

Human Reliability Analysis (HRA) in Surgery: identification and assessment of Influencing Factors

Rossella Onofrio and Paolo Trucco

To cite this paper:

Onofrio, R., Trucco, P., 2018. Human reliability analysis (HRA) in surgery: identification and assessment of influencing factors. *Safety Science* 110 (2018) 110–123
Doi:10.1016/j.ssci.2018.08.004

Abstract

Several authors encouraged the application of Human Reliability Analysis (HRA) to patient care, recognising the benefits it brought to high-risk industries in terms of human errors reduction, thanks to its anticipative approach. However, the literature on HRA in healthcare is still scanty, and shows significant gaps at methodological level. Only a very limited number of HRA studies implemented Influencing Factors (IFs) analysis in the quantification of the Human Error Probability, despite it represents a peculiar component and a well-established step in several HRA techniques. Furthermore, the need for a deep adaptation and translation of existing HRA techniques to the healthcare domain is emphasised by several authors. This paper aims at designing a taxonomy of Influencing Factors for HRA studies in surgery, and assessing their influence on surgeons' technical performance by means of structured elicitation of experts' judgement. An *ad hoc* taxonomy of 21 IFs has been validated through focus group and individual interviews, in Denmark and Italy; 215 questionnaires from expert surgeons in open and mini-invasive surgeries were then used for the quantification of the impact of each IF on surgeons' performance. An investigation on the possible influence of different surgical techniques revealed that only two out of the twenty-one IFs show different probability density functions (pdf) when considered in open or mini-invasive surgeries. The study offers an original and relevant contribution to the development of domain-specific knowledge and, as such, to fostering the diffusion of HRA studies in healthcare, and in surgery in particular.

Keywords: Human Reliability Analysis (HRA), Influencing Factors (IFs), Healthcare, Surgery.

1 Introduction

The benefits of transferring and applying risk analysis methods, traditionally implemented in high-risk industry, to healthcare services are fully recognised in the patient safety literature (Cagliano et al., 2011; Verbano and Turra, 2010; Lyons, 2009; Lyons et al., 2004; Vincent et al., 1998). Nevertheless, healthcare organizations are still characterised by a predominant reactive culture in their approach to patient safety. In fact, safety improvement actions are often taken *ex-post*, due to the occurrence of an adverse event or periodic audits and patient records reviews (Hudson 2003). The evidence that human error reduction in healthcare is still an urgent priority is underlined by a recent study (Makary and Daniel 2016) which estimated medical errors as the third biggest cause of death in hospitalizations in the United States. Many scholars who investigated patient safety issues in healthcare proposed the application of HRA (Deeter and Rantanen, 2012; Cagliano et al., 2011; Verbano and Turra, 2010; Lyons, 2009; Lyons et al., 2004),

recognising the contribution that this method brought to high-risk industries in terms of human error reduction and safety improvement, mainly thanks to its anticipative approach.

However, extant literature on HRA application in healthcare led to the identification of several methodological limitations (Deeter and Rantanen 2012). In particular, the identification and assessment of factors influencing human performance — the so-called Influencing Factors (IFs) — remains an issue of major concern in healthcare. In this regard, some scholars emphasized the need for a deep adaptation and translation of these techniques, originally designed for the specific needs of industrial contexts, to the healthcare environment (Pandya et al., 2017; Chadwick and Fallon, 2012). When it comes to the surgical setting, although human factors and ergonomics theories gave a great contribution to the identification of factors that could influence surgeons' performance, to the best of the authors' knowledge there is not much literature available about validated taxonomies of Influencing Factors (IFs) for the surgical context. Starting from the existing gaps in literature and from the general state-of-the-art in HRA theory, as reported in more detail in the Section 2, the aim of the present study is twofold: (i) designing a taxonomy of Influencing Factors for surgery; (ii) assessing the influence of the identified IFs on surgical performance by means of eliciting surgeons' perceptions. To this end, the following Research Questions are posed:

- RQ1: “What are the influencing factors (IFs) to be included in a taxonomy for HRA studies in surgery?”
- RQ2: “What is the perceived influence of the influencing factors (IFs) on surgeon's technical performance?”

The study builds on the results of a previous study reported in the literature that proposes a taxonomy of influencing factors in surgery, adopting literature review and field study research methodologies (Onofrio, et al., 2015)

The rest of the paper is organised as follows: Section 2 discusses the state of the art of HRA in healthcare; Section 3 presents the study methodology; Section 4 summarises and discusses the findings, in particular the taxonomy of IFs proposed for HRA studies in the surgical contexts and the quantification of their influence on surgeons' technical performance. Finally, Section 5 draws conclusions and suggests directions for future research.

2. State of the art

2.1 Literature search and review methodology

The present section aims at understanding how the HRA body of knowledge, mainly developed in the industrial sector, has been transferred or adapted for a proper application in the healthcare sector. To this purpose, a literature review methodology has been adopted for investigating the validity and suitability of HRA methods in healthcare, by analysing practical applications and discussing the most relevant theoretical and methodological contributions. The main outputs were: i) synthesising results into a summary of what is and is not known about HRA in healthcare; ii) identifying possible areas of controversy and key gaps in the literature; iii) identifying opportunities and formulating questions for relevant research.

The literature search was limited to the field of “Human Reliability Analysis” in healthcare. As such, the search fell at the intersection between medical, social science,

and engineering scientific literature, since HRA theory and practice was first developed in the industrial sector and only in recent years has been transferred and applied in healthcare settings. Thus, the scope of the analysis covered both domains. Different literature sources were thus explored: three databases covering medical literature (Pubmed, Cochrane and Embase) and three non-medical databases (Scopus, Isi Web of Science and IEEE). To complete the search for scientific and gray literature, Google Scholar was also used. The literature search process was organised as follows: i) screening process to identify descriptive articles and commentaries (examination of title and abstract); ii) insertion of relevant articles derived from Google Scholar; iii) selection of the papers matching the following inclusion criteria:

- Theoretical knowledge of HRA in healthcare;
- Applications of HRA in healthcare.

Finally, no time constraint was set to the literature search; eligible abstracts were selected and full articles were chosen for the complete detailed evaluation.

The results included a total of 288 references (from databases), from which 52 abstracts were selected as relevant papers about human error analysis in the context of healthcare risk management and patient safety. Since the primary objective of the study is investigating the documented HRA applications in different clinical settings and the types of HRA methodologies or techniques that are suitable for application in healthcare, 24 papers were finally selected for full review. Table 1 reports a summary description of the most relevant HRA applications in healthcare derived from the literature search and selection. They are classified according to: Clinical settings, Aim, Method and Results; given the relevance of influencing factors in HRA theory and applications, the set of human and organizational factors included in the reviewed studies, if any, are also reported.

Table 1. Synthesis of the scientific literature on HRA applications in healthcare

References	Clinical settings	Study aim	Method	Results	Human and Organizational Factors considered	
(Joice, Hanna, and Cuschieri 1998)	Endoscopic surgery: laparoscopic cholecystectomy	Technical Errors Documentation	Modified HRA – OCHRA-	Overall Error data (189 errors in 20 cholecystectomies)—error frequency within generic tasks—error frequency in the use of specific instruments	None	
(Malik, White, and Macewen 2003)	Endoscopic dacry-cystorhinostomy	Technical Errors Documentation	Modified HRA – OCHRA-	Recording of active skill- based errors in 12 operations but three different surgical trainees: the frequency of most procedural errors; the frequency of most execution errors; error frequency within generic tasks;	None	
(Alijani, Hanna, and Cuschieri 2004)	Laparoscopic cholecystectomy	Surgical procedures comparison and best practice identification	OCHRA	Errors Identification for Comparison intraoperative cardiac function, postoperative cognitive recovery, and surgical performance of l.c. with abdominal wall lift versus positive- pressure capnoperitoneum - Identification of the number of surgical error in 40 operations	None	
(Tang, Hanna, Joice, et al. 2004)	Laparoscopic cholecystectomy	Technical Errors Documentation	OCHRA	Identification of number of surgical technical errors—the stage of the operation in which errors are enacted most frequently and hazard zones (where these errors have serious consequences) in 200 laparoscopic cholecystectomies and 26 surgeons	None	
(Tang, Hanna, Bax, et al. 2004)	Laparoscopic pyloromyotomy	Technical error analysis	OCHRA	Examination of the performance of LPM by pediatric surgeons during the initial stages of their experience—Nature of consequential and inconsequential errors and their underlying modes in 50 cases of LPM (310 errors in 50 videotapes)—number of errors committed with different instruments—external error modes underlying errors committed with different instruments—error probability for different tasks—hazard zones of an operation—50 videotapes.	None	
(Inoue and Koizumi 2004)	Nursing practice	Error mechanism and influencing factors	EDIT model	Characterization of individual errors by evaluation of error type, performance shaping factors (direct threat), organizational factors (indirect threat). Error rates for module practices from incident reports in 6 hospitals	Yes	Direct threats (Performance Shaping Factors) and indirect threats (Organizational Factors)
(Tang et al. 2006)	Laparoscopic surgery	Application of OCHRA for competency analysis: evaluation of the construct and concurrent validity of OCHRA for competency at a specialist level		Assessment of trainees' operative and cognitive skills during laparoscopic training courses. Significant inverse correlations between the number of technical errors identified by OCHRA and the scores obtained by OSCE. Significant differences between trainees were observed in relation to both overall OSCE scores and OCHRA parameters. The performance of 60 trainees participating in 3-day essential laparoscopic skill training courses were assessed.	None	

(Continue)

