POSTPRINT VERSION. The final version is published here :
Gauthier, F., Chinniah, Y., Burlet-Vienney, D., Aucourt, B. et Larouche, S. (2018). Risk assessment in safety of machinery: Impact of construction flaws in risk estimation parameters. Safety Science (Vol. 109, p. 421-433).
doi: https://doi.org/10.1016/j.ssci.2018.06.024

Safety Science

Risk Assessment in Safety of Machinery: Impact of Construction Flaws in Risk Estimation Parameters

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

[1]Department of Industrial Engineering
Université du Québec à Trois-Rivières (UQTR),
C.P. 500 Trois-Rivières, (Québec), Canada, G9A 5H7
francois.gauthier@uqtr.ca

[2]Department of Mathematics and Industrial Engineering
Polytechnique Montréal, Montreal, Canada

[3]Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST)
Montreal, Canada

Abstract

In the risk assessment approach as defined in the International Standard ISO 12100 :2010, risk estimation is an essential step that allows machinery designers and users to determine the level of risk, and to identify the most critical hazardous situations. Previous studies allowed to demonstrate that the numerous qualitative tools proposed to estimate risks in safety of machinery take several forms, and that many of their features can significantly influence the level of risk obtained.

In this study, the impact of some of these features was assessed, and construction rules regarding the parameters used in risk estimation tools were validated through an experimental study involving several users mainly from the industry. Five potential construction flaws of the risk estimation parameters were analyzed. The experimental results show that when the users perceive a certain challenge in the utilization of a risk estimation parameter, they are usually able to associate it with the presence of the flaw affecting the parameter. The results also demonstrate quite clearly that the impact of the construction flaws in the parameters is not uniform. In addition to the presence of the flaws within these parameters, the results obtained suggest that the assessment of the probability of harm is a problematic aspect of the risk estimation process in safety of machinery that requires further research. These results could contribute to the improvement of the robustness and the reliability of the existing tools, and help to support the training actually given by the partners in the risk assessment field.

Key words: Risk Assessment; Risk estimation tools; Safety of machinery

* Corresponding author

1. INTRODUCTION

1.1 Context of the Research

Machine-related hazardous situations have resulted in serious accidents in industries [1, 2, 3]. In order to reduce these hazardous situations, machines must be designed or modified by integrating means of risk reduction. Without making a specific risk assessment, it is difficult to choose optimized means of risk reduction [4, 5, 6, 7]. Risk assessment is a series of steps used for examining the hazards associated with machinery. It can be divided into two phases, namely (i) risk analysis, and (ii) risk evaluation as explained in international standard ISO 12100 (2010) *Safety of machinery -- General principles for design -- Risk assessment and risk reduction* (Figure **1**) [8]. Risk analysis usually consists of three stages, namely (i) determining the limits of the machinery, (ii) hazard identification, and (iii) risk estimation. The risk assessment process is followed by the risk reduction process with an iterative approach (illustrated in broken lines). The risk assessment process comes to an end when the risk has been adequately reduced.

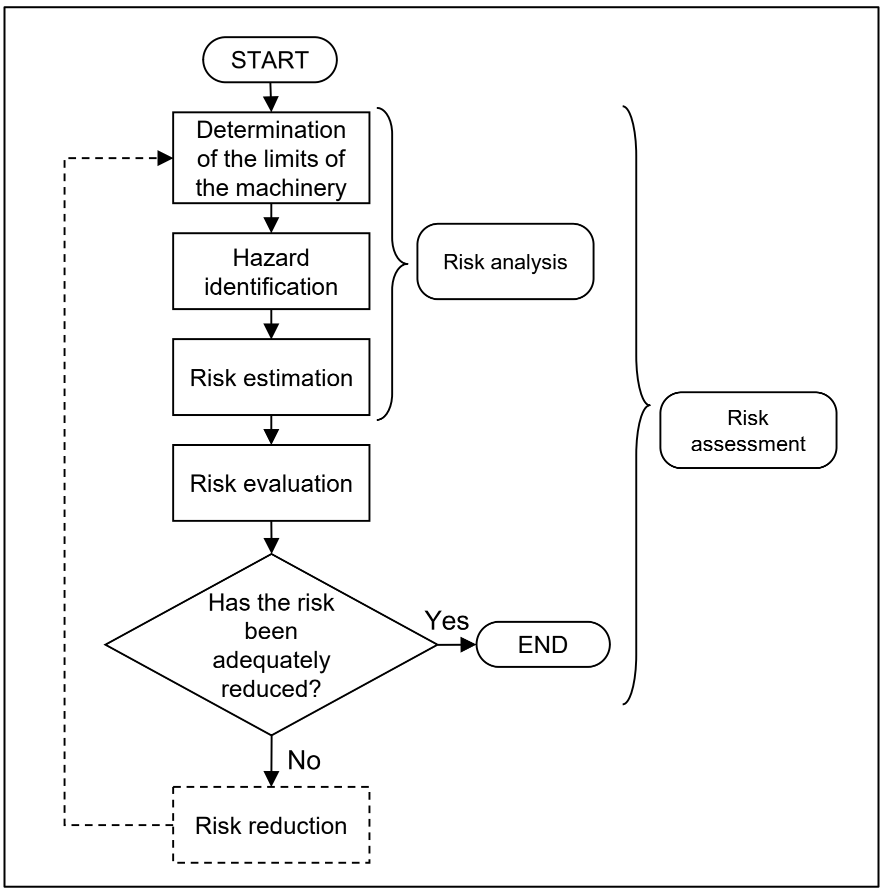

Figure 1. Simplified management of risk assessment based on standard ISO 12100.

This article puts emphasis on the risk estimation stage, which consists in estimating the inherent level of risk for each hazardous situation linked to the use of a machine. Figure 1 shows that risk estimation is the last step in the risk analysis process, then followed by the evaluation and the reduction of the risk. It is hence a critical stage for the prioritization of risk reduction activities. An incorrect estimation of the risk can lead to the implementation of insufficient or inadequate reduction measures on a machine.

1.2 Literature Review

According to Standard ISO 12100 :2010 [8], the risk related to a hazardous situation depends on a combination of the two following parameters: 1) the severity of harm, and 2) the probability of occurrence of that harm (in further text: probability of harm). The probability of harm can be established from a) the exposure of the person or persons to the hazardous phenomenon, b) the occurrence of a hazardous event, and c) the technical and human possibilities to avoid or limit the harm. Many risk estimation tools using different combinations of these parameters are proposed by organizations involved in the safety of industrial machines, and some companies have established their own methods and tools of analysis [9]. The primary objective of a risk estimation tool is to rank the different hazardous situations (scenarios) as per the risk indexes they represent in order to identify intolerable (unacceptable) risks and to prioritize interventions. Figure 2 presents an example of such tool, a two parameters risk matrix. In this example, both parameters use a three levels scale.

