A Methodology for Dynamic Human Reliability Analysis in Robotic Surgery

Rossella Onofrio and Paolo Trucco

Department of Management, Economics and Industrial Engineering - Politecnico di Milano, Via Lambruschini, 4/b - 20156, Milan, Italy

To cite this article:

"Onofrio, R., & Trucco, P. (2020). A methodology for Dynamic Human Reliability Analysis in Robotic Surgery. Applied Ergonomics, 88, 103150."

Highlights:

- A methodology for mesoergonomics investigation in surgery is proposed
- A modified HEART procedure is coupled with the DET modelling of the surgical procedure
- Monte Carlo method is used to assess the impact of Error Modes and Influencing Factors
- Uncertainty on patient outcome in Robot-Assisted Radical Prostatectomy is evaluated
- Team related factors are found to have the highest impact on patient outcome uncertainty

A Methodology for Dynamic Human Reliability Analysis in Robotic Surgery

Abstract

Surgery has changed significantly in recent years due to the introduction of advanced technologies, resulting in increased system complexity at the technical, human and organisational levels, which may lead to higher variability of patient outcome due to new error pathways. Current approaches towards a safer surgery are largely based on ex-post analysis of events and process monitoring (e.g. root cause analysis, safety checklists, safety audits). However, adopting a proactive approach enables the prior identification of critical factors and the design of safer sociotechnical systems, thanks to a multi-level (or mesoergnomics) perspective. In this paper, a methodology for performing mesoergonomics analysis of surgical procedures is proposed. It is a methodology for Dynamic Human Reliability Analysis in Robotic Surgery based on a modified version of human error assessment and reduction technique (HEART) integrated with a method for incorporating uncertainties related to the influence of personal and organisational factors on the execution of a surgical procedure. The pilot application involves a robot-assisted radical prostatectomy procedure, and the results reveal that team-related factors have the greatest impact on patient outcome variability.

Keywords: Human Reliability Analysis (HRA), HEART, Dynamic Event Tree (DET), Healthcare, Robotic Surgery.

1. Introduction

In recent years, minimally invasive surgery (MIS) has undergone remarkable development due to rapid technological innovation, and it has replaced traditional open, invasive surgery in several surgical procedures. Modern surgery seeks to not only remedy a patient’s condition but also minimise disruption and maximise the treatment (Chang et al., 2014; Hamad and Curet, 2010) and MIS grants less mental and physical impact on the patient and significantly reduces the potential for complications due to surgical wounds.

Robotic surgery, or robot-assisted surgery, is a specific type of MIS that allows doctors to perform many types of complex procedures with high precision, flexibility and control (Al-Naami et al., 2013). Robotic surgery is intended to overcome the limitations of laparoscopic surgery, such as 2D vision, inconsistent instrument movements, unnatural surgeon positions, dissociation between vision and instrument control and inability to perform micro-sutures. Using a computer and remote handling system, a surgeon can reproduce the movement of the human hand during surgery (Al-

Naami et al., 2013). The surgeon's experience remains fundamental for the assessment of the patient, the selection of information, and the execution of the operation. Proper diagnosis and assessment of the status, condition and risk class of the patient are crucial to ensure safe robotic surgical procedures. For some patients, robotic surgery is not suitable, unnecessarily expensive or riskier than traditional methods, while for others, in safe conditions, it may be more precise and effective than laparoscopic surgery (Al-Naami et al., 2013). Over the last decade, new robot-assisted surgical procedures have been developed in the fields of oncology, gynaecology, orthopaedics and maxillofacial, thoracic, paediatric, ophthalmologic and also cardiac surgery (Al-Naami et al., 2013).

However, this advanced technology may be associated with quality and safety issues due to its high degree of human–machine interaction and procedural complexity (Randell et al., 2016). Assessment of its clinical efficacy and patient safety implications is still early and has produced contrasting results and recommendations, with reports of several different limitations and problems associated with specific robotic surgical procedures. For example, some authors report problems due to poor communication between the operating surgeon and the rest of the surgical team, particularly the surgical assistant (Cao and Rogers, 2006b). Indeed, robotic surgery disrupts the existing workflow and changes the roles of every team member (Lai and Entin, 2005). Technical failures are reported as well, mainly due to system malfunction (e.g. instrumentation) or collision of the robotic arms with the patient, surgeon or each other (Binder et al., 2004). Taking a broader scope, a recent study from Onofrio and Trucco (2018) shows that the surgeon's technical performance is more vulnerable to personal and organisational risk factors in an MIS context than in an open surgery context. More specifically, the authors highlight the primary importance of team factors such as verbal interruptions and noise and ambient talk (Onofrio and Trucco, 2018).