References	Clinical settings	Study aim	Method	Results	Human and Organizational Factors considered	
(Gaubia et al. 2008)	Cataract Surgery	Technical errors analysis	OCHRA	Identification of frequency and pattern of technical errors observed during phacoemulsification cataract extraction by surgeons with varying levels of experience. Number of errors performed per task (228 errors)—nature of performed errors—hazard zones—surgical experience of operating surgeon. 33 consecutive phacoemulsification cataract operations performed by 33 ophthalmic surgeons	None	
(Cox, Dolan, and Macewen 2008)	Cataract surgery.	Technical Errors Documentation	OCHRA	Number of errors (84 errors)—probability and consequences of errors—steps necessary to reduce the incidence of such errors to as low as reasonably possible—16 consecutive phacoemulsifications performed by one surgeon	None	
(Talebpour et al. 2009)	Advanced Laparoscopic surgery	Proficiency gain curve assessment	OCHRA	Quantitative description of proficiency-gain curve for a laparoscopic operation and indication of the plateau stage when the individual surgeon attains maximal performance in the execution of a specific procedure Duration of tasks—total task time—total operative time—Economy of movement index—classification and incidence of errors for entire procedures for proficiency gain curve of surgeons.	Yes	Identified performance-shaping factors: concentration lapses (n = 1,321), misjudgements (n = 209), poor camera work (n = 193), fatigue (n = 128), and impaired coordination (n = 108).
(Castiglia, Giardina, and Tomarchio 2010)	Brachytherapy procedures	Human Error Assessment	Modified HEART on the basis of fuzzy set concepts: treezy computer program- fuzzy fault tree	The contribution of a single event to the overall probability of system failure, the contribution of a single event to the uncertainty in the probability of system failure. Impact of error—promoting factors on human error	Yes	Error Promoting condition considered (EPC): 1 - 17 EPCs considered
(Chadwick and Fallon 2012)	Radiotherapy treatment process	Safety barriers selection	Modified HEART	Prior identification of potential errors—Percentage contribution to unreliability and the specification of appropriate defenses against them, thus avoiding costs associated with adverse outcomes.	Yes	Error Producing Conditions of Williams classification (Williams, 1985) considered: 2-7-17-36.

(Continue)

References	Clinical settings	Study aim	Method	Results	Human and Organizational Factors considered	
(Deeter and Rantanen 2012)		test the suitability of a novel approach (CREAM) event analysis in healthcare by comparing and contrasting the results of the currently employed RCA method to those from CREAM.		Existing gap between CREAM analyses and the hospital's previous RCAs, where organizational and leadership factors were either minimized or ignored.	Yes	Common Performance Conditions (CPCs): (1) adequacy of organization, (2) working conditions, (3) adequacy of man-machine interface (MMI)/operational support, (4) availability of procedures/plans, (5) number of simultaneous goals, (6) available time, (7) time of day, (8) adequacy of training/experience, and (9) crew collaboration quality.
(Ahmed et al. 2013)	Bariatric Surgery Retrocolic laparoscopic roux- en-Y gastric bypass for obesity	Technical Errors Documentation	OCHRA	Systematic identification of the surgical technical errors leading to these complications. Identification of hazard zones—technical errors—prescription of “error reduction” mechanism	None	
(Miskovic et al. 2013)	Laparoscopic colorectal surgery	Technical Errors Documentation	CAT score and OCHRA	Identification of the consultant surgeons who can safely perform laparoscopic colorectal surgery before embarking on independent practice Moderate correlation between Cat and OCHRA because are methods based on different principles.	None	
(Ward et al. 2013)	Surgery Incident analysis in central venous catheterisation (CVC)	analysis of the risks surrounding guide wire use (in a surgical incident involving the accidental retention of a guide wire for central venous catheterisation (CVC))		“The nominal likelihood of failure was similar for each of the sub-tasks—approximately 0.01. This is around one order of magnitude greater than that measured through incident reporting over a six-year period in another hospital.”	None	

(Continue)

References	Clinical settings	Study aim	Method	Results	Human and Organizational Factors considered	
(Ghazanfar et al. 2014)		Compare the effect of dividing attention of novices and experts on a laparoscopic task performance	Task analysis according to HRA	“Primary outcome measures were correct auditory responses, task time, the number of surgical errors and instrument movements. Secondary outcome measures included error rate, error probability and hand specific differences. Non-parametric statistics were used for data analysis”	None	
(Mendez et al. 2014)		development and validation of an HD video-based teaching module (HDVM) to help instruct post-graduate otolaryngology trainees in performing neck dissection.	Observational Clinical Human Reliability Assessment (OCHRA) system.	“The total number of errors committed prior to exposure to the HDVM were 91. The total number of errors committed following exposure to the module were 41. This represented a 55% decrease in the number of errors committed pre-HDVM that was statistically significant.”	None	
(Foster et al. 2016)	Laparoscopic rectal surgery	Assessment of the technical performance of the surgical procedure and explore the validity and reliability of the method used (OCHRA).	OCHRA	A total of 335 errors identified, with a median 15 per operation. More errors were observed during pelvic tasks compared with abdominal ones.	None	

2.2 HRA in Healthcare

The literature shows a recent and growing interest in the HRA discipline in healthcare, with an increasing number of practical applications. The most prominent field of current HRA application is the operating room, in particular, endoscopic laparoscopic surgery, cataract surgery, and bariatric surgery. The first attempt to apply HRA in surgery dates back to 1998; it addresses endoscopic surgery, in particular, laparoscopic cholecystectomy (Joice, Hanna, and Cuschieri 1998). Other applications concern nursing practice and the radiotherapy treatment process.

It also emerges that the primary objective of documented HRA applications in surgery is the evaluation of surgeons' technical skills, to achieve different goals, such as performance analysis according to experience level, the proficiency gain curve of surgeons, and comparison between surgical procedures. As Malik et al. (2003) outline: "HRA may prove to be a useful training and assessment tool for the evaluation of technical competency in surgery." The validity of the tool for assessing competency at a specialist level in advanced laparoscopic surgery has been confirmed by Miskovic's studies (Miskovic et al., 2013, 2012).

Some of the HRA techniques adopted in healthcare have been largely adopted in industry (Kirwan 1994), some other were modified and adapted for a proper application in healthcare, and others claim to be HRA-like novel methods, specifically conceived for healthcare applications.

The interest in the potential benefits of HRA applications, especially in surgery, is clearly emerging and increasing in recent years as well (Miskovic et al., 2013, 2012; Cuschieri and Tang, 2010; Talebpour et al., 2009; Malik et al., 2003; Cuschieri, 2000). It is also worth noting that the method with the highest number of reported applications in surgery (12 applications documented in scientific literature) is the Observational Clinical Human Reliability Analysis (OCHRA), specifically developed for the healthcare domain. The main characteristics of this methodology were outlined directly by the authors (Alijani et al., 2004; Tang et al., 2004): detection of technical error (procedural / execution) and its underlying mechanisms, identification of the phases of the procedure with the highest expected frequency of errors and of the stages in which errors cause serious consequences to the patient. On the other hand, Ward et al. (2013) applied the Human Error Assessment and Reduction Technique – HEART - (Williams, 1985) to examine the potential for further occurrences of a surgical incident involving the accidental retention of a guide wire for central venous catheterization. Another clinical setting with documented HRA applications is the radiotherapy treatment. The methodology applied in the two studies is HEART; in one of the two a modification to the original method has been introduced, since fuzzy numbers were used to quantify the parameters of the model (Castiglia, Gardina, and Tomarchio 2010). The peculiarity of these studies is the consideration of the so-called Error-Promoting Condition (EPC), as the HEART technique suggests. In particular, the first 17 of 38 EPCs considered in the original HEART method (Williams, 1985) were used. The second study deals with critical nursing tasks in radiotherapy treatment (Chadwick and Fallon 2012). Furthermore, a recent contribution comes from the Risk and Human Reliability Group at the Paul Scherrer Institute (Switzerland). The study proposes a methodology to develop the foundations of a cognition-based human reliability analysis model for applications in radiotherapy (Pandya et al. 2017).

The extant literature also reveals that the attempts for a theoretical contribution to human error analysis in healthcare mainly come from Lyons' research group (Lyons 2009; Lyons et al. 2004). The authors pointed out that despite the availability of several

different error prediction approaches, such as Human Reliability Analysis methodologies, their current application in healthcare is still very limited (Lyons 2009; Lyons et al. 2004). The reasons lie in the major differences between the complexity of the healthcare operations when compared to the industrial ones. Indeed, healthcare contexts are characterised by a high variability of patient and operational conditions, and the quantification of Human Error Probability (HEP) becomes an even more difficult challenge. In addition, difficulties also stem from the lack of knowledge of methods and the lack of proper support to healthcare professionals in selecting the most appropriate technique (Lyons, 2009). The author also underlines that the systematic application of HRA techniques that might be used to support patient safety improvement under many perspectives are more detailed and more comprehensive than simple audits. This is the reason why the author proposed “a literature-based framework towards the development of a protocol for a novice user in the healthcare sector in selecting a predictive safety technique” (Lyons, 2009). The important contribution of Lyons' studies is about the identification of techniques suitable for application in the healthcare sector, according to the general steps of an HRA study, as identified by Kirwan (Kirwan 1994): data collection, task description, task simulation, human error identification and analysis, human error quantification.

Therefore, there seems to be an urgent need for fostering theoretical knowledge and practical expertise on HRA in patient safety research and among healthcare practitioners as well.

2.3 Types of Influencing Factors (IFs) included in HRA studies in healthcare

Throughout the qualitative and quantitative phases of HRA, many HRA methods make use of Influencing Factors (IFs)—i.e., those aspects of behaviour and context that impact human performance (Boring 2010). Several HRA techniques developed for application in high-risk industries provide a modelling approach and parameters to evaluate the influence of contextual factors on the reliability of human actions or decisions. The contextual factors may vary in nature, and in the number and models of impact mechanisms. In most cases, these factors come with some coefficients that modulate the human error probability (HEP).

