Semi-quantitative Approach in Risk Assessment: Inter and Intra-rater Reliability

Filipa Carvalho and Rui B. Melo

Faculdade de Motricidade Humana
Universidade de Lisboa
1499-002 Cruz-Quebrada, Portugal

ABSTRACT

The semi-quantitative approach used in occupational risk assessments is quite popular among safety practitioners, namely because it allies the advantages of both the quantitative and the qualitative approaches and overcomes some of their limitations. However it still rises some concerns as the sort of methods involved haven't been fully studied in order to assure high levels of confidence in the results they produce. This is a particularly relevant issue as the workers' health and safety may be compromised if risk assessment results are not reliable. Therefore, a comparative study on the reliability of matrix based risk assessment methods was accomplished, particularly focused on their stability and reproducibility. The Krippendorff's Alpha Coefficient (${}_K$) was chosen to assess inter-coder (reproducibility) and intra-coder (stability) reliability. In most cases, inter and intra coder reliability proved to be low (${}_K$ <0.6) for both the risk level and the intermediate variables of each assessment method. Consequently, different priority interventions may be suggested, depending on the analyst involved and/or the moment of the risk assessment accomplishment. The obtained ${}_K$ values for the three groups of analysts seem contrary to the literature as the risk assessment results appear not to be dependent on their levels of experience and expertise.

Keywords: Reliability, Assessment, Inter-coder, Intra-coder, Krippendorff's Alpha Coefficient

INTRODUCTION

The Risk Matrix method is a popular and common approach to evaluation in risk analysis. It is a semi-quantitative method, where probabilities and consequences are categorised, instead of using numerical values (Harms-Ringdahl, 2013). Risk Matrix is applied in many different areas, and consequently the terminology and methodology vary quite a lot. In Markowski & Mannan (2008) opinion one perspective is that it is a method or "mechanism to characterize and rank process risks" (as cited by Harms-Ringdahl, 2013). In other words, once the magnitude of the risk is obtained, priorities of assistance are established so that control measures can be put in place. Therefore, how a Risk Matrix is used will have a large impact on safety and system performance in numerous workplaces and systems. (Harms-Ringdahl, 2013).

There are many references in which the matrix is explained rather summarily, and the method is generally seen as a flexible, simple and efficient tool. However, the validity and reliability of risk matrices and other evaluation techniques have not been studied enough (Harms-Ringdahl, 2013). Carvalho and Melo (2007) that advocate that presently, it is common to use semi-quantitative risk assessment methods (SqtRAM) reinforce this opinion. In their opinions this kind of methods have proven to be, in most cases, the only available technique and the most suited to carry out this task, in Small and Medium Enterprises (SME). Like they defend, these methods present several advantages, including being generalists, user friendly and easy to apply. However, we cannot disregard the existing gap in terms of reliability of these applications. In other words, to be useful, these methods must prove to be reliable. "The importance of reliability rests on the assurance it provides that data are obtained independent of the measuring event, instrument or person. Reliable data, by definition, are data that remain constant throughout variations in the measuring process" (Krippendorff, 2004).

There are very few studies reflecting a concern about risk assessment's outputs when different methods are used, particularly methods relying on risk matrices. In Portugal, the few known studies reinforce the need of further scientific knowledge in this area to ensure the reliability of risk assessments (Branco, Baptista, & Diogo, 2007; Carvalho, 2007).

This paper reports the results of a study on the reliability of SqtRAM. Its use relies on coding categories, which are considered to be reliable if separate coding attempts end up with the content coded in a similar way.

It is useful to understand that, in general, reliability refers to the extent to which a test, experiment or measuring procedure gives the same result(s) on repeated trials or applications (Olsen, 2013). Krippendorff (2004) distinguishes three types of reliability: stability, reproducibility, and accuracy. Figure 1 shows the main characteristics of the types of reliability.

This study involved a comparative analysis of four SqtRAM, which were used to estimate and assess six risks identified in two tasks accomplished to produce Airbags. With this study we have assessed inter-method, inter-coder (reproducibility) and intra-coder (stability) reliability of four SqtRAM within the Occupational Safety and Health (OSH) scope. As the risk level estimation depends on the intermediate variables used by each SqtRAM, both types of variables were analysed.

In other words, this study was performed to pursue three principal objectives:

- to find if there is consistency among the risk level values obtained with different methods in each of the previously identified situations;
- to verify if risk level values are consistent (as well as those of the intermediate variables), when different analysts use the same method in each of the previously identified situations;
- to investigate if the analysts are consistent with themselves regarding the risk level values (as well as those of the intermediate variables), when using the same method in each of the previously identified situations, but in different moments.

METHODOLOGY

Stages of the Study

This study comprised four fundamental stages, which correspond to the stages of any risk assessment – Characterization of Work Situations, Hazard Identification, Risk Estimation and Risk Evaluation – as part of the risk management process (BSI, 2004; ISO/IEC, 1999; ISO, 2009; Suddle, 2009; van Duijne, van Aken, & Schouten, 2008). Figure 2 identifies and describes each of these stages.

A previous inspection to an airbag production unit allowed us to identify six risky work situations in two tasks (A: fabric cutting press; B: turning and folding Airbags), which were fully described and illustrated with pictures and videos. Additional data were provided concerning noise exposure, lighting conditions and the risk of developing musculoskeletal disorders whenever relevant. shows the relationship between the nature of the assessed risk, the respective task/situation and the adopted codes to identify the situation being analysed.