The adoption of a proactive and design-based approach to patient safety enables prior identification of the most critical factors and a more effective and ergonomic design of the system, thus allowing one to avoid the costs associated with adverse outcomes (Pandya et al., 2020). However, to the best of authors' knowledge, there is a lack of scientific literature addressing proactive analysis of human errors in robotic surgery. Furthermore, as argued by Karsh and colleagues (Karsh, 2014; Karsh and Brown, 2010), cross-level interactions in complex socio-technical systems should be taken into consideration to fully understand the causal mechanisms that link personal and organisational factors to patient safety. However, a proper implementation of the human factors engineering paradigm in healthcare (Karsh, 2006) requires the development of a robust methodological approach supported by suitable techniques.

With the aim of answering Karsh's call for enhancing human factors engineering application in healthcare, in this paper a new methodology is proposed to support multi-level mesoergonomic investigations in surgery (Karsh et al., 2014) (Figure 1). A modified version of the human error assessment and reduction technique (HEART; Williams, 1986) was developed and tested by applying it to a robot-assisted radical prostatectomy (RARP) procedure (Galfano et al., 2010). More specifically, a structured methodology for incorporating uncertainties related to the influence of personal, team, and organisational factors (so-called influencing factors, or IFs) on the surgeon's human error probability (HEP) is proposed; the cumulative effect of multiple IFs on the expected patient outcome is finally assessed by combining a dynamic event tree (DET) model of the surgical procedure and Monte Carlo simulation.

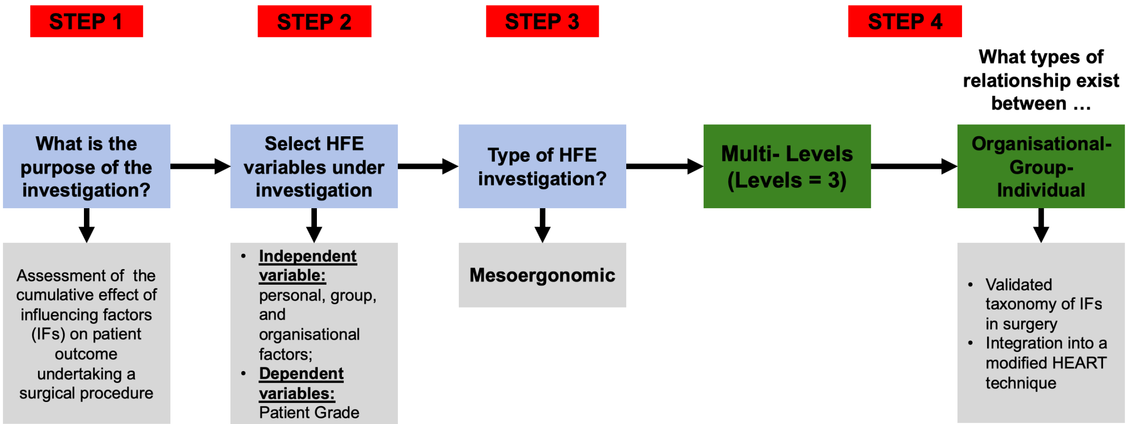

Figure 1. Overview of the proposed methodology according to the framework for mesoergonomics (Karsh et al., 2014).

The rest of the paper is organised as follows. Section 2 provides a detailed description of the original HEART technique. Section 3 briefly reports on the state-of-the-art on HEART applications in healthcare, and some recent advances that are relevant to the present study. Section 4 presents the modification of the HEART technique for applications in surgery. Section 5 introduces the proposed methodology for the Dynamic Human Reliability Analysis of a surgical procedure. Section 6 and 7 present a pilot application to a RARP procedure and discuss the main results. Finally, Section 8 draws conclusions and suggests directions for future research.