The terminology about context factors differs from one HRA technique to another. Some HRA techniques label these factors as "Performance Shaping Factors" (PSFs), such as the THERP technique (Swain and Guttman 1983), or “Common Performance Conditions (CPCs),” such as the CREAM technique (Hollnagel 1998), or again “Error Promoting Conditions (EPCs),” such as the HEART technique (Williams, 1985). Additionally, although some scholars attempted to develop a list of PSFs to be used across sectors (Kyriakidis et al. 2017), there is still not a commonly recognised and standardised set of IFs; despite considerable overlap between factors belonging to different HRA techniques, each method brings with it a slightly different emphasis and slightly different set of PSFs/CPCs/EPCs taxonomies. In the context of the present study, an “Influencing Factor (IF)” in the surgical context is defined as “a specific personal, environmental or organizational aspect that affects human performance and can change the likelihood of a surgeon’s error, having a negative effect on surgeons’ technical performance.”

In the majority of the quantitative HRA techniques, the estimation process of the Human Error Probability (HEP) of the task under investigation starts with a preliminary step where the basic HEP (the conditional or nominal probability) is estimated; secondly, the set of IFs, relevant for the specific task, is defined; finally, the effect of each IF on the

execution of the task, and the combined effects of these IFs, are generally assessed to modify the basic HEP according to this general formula:

$$HEP = NominalHEP * [\sum IF_i * W_i] \quad (1)$$

Where:

W_i = the weight factor of the i-th IF for the specific task;

$IF_i * W_i$ = the weighting effect for the i-th IF on the nominal HEP.

This procedure is used in many methods such as THERP (Swain and Guttman 1983), CREAM (Hollnagel 1998), HEART (Williams, 1985) and SPAR-H (Gertman et al. 2005).

As mentioned before, each specific technique offers its own list of factors and provides guidance on how to assess the status of an IF through direct measures or extrapolations. The determination of the states assumed by the IF can be evaluated by direct assessment methods, through experts' judgement elicitation, experience in the field, or empirical experiments. Furthermore, in some cases, the meaning of the IF is too abstract to be measured, it is then needed to obtain the status of the so-called "indirect IFs" (Boring, Griffith, and Joe 2007), by the combination of states assumed by other IFs. Experts' judgement elicitation is a key pillar of these techniques, nevertheless some of the existing HRA techniques have been criticised for being "judgemental and unstructured", since their strong dependence on a single 'expert assessor' (Kirwan 1994). In fact, the results of any analysis depend on the expert assessor's knowledge of the task under investigation, for which they may not have significant experience. This is particularly relevant in healthcare and in surgery. The reason lies in the fact that human factors or risk experts must deal with highly specialised medical domains, complex clinical tasks and processes, which cannot be neglected (Podofillini et al. 2010).

In the emerging stream of Human Reliability Analysis (HRA) in healthcare, a critical issue emphasised by many authors is the proper application or suitability of the existing taxonomies of Influencing Factors (IFs), since they were primarily designed and developed for industrial contexts (Ward et al. 2013; Chadwick and Fallon 2012; Tang, Hanna, Bax, et al. 2004; Joice, Hanna, and Cuschieri 1998). As a consequence, one relevant methodological flaw of the current way of implementing HRA in healthcare is that there are only a few HRA studies that make use of Influencing Factors (IFs), although influencing factors analysis has been raised by some authors as one of the most interesting dimensions of HRA for healthcare applications (Joice, Hanna, and Cuschieri 1998). This finding is not reflected in HRA theory and applications in the industrial sector, in which issues related to the modelling, selection, and quantification of IFs play a relevant role and still represents a hot research line in the discipline. From the conducted literature review, it also emerged that the currently documented applications of OCHRA - Observational Clinical Human Reliability Analysis (Cuschieri and Tang 2010; Tang et al. 2006; Tang, Hanna, Bax, et al. 2004) - are mainly focused on the assessment of surgeons' technical skills and their potential contribution to adverse events. Hence, personal, organizational, team and contextual factors, all of which contribute to the complexity of the operating theatre, are not addressed by the OCHRA method.

This methodological limitation of HRA applications in healthcare is particularly relevant in surgery, as underlined by the study of Deeter and Rantanen (2012). The authors, by replacing the Root Cause Analysis (RCA) with the CREAM application to the same adverse events, demonstrated how the use of a specific HRA technique, applied retrospectively, returns a thorough analysis of organizational and contextual factors, leading to a better understanding of causal mechanisms that are almost ignored or minimized by RCA.

A closer look at the few papers describing HRA applications in healthcare where human and organizational factors are included in the quantitative or qualitative analysis (Chadwick and Fallon 2012; Deeter and Rantanen 2012; Castiglia, Giardina, and Tomarchio 2010; Inoue and Koizumi 2004) revealed that all the authors dedicated particular attention to the choice of the most appropriate set of IFs (cf. Table 2). This modelling activity generally culminated in the decision to take into account only some of the factors covered by the selected HRA technique. Some authors (Chadwick and Fallon 2012) also highlighted that PSF (or IF) taxonomies incorporated in the most common HRA techniques have been developed and validated in industrial contexts, and as such are not fully applicable to the healthcare sector.

Table 2. Examples of Influencing Factors included in healthcare studies, according to the adopted HRA technique and the SHEL model (Edwards, 1972).

SHEL Component	HEART Error Promoting Conditions (EPCs)	EDIT Model Direct Threats (DTs) and Indirect Threats (IT)s	CREAM Common Performance Conditions (CPCs)
Software	EPC n.4-A means of suppressing or overriding information (Castiglia, Giardina and Tomarchio 2010);	DT n.3-Practice factor (Inoue and Koizumi 2004);	CPC n.4-availability of procedures/plans (Deeter and Rantanen 2012)
	EPC n.5-No means of conveying spatial or functional information to human operator (Castiglia, Giardina and Tomarchio 2010);	IT n. 4-Management of patient information (Inoue and Koizumi 2004);	CPC n.5-number of simultaneous goals (Deeter and Rantanen 2012)
	EPC n.9-Technique unlearning/one which requires application of opposing philosophy (Castiglia, Giardina and Tomarchio 2010);	IT n. 5-Management (Inoue and Koizumi 2004);	
	EPC n.11-Ambiguity in required performance standard (Castiglia, Giardina and Tomarchio 2010);	IT n. 6-Personal management (Inoue and Koizumi 2004);	
	EPC n.13-Poor, ambiguous or ill-matched feed-back (Castiglia, Giardina and Tomarchio 2010);		
	EPC n.16-Poor instructions or procedures (Castiglia, Giardina and Tomarchio 2010);		
	EPC n.17-Little or no independent checking or testing of output (Chadwick and Fallon 2012; Castiglia, Giardina and Tomarchio 2010);		
Hardware	EPC n.6-Poor system/human user interface (Castiglia, Giardina and Tomarchio 2010);	DT n.6-Machine factor (Inoue and Koizumi 2004);	CPC n.3-adequacy of man-machine interface (MMI)/operational support (Deeter and Rantanen 2012);
	EPC n.14-No clear/direct/timely confirmation of intended action from system (Castiglia, Giardina and Tomarchio 2010);		

(Continue)

SHEL Component	HEART Error Promoting Conditions (EPCs)	EDIT Model Direct Threats (DTs) and Indirect Threats (IT)s	CREAM Common Performance Conditions (CPCs)
Environment	<p>EPC n.2-A shortage of time available for error detection & correction (Chadwick and Fallon 2012; Castiglia, Giardina and Tomarchio 2010);</p> <p>EPC n.3- Noisy/confused signals (Castiglia, Giardina and Tomarchio 2010);</p> <p>EPC n.7. No obvious means of reversing an unintended action (Chadwick and Fallon 2012; Castiglia, Giardina and Tomarchio 2010);</p> <p>EPC n.8-Information overload Castiglia, Giardina and Tomarchio 2010);</p> <p>EPC n.12-Mismatch between perceived and actual risk (Castiglia, Giardina and Tomarchio 2010);</p> <p>EPC n.36. Task pacing caused by the intervention of others (Chadwick and Fallon 2012; Castiglia, Giardina and Tomarchio 2010);</p>	<p>DT n.1- Environmental and facilities factor (Inoue and Koizumi 2004);</p> <p>DT n.2-Work environment factor (Inoue and Koizumi 2004);</p> <p>IT n.1-National culture (Inoue and Koizumi 2004);</p> <p>IT n.2-Institutional culture (Inoue and Koizumi 2004);</p> <p>IT n. 3-Professional culture (Inoue and Koizumi 2004);</p>	<p>CPC n.1-adequacy of organization (Deeter and Rantanen 2012);</p> <p>CPC n. (2) working conditions (Deeter and Rantanen 2012);</p> <p>CPC n.6-available time (Deeter and Rantanen 2012);</p> <p>CPC n.7- time of day(Deeter and Rantanen 2012);</p>
Liveware	<p>EPC n.1- Unfamiliarity with novel or infrequent situation which is potentially important Castiglia, Giardina and Tomarchio 2010);</p> <p>EPC n.10-Transfer knowledge from one task to another (Castiglia, Giardina and Tomarchio 2010);</p> <p>EPC n.15-Inexperience (newly qualified but not an expert) (Castiglia, Giardina and Tomarchio 2010).</p>	<p>DT n.4-Personal factor(Inoue and Koizumi 2004);</p> <p>DT n.5-Team factor (Inoue and Koizumi 2004).</p>	<p>CPC n.8-adequacy of training/experience(Deeter and Rantanen 2012);</p> <p>CPC n.9-crew collaboration quality (Deeter and Rantanen 2012).</p>

2.4 Adapting HRA techniques for healthcare applications

Finally, when it comes to the adaptation of existing HRA techniques to the healthcare context, scholars pinpoint at least three issues that must be addressed for a proper integration of a taxonomy of IFs. First, there is an open debate about which IFs should be used in HRA and what is the appropriate number of IFs to be included in the analysis. In fact, “There is a considerable range in the number of IFs provided by individual HRA methods, ranging from single factor models such as time-reliability curves, up to 50 or more IFs in some current HRA models.” (Boring 2010). For example, the US Nuclear Regulatory Commission advocates 15 IFs in its HRA Good Practices (Kolaczkowski et al. 2005), while its SPAR-H method (Boring and Blackman 2007) espouses the use of 8 IFs, and its ATHEANA method (Barnes, Bley, and Cooper 2000) features an open-ended number of IFs. The apparent differences in the optimal number of IFs can be partly explained by referring to the level of detail of the human model or the different functions of IFs in HRA (Boring, 2010; Kim and Jung, 2003).