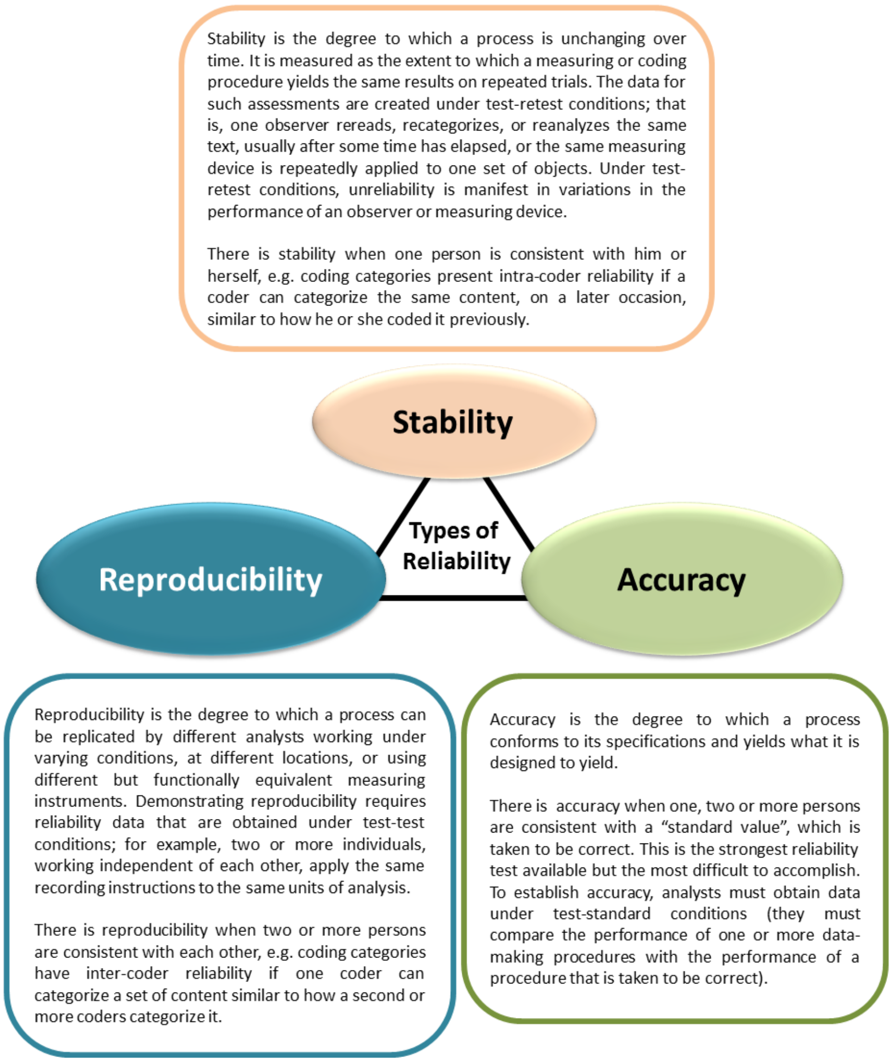


Figure 1 - Main characteristics of the types of reliability (Krippendorff, 2004).

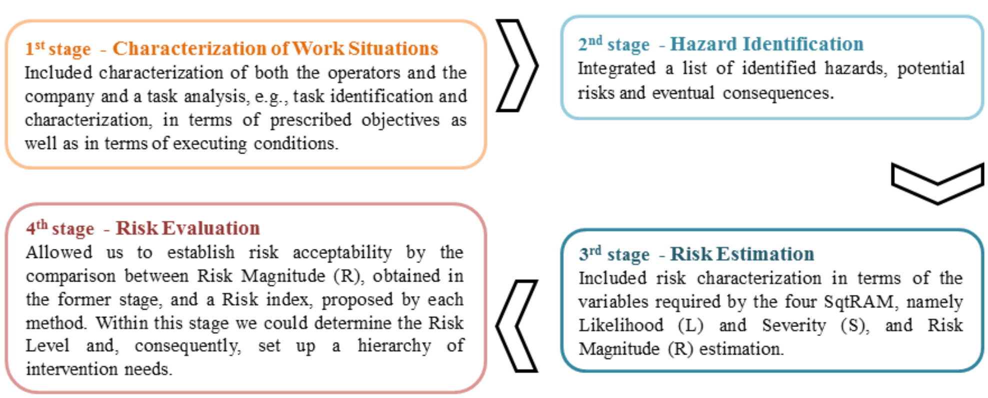


Figure 2– Identification and description of the four stages of the study.

Table 1 – Relationship between the nature of the assessed risk, the respective task/situation and the adopted codes to identify the situation being analysed.

Task	Nature of the assessed risk	code
A	Slipping, tripping and falling	**TA1**
	Noise Exposure	**TA2**
	Objects Falling on the feet	**TA3**
B	Mechanical contact – crushing	**TB2**
	Physical overload	**TB4**
	Hit by a fork-lift truck	**TB6**

Data collection

For data collection, we used different methods, tools and equipment, in accordance with the specificity of each stage of the study. Although this study focuses in the results obtained in stages 3 and 4, a global overview is provided.

