e., 1. Totally agree, 2. Somewhat agree, 3. Neither agree nor disagree, 4. Somewhat disagree, 5. Totally disagree) and explain their rating.

d. The researcher then saves the questionnaire and closes it.

3. Proceed in the same way with the other tools, starting at step 2.
4. Proceed in the same way with the other scenarios, redoing steps 1 to 3.

Applying the 20 parameters, divided into six categories, to the four scenarios should take an average of 2 hr, 45 min. The experimental principle is the same as for the tools (with the types of parameters replacing the tools). Each time subjects choose a parameter level, they are asked to rate how hard it was to make their choice (i.e., 1. Very easy, 2. Fairly easy, 3. Neither easy nor hard, 4. Fairly hard, 5. Very hard, 6. Does not apply) and explain their rating (Figure 18).

Section 3/7: Tool 33

3.1 Tool 33: Severity of injury or illness

- ☐ Moderate injury or illness
- ☐ Serious injury or illness
- ☐ Grievous injury or illness, or death

3.2 The descriptions, definitions and number of levels of this parameter made my choice:

- ☐ Very easy
- ☐ Fairly easy
- ☐ Neither easy nor hard
- ☐ Fairly hard
- ☐ Very hard
- ☐ Does not apply

3.3 Explain your answer (e.g., was the parameter definition clear, were the meanings of the different levels easy or hard to understand, was the number of levels adequate, were the terms used vague or clear, was sufficient information provided?)

____

____

____

Figure 18. Example of a questionnaire focusing on a specific parameter (severity of harm – tool 33)