Second, IFs are conceived in terms of the negative effects they might induce on human performance, even though a greater emphasis has been recently given to possible ways of modelling IFs that also take into account their potential influence on enhancing human performance, i.e. reducing the HEP (Hallbert et al., 2006).

Third, the terminology of the IFs is strictly connected to the meaning given in a specific context. For this reason, the assessments provided by different analysts could differ because of subjective interpretations of the meaning of IF labels and descriptions under different operational contexts (Kim and Jung, 2003), undermining the transparency and reliability of several HRA studies.

3 Study Methodology

The research methodology followed a multi-phase and mixed method approach. Focus groups, individual interviews and questionnaire methods were used in each phase, depending on the specific objective. Given the complex nature of contemporary social science, there is an increasing consensus around the benefits that were identified in the literature about mixed-method approaches to address different research settings and problems (Johnson, Onwuegbuzie, and Turner 2007). As stated by Johnson et al. (2007): “This type of research should be used when the nexus of contingencies in a situation, in relation to one’s research question(s), suggests that mixed methods research is likely to provide superior research findings and outcomes.” In the context of the present study, this property fits with the need of deep adaptation and translation of the well-structured HRA discipline in industry to a new application domain, where experience is still limited and literature is scanty. In particular, a qualitative research approach was adopted to explore the role of IFs, specific to the surgical environment, and design the final taxonomy. Then, a quantitative approach was used to assess the influence of IFs on surgeons’ technical performance, according to the surgeons’ perception.

3.1 Designing a taxonomy of IFs for HRA application in surgery

The design of IF taxonomy items, to be implemented in the final taxonomy of IFs for HRA application in surgery, was performed involving healthcare professionals in Italy and Denmark. Starting from the preliminary list of IF categories proposed by Onofrio et

al. (2015), focus group methodology (Krueger and Casey 2000) and individual semi-structured interviews were applied aiming at capturing the ways in which surgeons and anaesthetists conceptualize, cluster and articulate IFs.

In particular, the focus group held in Denmark (in March 2015) involved only surgeons (n = 7), whereas the individual semi-structured interviews involved both surgeons (n=3) and anaesthetists (n = 9); in Italy, semi-structured interviews involved surgeons only (n=9). The size of the focus group (7 participants) obeys to the common standard of four (minimum) to twelve (maximum) participants per group (Krueger and Casey 2000). The participants were selected through consultation with the manager of the surgical department of the Hillerød Hospital (Copenhagen, DK). The informal nature of this methodology facilitated an interactive discussion and reflection about the surgeons' perception of factors that shape surgical performance.

An interview guide was created for both the focus group session and individual semi-structured interviews. It focuses on a brief presentation of the study and shows the participants an A4 paper including a table with the list of the categories of factors identified from (Onofrio, et al., 2015). The set of IF items was developed and turned into a checklist to gather data from surgeons about each factor category and assure the traceability of answers. Both the guide and the checklist were originally implemented in English; in Denmark they were proposed in English, and the focus group presentation and discussion took place in English as well. However, the presence of a mother tongue moderator offered the opportunity of a translation into Danish during the discussion, when requested by some attendees. For their use in Italy, the two documents were translated into Italian and validated through a process of back translation.

Content-wise, the discussion was organised following the criteria suggested by Kim and Jung (2003) for the selection and structuring of IFs items:

- Including factors to assess the overall task context in a comprehensive way;
- Avoiding overlap between factors;
- Selecting the factors that directly affect human error occurrences;
- Selecting the factors that could be reflected in human error/reliability analysis;
- Selecting the factors that are observable in practice;
- The terminology should be as specific and practical as possible so that the meanings of the terms might be easily understood by practitioners.

The focus group session lasted about two hours and was led by two moderators, who covered all the IF categories in the interview guide. The discussion was primarily focused on labels, definitions, meanings of the factors to validate items and descriptors of influencing factors. In order to identify and understand the relevant IF items, when an interviewee was explicitly referring to the specific IF item, the interviewers marked the checklist. When some IF items were not explicitly mentioned, the interviewers asked additional questions about the missing ones. The ending of each interview occurred when the participants had no further thoughts on IFs.

Each of the individual semi-structured interviews lasted around 30 minutes of discussion. All the interviews were audio-recorded and subsequently transcribed verbatim. Data analysis steps were discussed continuously among the authors, and when possible, with other scholars, ensuring the reliability of data analysis.

The information collected from the focus group and individual interviews was analysed using a uniform approach based on transcription, validation, and integration. In particular, the coding technique was used to examine data line by line, or paragraph by paragraph, for significant events, experiences, feelings, and so on, that were then denoted as concepts (Corbin and Strauss 2014). Comparative tables facilitated the analysis of

information (Engle 1999). The analysis started with the first order codes, to identify informants' views of the Influencing Factors on surgeons' performance. The second step consisted in organizing and grouping the first-order codes into categories, highlighting similar characteristics shared among them to be further investigated.

3.2 Questionnaire method for the quantification of the impact of IFs on surgical performance through experts' judgements elicitation

In line with the HRA approach, IFs are those factors that negatively affect the human performance and their influence needs to be quantified accordingly. To this end, the objective of the second part of the study was the assessment of the impact of the proposed 21 IFs on surgeons' performance (the IFs investigated in this study are reported and discussed later in section 4.1, cf. Table 3). The assessment was based on experts' judgement elicitation process, to overcome the limitations in using the judgements of a single expert assessor during a specific HRA application, as it is in the majority of the HRA applications. The practical aim was to assess the degree of influence of an IF on the operator's performance rated on a scale from 0 to 10. As a matter of example, in HEART technique (Williams, 1985) this parameter is referred to as the "proportion of affect" (PoA). It can also be viewed as the conditional probability that the IF negatively affects the operator's performance, given it is present.

The questionnaire method was chosen to get responses from a large number of experts, and the data gathered may therefore be seen to generate findings with higher consistency and potential for generalisation. The questionnaire method is also suitable because willing respondents (Italian surgeons) who are in a position to provide meaningful data about a topic can be identified, and the purpose of our descriptive research was to generate a profile of the sample.

The target population of the analysis refers to the Italian surgeons. In order to reach out to expert surgeons and collect the opinion of a large number of experts, three Surgical Societies have been involved ("Associazione Chirurghi Ospedalieri Italiani – ACOI"ⁱ, "Società Italiana di Chirurgia – SIC"ⁱⁱ, and "Società italiana di Urologia – SIU"ⁱⁱⁱ) which endorsed the study. Indeed, a convenience sampling strategy was adopted, specifically the non-probability sampling method, which is based on data collection from a population of members who were conveniently willing to be involved in the study.

The questionnaire was divided into three sections: the first one was dedicated to the instructions, the second one was dedicated to capturing the surgeon's perception about the influence of each of the IFs on her/his technical performance, when present. In detail, the question repeated for each one of the identified Influencing Factors (X) was:

"Please provide your subjective estimate of how often the factor X has had a negative influence on your own performance and on the risk of adverse patient outcome in the surgical operations in which you have participated during the last year?"

Please indicate your estimate by a number from 0 and 10 where:

0 = the factor has had no influence at all, if present;

10 = the factor has had a negative influence in every operation, if present".

Finally, the third one was dedicated to the collection of demographic information: the prevalent surgical technique used in the last year, with distinction between open, and

minimally invasive surgery (i.e. laparoscopic and robotic surgery); position/title of the surgeon; surgical specialty; years of professional experience; the surgical role during the last year of surgical activity, if involved in robotic surgery (i.e., first operator in console, or second operator at the operating table); the total number of surgical interventions performed to date (approximate number). Finally, two open questions were included: one to stimulate suggestions for possible additional factors not listed in the questionnaire; and a second one dedicated to further general comments, if any.

3.2.1 Implementation Strategy

The questionnaire was edited in English and then translated into Italian. Once the draft questionnaire had been designed a pilot test has been run with a group of surgeons of the Urology Department of a leading Italian hospital.

The main reasons that suggested giving preference to an online administration of the questionnaire are: i) most of the surgeons in Italy have easy access to high speed internet; ii) the questionnaire is short (28 closed-ended questions and 2 optional questions / about 15 mins); iii) the interaction with surgeons out of the working hours is extremely difficult. Typeform © software package was used to implement the questionnaire online.

The questionnaire was administered in collaboration with the three Surgical Societies that endorsed the study and promoted the administration of the questionnaire through their institutional websites and electronic mailing lists. E-mails inviting surgeons to answer the online questionnaire included a cover letter, a link to the questionnaire, and instructions for respondents. Two reminders were sent after one and two months from the first launch.

3.2.2 Data analysis

The questionnaire allowed us to estimate the relative rank that surgeons assign to factors that negatively impact surgeons' performance. Data were recorded in MS Excel, and subsequently analysed with the Statistical Package for the Social Sciences (IBM – SPSS Statistics v23.0).

Data analysis has been organised into multiple steps. Two preliminary steps were the descriptive statistics of the data and the normality test of the pdf of the scores of the 21 variables.

Direct logistic regression was performed to assess the impact of five factors on the likelihood that respondents would report higher values (median value) of the negative influence of the factor perceived as the most influential factor by surgeons. The model contained five independent variables (age, job position, number of interventions, professional age, and surgical speciality).