2. Overview of HEART technique: the original method and recent advances

The original form of HEART technique (Williams, 1986) includes 11 steps:

1. Determine the task or scenario under analysis.
2. Assign a Nominal Human Unreliability to the task.
3. Identify relevant Error Promoting Conditions (EPCs).
4. Take the first/next relevant EPC.
5. Determine the Assessed Proportion of Affect (PoA).

6. Calculate the Assessed EPC Affect.
7. Consider any other EPC and repeat steps 4,5,6.
8. Calculate the Assessed Nominal Likelihood of Unreliability (ANLU).
9. Calculate the relative Percentage Contribution to Unreliability (%CU).
10. Determine remedial measures.
11. Consider the presence of any other more tasks/scenarios for analysis.

Step 1 involves the identification of the task under analysis. Then, a nominal human unreliability (NHU) is assigned to the task (Step 2), referring to the eight generic task types (GTTs) as reported in Williams (1986). If none of these eight task descriptions fits the task under analysis, the following values can be used, as recommended by Williams (1986):

- Proposed NHU: 0.03;
- Percentile bounds (5–95th): 0.008-0.11.

In Step 3, the assessor chooses the main EPCs that influence the operator's performance, ensuring not to count EPCs twice by overlaying them on generic tasks. The assessor then determines the assessed "proportion of affect" (PoA) for each selected EPCs (Steps 4 and 5), which results in a measure of the EPC's effect magnitude, rated on a scale from 0 to 1. The multiplier factor associated with each EPC is defined by Williams (1986) as the "maximum predicted nominal amount by which unreliability might change going from good conditions to bad". If an analyst determines that many EPCs are applicable, the model will tend towards further unreliability (pessimism). The set of general formulae used to evaluate the error probability of each critical task (Steps 6–9) is as follows:

$$\mathrm{AssessedEPCAffect}_i = [(\mathrm{EPCMultiplier}_i - 1) * \mathrm{PoA}_i] + 1 \qquad (1)$$

$$\mathrm{ANLU} = \mathrm{NHU} * \prod_{i=1}^{n} \mathrm{AssessedEPCAffect}_i \qquad (2)$$

$$\%\mathrm{CU} = \frac{\mathrm{AssessedEPCAffect}_i}{(\mathrm{NHU} + \sum \mathrm{ni=1}\ \mathrm{AssessedEPCAffect}_i)} \qquad (3)$$

These formulae calculate (1) the assessed effect of the i-th EPC; (2) the ANLU, and (3) the % CU (Contribution to Unreliability) of the i-th EPC, respectively. The CU enables the assessor to rank the EPCs in terms of their gross affect on unsuccessful task completion.

With reference to the list of EPCs originally proposed by Williams (1986), it is important to notice that lexical precision is crucial for obtaining reliable and reproducible applications, since the quantitative assessment of ANLU is highly dependent on the multipliers suggested for each EPC. The literature offers evidence of possible inconsistencies induced by difficulties in translating the original industry-oriented terminology of EPCs into terms that are meaningful for applying HEART in specific domains and in healthcare in particular (Chadwick and Fallon, 2012; Pandya et al. 2017).

2.1 Revisions of the list of EPCs and multipliers for application in specific domains

The original HEART technique already received some modifications in the past, to overcome limitations and to better define the set of EPCs to specific domains. In particular, Nuclear Action Reliability Analysis (NARA; Kirwan et al., 2005, and 2016) is a modified version of HEART for HRA studies in nuclear power plant (NPP) operations; it was developed using updated data from the HRA analysis of UK nuclear power plants (Kirwan et al. 2016). On the other hand, Controller Action Reliability Analysis (CARA; Kirwan, 2007) is a modification of the original HEART technique specifically developed for HRA studies in air traffic management (ATC). CARA was developed based on the results of adapting HEART to different domains, such as railway transport (Kim et al., 2006) and nuclear (Kirwan et al., 2005, and 2016) industries. Modifications were introduced to take into account specific aspects of the context and mainly resulted in a tailored list of GTTs and a specific taxonomy of EPCs.

A detailed description of NARA and CARA is beyond the scope of this paper, but the selection process of suitable EPCs for the healthcare sector is relevant to the current study. Thus, the EPC taxonomies developed in NARA and CARA as well as the multipliers originally proposed by Williams (1986) were used for guidance as it will be fully described in Section 4.