On the first two stages, data collection relied on free and systematized observations and made use of video recording, documental research, analysis grids and questionnaires specifically developed for this purpose. A single analyst conducted these stages. We tried to collect and to analyse useful data to provide the analysts with a complete characterization of the analysed work situations. In this way, all analysts could perform stages 3 and 4 - Risk estimation and Risk evaluation – relying on the same exact information.

gives us an overview of the used methods/tools and equipment, their purposes and some of their particular characteristics.

Table 2 – Overview of the methods/tools and equipment used, their purposes and some of their particular characteristics.

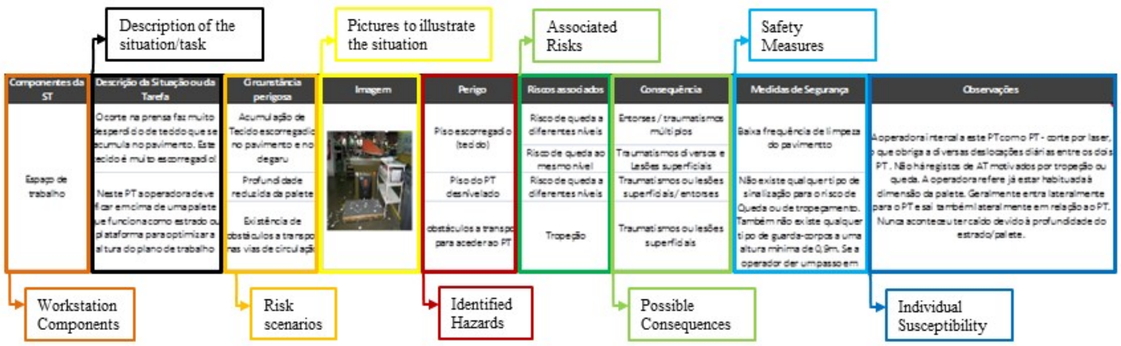
Methods Tools Equipment	Purposes	Particular characteristics	Comments
Documental research	Hazards associated with work activities can be present as a result of any or a combination of the following: substances, machinery/processes, work organization, tasks, procedures and people and circumstances in which the activities take place, including the physical aspects of the plant and/or premises (Gadd, Keeley, & Balmforth, 2003).	The characterization of the work situations began with the analysis of the information included in relevant sources such as: ▪ Legislation and Standards with relevant impact to this particular activity sector; ▪ Internal workplace accident's report; ▪ European data on workplace accidents within this activity sector; ▪ Environmental risk assessment reports; ▪ Work procedures; ▪ Occupational diseases' reports; ▪ Internal ill health and incident data; ▪ Tools, equipment and materials characteristics, ...	At this level, the analysis was focused on the socio-demographic and organizational aspects, as well as on the possible causes of workplace accidents. These analyses resulted in relevant information to be included in both the questionnaire and the analysis grid specifically developed for characterization of the workstations.

Table 2 (cont.) – Overview of the methods/tools and equipment used, their purposes and some of their particular characteristics.

Methods Tools Equipment	Purposes	Particular characteristics	Comments
Questionnaire	The participation of the employees in risk analysis is helpful, as they know how the work is actually performed. Therefore, the questionnaire is a useful tool to collect data on how safely or risky the employees work.	It was organized in three parts: ▪ Part A focused on the operators' characterization which involved indicators such as: age; gender; experience; safety knowledge; life habits and health complaints (identification of body regions, pain and discomfort levels, etc..). ▪ Part B was concerned with the operator's perception of its work conditions. It involved questions about personal protective equipment (PPE) characteristics; equipment use and characteristics of hand tools. ▪ Part C involved a set of questions to provide the information required for the QEC (Quick Exposure Check) tool (a method to assess musculoskeletal disorders risk).	Some of the questions we have included in the questionnaire were: ▪ Have you had OSH training? If Yes, in which domains? ▪ Do you usually feel pain or physical discomfort? If Yes, indicate the affected regions in the body map figure. ▪ Which postures do you adopt more often to get the job done? ▪ How do you evaluate the available PPE? ▪ How do you evaluate the available tools? ▪ How do you evaluate the work environment?
Analysis Grid	To gather all relevant information in a unique document.	This grid was organized by topics that represent the principal items influencing the working conditions. The following list summarizes the main topics: A - Organizational conditions; B - Equipment / Machinery / Tools; C - Manual materials handling; D - Postures and related issues; E - PPE; F - Building characteristics; G - Electricity hazards; H - Complementary information used in QEC tool.	Previous free observations were conducted in order to develop an analysis grid to be used in the following systematized observations.
Images collection	Video recording was the chosen technique to collect images related to work activity.	A digital camera SONY Handycam, HDR-SR10 model, was used.	Previous verbal consent of the operators involved was obtained.
Sound meter	Noise, light and thermal variables were measured to provide a better characterization of the workplace.	Noise was measured with a Bruel & Kjaer Sound meter, 2260 model.	The device was carefully placed near the operator's ear, and was subjected to verification in the workplace before each series of measurements. Both Continuous A-Weighted Sound Pressure Level (dB (A)) and Maximum Peak Level (dB (C)) were measured.
Lux meter		The illuminance level was assessed with a digital Krochmann lux meter, 106E model.	The device was strategically put on the surface of the workstations.
Psychrometer		Dry and wet air temperatures were assessed with a THIES sling psychrometer, 450 model.	Air humidity was computed from these two variables.