Probability of harm	**Severity of harm**		
	Slightly harmful	**Harmful**	**Extremely harmful**
Highly unlikely	Trivial	Tolerable	Moderate
Unlikely	Tolerable	Moderate	Intolerable
Likely	Moderate	Substantial	Intolerable

Figure 2. Example of a two parameters estimation tool

Etherton [10] confirmed that the results of risk estimation are obtained by gathering and analyzing the qualitative information related to the severity of the injuries or harm to the health, and also to the probability of occurrence of the events that could lead to those harms. However, some experts in machinery risk estimation observed that "*the tools used in different European countries to assess the risks related to a machine, when such methods exist, can give different results, even contradictory. In some situations, they can lead to different levels of safety for a given machine…"* [11]. A certain difference in the risk estimation results can be considered as "normal," hence tolerable, but a gap that is too important can eventually lead to the implementation of inappropriate risk reduction measures (insufficient or excessive) [12]. Abrahamsson [13] emphasized the fact that some potential users of the risk estimation tools give them little credibility and regard them as unusable. This same author actually attempted to validate different risk estimation tools, particularly in the context of chemical exposure at work [14]. His research was exclusively oriented towards the analysis of the variables related to the tool (e.g. model, parameters), without analyzing the other variables that can impact risk estimation (e.g. prior training, features of the people conducting the risk assessment). He concluded that the uncertainty is inherent to risk estimation, but that the guidelines specific to various industrial sectors could help improve this process.

The utilization of common risk estimation tools in the field of safety of machinery requires the interpretation of information that is often of a qualitative nature, usually using an ordinal scale as defined by Stevens [15]. Yet, many risk estimation tools are not accurate or detailed enough [16]. For instance, a verbal qualitative scale of the type *Highly unlikely, Unlikely* and *Likely* is used in certain tools to determine the probability of harm. Without any other explanation, what is the exact meaning of *Unlikely*? This type of construction can lead to bias within the estimation process, and can significantly affect the final result [17-21]. Beyth-Maron [22] reported the results of an experiment conducted in an organization specialized in professional forecasting, with people accustomed to do verbal assessment of probabilities ("probable," "possible," etc.). This study highlighted the communication issues due to the verbal terms used to express the probability. It also revealed that there is a great variability between the interpretations of the verbal terms used to express the probability, as Hubbard & Evans [23] confirmed it.

Nevertheless, and despite the issues inherent to ordinal qualitative scales described in the literature [23-26], it is worth considering the intensive use of these scales to assess risks in the fields where quantitative data is not readily available, as it is the case in safety of machinery. There are also numerous benefits linked to the utilization of ordinal qualitative scales in risk estimation tools, such as providing a simple approach and a systematic framework for the assessment of hazardous situations [27]. Consequently, as many authors suggest it, some research needs to be carried out in order to better characterize the conditions in which they are more likely to be useful or detrimental for the decision-making process in risk management [20, 28, 29].

Considering i) the increasing use of qualitative risk estimation tools in the field of safety of machinery, ii) the great diversity of these tools, and iii) the significant gap between the results they generate, Paques and Gauthier have launched in 2004 a research program aiming at deeply analyzing the features of the tools suggested in the literature or used in the industry [30]. Two prior studies conducted in this research program showed that the numerous tools allowing to conduct risk estimation are of very diverse types, and that many of their characteristics can considerably influence the level of risk obtained [16, 30, 31]. These studies also demonstrated that the tools showed significant differences regarding the risk estimation for the same hazardous situation. The field of application of the tool, its configuration and the details of its parameters seem to be the main contributing factors attributed to this variability in the results.

During these researches, flaws and bias in the construction of those tools, likely to influence the result in certain circumstances, have been identified [31]. A series of construction rules for the risk estimation parameters aiming at eliminating these flaws has also been suggested [16,32]. These flaws and construction rules are shown in Table 1.

These rules could potentially solve the variability issues observed in risk estimation. They could also guide the users to make a choice or help in the improvement of the existing risk estimation tools. However, additional research, through an experimental study involving users from the industry, was necessary in order to confirm the impact of the flaws of the risk estimation tools, and to validate the construction rules suggested.

Table 1. Flaws and construction rules of the risk estimation parameters [27]

Label of the flaw	Construction rules suggested
No definition of the range of exposure (probability parameter only)	Defining the probability parameters related to the range of exposure
Poor definition of the levels	Avoiding the use of unique or vague terms to define the levels of the parameters
Inconsistent definitions of the different levels	Avoiding the use of the same term or expression in the description of two levels of a parameter
Inadequate number of levels	Using between three and five levels for the severity of harm parameter
	Using between three and five levels for the probability of harm parameter
Gap between the levels	No discontinuity or gap between the levels of the parameter

1.3 Goals of the Research

In essence, risk estimation tools aim at distinguishing in a qualitative manner the most critical risks from the less critical ones. These tools need to be designed in a way that the level of risk obtained by different users when applying a tool to a given scenario is more or less similar. Moreover, the levels of risk obtained by different tools for the same scenario should show a certain similarity. Finally, when different tools are applied to different scenarios, they should allow to classify the risks in the same order.

The underlying hypothesis of this study is that a variability that is too important in risk estimation can be attributed to the presence of flaws or bias in the construction of the parameters used in the risk estimation tools. In other words, the flaws or bias in the parameters increase the scattering of the results (i.e. the selected level of the parameter or the estimated level of risk) among the users.

In a previous paper, the impact of flaws in the architecture of the risk estimation tools were addressed [33]. In this second part of the study, the impact and the grounds of the previously identified [16,32] flaws affecting the risk estimation parameters are analyzed. It is understood that these flaws in the parameters, or the non-compliance with the construction rules, increase the scattering of the experimental results, or influence the selected level of the parameter.

2. METHODS

2.1 Preparatory Phase

2.1.1 Selection and Preparation of Scenarios Involving Hazardous Situations

Four scenarios involving hazardous situations related to industrial machinery have been selected among 20 scenarios developed and validated during a previous study [16]. They were chosen because they include different levels of risk. The retained scenarios, coded with the letters A, G, M, and S, are summarized in table 2. The table also presents, for each scenario, the reference levels. These reference levels correspond to the severity of harm level, the probability of harm level, and the risk level for each of these four scenarios. Because of its qualitative nature, risk estimation in safety of machinery cannot be clearly quantify. Thus, these reference levels derive from the average result of the application of 31 risk estimation tools, and validated by a team of six researchers from Canada and the United Kingdom [16].

For the purpose of this experimentation, each scenario included an image of the machine, a description of the activity, the hazardous phenomenon, the hazardous situation, and the hazardous event; as well as data concerning the probability of the hazardous event, the possible harm, the exposure, and the possibility of avoidance. This segmentation of the information used to describe a scenario is based on the components of the accidental process and the risk, according to standard ISO 12100: 2010 [8].

Table **3** shows, as an example, the Scenario S (robot), as presented to the participants during the experiment.