3. State-of-the-art applications of HEART in healthcare

Since the publication of *To Err is Human* (Kohn et al., 1999), a key concern of healthcare professionals has been investigation of human error in different clinical settings. Patient safety and clinical risk management researchers have sought to learn from high-risk industries how to reduce human error, particularly how the body of knowledge on system safety engineering can be used to enhance patient safety (Verbano and Turra, 2010). There has also been growing interest in human reliability analysis (HRA) in the healthcare sector. The most prominent field in which HRA is applied is surgery, especially endoscopic laparoscopic, cataract and bariatric surgery. The first attempt to apply HRA-like methods in surgery was made by Joice et al. (1998), who introduced observational clinical human reliability analysis (OCHRA) as a way to investigate endoscopic surgery. Other more recent applications address nursing practice (Inoue and Koizumi, 2004) and the radiotherapy treatment process (Chadwick and Fallon, 2012).

According to literature, the HRA methods and techniques that have been applied in healthcare include observational clinical human reliability analysis (OCHRA; Joice et al., 1998); error type, direct threat and indirect threat (EDIT; Inoue and Koizumi, 2004); objective structured clinical examination (OSCE; Tang et al., 2006); modified versions of HEART (Castiglia et al., 2010; Chadwick and Fallon, 2012; Pandya et al., 2017; Ward et al., 2013); the Cognitive Reliability and Analysis Method (CREAM; Deeter and Rantanen, 2012) and competency assessment tool (CAT) score (Miskovic et al., 2013). Some are mainly adopted in industrial settings and represent a best practice (Kirwan, 1994). Others have been modified and adapted for healthcare applications, and some are novel HRA-like methods that were specifically developed for healthcare applications (Onofrio and Trucco, 2018).

HEART is one of the most well-known HRA techniques with documented applications in healthcare (Lyons et al., 2004, 2009; Williams, 1986). A closer look at the papers describing HEART applications in healthcare reveals that there is general consensus regarding the opportunity of applying HEART in healthcare, as it is a quick and useful method for analysing different healthcare processes and treatments (Castiglia et al., 2010; Chadwick and Fallon, 2012; Lyons, 2009; Lyons et al., 2004). Nevertheless, the same authors underline that the HEART technique may not be suitably developed or generic for application to all healthcare tasks, and accordingly they proposed a series of modifications to properly take into account the peculiarities of the healthcare context (Chadwick and Fallon, 2012).

Table 1 summarises the main characteristics of HEART applications in healthcare documented in the extant literature. The clinical setting, objectives of the study, modifications applied to the traditional HEART technique and EPCs considered are listed.

Table 1. Applications of HEART in the Healthcare Context

Reference	Clinical setting	Aim of the study	Modifications to the original HEART method	Results
(Pandya et al., 2017)	Radiotherapy treatment process	Development of a methodology	Cognition-based human reliability analysis model involving mapping of task types and performance-influencing factors.	Methodology to develop a GTT-PIF structure (Generic Task Types - Performance Influencing Factor) as the causal mapping foundation for a new HRA method based on GTT and PIF.
(Chadwick and Fallon, 2012)	Radiotherapy treatment process	Safety barrier selection	Modified HEART: Participative team approach	Prior identification of potential errors by determining percentage contribution to unreliability and the appropriate defences against errors. Thus, the costs associated with adverse outcomes are avoided.
(Castiglia et al., 2010)	Brachytherapy procedures	Human error assessment	Fuzzy set of concepts, including the Treezy computer program and fuzzy fault tree	The contribution of a single event to the overall probability of system failure, the contribution of a single event to uncertainty in the probability of system failure, and the impact of error (promoting factors affecting human error).
(Ward et al., 2013)	Accidental retention of a guide wire for central venous catheterisation (CVC) inside a patient's venous system	Human error assessment and evaluation of the applicability of the technique	HEART	The nominal likelihood of failure was similar for each of the sub-tasks (approximately 0.01), which is around one order of magnitude greater than that measured through incident reporting over a six-year period in another hospital.

An investigation into a surgical incident involving accidental retention of a guide wire for central venous catheterisation inside a patient's venous system (Ward et al., 2013) is the first example of a healthcare application of a modified version of HEART. Three critical sub-tasks were analysed independently by a team comprised of a safety engineer, a human factors expert and a medical student. Only 12 EPCs, rather than the original 38, were considered in the analysis. Ward et al. (2013) underlined the pros and cons of the HEART original technique. The researchers also pointed out some difficulties encountered in the study, including those regarding the correct interpretation and translation of original EPC descriptions in the healthcare context, the lack of accurate data, and integration of the high level of variability of many contextual factors (Ward et al., 2013).