On the second stage, a worksheet was developed taking into account the identified hazards, the potential risks and the possible consequences. It presented nine fields: Workstation Components; Description of the situation/task; Risk scenarios; Pictures to illustrate the situations; Identified Hazards; Associated Risks; Possible Consequences; Safety Measures and Individual Susceptibility. To complete this stage this sheet was filled out with the information collected in the previous stage. This procedure allowed gathering all information in a unique document.

Figure 3 illustrates the worksheet developed for this stage of the study.



The Quick Exposure Check (QEC) observational tool was integrated in our tools (Part C of the Questionnaire and topic H of the Analysis Grid) because it was our intention to compare the results with those of MASqt. However, this paper does not report this part of the study.

For the third stage, we developed an on-line questionnaire, with LymeSurvey® software, so that analysts could perform risk estimation from wherever and whenever they would consider it appropriate. This on-line questionnaire was organised in three parts:

- the 1st part presented the rules and conditions to take part of the study;
- the 2nd part described the tasks under analysis and the situations to be assessed resorting to the 4 SqtRAM. All variables estimated by each method were well described so that the analysts would just select the option they considered correct;
- a 3rd part addressed to the analysts' characteristics, namely: age, gender, academic background, OSH professional experience, knowledge about the SqtRAM approach, perception about the ease/difficulties when using this kind of approach, and the analysts' perception of SqtRAM's reliability.

With this questionnaire, we ensured that all analysts would possess the same information about the situation under analysis to estimate the variables involved in each method (Likelihood, Severity, and so on). It included 42 questions of which 6 depended on previous answers. The estimated time to fill out the questionnaire was around 30 minutes.

For the 4th stage we used SPSS (Statistical Package for Social Sciences – version 20) to automatically generate, for all assessments, the main studied variable – Risk Level.

To make it possible to perform the intra-coder reliability assessment, the two last stages were repeated with a lag time of 5 months.

Participants

To complete the last two stages of a risk assessment (Risk Estimation and Risk Evaluation) 81 analysts were invited. The selection criteria to include them in the study comprised their:

- Academic background;
- OHS practice experience;
- Expertise in using this type of methods.

With these specifications, we organized the analysts in groups as presented in Table 3.

Table 3 – Analysts' groups, respective dimension (N) in each round of the study and associated code.

	Code	Sample group	N	
			1st round	2nd round
Academic background	**SA_A1**	Ergonomists	19	16
	SA_A2	OHS practitioners	5	5
	SA_A3	Engineers	14	12
OHS practice experience	**SA_B1**	With practical experience	28	23
	SA_B2	With no practical experience	16	16
Expertise	**SA_C1**	Experts	15	2
	SA_C2	Non-experts	29	27

Only 44 (26 women and 18 men) out of the 81 invited analysts (response rate: 54%) agreed to participate in the 1st risk assessment round. Five months later, in the 2nd risk assessment round, only 39 analysts (23 women and 16 men) responded in good time to integrate the intra-coder reliability assessment. All analysts were familiarised with SqtRAM (either theoretically or in practice).

Semi-quantitative risk assessment methods

A set of four different methods were selected from those described in international references and those used in particular organisations. These four SqtRAM can be divided into two categories (Carvalho & Melo, 2007): a) simple matrix methods (MMS 3x3 and BS8800), which resort to the use of two single variables (Likelihood/Frequency and Severity/Consequence) to compute Risk Magnitude; b) complex matrix methods (MMCP and WTF), which rely on three or more variables. In addition to the previous variables we can find Exposure factor, Procedures and safety conditions or Number of persons exposed or affected.

The selection of these particular methods was based on the following criteria: a) all methods should have a Risk Index scale with five levels; b) two methods belonging to each of the above described categories should be included; c) Risk Magnitude (R) should result from a pre-established combination of the intermediate variables (ex: Likelihood and Severity), and d) the use of the same label to identify variables taking part in Risk Estimation was not compulsory (ex: Likelihood or Frequency; Severity or Consequence). Table 4 shows an outline of the methods' main characteristics.

Statistical analysis

For data processing, we resorted to the Statistical Package for the Social Sciences (SPSS – version 20). The nonparametric Friedman test and the Krippendorff's Alpha Coefficient ($_K$) were the statistical techniques used.

Whenever the Friedman test outputs lead to H_0 rejection, a posterior analysis was performed to identify the pairs that rendered those results. A significance level of 0.05 was adopted as a criterion to reject the null hypothesis. For the pairs of methods not revealing differences between Risk Level values, the $_K$ was used to find if there was agreement.

Krippendorff's Alpha Coefficient ($_K$) was chosen to assess inter-method, inter-coder (reproducibility) and intra-coder (stability) reliability. The selection of $_K$ was based on the following criteria: it handles ordinal variables, multiple coders ($k \geq 3$) and missing data. $_K$ is a statistical measure of the extent of agreement among coders, and is regularly used by researchers in the area of content analysis (Gwet, 2011).

Table 4 - Main characteristics of the used methods.