Subsequently, two surgical groups (open vs mini-invasive surgery) have been compared through non parametric test. Finally, through fitting analysis, it was possible to investigate if the pdf of each IF variable can be assumed as a specific curve. Eliciting a probability distribution rather than asking an expert for a single point estimate allows an expert to express her/his assessment on the uncertainty about the parameter; then, through sensitivity analysis techniques, the consequences of this uncertainty can be investigated (Saltelli, 2002).

4 Results and discussion

4.1 A taxonomy of IFs for surgery

The first type of findings concerns the level of awareness of those factors as influential on surgeons' performance. In order to achieve the goal, the 11 IF categories introduced in Onofrio et al. (2015) have been initially considered. 10 categories were identified in the scientific literature on human factors in healthcare, and include: Noise and background talk not related to the task; Safety culture and safety climate; Standardisation; Communication and teamwork; Experience and team training; Leadership; Staffing and team member familiarity; Workload; Equipment HMI and space design; one additional factor was identified from a field study in Italy, namely Distraction.

All the IF items proposed for evaluation were confirmed as relevant by interviewees. During the focus group, one could never observe any disagreement and opposite views on the factors discussed with the informants. Whenever an informant brought attention to one of the influencing factors, the others supported the relevance of the mentioned factor with a variety of outstanding example cases. This also happened during the individual interviews, where none of the indicated IFs had been defined as inappropriate or negligible.

The discussion during the interviews and focus group was also targeted to reaching a good agreement among respondents on labels, definitions, and precise meanings of each single IF. Twenty-one IF items resulted from the selection by the focus group and individual semi-structured interviews. Table 3 provides the final taxonomy of IFs and their respective labels and definitions.

Table 3. Validated list of IFs for HRA applications in surgery.

SHEL Component	IF category (Onofrio et al., 2015)	IF items
Environment	Noise & back-ground talk not related to the task	<u>Noise and ambient talk.</u> Continuous or sudden noise; team members talking in the background or coming and going and moving around in a noisy way.
		<u>Music.</u> Presence of background music in the operating room.
	Distractions	<u>Noisy use of social media.</u> Team members talking about and obtrusively sharing social media content.
		<u>Verbal interruptions.</u> Verbal Interruptions that are either untimely or not patient relevant.
	Safety Culture and Safety Climate	<u>Poor management of errors and threats to patient safety.</u> Failure to share information promptly and openly about errors and threats to patient safety.

SHEL Component	IF category (Onofrio et al., 2015)	IF items
Software	Standardisation	<p><u>Poor guidelines, procedures or checklists.</u> Guidelines, procedures or checklists are inadequate: lacking, too complex, or not at right level.</p> <p><u>Improper use of procedures and checklists.</u> The improper use, or non-use, of the WHO checklist (or similar), protocols and procedures.</p>
Hardware	Equipment, HMI and space design	<p><u>Inadequate ergonomics of equipment and work place.</u> Equipment and workplace not designed to optimize usability and reduce operator fatigue and discomfort</p> <p><u>Poor use of technology:</u> Lack of ability to use relevant technology.</p>
Liveware	Communication and team-work	<p><u>Rude talk and disrespectful behaviours.</u> Derogatory remarks, behaviours showing a lack of respect of OR team members, shouting and harsh tones of voice.</p> <p><u>Unclear or failed communication.</u> Communication that should have been given wasn't or was inadequate or was misunderstood and not corrected.</p> <p><u>Poor coordination.</u> Failure in coordinating team activities; failure to anticipate the needs of the lead surgeon or lead anaesthetist (surgeon at the console in robotic surgery).</p> <p><u>Poor decision making.</u> Failure to consider, select and communicate options; inadequacy or delay in implementing and reviewing decisions.</p> <p><u>Poor situation awareness.</u> Failure to gather and/or to integrate information or failure to use information to anticipate future tasks, problems, and states of the operation.</p>
	Experience and Team Training	<p><u>Lack of experience of surgical team colleagues.</u> Lack of experience of within surgical team, with the surgical procedure or technology.</p> <p><u>Lack of experience of anaesthetics team colleagues.</u> Lack of experience within anaesthetic team, with the anaesthetic procedure or technology</p>
	Fatigue	<u>Fatigue.</u> Mental fatigue or physical fatigue.

SHEL Component	IF category (Onofrio et al., 2015)	IF items
	Leadership	<u>Poor leadership.</u> Failure to set and maintain standards or to support others in coping with pressure.
	Staffing and team member familiarity	<u>Team member unfamiliarity.</u> Team members unfamiliar with each other and each other's competencies.
	Workload	<u>Time pressure.</u> Psychological stress resulting from experiencing a need to get things done in less time than is required or desired.
		<u>Perioperative Emotional Stress.</u> Stress induced by factors not directly related to the team, or to the characteristics and evolution of the intervention, e.g., responsibility for the budget and for other hospital objectives, organizational problems of the department, other critical patients, lawsuits.

Many of the 21 IFs identified are not included in the existing IF taxonomies of HRA techniques. This represents a contribution to existing knowledge on HRA taxonomies in a different domain in order to investigate at a lower level of classification the contingent and detailed factors of the specific domain.

Beside the validation of the IF items in each single category, some useful information about the perception of the influence of the IFs on surgeon performance also emerged from the interviewees. In particular, interviewed surgeons and anaesthetists clearly revealed a personal ranking among factors, by openly claiming that some factors have higher negative influence than other. A second relevant aspect is the difference between surgical contexts: the perceived impact of some IFs vary when considering open vs. laparoscopic vs. robotic surgeries.

The results also contributed to raising awareness about the need for rich qualitative analyses within HRA applications in healthcare to produce qualitative insights on performance (Pandya et al., 2017). One example is the meaning of the valence of a single factor which may limit or magnify performance, such as the presence of background music in the operating room (IF item: music). Results also have relevant implications for improving patient safety in surgery: increasing awareness of those IFs with the potential for impairing surgeons' technical performance and its variability may prevent compromises in safety tasks and instead allows to give higher priority to plan, diagnosis, and the use of checklists. However, the assessment of possible positive effects associated to some IF categories was beyond the scope of the present study, where only the negative influence of specific IF items has been assessed, for the sake of better HRA applications in surgery.

4.2 Quantification of the impact of IFs on surgeons' performance

A convenience sample of surgeons provided anonymous answers to the questionnaire. The three Surgical Societies promoted the survey leveraging on their institutional mailing

lists or other lists of suitable contacts (Table 4). On the one hand, the use of three independent channels maximised the access to the target population, but, on the other, it did not allow for a full control over the precise number of the recipients reached out. The estimated number of reliable contacts in the three mailing lists is higher than 2,500, in total.

Table 4. Composition of the sample.

	ACOI	SIC	SIU
Type of recipients	Surgeons, ACOI associate members	Surgeons, SIC associate members	Surgeons, enrolled in a SIU training course
Estimated number of reliable contacts in the used mailing list	> 1,500	> 1,000	15
Number of usable questionnaires		215	
Estimated minimum response rate		8.6%	

Eventually, 215 anonymous questionnaires were collected, with an estimated minimum response rate of 8.6%; 105 questionnaires came from surgeons working in open surgery (49%), and 110 came from surgeons operating primarily in mini-invasive surgery (51%). Table 5 summarises other demographic characteristics of the respondents.

Table 5. Demographic characteristics of respondents.

Number of respondents: 215
Age of respondents: from 26 to 70, mean: 51; standard deviation: 11.83
Years of experience: from 1 to 46, mean: 23; standard deviation: 11.86
Total number of surgical interventions: 10 to 25000
Surgical settings:
- Open surgery: 105 (49%)
- Mini-Invasive surgery: 110 (51%)

Table 6 shows the cumulative descriptive profiles about the perceived influence of IFs (215 respondents). For 11 IFs, out of the 21, the perceived influence has a mean rating equal to or higher than the median of the entire sample (4.00); they are: IF n. 7 – *rude talk and disrespectful behaviours* (5.07); IF n. 16 – *time pressure* (4.74); IF n. 15 – *fatigue* (4.51); IF n. 5 – *poor management of errors and threats to patient safety* (4.29); IF n. 10 – *poor coordination* (4.29); IF n. 14 – *lack of experience of anaesthetics team colleagues* (4.26); IF n. 9 – *unclear or failed communication* (4.26); IF n. 3 – *noisy use of social media* (4.20); IF n. 17 – *perioperative emotional stress* (4.06); IF n. 13 – *lack of experience of surgical team colleagues* (4.03); and IF n. 1 – *noise and ambient talk* (4.01).

Table 6. Descriptive statistics of 21 IFs.

IFs							
	1 NOISE AND AMBIENT TALK	2 MUSIC	3 NOISY USE OF SOCIAL MEDIA	4 VERBAL INTERRUPTIONS	5 POOR MANAGEMENT OF ERRORS AND THREATS TO PATIENT SAFETY	6 POOR GUIDELINES, PROCEDURES AND CHECKLISTS	7 RUDE TALK AND DISRESPECTFUL BEHAVIOURS
Mean	4,01	1,26	4,20	3,80	4,29	3,36	5,07
Median	4,00	0	4,00	4,00	4,00	2,00	5,00
Mode	4	0	0	0	0	0	8
Variance	8,416	4,051	11,887	8,216	9,943	10,892	12,193
Minimum	0	0	0	0	0	0	0
Maximum	10	9	10	10	10	10	10

IFs							
	8 IMPROPER USE OF PROCEDURES AND CHECKLISTS	9 UNCLEAR OR FAILED COMMUNICATION	10 POOR COORDINATION	11 POOR DECISION MAKING	12 POOR SITUATIONAL AWARENESS	13 LACK OF EXPERIENCE OF SURGICAL TEAM COLLEAGUES	14 LACK OF EXPERIENCE OF ANAESTHETICS TEAM COLLEAGUES
Mean	3,70	4,26	4,29	3,71	3,86	4,03	4,26
Median	3,00	4,00	4,00	3,00	3,00	3,00	4,00
Mode	0	0	0	0	0	0	0
Variance	10,635	10,270	10,936	9,064	10,575	10,877	9,892
Minimum	0	0	0	0	0	0	0
Maximum	10	10	10	10	10	10	10

(Continue)

	IFs						
	15 FATIG UE	16 TIME PRESS URE	17 PERIOPERA TIVE EMOTIONA L STRESS	18 POOR LEADER SHIP	19 TEAM MEMBER FAMILIA RTY	20 POOR USE OF TECHNOL OGY	21 INADEGU ATE ERGONO MICS OF EQUIPME NT AND WORK PLACE
Mean	4,51	4,74	4,06	3,40	2,90	3,73	3,75
Media n	5,00	5,00	3,00	3,00	2,00	3,00	3,00
Mode	2	8	0	0	0	0	0
Varian ce	9,588	10,279	10,209	10,129	9,187	11,214	10,535
Minim um	0	0	0	0	0	0	0
Maxim um	10	10	10	10	10	10	10

When all the surgical settings are taken into consideration (i.e. both open and mini invasive surgeries), “*Rude talk and disrespectful behaviours*” is the IF perceived by surgeons as the one with the most negative impact on their technical performance. This finding has been further discussed with some of the involved surgeons to achieve a better understanding of the underlying causes and mechanisms of influence. All the contacted surgeons agreed with the fact that “*Rude talk and disrespectful behaviours*” paralyse the working activity of all the surgical team on which the outcome of the surgical intervention depends.