Castiglia et al. (2010) used a modified version of HEART to assess the impact of EPCs on the exposure of radiological medical operators working at a high dose rate (HDR) brachytherapy irradiation facility. In this study, the HEART technique was modified to better address the uncertainties encountered when choosing and assessing EPCs by making use of fuzzy sets.

Chadwick and Fallon (2012) adopted the HEART technique to analyse a critical nursing task, 'record abnormal blood results', in a radiotherapy treatment process. The aim of this task was to determine the main factors influencing correct completion of the task. The researchers underlined that the traditional HEART technique has some limitations. The modifications they proposed called for an assessment team, instead of a unique assessor, and a new way of evaluating the strength of EPCs' impact based on experts' judgments. Accordingly, some steps of the traditional methodology were modified, and the set of EPCs considered in the study was limited to the following: shortage of time available for error detection and correction (%CU: 49%); no obvious means of reversing an unintended action (31%); little or no independent checking or testing of output (14%) and task pacing caused by the intervention of others (6%) (Chadwick and Fallon, 2012).

In a more recent HRA study of HDR radiotherapy, Castiglia et al. (2015) integrated the HEART and technique for human error-rate prediction (THERP) techniques to assess human errors leading to radiological over-exposure of patients. The THERP technique was used to draw the event tree of errors in two different tasks (i.e. computation of dose distribution and textual documentation of dosimetry details). These tasks were divided into subtasks, and it was determined whether the task was performed correctly. The stages of the task were reported in a logical order that allowed for more accurate error assessment. For each subtask, the HEP was obtained using fuzzy HEART. The following EPCs were considered in the study: little or no independent checking or testing of output; mismatch between perceived and actual risk; information overload; transfer of knowledge from one task to another; poor, ambiguous or ill-matched feedback and ambiguity of the required performance standard (Castiglia et al., 2015).

With reference to the abovementioned limitations of the HEART technique, when applied to healthcare procedures, many authors argue that it is difficult to express complete confidence in the reliability and replicability of results, and they confirm the need for further modification and enhancement to support the applicability of HEART in healthcare after its validity was established in several industrial domains (Kirwan, 1994). This is why the HEART technique has been chosen in the present study as the reference for developing an ad hoc HRA methodology to be used in surgery.

4. Revision of the HEART technique for application in surgery

In line with previous studies, the modified HEART technique proposed by Chadwick and Fallon (2012) for healthcare applications was employed as a reference. A participative team approach was adopted rather than an involving just a single external expert assessor, as required by the original HEART method (Williams, 1986). A participative team approach allows one to gather all the necessary information through interviews and questionnaires administered to the surgeons involved in the study. Expert surgeons are directly involved in selecting the appropriate NHU category, associating IFs with tasks selected from the surgical taxonomy (Onofrio and Trucco, 2018) and assessing the corresponding assessed PoA. The PoA coefficient is used to determine the extent to which each identified EPC affects the surgeon's performance and was rated on a scale from 0 to 100. Table 2 presents the main differences between the traditional HEART technique and the modified version adopted in this study.

Table 2. Modifications to the original HEART technique.

Proposed modifications to HEART	***Original HEART (Williams, 1986)***	***Rationale***	
Specific taxonomy for surgical context: 21 IFs (Step 3 of traditional HEART)	Taxonomy of 38 EPCs focused on industrial working conditions	In-depth investigation of the surgical context through a validated IF taxonomy designed for surgery (Onofrio and Trucco, 2018)	*4.1*
Assessor team is asked to assess the amount of PoA (PoA*) attributed to the EPC (Step 3 of traditional HEART)	None	Mapping analysis of the two taxonomies	
Participative team composed of three surgeons (Steps 3–5 of traditional HEART Chadwick and Fallon, 2012)	Single assessor	Reduce subjectivity, as the original method is heavily based on the experience of a single assessor, in line with HRA applications in healthcare (Chadwick and Fallon, 2012)	
Rating scale of 0–100 is used to obtain PoA values for each EPC (Step 5 of traditional HEART)	Calculation of PoA rated on a scale from 0 to 1	Increase the precision with which the uncertainty of the EPC factors can be estimated	

Selection and calibration of EPCs for healthcare

In line with the approaches adopted in NARA and CARA for achieving a tailored modification of the original HEART technique, we preliminary adopted the full set of IFs proposed by Onofrio and Trucco (2018) for the healthcare domain. Table 3 reports the cross-matching of the taxonomy of 21

IFs with the components of Karsh’s sociotechnical model (Karsh, 2006) and the original HEART taxonomy of 38 EPCs.