Method (code)	Main characteristics
3x3 simple matrix method (MMS3x3)	It's a simple method using a (3x3) Risk Estimation Matrix resorting to two variables, Gravity/Severity (G/S) and Likelihood (L), both expressed in scales with three levels. It integrates a Risk Index scale with 5 levels, to prioritize intervention.
BS 8800 simple matrix method (BS8800)	It's a method that was introduced in BS 8800. Its matrix resorts to two variables: Gravity/Severity (G/S) expressed in a 3 levels' scale, and Likelihood (L), presenting a scale of 4 levels. The Risk Index integrates 5 levels of intervention priority.
P - complex matrix method (MMCP)	Besides Frequency (F) and Severity (S), it integrates two more variables: Procedures and safety conditions adopted (Ps) and the Number of persons exposed or affected (N). All variables (F, S, Ps and N) are expressed in a 5 levels' scale. On its turn, Risk Magnitude varies between 1 (very bad) and 625 (very good). This method integrates a Risk Index scale with 5 intervention priority levels.
William T. Fine method (WTF)	Risk estimation that resorts to the knowledge of 3 variables, here designated as Consequence factor (C), Exposure factor (E) and Probability factor (P). Each variable (C, E and P) is assessed in a 6 levels' scale. The Risk Magnitude scale varies between 0.05 (optimal situation) and 10000 (worst situation). This method integrates a Risk Index scale with 5 levels of intervention priorities.

Coders were grouped according to Table 3 in order to investigate the relative importance of academic background; OHS practice experience and expertise level regarding the SqtRAM use, on both inter-coder and intra-coder reliability. Therefore, we had 8 groups of analysts, including the complete sample.

The software application KALPHA (macro for SPSS) was used to compute $_K$ (Hayes, 2005). KALPHA macro was set-up to bootstrap 1000 samples of hypothetical values of $_K$, leading to the estimation of $_K$ true, for a 95% confidence interval. The agreement percentage (%Agr) was also computed in some cases, to provide a better interpretation of the results.

The following assumptions were adopted: a good agreement level was assumed for $_K > 0.8$, an acceptable agreement level was assumed for $0.6 \leq {}_K < 0.8$ and a low agreement level was assumed for $_K < 0.6$ (Krippendorff, 2004, 2007). The same criteria were used to evaluate the %Agr.

Bearing in mind that this study attempts to assess individual risk, we also tried to identify the most protective methods, as far as the worker's protection is of concern. Mean rank values were used to rank methods and it was assumed that more protective methods present lower values.

RESULTS AND DISCUSSION

Considering the number of assessed situations, six tests were carried out to find if the Risk Levels obtained with different methods were identical for each of the previously identified situations (risks/consequences) in each task. Figure 3 shows the results obtained in each situation with each method and in each round.

Statistically significant differences were observed ($p<0.05$) among Risk Levels obtained with different methods for each of the previously identified situations (risks/consequences). From the point of view of the workers protection, the most powerful methods were, in a consistent way, WTF and BS8800, assuming the 1st position of the Mean Rank all the times (Figure 4).

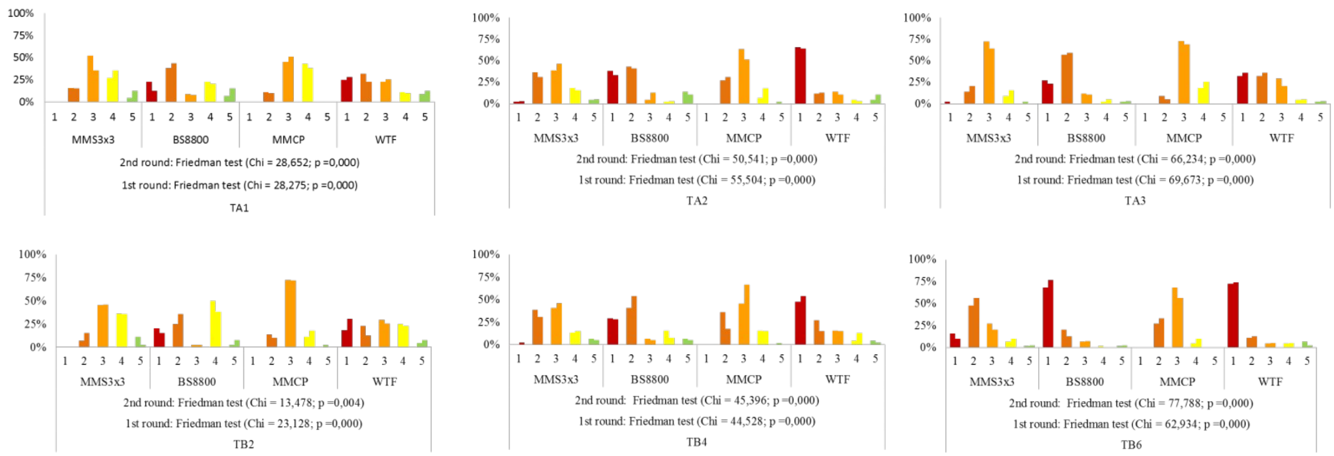


Figure 3 – Risk Level frequency (%) and Friedman test outputs obtained in each situation (TA1, TA2, TA3, TB2, TB4 and TB6) with each method in each round of the study.