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

Scenario	Title	Summary of the work activity	Reference levels		
			Severity of harm	Probability of harm	Risk level
A	Punching machine with mobile table	Functional demonstration of a punching machine during a trade show	Low	Mid-high	Low
G	Automated guided vehicle	Moving of an automated guided vehicle (without a driver) across a plant by following a route painted in yellow on the ground.	Mid-Low	Mid-High	Mid-low
M	Rewinder (paper machine)	Finishing tasks on a rewinder (removing of the irregular parts of the roll) while the rewinder is switched on manual mode	Mid-high	Mid-Low	Mid-high
S	Robot	Tool changing by a worker on a numerically controlled lathe. Presence of a robot that supplies the lathe with metal pieces	High	Mid-low	High

Table 3. Description of Scenario S (Robot)

Robot	
Activity	Tool changing by a worker on a numerically controlled lathe used to process metal parts. A robot supplies the lathe with metal pieces to be processed, and then removes them once processed.
Hazardous phenomenon	Movement of the robot towards the worker.
Hazardous situation	The worker is on the trajectory of the robot. The robot is switched on, and stays in the stand-by position.
Hazardous event	The worker is hit by the robot. The robot received a start command caused by a failure of the programmable logic controller (PLC) that commands it.
Probability of the hazardous event	The robot is controlled by a standard PLC and not by a safety PLC.
Possible harm	Multiple fractures, concussion, death.
Exposure	Intervention of 10 minutes, 2 times per 8-hour working shift.
Possibility of avoidance	The worker turns his back to the robot and wears a hearing protector. The extremity of the robot moves very quickly (approximately 2 m/s). The robot has a visual and audible warning. No previous or recent past failures listed.

2.1.2 Selection and Preparation of the Risk Estimation Parameters

The two main risk estimation parameters advised by Standard ISO 12100: 2010 [8] have been considered:

- Severity of harm (S)
- Probability of harm (Ph)

The aim of the experiment was to test the impact of the flaws shown in Tab on those two types of parameter. Eleven combinations of the parameters/flaws were selected from a previous study [16]. The origin of each parameter retained, the description of its levels, as well as the flaw identified for each are shown in Table 4 and Table 5. For comparative purpose, a parameter considered with no identified flaw (labelled “no flaw”) was also added to the list for each case.

Table 4. Severity of harm parameters tested and associated flaws

Flaw	Descriptors of the levels of the parameter as found in the tools	Original tool reference number from [13]	Reference
Poor definition of the levels	– Moderate injury or illness – Serious injury or illness – Death / grievous injury or illness	#33	Main (2004) pp. 155-157 [34]
Poor definition of the levels	– 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	#55	Company X [35]
Inconsistent definitions of the different levels	– 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	#66	IEC 62278 : 2001 [36]
Inadequate number of levels	– 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, major musculoskeletal troubles (MST), fatalities	#91	ISO 14121-2:2007 [37]
Gap between the levels	– 1) Minor: means that the consequences are not very serious – 2) Significant: means that works has to stop, first aid is really needed – 3) Disastrous: means that there has been a very serious accident (someone has been scarred for life, blinded or even killed)	#102	Gondar (2000) [38]
No flaw	– No harm – Low: trivial harm with no permanent results – Middle: serious harm with no permanent results – High: serious harm with permanent results, death	#69	Görnemann (2003) [39]

Table 5. Probability of harm parameters tested and associated flaws

Flaw	**Levels of the parameter**	**Original tool reference number from [13]**	**Reference**
No definition of the range of exposure	– Very unlikely: could happen, but probably never will – Unlikely: could happen, but rarely – Likely: could happen occasionally – Very likely: could happen frequently	#89	The Metal Manufacturing and Minerals Processing Industry Committee (2002) [40]
Poor definition of the levels	– Remote – Improbable – Possible – Probable – Likely	#7	Raafat, H. (1995) [41]
Inconsistent definitions of the different levels	– Improbable – so unlikely that the probability is close to zero – Remote – unlikely, though conceivable – Possible – could occur sometime – Probable – no surprise, will occur several times – Likely / frequent – occurs repeatedly / event only to be expected	#6	Kazer, BM. (1993) [42]
Gap between the levels	– Low – very seldom or never occurs – Medium – reasonably likely to occur – High – certain or near certain to occur	#34	Main (2004) p. 164-165 [34]
No flaw	– F- Highly improbable – probability cannot be distinguished from zero – E- Improbable – very unlikely to occur in the life cycle – D- Remote – unlikely, but may possibly occur in the life cycle – C- Occasional – likely to occur at least once in the life cycle – B- Probable – likely to occur several times in the life cycle – A- Highly probable – likely to occur frequently in the life cycle	#41	ISO/TS 14798 : 2006 [43]

2.1.3 <u>Selection of the Participants for the Experimentation</u>

The participants in the study had a sufficient level of knowledge and experience in machinery safety and risk analysis related to machines. This information was verified during the recruitment process through a questionnaire aiming at identifying the participant. A sample of 25 participants was selected. These participants were equally divided between (i) advisors from industry associations in OSH, (ii) maintenance staff or safety practitioners in enterprise, and (iii) engineers specialized in safety of machinery.

2.2 Experimental Protocol

The experimental protocol was developed, then tested under real-life conditions by two members of the research team and two potential participants. All the data collection tools were validated by two members of the research team who did not participate in their creation in order to ensure the functionality of the system and the accuracy of the wording of the tools. The experiments were carried out at the workplace of the participants. The researcher conducting the experimentation had a grid developed with Microsoft Excel© in which a URL for each of the cases of the experiment was available (i.e. parameter X/scenario Y; 11 parameters tested on 4 scenarios). Each URL directed to an on-line questionnaire. The participant was accompanied at all time by the researcher, and the average duration of the experimentations was about three hours. In order to limit bias, the order of the questionnaires was random from a participant to another. The scenarios were presented one by one in the series of questions.

2.3 Analysis of the Experimental Results

The impact of the different flaws on the risk estimation parameters was analyzed from the results gathered during the experiment, i.e. for the application of each parameter to each scenario:

1. The level of the parameter chosen by the participant;

2. The difficulty of making that choice, as indicated by the participant;

3. The comments expressed by the participant.

2.3.1 The Level of the Parameter Chosen by the Participant

The levels chosen during the application of each parameter to each scenario by the 25 participants were broken down into percentages according to the different scales of each parameter. The mode was then established for each case. The mode is defined here as the level of a given parameter chosen by the highest number of participants for a given scenario. The mode percentage is hence the proportion of the 25 participants who answered the same level for a given parameter, applied to a given scenario. This percentage allowed to provide an indication regarding the *convergence* (or divergence), that is the repeatability of the participants' answers depending on each flaw and each parameter (inter-participant repeatability). Thus, the higher the mode percentage is, the better is the convergence of the results of the application of a parameter. The mode percentage being affected by the number of levels of each parameter, only the values lower or equal to 60% were considered as an indication of a poorer convergence of the results.

2.3.2 The Difficulty of Making that Choice, as Indicated by the Participant

An analysis regarding the facility to apply each parameter was also carried out. During the experiment, the participants had to indicate, on a scale from 1 to 5, the level of difficulty they experienced with respect to the application of each parameter to each scenario, according to the statement: "The descriptions, the definitions or the number of levels of this parameter made my choice…" The number of participants who chose the levels of difficulty 4 (*quite difficult*) or 5 (*very difficult)* accounted for as an indicator of the level of difficulty.

2.3.3 The Comments Expressed by the Participant

In addition, the comments written by the participants in each occasion were recorded and classified in three categories:

1. General comments: Neutral or positive comments indicating for instance the facility to apply the parameter in a given situation or highlighting a feature of a parameter. Example: "*No overlaps, simple and clear description*."