Table 3. Mapping of EPCs with specific IFs for surgery applications (Onofrio and Trucco, 2018).

Taxonomy of IFs in surgery *(Onofrio and Trucco, 2018; Trucco et al., 2017)*			***Components of Karsh's sociotechnical model*** *(Karsh, 2006)*	***HEART EPCs*** *(Williams, 1986)*	
No.	**Description**	**Multiplier**	**Group –** *Sub-group*	**No.**	**Description**
1	Noise and ambient talk	10,0	**Team/group/unit department -** *Noise, temperature, lighting*	3	Low signal-to-noise ratio
2	Music	10,0	**Team/group/unit department -** *Noise, temperature, lighting*	3	Low signal-to-noise ratio
3	Noisy use of social media	9,8	**Team/group/unit department -** *Noise, temperature, lighting*	3 4	Low signal-to-noise ratio Means of suppressing or overriding information or features that are too easily accessible
4	Verbal interruptions	1,0	**Team/group/unit department -** *Noise, temperature, lighting*	36 37	Task pacing caused by intervention of others Additional team members beyond those necessary to perform a task
5	Poor management of errors and threats to patient safety	9,8	**Organization factors -** *Organisational policy/priorities*	2 7 12 18	Shortage of time for error detection and correction No obvious means of reversing an unintended action Mismatch between perceived and real risk Conflict between immediate and long-term objectives
6	Poor guidelines, procedures or checklists	2,0	**Organization factors -** *Organisational policy/priorities*	26	No obvious way to keep track of progress during an activity
7	Rude talk and disrespectful behaviours	5,0	**Team/group/unit department -** *Noise, temperature, lighting*	16 13	Low quality of information conveyed by procedures and person–person interaction Poor, ambiguous or ill-matched system feedback
8	Improper use of procedures and checklists	5,8	**Organization factors -** *Organisational policy/priorities*	16 32 11 9 21 14	Low quality of information conveyed by procedures and person–person interaction Inconsistent meaning of displays and procedures Ambiguity in the required performance standards Need to unlearn a technique and apply one that requires application of an opposing philosophy Incentive to use other, more dangerous procedures No clear, direct and timely confirmation of an intended action from the portion of the system over which control is to be exerted

9	Unclear or failed communication	6,8	**Team/group/unit department -** *Noise, temperature, lighting*	8 5	Channel capacity overload, particularly one caused by simultaneous presentation of non-redundant information No means of conveying spatial and functional information to operators in a form that they can readily assimilate
10	Poor or lacking coordination	4,5	**Individual -** *Skills, knowledge, training, education*	10 25	Need to transfer specific knowledge from task to task without loss Unclear allocation of function and responsibility
11	Poor decision-making	1,9	**Individual -** *Skills, knowledge, training, education*	25 17	Unclear allocation of function and responsibility Little or no independent checking or testing of output
12	Poor situational awareness	20,0	**Individual -** *Skills, knowledge, training, education*	1	Unfamiliarity with a situation that may be important
13	Lack of experience	8,0	**Individual -** *Skills, knowledge, training, education*	15	Inexperienced operator
14	Lack of experience of anaesthetics team	8,0	**Individual -** *Skills, knowledge, training, education*	15	Inexperienced operator
15	Fatigue	1,3	**Individual** - *Needs, biases, beliefs, mood*	35 22	Disruption of normal work/sleep cycles Little opportunity to exercise mind and body outside of work
16	Time pressure	11,0	**Team/group/unit department -** *Time and sequence demands*	2	Time shortage (based on the description in Williams, 1986)
17	Emotional perioperative stress	4,0	**Organization factors -** *Organisational policy/priorities*	29 22	High-level emotional stress. Little opportunity to exercise mind and body outside of work
18	Poor leadership	1,6	**Individual -** *Skills, knowledge, training, education*	24	Need for absolute judgements that are beyond the capabilities or experience of an operator
19	Team member familiarity	5,0	**Team/group/unit department -** *Task demands, complexity, difficulty*	16	Low quality of information conveyed by procedures and person–person interaction.
20	Poor use of technology	6,4	**Team/group/unit department** - *Technology functions/features*	6 20 19	Poor system/human user interface Mismatch between the education level of an individual and the requirements of the task Lack of diversity of information input for veracity checks
21	Inadequate ergonomics of equipment and workplace	6,1	**Team/group/unit department** - *Technology functions/features*	33 23	Poor or hostile environment. Unreliable instrumentation.