Finally, direct logistic regression was performed to identify possible demographic characteristics of the surgeons correlated to higher estimated impacts of the 11 most critical factors (i.e. those with mean rating > 4.00). Table 7 and Table 8 report the independent variables and the results, respectively, obtained for IF n. 7 “*Rude talk and disrespectful behaviours*.” The model contained five independent variables: age, job position, number of interventions, professional age, and surgical speciality. The five univariate logistic regressions were not statistically significant for all the 11 tested IFs.

Table 7. Independent variables

Independent variables	Levels of analysis		
	Age (3 levels)	≤ 40 years	$40 < x \leq 55$ years
Job position (2)	Surgeon	Director	
Number of interventions (3 levels)	≤ 100	$100 < x \leq 1000$	> 1000
Professional age (3 levels)	$0 \leq x \leq 9$	$10 \leq x \leq 30$	≥ 30
Surgical typology (3)	General surgery	Urgency	Specialized

Table 8. The logistic regression - Type 3 Analysis of Effects

Effect	Wald		
	DF	Chi-Square	Pr > ChiSq
Job position	1	1.6226	0.2027
Age	2	1.1154	0.5725
Professional age	2	0.1756	0.9159
Number of Interventions	2	0.1010	0.9508
Surgical typology	2	0.0921	0.9550

The next step of the analysis on the relative weight of the influence of IFs on surgeons' performance was focused on the comparison between different surgical techniques and settings, i.e. open vs mini-invasive surgery. The Mann-Whitney Wilcoxon test (nonparametric test) was used to this end. There was a significant ($p < .05$) difference in the perceived influence of IFs in the two surgical contexts for two IFs only, namely: i) IF n. 1 *noise and ambient talk* ($p=0.049$); ii) IF n. 4 *verbal interruptions* ($p=0.030$). The descriptive statistics are reported in Table 9 and graphically shown in Figure 1.

Table 9. Descriptive statistics about the perceived influence of IFs resulted significantly different in the two surgical contexts: open vs mini-invasive surgeries.

Surgical Approach		Statistics	Error std.		
1. Noise and ambient talk	Open	Mean	3.66	.291	
		95% the confidence interval for the average	Inferior limit	3.08	
			Superior Limit	4.23	
		Mean cut to the 5%	3.52		
		Median	3.00		
		Variance	8.862		
		Dev. std.	2.977		
		Min	0		
		Max	10		
		Mini-Invasive	Mean	4.35	.267
95% the confidence interval	Inferior limit		3.83		

		for the average	Superior Limit	4.88
		Mean cut to the 5%		4.28
		Median		4.00
		Variance		7.827
		Dev. std.		2.798
		Min		0
		Max		10
4. Verbal Interruptions	Open	Mean		3.36 .264
		95% the confidence interval for the average	Inferior limit	2.84
			Superior Limit	3.89
		Median		3.00
		Variance		7.329
		Dev. std.		2.707
		Min		0
	Max		10	
	Mini-Invasive	Mean		4.25 .282
		95% the confidence interval for the average	Inferior limit	3.69
			Superior Limit	4.80
		Median		4.00
		Variance		8.756
		Dev. std.		2.959
Min			0	
Max		10		

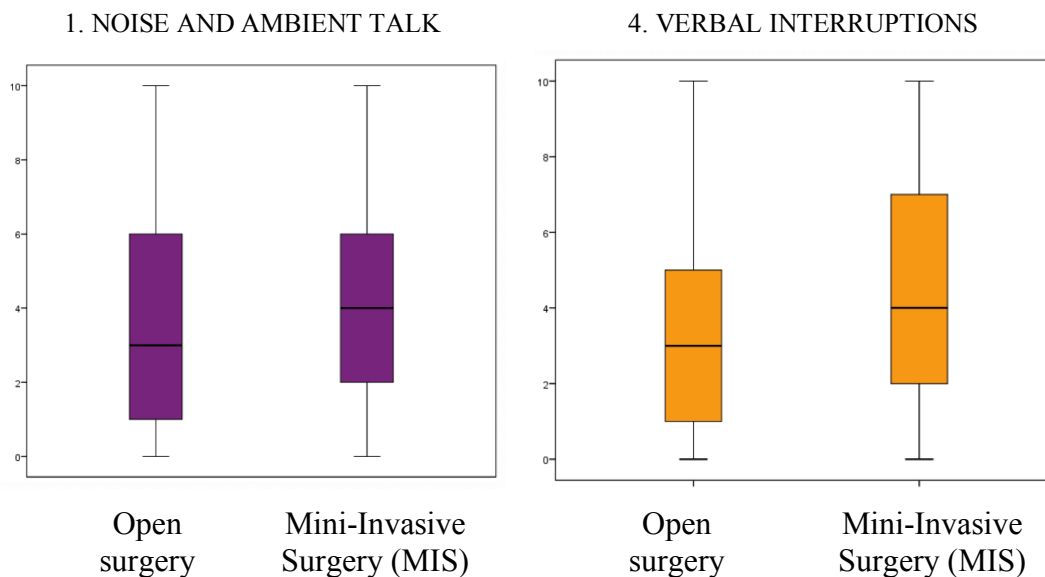


Figure 1. Descriptive profiles of the perceived influence of IFs resulted significantly different in the two surgical context.

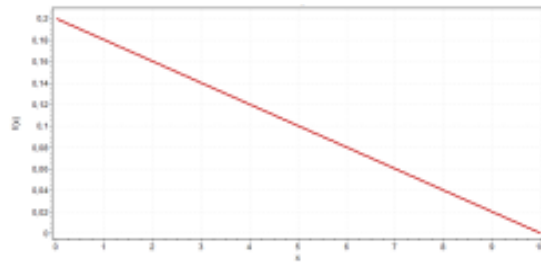
Surgeons perceive the negative influence of these two factors as stronger in mini-invasive surgery than in open surgery. As for IF n. 1 “*noise and ambient talk*,” the first and third quartile are similar for both contexts, whereas the first and third quartile of the perceived influence of IF n. 4 “*verbal interruptions*” are higher for the mini-invasive setting. Values of the median in mini invasive surgery are higher (1.00 point) than in open surgery for both IFs.

Despite the descriptive statistics of the two IFs are similar, the Mann-Whitney Wilcoxon test supports the hypothesis of taking the two aggregate perceptions as different populations and pdfs. This result has significant implications in terms of ergonomic and safety design requirements of operating rooms and organisational set-ups in surgery. Indeed, results indicate that surgeons operating in MIS have the perception of being relatively more vulnerable to multiple sources of disturbance, and it might be due to the fact that the cognitive workload and the level of coordination among the team members (e.g. first and second operators) in MIS are higher compared to the traditional open surgery setting.

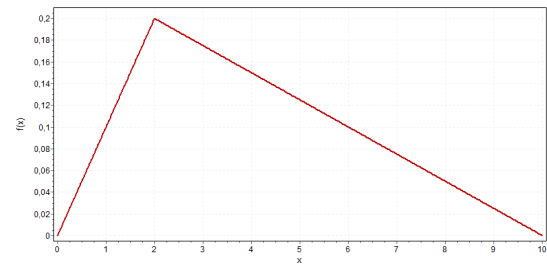
Finally, the fitting analysis was performed. The Goodness-of-Fit tests were executed for four common pdfs – namely, the lognormal, Weibull, gamma, and beta pdf – for all the 21 IFs, and none of the hypotheses tested acceptable at $p < 0.05$. Then, in order to arrive at a pdf usable in the context of HRA studies, a triangular distribution has been assumed, which is frequently used in risk analysis to represent epistemic uncertainties related to experts’ evaluations (Kotz and Van Dorp, 2006; Johnson, 1997). The triangular distribution has three parameters: the lower limit a , the upper limit b , and the mode m . The probability density at $x=m$ is determined by imposing the area of the triangle to 1.

Accordingly, the modal value (most likely value) has been calculated in the interval between the observed values (for each IF: a =minimum value observed; b =maximum value observed) considering that 0 and 10 are the extremes of the scale used in the questionnaire.

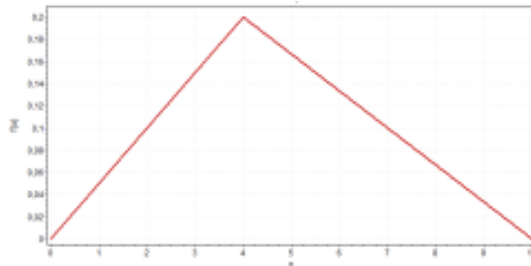
As shown by the triangular probability distribution functions reported below in Figure 2, 0 is the most likely value for 17 IFs; 2 is the modal value for the influencing factor n.15 “fatigue”; 4 is the modal value for influencing factor n.1 “Noise and ambient talk”; 8 is the modal value for the influencing factor n.7 “Rude talk and disrespectful behaviours” and the influencing factor n.16 “Time Pressure”.