2. Negative comments related to the flaw involved: Comments expressed by the participants that directly point out the studied flaw in the parameter. The number of negative comments written by the participants related to each flaw was used as an indicator for the perception and the impact of the flaws under study. The nature of these comments also allowed to better understand the impact of the different flaws in the risk estimation process. In order to identify

the comments linked to the different flaws, linguistic tables were developed for each flaw. These tables provided key words or expressions, allowing to link the comment to the flaw. For example, the underlined expression in the comment "*There is a gap that is not covered <u>between the two</u> levels*" was liked to the flaw "Gap between the levels".

3. Negative comments without any link to the flaw involved: Comments indicating a challenge or pointing out a negative aspect of the application of the parameter, without any link to the flaw under study. These comments were analyzed in order to identify the presence of other issues (regarding the construction of the parameters) that could have been raised by the participants.

From these data, the impact of the presence of the different flaws within these risk estimation parameters was globally and separately assessed for the two types of parameters under study. Cross analyses of these data were also conducted.

3. RESULTS

3.1 Overall Convergence of the Results

The analysis of the overall convergence of the results was established according to the mode percentage of the participants' answers, as defined in section 2.3.1.

Table 6 shows the mode percentage obtained for the two studied parameters in accordance with the different flaws that they present and for all four scenarios. An analysis of the table demonstrates primarily that the severity of harm parameters generated fairly consistent results from a participant to another, with an average mode percentage of 83%. On average, 21 of the 25 participants showed the same result regarding the application of the severity of harm parameters for all the studied scenarios. However, the probability of harm parameters offered a poorer performance in terms of convergence of the results, with an average mode percentage of 49%. For these parameters, it seems that it was more challenging for the participants to establish a consensual level.

Table 6. Mode percentage obtained for each type of parameter depending on the flaws

Flaws / Parameters	S	Ph	Avg. by flaw
No definition of the range of exposure	-	47%	47%
Poor definition of the levels	86%	47%	67%
Inconsistent definitions of the different levels	70%	51%	61%
Gap between the levels	86%	52%	69%
Inadequate number of levels	90%	-	90%
Average per parameter	**83%**	**49%**	**66%**
No flaw	**89%**	**46%**	**68%**

3.2 Overall Analysis of the Difficulty of Application

The following tables present, for all the scenarios, the overall results related to the difficulty of application of the parameters in the presence of the studied flaws. Table 7 displays the average number of participants who indicated, for each of the four scenarios, that it was *quite difficult* or *very difficult* (answers 4 or 5) to make a choice of level for each type of parameter, according to the studied flaws. As can be seen, the flaw “Poor definition of the levels” seems to have caused a greater challenge to the participants, with an average of 6.6 participants who indicated a level of difficulty of 4 or 5 in each occasion. The flaw “Inconsistent definitions of the different levels” follows, with an average of 6.3 participants.

Table 7. Average number of participants who indicated that it was *quite difficult* or *very difficult* to make a choice of level for each type of parameter according to the studied flaws

Flaws / Parameters	S	Ph	Avg. by flaw
No definition of the range of exposure	-	4.3	4.3
Poor definition of the levels	5.8	7.3	6.6
Inconsistent definitions of the different levels	7.5	5.0	6.3
Gap between the levels	2.8	6.5	4.7
Inadequate number of levels	2.3	-	2.3
Average per parameter	**4.6**	**5.8**	**5.2**
No flaw	**2.3**	**6.8**	**4.6**

Table 8 shows the number of negative comments related to the flaw submitted by the participants in each case. Figure 3 illustrates the relationship between the number of participants who indicated a level of difficulty of 4 or 5 and the number of negative comments related to the flaws. It can be noted that the correlation between these two results is relatively good, with $R^2 = 0.87$. Thus, when the participants had trouble to make their choice, they were generally able to attribute this difficulty to the presence of the flaws.

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

Flaws / Parameters	S	Ph	Avg. by flaw
No definition of the range of exposure	-	1.0	1.0
Poor definition of the levels	10.6	13.0	11.8
Inconsistent definitions of the different levels	11.3	9.3	10.3
Gap between the levels	5.0	6.0	5.5
Inadequate number of levels	5.8	-	5.8
Average per parameter	**8.2**	**7.3**	**7.8**

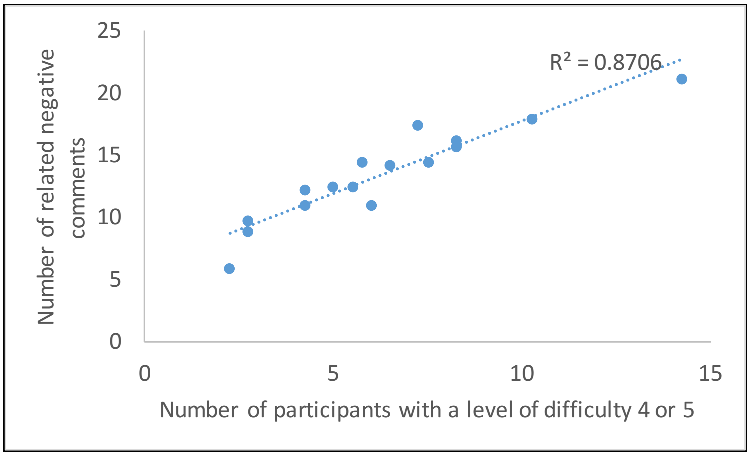

Figure 3 – Relationship between the number of participants who indicated a level of difficulty of 4 or 5 and the number of negative comments related to the flaws

In-depth analyses of these general trends in the two types of parameter (S and Ph) are addressed in the following sections.

3.3 Analysis of the Impact of the Flaws on the Severity of Harm Parameters (S)

Table 9 presents the date for the severity of harm parameters. Six severity of harm parameters were used to assess the impact of the different flaws. The flaw "Poor definitions of the levels", being very common within the risk estimation tools, was assessed with two parameters.

3.3.1 Impact of the Flaws on the Convergence of the Results

As mentioned previously, the convergence of the results for this parameter is good, with an average mode percentage of 83%. This can indicate that this parameter is relatively solid, and that the different flaws studied have little impact on the determination of the severity of harm level. Nonetheless, four cases draw the attention in Table 9, with a mode percentage lower or equal to 60%.

The first two cases relate to the flaw "Inconsistent definitions of the different levels" found in the severity of harm parameter of the tool #66, applied to the scenarios A and G (Table 9). These two scenarios show the lowest severity of harm reference levels. For this parameter, the expression "*minor injury*" is used in the first two levels of the severity of harm parameter of the tool #66 (ref. Table 4). This inconsistency in the definitions of the levels seems to have led to a greater answer scattering for the scenarios where these levels are potential options, i.e. scenarios with a lower level of severity.