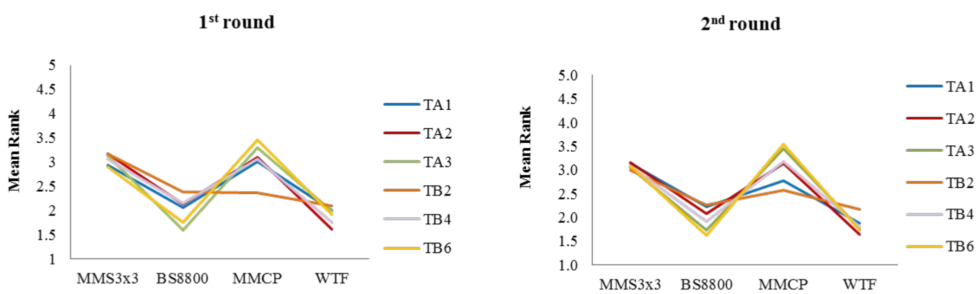
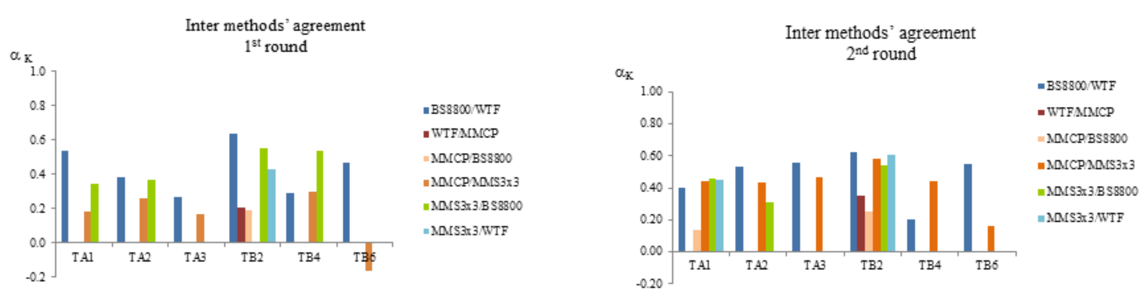


Figure 4 – Mean Rank obtained with each method in each situation (TA1, TA2, TA3, TB2, TB4, TB6) in both rounds.

Despite the previously found differences, the posterior analysis revealed no significant differences (p>0.05) between the results of some pairs of methods. Consequently, the Krippendorff's Alpha Coefficient was used to verify the consistency of the results. Regarding the inter methods' agreement (Figure 5), the Krippendorff's Alpha Coefficient results revealed low concordance ($_K$ < 0.6) in most cases (only 3 cases revealed $_K$ > 0.6). However, it appears that the assessed risk type/nature influences the results, and reinforces that the choice of a risk assessment method should be a rigorous process (Carvalho, 2007).



It is worth to remind that this paper focuses on the last two stages of risk assessment: risk estimation and risk evaluation. Hence the reported results concern the reliability associated to the inter and intra coder consistency.

Considering the obtained values of $_K$ (0≤ $_K$ <0.3) for the inter-coder assessment (Figure 6), it is possible to verify that there is a low inter-coder agreement level ($_K$ <0.6). Therefore, it is appropriate to state that risk assessments made by different analysts with the same method may lead to different intervention priorities.

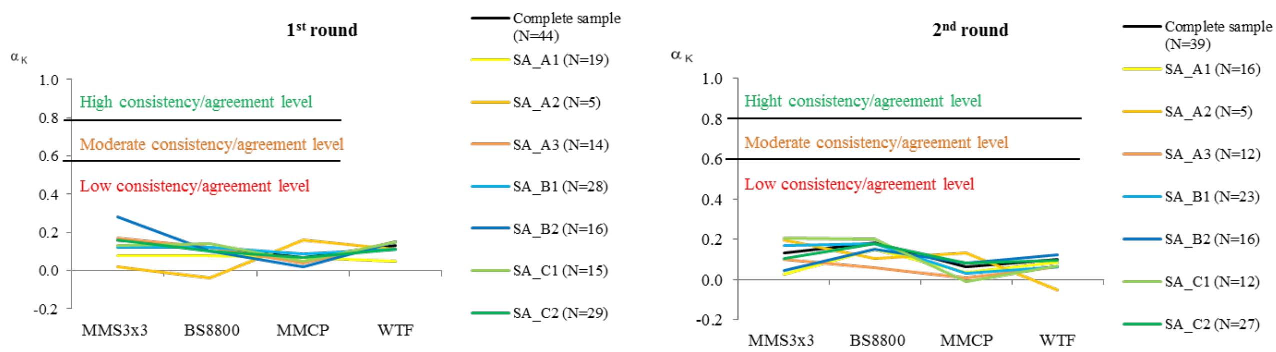


Figure 6 - $_K$ values obtained for the inter-coder reliability assessment of risk level variable: left - 1st round of the study; right - 2nd round of the study.

The analysts' academic background (SA-A1; SA-A2; SA-A3), their OHS practice experience (SA-B1; SA-B2) and their level of expertise in using these SqtRAM (SA-C1; SA-C2) did not influence the results. This fact contradicts the disadvantage pointed to this type of approach, particularly its results dependency on these two last factors (Marhavilas, Koulouriotis, & Gemeni, 2011; Marhavilas & Koulouriotis, 2008; Reniers, Dullaert, Ale, & Soudan, 2005). Likewise, it is possible to highlight that the method referenced as the most popular (MMS 3x3) among the experts' group (SA-C1) did not show relevant differences in the results of $_K$.

Despite the low level of agreement revealed by the $_K$ values, the inter-coders reliability evaluation for the intermediate variables used in each SqtRAM revealed a constant behaviour over the two evaluation rounds (Figure 7). It is worth to highlight: a) the similarity of the obtained results for all groups of analysts; b) the fact that three variables (Gravity (G), Severity (S), and Consequence factor (C)) delivered values of α_K: 0.25< $_K$ <0.5. For these variables, the %Agr revealed high level consensus (%Agr >0.8) when using methods MMS3x3 and BS8800 to assess TA2, TA3 and TB4. These results suggest that the clarity of descriptors can minimize the uncertainties of analysts, which results in higher levels of consensus.