According to the findings of Trucco et al. (2017), the EPCs traditionally used in HEART method are not able to fully capture and explain the relevant organisational and personal IFs in a surgical context. In particular, organisational EPCs capture the 87.2 % of the scope of the organisational IFs; and the personal EPCs capture the 83.2 % of the scope of personal IFs (Trucco et al., 2017). As previously mentioned, the assessed effect of the i-th EPC is calculated based on the EPCs multipliers (cf. par 3) as the "maximum predicted nominal amount by which unreliability might change going from good conditions to bad" (Williams, 1986). These multipliers were originally proposed for general nuclear power plant applications by Williams (1986) and later modified in the NARA and CARA methods for domain-specific applications. Since the cross-matching between IFs and EPCs is many-to-many, a prior alignment and weighting must be performed to calibrate IFs multipliers starting from the prior EPCs multipliers. To this end, the multiplier values of HEART, NARA and CARA were taken into consideration. Full details on the method adopted for calibrating IFs multipliers can be found in Trucco et al. (2017), here below a brief description of the overall approach and few explanatory examples are reported.

Where possible, the multipliers used in the CARA method were preferred because the ATC context has more similarities with the surgical context than NPP. Generally, these similarities include workplace ergonomics; the centrality of the operator in execution of the procedure with respect to technology; the absence of actual technological barriers preventing accidents (in NPP, specific instruments monitor and filter human behaviours, correcting dangerous system states in a completely autonomous way, while this is not the case in surgery and ATC); and the possibility of directly, and most of the time personally, verifying the empirical state of the system through visual inspection.

First, the span of coverage and the degree of matching between Williams' (1986) taxonomy of EPCs and the surgical taxonomy of IFs were assessed by surgeons as reported in Trucco et al. (2017). For example, EPC13 (i.e. poor, ambiguous or ill-matched system feedback), which is involved in the evaluation of IF7 (i.e. rude talk and disrespectful behaviours) was assigned a relative weight of 30%, whereas EPC16 (i.e. low quality of information conveyed by procedures and person–person interaction) was assigned a relative weight of 70%. EPC13 has a multiplier of 4 according to HEART and values of 4 and 5 in the NARA and CARA methods, respectively. EPC16 is the only one linked to IF19 (i.e. team member familiarity), but surgeons believe that it has a relative weight of 88%. EPC29 (i.e. high-level emotional stress) covers 70% of IF17 (i.e. emotional perioperative stress). This is the EPC with the largest difference between multipliers, especially between the traditional HEART and the CARA methods, which is reasonable since the contexts of the methods are completely different, mainly because of the differential role of technology; while in NPP human behaviour, and occasionally errors, are mediated by instruments, this is not possible in ATC. Based on this evidence, it is apparent that emotional and personal aspects have much more impact on the probability of success of ATC tasks and that this context is most similar to surgery.

The multiplier suggested by CARA was taken for EPC29. For all other EPCs, there was no need to choose between potential multipliers since the values were the same according to all the methods or only one method pertained to the EPC. Finally, by weighting the multipliers of the EPCs showing partial or full coverage, a new set of multipliers for IFs was proposed, as reported in Table 3.

5. A Methodology for Dynamic Human Reliability Analysis in Surgery

The proposed methodology is organised into three main steps as shown in Figure 2.