Table 9. Analysis of the severity of harm parameters

	Flaws	No definition of the range of exposure	Poor definition of the levels	Inconsistent definitions of the different levels	Inadequate number of levels	Gap between the levels	No flaw
	Original tool reference number from [13]	#33	#55	#66	#91	#102	#69
	Number of Levels	3	4	4	2	3	4
Scenario A	Mode %	100%	80%	60%	100%	56%	92%
	Nbr of participants who had difficulty to answer	2	5	8	0	4	1
	Nbr of negative comments related to the flaw	4	14	19	3	4	0
	Total nbr of negative comments	13	17	19	3	12	6
Scenario G	Mode %	68%	64%	56%	60%	88%	68%
	Nbr of participants who had difficulty to answer	9	12	16	9	6	8
	Nbr of negative comments related to the flaw	14	17	17	11	7	0
	Total nbr of negative comments	18	19	22	12	13	14
Scenario M	Mode %	88%	100%	92%	100%	100%	96%
	Nbr of participants who had difficulty to answer	9	1	4	0	1	0
	Nbr of negative comments related to the flaw	13	6	6	6	6	0
	Total nbr of negative comments	15	9	11	6	8	4
Scenario S	Mode %	100%	87%	74%	100%	100%	100%
	Nbr of participants who had difficulty to answer	2	6	2	0	0	0
	Nbr of negative comments related to the flaw	3	14	3	3	3	0
	Total nbr of negative comments	5	20	6	3	3	2

The third case is observed for the flaw “Inadequate number of levels” of the tool #91, applied to the Scenario G. This parameter has only two levels: “*S1 - slight injury*” and “*S2 - serious injury”* (ref. Table 4). These levels do not really fit with the possible harm “*contusion, simple fracture*” of the Scenario G, that falls in-between. Hence, 40% of the participants opted for the level S1, and 60% for the level S2.

The fourth case involves the flaw “Gap between the levels” in the severity of harm parameter of the tool #102, applied to the Scenario A. For this parameter, the first level of severity is defined by “*Minor: means that the consequences are not very serious,*” while the second level is expressed by “*Significant: means that works has to stop, first aid is really needed*” (ref. Table 4). The severity reference level is low for the Scenario A, with “*Ecchymosis, cuts*” as information provided on the possible harm. In this case, it seems that it could be the lack of a gap between those two contiguous levels that would have influenced the participants’ answers: “*Ecchymosis, cuts*” could have been considered as a harm with a low level of severity, but that could also require a first aid intervention.

For this same severity of harm parameter (tool #102), there is a significant gap between the two higher levels of its scale, which goes from a first aid intervention to an injury causing permanent harm. The presence of this flaw seems to have promoted a better convergence of the results for the scenarios with a high level of severity (scenarios M and S).

3.3.2 Impact of the Flaws on the Difficulty to Apply the Severity of Harm Parameters

In addition to the mode percentage, Table 9 also shows the number of participants who indicated that it was *quite difficult* or *very difficult* to make their choice of level for each case, as well as the number of negative comments submitted by the participants. In a few cases, almost the third of the participants (8 out of 25) found it *quite difficult* or *very difficult* to make their choice. Two cases draw the attention concerning the flaw "Inconsistent definitions of the different levels" within the severity of harm parameter of the tool #66. The scenarios A and G generated a relatively low mode percentage, as discussed previously (see 3.3.1). The challenge faced by the participants hence caused a more significant scattering of their answers. This observation incidentally seems to be supported by the comments made by the participants. As Table 9 illustrates it, 19 participants submitted negative comments related to this flaw for the Scenario A, and 17 for the Scenario G. Among these comments are the following:

- *Between insignificant and marginal, the gap is thin. They overlap.*

- *Two choices of minor injuries because of the choice "and/or."*

- *Difficulty to make the distinction between negligible and marginal. The notion of "possible" does not necessarily help me in my selection.*

The participants also noted that this parameter associates a notion of probability to the determination of the severity by the use of the word "*possible*" in the first level (ref. Table 4), as well as the presence of the notion of environmental harm. These elements seemed to add to the inconsistency in the definitions of the levels. It can hence be noted that the presence of this flaw had an influence on the convergence of the results. Many participants attributed their difficulty to make a choice to the presence of this flaw in the parameter.

The presence of the flaw "Poor definition of the levels" within the severity of harm parameter of the tool #33 gave rise to many negative comments. For instance, for the Scenario M, 9 participants indicated that they had difficulty making their choice of level, and 13 participants wrote a comment pointing out this flaw. The Scenario M (Rewinder) indicates "*Partial or total amputation of the upper limbs*" as information on the possible harm. For the severity of harm parameter of the tool #33, the second level of severity is defined by "*Serious injury or illness*" while the third level is defined by "*Death / grievous injury or illness.*" Although the mode percentage is relatively good (88%), it can be assumed that the participants might have had difficulty to distinguish if the amputation was a "*serious injury*" or a "*grievous injury*." This hypothesis seems to be confirmed by the negative comments related to the flaw submitted by the participants:

- *Arbitrary choice between* serious *or* grievous*.*

- *Interpretation of* serious *and* grievous*. According to me, an amputation is really serious… An interpretation is necessary.*

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

- *Defining serious or grievous illness, is an amputation grievous?*

- *Difference between serious and grievous?*

The presence of the flaw "Poor definition of the levels" in the severity of harm parameter of the tool #55 also lead to many negative comments. The participants often noted the lack of details regarding the definition "*Marginal: minor injury or occupational illness"* versus "*Critical: severe injury or occupational illness*." Once again, the vague and imprecise definitions of the levels was pointed out by many participants.

Table 9 also shows that it is for the Scenario G that the participants indicated having had the most difficulty to make their choice. These difficulties are also reflected in the mode percentage that is systematically the lowest of the four scenarios. The scenario G (Automated guided vehicles) indicates "*contusion, simple fracture*" as possible harm. This statement, combined with the comments made by the participants, suggests that the flaws in the severity parameters could have a more significant impact on the determination of the level of severity of medium harm like the simple fracture (scenario G) in comparison for really low harm (for instance, the ecchymosis in the Scenario A) or really high harm (for example, the amputation in the Scenario M and the death in the Scenario S).

Overall, the flaw "Inadequate number of levels" found in the tool #91 had a low impact on the convergence of the results (except for scenario G) or the level of difficulty perceived by the participants. It is also for this flaw that the total number of negative comments submitted by the participants is the lowest, with 24 comments against 43 in average for the other flaws. However, almost all of these negative comments (23 out of 24) specifically point out this flaw, for example:

- *It is good because there are examples. However, there are not enough categories.*

- *I would have liked to have more levels (I like the fact that there were examples).*

- *Not enough levels.*

- *Not good: simple fracture is on the same level as death. Not enough levels (too Boolean).*

- *One level missing between the two; the fracture is not addressed.*

- *Not enough levels, choice by default.*

The participants could easily make a choice of level, but many of them thought this choice was constraining and uncomfortable, given that the parameter included only two levels.

3.3.3 Observations Regarding the Impact of the Flaws on the Severity of Harm Parameter

The observations and analysis presented in the previous sections bring forward the following statements regarding the impact of the construction flaws on the severity of harm parameter:

1. The severity of harm parameter seems relatively solid, and the convergence of the results is little affected by the presence of the flaws in most cases.

2. The number of negative comments expressed by the participants related to the different flaws represents 67% of all the negative comments (173 out of 260). This seems to indicate that despite the good convergence of the results, the participants were generally able to identify the flaws and to determine their negative influence on the choice of the severity of harm level.

3. The flaws do not have a uniform impact on the application of the severity of harm parameters. The nature of the flaw, its position on the severity scale and the possible harm seem to influence the impact of the flaw. For instance, some flaws could have no effect for the lowest and the highest harms, but impact more the level of severity of medium harm, and vice versa.

4. The flaws "Poor definition of the levels" and "Inconsistent definitions of the different levels" seem to have had the most impact on the participants. These flaws gave rise to 174 out of the 260 negative comments submitted by the participants, and 130 of these comments specifically pointed out these flaws.