For the other variables (Likelihood (L); Procedures and safety conditions (Ps); Number of people exposed or affected (N); Exposure factor (E) and Probability factor (P)) the $_K$ values were slightly lower.

The obtained values of $_K$ for the intra-coder reliability evaluation (Figure 8) concerning the Risk Level variable show higher agreement levels (0.2 ≤ $_K$ <0.6) compared to the inter-coder reliability evaluation, highlighting methods MMS3x3 and WTF.

The analysts' professional experience and their level of expertise did not produce relevant differences in the results of $_K$. As for their academic background, some inconsistency (min $_K$ = - 0.07 and max $_K$ = 0.62) was found in the OHS group (SA-A2), whereas greater consistency and higher agreement level was found within the Engineers group (SA-A3): min $_K$ = 0.30 and max $_K$ = 0.45.

Intra-coders reliability evaluation (Figure 9) for the intermediate variables used in each SqtRAM came to reinforce the trend already found, highlighting Gravity (G), Severity (S) and the Consequence factor (C) as the ones with higher levels of agreement.

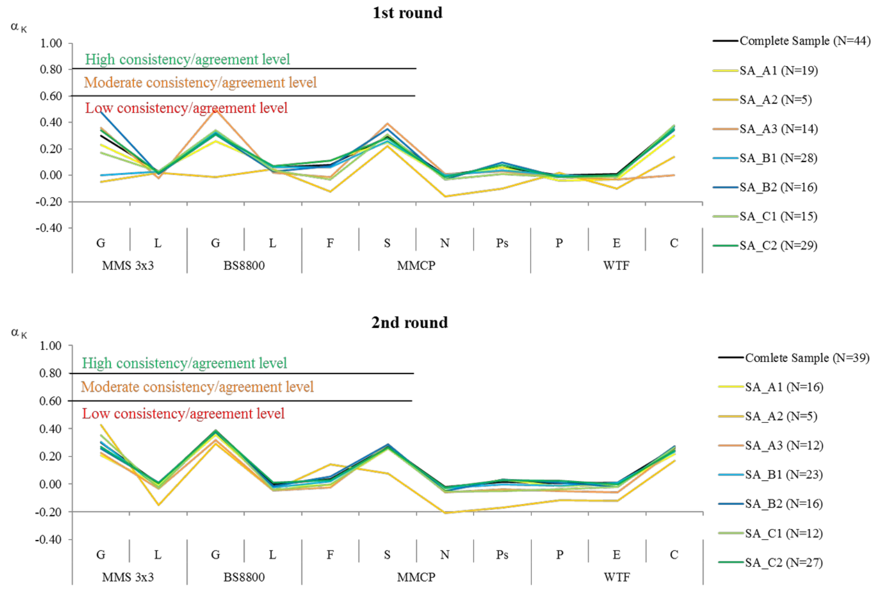


Figure 7 - $_K$ values obtained for the inter-coder reliability assessment of the intermediate variables: top – 1st round of the study; bottom – 2nd round of the study.

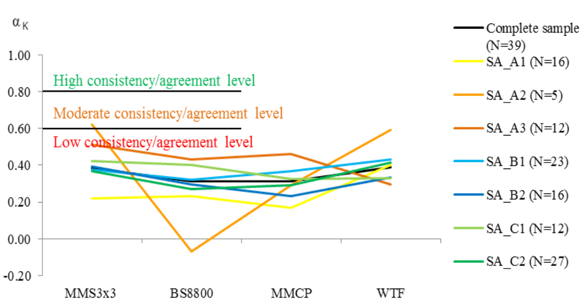


Figure 8 - $_K$ values obtained for the intra-coder reliability assessment of the risk level variable.

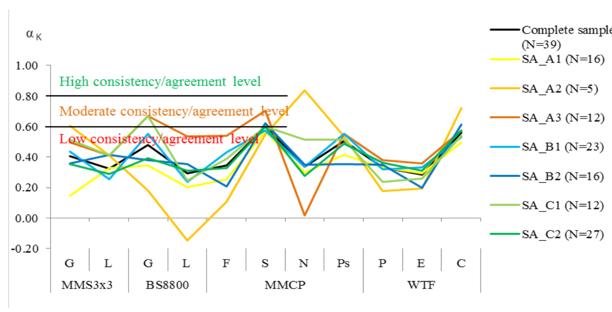


Figure 9 - $_K$ values obtained for the intra-coder reliability assessment of the intermediate variables.

CONCLUSIONS

The preliminary results of this study revealed that statistically significant differences ($p<0.05$) and low agreement levels ($_K$ <0.6) were found among Risk Levels obtained in all 6 work situations with the 4 methods.

Similarly, the $_K$ results also revealed inter methods low agreement. Therefore, it seems reasonable to conclude that, despite each method's intrinsic validity its choice should not be arbitrary. Nevertheless, there was some consensus in the power revealed by the methods regarding the protection they confer to the workers. The WTF and BS8800 methods were the most powerful.