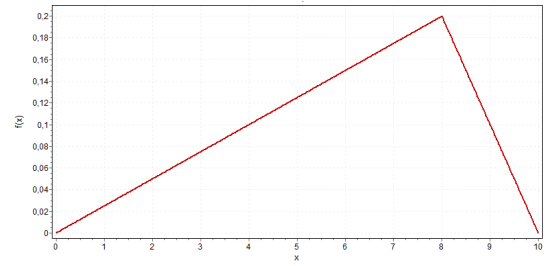
a) triangular distribution ($a=0; b=10; m=0$) for IFs: 3; 4; 5; 6; 8; 9; 10; 11; 12; 13; 14; 18; 19; 17; 21; 20.



b) triangular distribution ($a=0; b=10; m=2$) for IF: 15.



c) triangular distribution ($a=0; b=10; m=4$) for IF: 1.



d) triangular distribution ($a=0; b=10; m=8$) for IFs: 7; 16.

Figure 2 (a, b, c, d). Triangular distributions of the 21 IFs

5 Conclusions

The review of extant literature on HRA in healthcare has shown significant gaps at the methodological level. Indeed, only a very limited number of HRA studies implemented Influencing Factors analysis in the quantification of the HEP, despite it represents a peculiar component and a well-established step in several HRA techniques for their application in industrial environments. Moreover, the taxonomies adopted in the most common HRA techniques, such as THERP and HEART, have been developed and validated in the industrial domain; many researchers highlighted that their unmodified adoption in the healthcare context is problematic and undermines the reliability of results.

As a methodological contribution to the need for adapting existing HRA techniques for applications in healthcare, and surgery in particular, the present study proposes a validated ad hoc taxonomy of IFs, leveraging the elicitation of surgeons' perception. The new taxonomy supports a more coherent and transparent selection of IFs to be implemented in HRA applications in surgery. Additionally, the quantification of their impact on the HEP associated with surgical tasks has been proposed in the form of a triangular pdf for each single IF. Indeed, knowing the pdf of the perceived influence of an IF helps the analyst in reaching a better estimation and increasing the reliability of the study, when compared to the common practice of referring to a single assessor (Kirwan et al., 1997). Furthermore, it would be possible to directly use the pdf of the selected IFs as inputs in a stochastic HRA simulation to account for the uncertainties on experts' perceptions. These results contribute to the diffusion and enhancement of HRA in healthcare, improving the quality of future HRA applications in surgery, where the diffusion of advanced technological solutions adds further complexity and, consequently, possible new error pathways.

Beyond the novelty at the methodological level, the study also shed light on the most relevant influencing factors that impair the surgeon's performance, increasing the

probability of surgical errors. In particular, “*Rude talk and disrespectful behaviours*” is the IF perceived by surgeons as the one with the highest negative impact. This result has no prior evidence in literature, and is worth of further investigation for achieving a deeper understanding of the underlying mechanisms. More in general, the pdfs of the 21 IFs on surgeon’s performance can be used as reference for setting safety priorities in different surgical procedures. Particular attention should be paid to the minimally invasive and robotic surgeries, where a higher level of vulnerability of surgical teams to different sources of disturbance has been highlighted in the present study. Indeed, the HEP in mini-invasive surgical procedures was assessed relatively more vulnerable to “*noise and ambient talk*” (IF n.1) and “*verbal interruptions*” (IF n. 4).

Furthermore, no statistical correlation has been found between the estimated negative effect of the IFs on the HEP and the surgeon’s profile. It can be argued that the age and the technical expertise (i.e. total number of interventions) of the surgeon do not explain the differences in the perceived negative effects of any of the considered IFs. Also these findings are worth of further investigation, maybe focusing on those that already proved to be dependent on the surgical setting, and leveraging on a larger sample.

The study suffer from some limitations. First, the proposed taxonomy of IFs cannot be directly integrated into any one of the existing HRA techniques. New research endeavours are needed for matching or aligning the proposed taxonomy with those adopted in the most frequently used HRA techniques. A specific set of multipliers - i.e. the maximum predicted nominal amount by which unreliability may increase (Williams, 1985) – needs also to be established before using the new IFs when calculating the nominal HEP of a generic surgical procedure.

Moreover, the analysis has been limited to mini-invasive surgery (MIS) without differentiating between laparoscopic surgery and robotic surgery; the present study was not able to test if some IFs have different impacts on the HEP when moving from one technique to the other.

The study has relevant practical implications as well. Indeed, while ongoing training programmes clearly focus on granting the technical proficiency of the surgeon, they are not able to reduce surgeon’s vulnerability to the most critical IFs. In this regard, the effectiveness of the most common training practices seems to be limited to the increase in surgeons’ awareness. Finally, when it comes to the comparison of different surgical techniques, results show that surgeons implementing mini-invasive surgery are more likely to experience detrimental effects of IFs on performance than surgeons involved in open surgery.

Overall, the study offered a clear demonstration of the importance of developing specific knowledge for fostering a proper application of HRA in healthcare. In this regard, it calls for a new research agenda.

Acknowledgements

The authors wish to thank Prof. Henning Boje Andersen (DTU Management Engineering Department - Denmark) for his contributions to the development of the IF Taxonomy and the support in carrying out the focus group. The research that is presented in this paper originates from a development process that was carried on by continuously interacting and collaborating with healthcare professionals (surgical departments in Italy and Denmark). Moreover, as mentioned in the paper, the research was conducted in collaboration with three Italian surgical societies (*Associazione Chirurghi Ospedalieri Italiani ACOI*; *Società Italiana di Chirurgia SIC*; *Società Italiana di Urologia SIU*), to

whom the authors are extremely thankful. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

References

Ahmed, K., Ashrafian, H., Harling, L., Patel, V.M., Rao, C., Darzi, A., Hanna, G.B., 2013. Safety of training and assessment in operating theatres--a systematic review and meta-analysis. *Perfusion*, Vol. 28 No. 1, pp. 76–87. Doi:10.1177/0267659112460882.

Alijani, A., Hanna, G.B. and Cuschieri, A., 2004. Abdominal Wall Lift Versus Positive-Pressure Capnoperitoneum for Laparoscopic Cholecystectomy. *Annals of Surgery*, Vol. 239 No. 3, pp. 388–394. Doi:10.1097/01.sla.0000114226.31773.e3.

Barnes, M.J., Bley, D. and Cooper, S., 2000. Technical basis and implementation guidelines for a technique for human event analysis (ATHEANA). NUREG-1624.

Boring, R.L., 2010. How Many Performance Shaping Factors are Necessary for Human Reliability Analysis ?. No. INL/CON-10-18620. Idaho National Laboratory (INL).

Boring, R.L. and Blackman, H.S., 2007. The origins of the SPAR-H method's performance shaping factor multipliers, 2007. IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting, Ieee, pp. 177–184.

Boring, R.L., Griffith, C.D. and Joe, J.C., 2007. The Measure of human error: Direct and indirect performance shaping factors. IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting, Ieee, pp. 170–176.

Cagliano, A., Grimaldi, S. and Rafele, C., 2011. A systemic methodology for risk management in healthcare sector. *Safety Science*, Vol. 49 No. 5, pp. 695–708. Doi: 10.1016/j.ssci.2011.01.006.

Castiglia, F., Giardina, M. and Tomarchio, E., 2010. Risk analysis using fuzzy set theory of the accidental exposure of medical staff during brachytherapy procedures. *Journal of radiological protection : official journal of the Society for Radiological Protection*. Vol. 30 No. 1, pp. 49–62. Doi: 10.1088/0952-4746/30/1/004.

Chadwick, L. and Fallon, E.F., 2012. Human reliability assessment of a critical nursing task in a radiotherapy treatment process. *Applied ergonomics*, Vol. 43 No. 1, pp. 89–97. Doi: 10.1016/j.apergo.2011.03.011.

Corbin, J. and Strauss, A., 2014. *Basics of qualitative research: Techniques and procedures for developing grounded theory*, Sage publications.

Cox, a, Dolan, L. and Macewen, C.J., 2008. Human reliability analysis: a new method to quantify errors in cataract surgery. *Eye (London, England)*, Vol. 22 No. 3, pp. 394–7. Doi: 10.1038/sj.eye.6702648.

Cuschieri, A., 2000. Human reliability assessment in surgery--a new approach for improving surgical performance and clinical outcome. *Annals of the Royal College of surgeons of England*, pp. 83–87.

Cuschieri, A. and Tang, B., 2010. Human reliability analysis (HRA) techniques and observational clinical HRA. *Minimally invasive therapy & allied technologies: MITAT : official journal of the Society for Minimally Invasive Therapy*, Vol. 19 No. 1, pp. 12–7. Doi: 10.3109/13645700903492944.

Deeter, J. and Rantanen, E., 2012. Human Reliability Analysis in Healthcare. *Proceedings of the 2012 Symposium on Human Factors and Ergonomics in Health Care*, Human Factors and Ergonomics Society, pp. 45–51. Doi: 10.1518/HCS-2012.945289401.008.

Edwards E., 1972. *Man and machine: systems for safety*, in proceedings of British Airline pilots association Technical symposium. British airlines pilots association, London.

Engle, M., 1999. *Qualitative Data Analysis: An expanded Sourcebook* (2nd Ed.) Matthew B. Miles and A. Michael Huberman. Thousand Oaks, CA: Sage publications. 1994. *American Journal of Evaluation*, Vol. 20 No. 1, pp. 159–160. Doi: 10.1016/S1098-2140(99)80125-8.