5. The flaw "Gap between the levels" shows variable effects depending on the case. On the one hand, a gap too important between the levels of a parameter had little impact on the convergence of the results and on the difficulty to answer, as the results of the application of the tool #102 suggest. Nevertheless, this flaw was raised by the participants who made 20 comments pointing it out directly. At the reading of these comments, it seems that this gap directs the choice of the participants towards a specific level in each case. However, in the majority, the participants mentioned that there is a level missing on the scale, which is translated by a certain discomfort in the selection process. On the other hand, the lack of a gap between two levels can have a notable effect on the convergence of the results. Incidentally, still regarding the tool #102, 9 of the participants made a comment supporting this interpretation during its application to the Scenario A, for which a mode percentage of only 56% was obtained.

6. The inclusion of only two levels to establish the severity of harm can make it easier for the users to make a choice. However, this situation seems to generate a certain discomfort. Hence, this flaw has a similar effect as would a too important gap between two levels.

3.4 Analysis of the Impact of the Flaws on the Probability of Harm Parameters (Ph)

Table 10 shows the data for the probability of harm parameters. Five probability of harm parameters were used to assess the impact of the different flaws.

3.4.1 Impact of the Flaws on the Convergence of the results

As mentioned previously, the convergence of the results for this parameter is low, with an average mode percentage of 49%. Even the parameter of the tool #41, presumably with "no flaw," did not allow to obtain a better convergence of the results. These results are actually similar for the four studied scenarios, and indicate that this parameter presented a great challenge to the participants. It is hence difficult to establish the specific impact of the flaws on these parameters from the mode percentage.

Table 10. Analysis of the probability of harm parameters

	Flaws	No definition of the range of exposure	Poor definition of the levels	Inconsistent definitions of the different levels	Gap between the levels	No flaw
	Original tool reference number from [13]	#89	#7	#6	#34	#41
	Number of Levels	4	5	5	3	6
Scenario A	Mode %	48%	36%	44%	44%	68%
	Nbr of participants who had difficulty to answer	3	6	4	7	6
	Nbr of negative comments related to the flaw	1	16	12	5	0
	Total nbr of negative comments	10	17	14	14	18
Scenario G	Mode %	60%	40%	44%	52%	40%
	Nbr of participants who had difficulty to answer	4	4	2	5	5
	Nbr of negative comments related to the flaw	1	11	7	5	0
	Total nbr of negative comments	11	17	11	12	12
Scenario M	Mode %	46%	63%	63%	63%	38%
	Nbr of participants who had difficulty to answer	5	10	5	8	7
	Nbr of negative comments related to the flaw	0	13	10	9	0
	Total nbr of negative comments	12	20	13	19	14
Scenario S	Mode %	33%	50%	54%	50%	38%
	Nbr of participants who had difficulty to answer	5	9	9	6	9
	Nbr of negative comments related to the flaw	0	12	8	5	0
	Total nbr of negative comments	9	16	12	12	12

3.4.2 Impact of the Flaws on the Difficulty to Apply the Probability of Harm Parameters

Regarding the difficulty to apply the parameter, Table 10 shows five cases for which almost the third of the participants (8 out of 25) found it *quite difficult* or *very difficult* to make their choice. These five cases were observed for the scenarios M and S. Two of the cases are related to the flaw "Poor definition of the levels" found within the probability of harm parameter of the tool #7. As shown in Table 4, this parameter only uses qualitative terms, without any other indications, to define the levels of its scale. Thirteen participants wrote negative comments directly implicating this flaw for the Scenario M, and 12 for the Scenario S. Among the comments are the following:

- *Defining better the descriptions of the levels.*
- *It is vague, we need to interpret. Hesitation between possible and probable… it is the same thing.*

- *The probability parameter is not clear (probability of harm versus probability (risk)). Level: We based ourselves on words… no explanation, no clarification, no scale.*

- *The description is not sufficient to allow the choice of level (possible vs probable) for a communication error.*

- *No definition; we need to interpret.*

The scenario S involves two other cases. The first case relates to the parameter of the tool #41, considered with "no flaw." An analysis of the comments submitted by the participant reveals that it uses the notion of "life cycle" to define the range of exposure allowing to establish the probability. Many participants indicated that this aspect of the definition made their choice more difficult, for example:

- *I don't like "life cycle" because it does not mean anything to me.*

- *The notion of cycle confuses things.*

- *The notion of life cycle becomes very difficult to judge in a scenario like this one.*

- *More precise, but defining the life cycle and the training of the worker.*

- *What life cycle (machine or range of exposure)?*

- *I am not certain about the life cycle of the equipment.*

Consequently, although this parameter was considered with "no flaw," yet it presented an issue in its construction.

The second case relates to the flaw "Inconsistent definitions of the different levels" found in the probability of harm parameter of the tool #6. For this parameter, 54% of the participants chose the level "*Remote – unlikely, though conceivable*," and 33% chose the level below "*Improbable – so unlikely that probability is close to zero"* (ref. Table 5). When analyzing the 12 comments submitted by the participants, it can be noted that many of them might have been influenced by the presence of the flaw, for example:

- *Hesitation between remote and improbable. The definitions are really similar.*

- *Detailed description of the levels, but vague, similar words that don't mean anything… I hence based myself on my own perception. The description of the parameter gives room for judgment (choosing between two terms with a slightly different meaning).*

- *The descriptions do not allow to clearly choose between remote and improbable. In doubt I take the riskiest one.*

- *The definition of remote is similar to the one of improbable.*

A last case relates to the flaw "Gap between the levels" found within the probability of harm parameter of the tool #34, applied to the scenario M. For this parameter, 8 participants indicated that they had a certain difficulty to make their choice, and 9 participants wrote negative comments related to this flaw. Regardless of the acceptable mode percentage (63%), many comments show uncertainty with respect to the presence of this flaw:

- *There are not enough levels. Hesitation between low and medium. The gap between never and reasonable is huge. What does reasonable mean? Result tinged by my experience.*

- *Not enough categories. The terms are not clear, lack of explanation.*

- *Not enough levels. Not enough level gradation. Choice made by default to the closest.*

- *According to me, there is at least one level missing between low and medium. This probability is not non-existent, but neither medium.*

It is, however, important to note that many of these comments refer to the fact that it sometimes exists a significant gap between the levels, but that this gap is due to the lack of levels in the scale of this parameter, which includes only three levels.

3.4.3 Observations Regarding the Impact of the Flaws on the Probability of Harm Parameters

The observations and analysis presented in the previous sections bring forward the following statements regarding the impact of the construction flaws on the probability of harm parameter:

- The probability of harm seems difficult to be established by the participants. The low convergence of the results suggests that this parameter represents a challenge in most cases.

- Although the specific impact of the flaws in this parameter is difficult to establish, the participants were often able to identify these flaws.

- The flaws "Inconsistent definitions of the levels" and "Poor definition of the levels" were the ones identified the most by the participants (number of related negative comments). These flaws gave rise to 89 out of the 115 negative comments related to these flaws. This suggests that despite the fact that the participants did not always indicate that these flaws may have caused them trouble during the selection process, they were able to recognize their presence and their potential impact.