Considering that this study focused on the reliability of the two last stages of risk assessment (risk estimation and risk evaluation) the preliminary results revealed that the use of SqtRAM, to accomplish the requirements of the legislation, should be done with caution, as most of the inter and intra-coders assessments showed low levels of consistency. This means that the use of risk matrices within the OHS scope revealed low reliability, either in terms of stability or in terms of reproducibility.

Apparently, there are no relevant differences among risk assessment results obtained by individuals with different levels of OHS practice experience and different levels of expertise in using risk matrices. Such evidence seems to be opposite to that found in several scientific references on the use of this type of approach. In most cases, inter and intra-coder reliability proved to be low (κ <0.6) for both the risk level and the intermediate variables of each assessment method. Consequently, different priority interventions may be suggested, depending on the analyst involved and/or the moment of the risk assessment accomplishment.

It was still possible to register similar κ values among the analyses concerning the intermediate variables, which reveals that some of them seem to be easier to estimate than others. In fact, Gravity, Severity and Consequence factor were the three variables for which the agreement was higher (%Agr>80%) and therefore variables such as Likelihood, Frequency or Probability factor require improvements regarding their coding process.

Bearing in mind the advantages attributed to SqtRAM, we consider that its reliability is relevant and should continue to be investigated, in order to find and define adequate criteria to support the risk assessment methods' selection process. We would like to point out the lack of studies regarding this subject. In our opinion, this kind of studies strengthens risk assessments carried out and, consequently, their objectives: the improvement of Health and Safety at Work.

REFERENCES

Branco, J. C., Baptista, J. S., & Diogo, M. T. (2007). Comparação da avaliação dos riscos por dois métodos correntemente utilizados na Industria Extractiva. In P. Arezes, J. S. Baptista, M. Barroso, A. Cunha, R. Melo, A. S. Miguel, & G. P. Perestrelo (Eds.), *Colóquio Internacional Segurança e Higiene Ocupacionais - SHO2007* (pp. 177–181). Guimarães: Sociedade Portuguesa de Segurança e Higiene Ocupacionais (SPOSHO).

BSI. (2004). BS8800 - Occupational health and safety management systems - Guide. UK: British Standard Instituton.

Carvalho, F. (2007). *Estudo Comparativo entre diferentes métodos de avaliaçao de risco, em situaçao real de trabalho*. Master Thesis in Ergonomia na Segurança no Trabalho (not published), Cruz-Quebrada.

Carvalho, F., & Melo, R. B. (2007). Comparação entre diferentes métodos de avaliação de risco, em situação real de trabalho. In C. Guedes Soares, A. P. Teixeira, & P. Antão (Eds.), *Riscos Públicos e Industriais* (Vol. 2, pp. 853–868). Lisboa: Edições Salamandra.

Gadd, S., Keeley, D., & Balmforth, H. (2003). *Good practice and pitfalls in risk assessment*. Shelfield, UK: Health & Safety Executive.

Gwet, K. L. (2011). On The Krippendorff ' s Alpha Coefficient. Retrieved February 1, 2012, from http://agreestat.com/research_papers/onkrippendorffalpha.pdf

Harms-Ringdahl, L. (2013). *Guide to safety analysis for accident prevention* (1st ed.). Stockholm, Sweden: IRS Riskhantering AB.

Hayes, A. F. (2005). An SPSS procedure for computing Krippendorff's alpha [Computer software]. Retrieved January 12, 2012, from http://www.afhayes.com/spss-sas-and-mplus-macros-and-code.html

ISO. (2009). *ISO 31000 - Risk management - principles and guidelines* (1st ed.). Geneva: International Organization for Standardization.

ISO/IEC. (1999). *GUIDE 51 Safety aspects — Guidelines for their inclusion in standards* (2nd ed.). Geneva: International Organization for Standardization & International Electrotechnical Commission.

Krippendorff, K. (2004). *Content Analysis: An Introduction to its Methodology.* (2nd ed.). Thousand Oaks, California: Sage Publications.

Krippendorff, K. (2007). Testing the reliability of content analysis data: What is involved and why. Retrieved February 3, 2012, from http://www.asc.upenn.edu/usr/krippendorff/dogs.html.

Marhavilas, P. K., & Koulouriotis, D. E. (2008). A risk-estimation methodological framework using quantitative assessment techniques and real accidents' data: Application in an aluminum extrusion industry. *Journal of Loss Prevention in the Process Industries*, *21*(6), 596–603.

Marhavilas, P. K., Koulouriotis, D. E., & Gemeni, V. (2011). Risk analysis and assessment methodologies in the work sites: On a review, classification and comparative study of the scientific literature of the period 2000–2009. *Journal of Loss Prevention in the Process Industries*, *24*(5), 477–523.

Olsen, N. S. (2013). Reliability studies of incident coding systems in high hazard industries: A narrative review of study methodology. *Applied Ergonomics*, *44*(2), 175–84.

Reniers, G. L. L., Dullaert, W., Ale, B. J. M., & Soudan, K. (2005). The use of current risk analysis tools evaluated towards preventing external domino accidents. *Journal of Loss Prevention in the Process Industries*, *18*(3), 119–126.

Suddle, S. (2009). The weighted risk analysis. *Safety Science*, *47*(5), 668 – 679.

van Duijne, F. H., van Aken, D., & Schouten, E. G. (2008). Considerations in developing complete and quantified methods for risk assessment. *Safety Science*, *46*(2), 245–254.