Foster, J.D., Miskovic, D., Allison, A.S., Conti, J.A., Ockrim, J., Cooper, E.J., Hanna, G.B., et al., 2016. Application of objective clinical human reliability analysis (OCHRA) in assessment of technical performance in laparoscopic rectal cancer surgery. *Techniques in Coloproctology*, Springer International Publishing. Vol. 20 No. 6, pp. 361–367.

Gauga, V., Tsangaris, P., Tossounis, C., Mitra, A., McLean, C. and Saleh, G.M., 2008. Human Reliability Analysis of Cataract Surgery. *Archives of ophthalmology*. Vol. 126 No. 2, pp. 173–7. Doi: 10.1001/archophthalmol.2007.47.

Gertman, D., Blackman, H., Marble, J., Byers, J. and Smith, C., 2005. The SPAR-H human Reliability Analysis Method, Idaho National Laboratory, US Nuclear Regulatory Commission.

Ghazanfar, M.A., Cook, M., Tang, B., Tait, I. and Alijani, A., 2014. The effect of divided attention on novices and experts in laparoscopic task performance. *Surgical endoscopy*. Doi:10.1007/s00464-014-3708-2.

Hallbert, B., Boring, R., Gertman, D., Dudenhoefter, D., Whaley, A., Marble, J. Joe, J., 2006. Human event repository and analysis (HERA) system overview, NUREG/CR-6903, Volume 1. US Nuclear Regulatory Commission, Washington, DC.

Hollnagel, E., 1998. *Cognitive Reliability and Error Analysis Method (CREAM)*, Cognitive Reliability and Error Analysis Method (CREAM), Elsevier.

Hudson, P., 2003. Applying the lessons of high risk industries to health care. *Quality and safety in health care*.

Inoue, K. and Koizumi, A., 2004. Application of human reliability analysis to nursing errors in hospitals. *Risk analysis : an official publication of the Society for Risk Analysis*. Vol. 24 No. 6, pp. 1459–73. Doi: 10.1111/j.0272-4332.2004.00542.x.

Johnson, D., 1997. The triangular distribution as a proxy for the beta distribution in risk analysis. *Royal Statistical Society: Series D*. Vol. 46 No. 3, pp. 387–398.

Johnson, R., Onwuegbuzie, A.J. and Turner, L., 2007. Mixed methods research. *Journal of mixed methods research*. Vol. 1 No. 2, pp. 112–133. Doi: 10.1258/135581907781543085.

Joice, P., Hanna, G.B. and Cuschieri, A., 1998. Errors enacted during endoscopic surgery--a human reliability analysis. *Applied ergonomics*. Vol. 29 No. 6, pp. 409–14.

Kim, J.W. and Jung, W., 2003. A taxonomy of performance influencing factors for human reliability analysis of emergency tasks. *Journal of Loss Prevention in the Process Industries*. Vol. 16 No. 6, pp. 479–495. Doi: 10.1016/S0950-4230(03)00075-5.

Kirwan, B., et al., 1997. The validation of three human reliability quantification techniques e THERP, HEART and JHEDI: part II e results of validation exercise. *Applied Ergonomics*. 28, 17e25. Doi:10.1016/S0003-6870(96)00045-2.

Kirwan, B., 1994. *A Guide To Practical Human Reliability Assessment*, A Guide To Practical Human Reliability Assessment. CRC Press, Doi:10.1016/0925-7535(95)90078-0.

Kolaczowski, A., Forester, J., Lois, E. and Cooper, S., 2005. Good Practices for Implementing Human Reliability Analysis (HRA) (NUREG-1792). Nureg-1792.

Kotz, S. and Van Dorp, J.R., 2006. A novel method for fitting unimodal continuous distributions on a bounded domain utilizing expert judgment estimates. *IIE Transactions*. Vol. 38 No. 5, pp. 421–436. Doi: 10.1080/07408170500517267.

Krueger, R. and Casey, M.A., 2000. Focus groups: A practical guide for applied research. *Review Literature And Arts Of The Americas*. Vol. 22, pp. 129–152. Doi: 10.1002/j.1556-6678.2007.tb00462.x.

Kyriakidis, M., Kant, V., Amir, S. and Dang, V.N., 2017. Understanding human performance in sociotechnical systems – Steps towards a generic framework. *Safety Science*, Elsevier Ltd, No. July. Doi:10.1016/j.ssci.2017.07.008.

Lyons, M., 2009. Towards a framework to select techniques for error prediction: Supporting novice users in the healthcare sector. *Applied Ergonomics*. Vol. 40 No. 3, pp. 379–395. Doi: 10.1016/j.apergo.2008.11.004.

Lyons M., Adams, S., Woloshynowych, M., Vincent, C., 2004. Human reliability analysis in healthcare: A review of techniques. *International Journal of Risk & Safety in Medicine*, Vol. 16 No. 4, pp. 223–237.

Makary, M. and Daniel, M., 2016. Medical error — the third leading cause of death in the US. *Bmj*, Vol. 2139 No. May, pp. 1–5. Doi: 10.1136/bmj.i2139.

Malik, R., White, P.S. and Macewen, C.J., 2003. Using human reliability analysis to detect surgical error in endoscopic DCR surgery. *Clinical Otolaryngology and Allied Sciences*, Vol. 28 No. 5, pp. 456–460. Doi: 10.1046/j.1365-2273.2003.00745.x.

Miskovic, D., Ni, M., Wyles, S.M., Kennedy, R.H., Francis, N.K., Parvaiz, A., Cunningham, C., 2013. Is Competency Assessment at the Specialist Level Achievable? A Study for the National Training Programme in Laparoscopic Colorectal Surgery in England. *Annals of surgery*, Vol. 257 No. 3, pp. 476–482. Doi: 10.1097/SLA.0b013e318275b72a.

Miskovic, D., Ni, M., Wyles, S.M., Parvaiz, A. and Hanna, G.B., 2012. Observational clinical human reliability analysis (OCHRA) for competency assessment in laparoscopic colorectal surgery at the specialist level. *Surgical endoscopy*, Vol. 26 No. 3, pp. 796–803. Doi: 10.1007/s00464-011-1955-z.

Onofrio, R., Trucco, P. and Torchio, A., 2015. Towards a Taxonomy of Influencing Factors for Human Reliability Analysis (HRA) Applications in Surgery. *Procedia Manufacturing*, Vol. 3, pp. 144–151. Doi: 10.1016/j.promfg.2015.07.119.

Pandya, D., Podofillini, L., Emert, F. and Lomax A.J., and Dang V.N., 2017. Developing the foundations of a cognition-based human reliability analysis model via mapping task types and performance-influencing factors: Application to radiotherapy. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, Vol 232, Issue 1, pp. 3 – 37. Doi:10.1177/1748006X17731903.

Podofillini, L., Dang, V., Zio, E., Baraldi, P. and Librizzi, M., 2010. Using expert models in human reliability analysis—a dependence assessment method based on fuzzy logic. *Risk analysis: an official publication of the Society for Risk Analysis*, Vol. 30 No. 8, pp. 1277–97. Doi: 10.1111/j.1539-6924.2010.01425.x.

Saltelli, A., 2002. Sensitivity analysis for importance assessment. *Risk Analysis*, Vol. 22, pp. 579–590.

Swain, A.D. and Guttman, H.E., 1983. *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications Final Report*. NUREG/CR-1278 SAN80-0200 RX, AN.

Talebpour, M., Alijani, A., Hanna, G.B., Moosa, Z., Tang, B. and Cuschieri, A., 2009. Proficiency-gain curve for an advanced laparoscopic procedure defined by observation clinical human reliability assessment (OCHRA). *Surgical Endoscopy and Other Interventional Techniques*, Vol. 23 No. 4, pp. 869–875.

Tang, B., Hanna, G.B., Bax, N.M. and Cuschieri, A., 2004. Analysis of technical surgical errors during initial experience of laparoscopic pyloromyotomy by a group of Dutch pediatric surgeons. *Surgical endoscopy*, Vol. 18 No. 12, pp. 1716–20. Doi: 10.1007/s00464-004-8100-1.

Tang, B., Hanna, G.B., Carter, F., Adamson, G.D., Martindale, J.P. and Cuschieri, A., 2006. Competence assessment of laparoscopic operative and cognitive skills: Objective Structured Clinical Examination

(OSCE) or Observational Clinical Human Reliability Assessment (OCHRA). World journal of surgery, Vol. 30 No. 4, pp. 527–34. Doi: 10.1007/s00268-005-0157-z.

Tang, B., Hanna, G.B., Joice, P. and Cuschieri, A., 2004. Identification and categorization of technical errors by Observational Clinical Human Reliability Assessment (OCHRA) during laparoscopic cholecystectomy. Archives of surgery, Vol. 139 No. 11, pp. 1215–1220.

Verbano, C. and Turra, F., 2010. A human factors and reliability approach to clinical risk management: Evidence from Italian cases. Safety Science, Vol. 48 No. 5, pp. 625–639. Doi: 10.1016/j.ssci.2010.01.014.

Vincent, C., Taylor-Adams, S. and Stanhope, N., 1998. Framework for analysing risk and safety in clinical medicine. BMJ, Vol. 316 No. April, pp. 1154–1157. Doi: 10.1136/bmj.316.7138.1154.

Ward, J., Teng, Y.C., Horberry, T., and Clarkson, P.J., 2013. Healthcare human reliability analysis – by HEART. In M. Anderson (Ed.), 'Contemporary Ergonomics and Human Factors 2013, pp. 287–288. UK: Taylor and Francis. Doi: 10.1201/b13826-62.

Williams, J.C., 1985. HEART - A proposed method for achieving high reliability in process operation by means of human factors engineering technology in Proceedings of a Symposium on the Achievement of Reliability in Operating Plant, Safety and Reliability Society (SaRS), NEC, Birmingham.

i <http://www.acoi.it/>

ii <http://www.sicitalia.org/>

iii <http://www.siu.it/>