- The flaws "Gap between the levels" and "No definition of the range of exposure" were less often recognized by the participants in the studied parameters.

- The notion of "life cycle," used to establish the range of exposure for the parameter of the tool #41 does not seem to be suitable according to many participants.

- The flaw "Gap between the levels" seems to be perceived in the same way as the flaw "Inadequate number of levels" when the number of levels of the scale is small. Moreover, in the presence of this flaw, the participants mentioned in majority that there are not enough levels in the scale, which results in a discomfort during the selection process.

4. DISCUSSION

The objective of the first study was to confirm, through a practical experimentation, the real and perceived impact of the flaws and bias within the configuration of the parameters of the risk estimation tools. Ultimately, this in-depth knowledge aims at suggesting more robust and more reliable configurations, and at defining clear criteria for the assessment of the existing tools or for the development of new tools.

Five of the potential flaws were analyzed:

1. No definition of the range of exposure;
2. Poor definition of the levels;
3. Inconsistent definitions of the different levels;
4. Inadequate number of levels; and
5. Gap between the levels.

The first general observation relates to the capacity of the participants to recognize the flaws found in the construction of the risk estimation parameters. Actually, the experimental results show that when the participants perceive a certain difficulty to apply a parameter, they are generally able to associate this difficulty to the presence of the flaw. This observation is supported by the strong correlation between the number of related negative comments and the level of difficulty expressed (figure 3).

However, the perception of the participants regarding the difficulty to apply the different parameters does not seem to be linked to the convergence of the results. Even when the level of difficulty perceived is low, the convergence of the levels chosen by the participants can sometimes be poor. The participants then choose different levels according to their individual understanding of the parameter construction and the given hazardous situations. Thus, some flaws can impact the risk estimation process even though their presence and influence is not necessarily identified by the users. It is particularly the case for the flaw “No definition of the range of exposure.”

Besides, the results also indicate clearly enough that the impact of the construction flaws of the parameters is not uniform. Important variations can be observed, not only with respect to the type of parameter, but also depending on the hazardous situation scenario analyzed. The severity of harm parameters seem relatively solid, and allow to obtain a good consensus among the users despite the presence of flaws. However, the probability of harm parameters are definitely less solid, which is usually translated by a poorer convergence. The nature of the flaw, its position on the scale of the parameter and the studied scenario also influence the impact of the flaw. This can be manifested as much in the convergence of the results than on the level of difficulty or the number of related negative comments. For instance, some flaws have no impact regarding the determination of the severity of harm when the potential harm is high, but impact more the selection of the severity levels for low or medium harm. The very variable impact of the flaw “Gap between the levels” notably allowed to make this observation. This could explain the behaviour of some risk estimation tools analyzed during a previous study, which clearly overestimated or underestimated the risk of some of the 20 hazardous situation scenarios studied [16].

The next sections discuss each of the five construction flaws of the risk estimation parameters studied.

4.1 No Definition of the Range of Exposure

This flaw is related to the probability of harm parameter (Ph). For the reasons previously mentioned (ref. section 3.4.1), it is difficult to make a conclusion regarding the specific impact of the flaws that apply to the probability of harm parameter. On one hand, of all the probability of harm parameters studied, only the parameter of the tool #41 (presumed with "no flaw") included an indication regarding the range of exposure. On the other hand, the convergence of the results is poor in every case. Moreover, the notion of "life cycle," used to establish the range of exposure in the parameter of the tool #41 was judged confusing by many of the participants.

Consequently, these results can indicate that:

- The notion of "life cycle," used to determine the range of exposure for the parameter, lacks of precision to be useful;
- The participants prefer defining themselves, and in a qualitative way, the range of exposure in the probability of harm estimation; or
- The lack of information on the range of exposure is a flaw that can influence, even bias, the level selection process of the participant, even though this influence is not necessarily perceived negatively.

This last hypothesis could be one of the explanations for the low convergence rate (mode percentage) of the results for the probability of harm parameters. Linked to this hypothesis, an exhaustive analysis of the 412 comments submitted by the participants (for all the studied probability of harm parameters applied to the four scenarios) only allowed to find 3 comments stating the lack of indication regarding the range of exposure. Clearly, the participants did not perceive this lack of information as an influential factor in the risk estimation process.

More in-depth studies are necessary in order to better understand the intellectual path leading to the qualitative estimation of the probability of harm [18,19].

4.2 Poor Definition of the Levels

With an average of 6.6 participants who indicated a level of difficulty of 4 or 5 for each application (in comparison with 5.2 for the totality of the flaws) and an average of 11.8 related negative comments expressed for each application (against 7.8 for the totality of the flaws), the flaw "Poor definition of the levels" is the one that had the most impact on the risk estimation parameters. In the light of those experimental results, the impact of this flaw can be confirmed. Both the quantitative results and the qualitative analysis of the comments made by the participants attest the considerable negative aspect of this flaw in the risk estimation process. Its impact is obvious on the two types of parameter.

Consequently, the levels of the parameters must be defined adequately to reinforce the risk estimation process. Some tools use only figurative terms or expressions (e.g. "possible" or "probable" for the parameter Ph) to define the different levels of their parameters. Yet, this leaves the users a lot of room for interpretation. Does "possible" bear the same meaning for all the users? What does "frequent enough" exactly mean? Given the lack of precision of the terms chosen, anyone that uses such tools can attribute to each level a different interpretation that differs from one to another. This interpretation issue is mitigated when a detailed definition is used. Combined with figurative terms and expressions, detailed definitions can provide the user with a better structure in which he can work, diminishing the level of difficulty and encouraging a greater convergence of the risk estimation results. These observations confirm those of other authors [18-21].

4.3 Inconsistent Definitions of the Levels

The experimental results confirm the significant impact of this flaw on the risk estimation parameters, with an average of 6.3 participants who indicated a level of difficulty of 4 or 5 for each application (against 5.2 for the totality of the flaws), and an average of 10.3 related negative comments submitted for each application (in comparison with 7.8 for the totality of the flaws).

As for the mode percentage (inter-participant convergence), this flaw generated the worst result (61%), even affecting the severity of harm parameter, which was merely impacted by the other flaws. The qualitative analysis of the comments written by the participants also confirms the negative impact of the inconsistent definitions of the parameter levels on the risk estimation process.

Thus, even though they define their levels in a relatively detailed manner, some risk estimation parameters use terms that are inappropriate, confusing or too close semantically. The examples of the parameters of the tools #6 and #66 (see Table 4 and Table 5) illustrate well the different forms that can take this construction flaw.

In order to avoid confusing some users, it is desirable to provide precise and complete definitions, and to eliminate any possible ambiguity in the designation of the different levels. The consistency of the terms used to express the gradation of the levels defined by a given parameter is also important [18,19,44-46]. As an example, the use of the expression « *minor injury* » in the first two levels of the severity of harm parameter scale of the tool #66 should be avoided. No matter which parameter is involved, its levels should show a progression from the lower to the higher, and the terms used should reflect this progression in a way to help the user to clearly distinguish the levels from one another, and to select the level that corresponds to the risky situation that needs to be estimated.

4.4 Inadequate Number of Levels

This flaw was evaluated only for the severity of harm parameter (S). The experimental results suggest that the use of only two levels to establish the severity of harm facilitates the users' choice, but that this choice is sometimes perceived as uncomfortable and constraining. For the tool #91, the binary aspect of the definition of the potential severity, based on its reversibility, seems to have indisposed many participants. With this parameter, an irreversible injury (e.g. loss of a fingertip) is considered on the same level as the death of a worker. In this case, it is possible that the user would not be able to choose the appropriate level easily. When the number of levels of a parameter is inadequate, some levels tent to cover too many different situations, if not extreme ones.

Consequently, although the low number of experimental results do not allow to make a conclusion with certainty, the question arises on what would be the impact of this type of construction on the perception and on the good deployment of the risk estimation process. Therefore, the recommendation deriving from a previous study, to the effect that the risk estimation parameter should normally include 3 to 5 levels, remains a cautious approach to be privileged. Standard

ISO14121-2 :2007 [37] indicates that the parameters should comprise a minimum and maximum number of levels, without specifying the number.

However, it is also possible that this feature does not represent a flaw for all the types of parameters used within the risk estimation.

4.5 Gap Between the Levels

The results obtained for this flaw show an average of 4.7 participants who indicated a level of difficulty of 4 or 5 for each application (against 5.2 for the totality of the flaws), and an average of 5.5 related comments submitted for each application (in comparison with 6.9 for the totality of the flaws). These experimental results tend to indicate that the impact of the flaw “Gap between the levels” is variable. Its impact can be practically non-existent on the convergence (the mode percentage) when this gap is really important: it directs or even forces the choice of the participants towards a specific level for each case. It is therefore perceived in the same way that the flaw “Inadequate number of levels.” The gap between the levels can actually be more significant when the scale of a parameter comprises few levels.

The position of the flaw on the parameter scale, combined with the information related to the scenario under study, also influences its impact. When a scenario involves a situation that meets the gap within the parameter scale, the level of difficulty perceived and the comments demonstrate the impact of the flaw. Otherwise, its impact can be non-existent.

One can then conclude that this flaw can have an impact on the risk estimation process, but that this impact can be low or important depending on the situations. It is still important to be attentive to the presence of this flaw since it can lead to an underestimation or an overestimation of certain scenarios in particular. Moreover, even if an important gap between two levels can promote a better convergence of the results, this result generates an evident discomfort among the users, which can affect negatively the perception, as well as the good deployment of the risk estimation process.

5. CONCLUSION

This study allowed to confirm the impact of the flaws in the parameters of the risk estimation tools used in safety of machinery. The results show that these flaws can be translated into a low convergence of the risk levels obtained for the same hazardous situation by many participants, and the dissatisfaction of the participants with respect to the performance and the accuracy of these tools. In the majority of the cases, the participants are able to identify a flaw when it increases the level of difficulty regarding the choice of level corresponding to a given situation.

The results also show that the impact of the construction flaws in the parameters is not consistent. The nature of the flaw, its position on the parameter scale and the studied scenario influence the impact of the flaw with respect to identifying the level of a parameter. The severity of harm parameter is relatively solid against the different flaws. The probability of harm parameter is clearly more affected. In addition to the presence of the flaws within those parameters, the results obtained suggest that the assessment of the probability is problematic in risk estimation, and that a particular attention must be paid to it.

These experimental results confirm the construction rules (ref. : Table 1) brought forward in a previous study [16]. These rules will allow to address some of the issues linked to the significant variability in the risk estimation process. The results obtained here could hence contribute to the improvement of the robustness and the reliability of the existing tools, and help to support the training actually given by the partners in the risk assessment field.

Acknowledgement

This study was done in collaboration with the HSE's Health and Safety Laboratory in Great Britain and the authors would like to acknowledge the much appreciated involvement of Nicola Stacey, Nicola Healey and Colleen Butler.

This work was supported by the Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST) [grant number 2010-0036].

6. REFERENCES

1. John R. Etherton, J.R. Myers, M.L. 1990. Machine safety research at NIOSH and the future directions, International Journal of Industrial Ergonomics, 6(2) 163-174.
2. Backström, T., Döös, M. 2000. *Problems with machine safeguards in automated installations*. International Journal of Industrial Ergonomics, 25(6), 573-585.
3. Lind, S. 2008. *Types and sources off fatal and severe non-fatal accidents in industrial maintenance*, International Journal of Industrial Ergonomics 38(11-12) 927-933.
4. Lyon, B. & Hollcroft, B. 2012. *Risk Assessments top 10 pitfalls and tips for improvement*. Professional Safety, 57, (12) 28-34.
5. Hughes, P. & Ferrett, E. 2005. Introduction to Health and Safety at Work, 2nd Edition, 2nd ed. Oxford, Elsevier Butterworth-Heinemann.
6. Main, B.W. 2012. Risk Assessment Challenges and Opportunities, Ann Arbor, USA, Design Safety Engineering inc.
7. Pickering, A. & Cowley, S.P. 2010. *Risk Matrices: Implied accuracy and false assumptions*. Journal of health and safety research and practice, 2, (1).
8. ISO 12100 (2010). Safety of machinery -- General principles for design -- Risk assessment and risk reduction, International Standard.
9. Paques, J.-J. and 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.
10. Etherton, J. (2007). *Industrial Machine Systems Risk Assessment: A Critical Review of Concepts and Methods*, Risk Analysis, Vol. 27, No. 1, pp. 71-82.
11. 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.
12. 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.
13. Abrahamsson, M. (2000). Treatment of uncertainty in risk based regulations and standards for risk analysis, Report 3116, Lund University, Sweden, 82 pages.
14. Abrahamsson, M. (2002). Uncertainty in quantitative risk analysis- characterisation and methods of treatment. Department of Fire Safety Engineering, Lund University, Lund, Sweden.
15. Stevens, S.S. (1946). *On the theory of scales of measurement*. Science, Vol. 103, No. 2684, pp. 677-680.
16. Chinniah, Y., Gauthier, F., Lambert, S. and Moulet, F. (2011). Experimental Analysis of Tools Used for Estimating Risk Associated with Industrial Machines, Rapport R-684, IRSST #099-661, 77 p.

17. 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

18. Carey, J.M. and 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.

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

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

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

22. 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.

23. 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.

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

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

26. 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.

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

28. 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.

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

30. Paques, J.-J. and 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.

31. Paques, J-J., Gauthier, F., Perez, A., Charpentier, P., Lamy, P. and David, R. (2005b). Bilan raisonné des outils d'appréciation des risques associés aux machines industrielles, Rapport R-459, IRSST #099-343, 64 p.

32. 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.

33. Chinniah, Y. Gauthier, F. Aucourt, B., Burlet-Vienney, D. (2018) *Validation of the Impact of Architectural Flaws in Six Machine Risk Estimation Tools*, Safety Science, Vol. 101, pp 248-259.

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

35. Company X (1997). Tableau d'analyse de risque (sans titre), Document interne.

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

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

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

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

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

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

42. Kazer, BM. (1993) *Risk Assessment: A Practical Guide*, The Health and Safety Practionner, Institution of Occupational Safety and Health.

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

44. 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

45. Willquist, P. & Torner, M. 2003. *Identifying and analyzing 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.

46. Mosteller, F. & Youtz, C. 1990. *Quantifying probability expressions*. Stat Sci (5) 2-12.