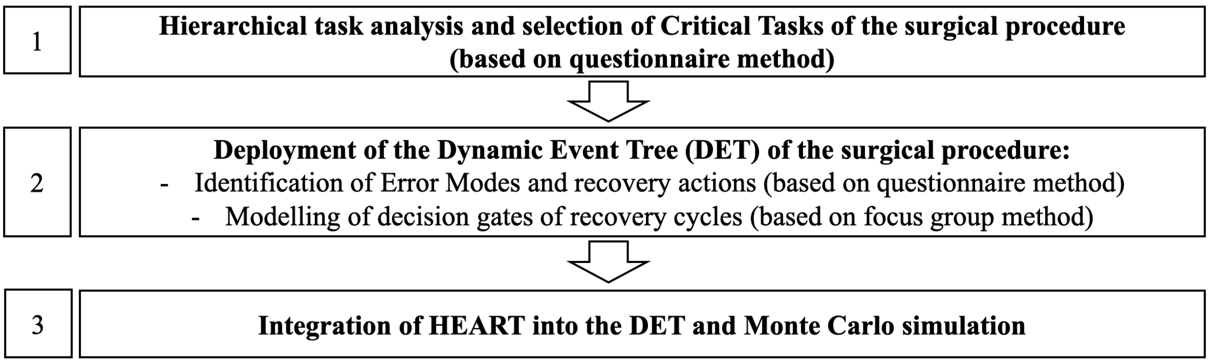

Figure 2: A three-steps methodology for Dynamic Human Reliability Analysis in surgery.

5.1 Hierarchical task analysis and selection of critical tasks of the surgical procedure

The starting point of this study is Hierarchical task analysis (HTA) and its validation by a team of surgeons or complete surgical team. HTA should be based on various data sources, such as previous studies, training manuals and the technical protocol for the operation (e.g. Galfano et al., 2010). To confirm the accuracy of the collected information, sessions of direct observation should be planned to achieve a deeper understanding of the entire procedure and the personal or team dynamics in the operating room.

Among the different types of task analysis documented in the literature (Lane at al. 2006; Tang et al., 2006; Huang and Gramopadhye, 2014), hierarchical task analysis, which considers the actions of the operator in a chronological sequence, is the most suitable for studying surgical procedures (Cuschieri, 2000). This analysis should include all the steps of the surgical procedure, all the equipment used, the surgeon's technical experience with the procedure and the plausible error modes (EMs) for all the identified critical tasks. Relying on medical evidence and the surgeons' judgment, the most critical tasks of the surgical procedure should be selected.

5.2 Deployment of the Dynamic Event Tree (DET) of the surgical procedure

Task analysis serves as the reference for designing the DET of the surgical procedure under investigation. Basically, the DET technique identifies branching points (also called nodes) in the evolution of the procedure at which stochastic events occur, records the state of the system at each branching point and successively simulates all branches, each of which represents a possible outcome of the stochastic event, including the error recovery actions. Simulating the evolution of a scenario in this way enables exploration of all possible occurrences of the process variables (Rao et al., 2015, 2009). In order to develop the DET model, the stochastic events in this study refer to "Error Modes" of each surgical task identified. The identification of possible EMs in the execution of the critical tasks and the corresponding recovery paths was based on surgeons' experience and judgements, which were obtained using a paper-based questionnaire.

5. 3 Integration of HEART into the DET and Monte Carlo simulation

In a recent study, Onofrio and Trucco (2018) proposed a novel validated taxonomy of IFs in surgery. Using a survey method, they used surgeons' judgements to quantify the perceived influence of each IF on the surgeon's technical performance. Despite a documented degree of uncertainty, the results show, on average, that some IFs are more likely to have a stronger effect on robotic surgery than open surgery. The revised HEART procedure was applied to enable one to determine the influence of uncertain EPCs, analyse their consequences on the evolution of the surgical procedure and, eventually, assess the cumulative effect on the patient's outcome. This procedure was developed by integrating a modified version of HEART and DET into a unique simulation tool. The modified HEART method was integrated with Monte Carlo method to estimate the surgeon's unreliability for a fixed sequence of critical tasks by applying the method to evaluate nodes of the DET. The following steps should be followed:

- Initialisation of the assessed PoA, which measures the magnitude of effect of each IF.
- Initialisation of the ANLU for critical tasks.
- Identification of possible EMs in each critical task and evaluation of the probability of each branches by adopting a linear additive model and modified HEART algorithm.
- Identification of the expected patient outcomes for all branches. In the present study, the Clavien-Dindo classification of surgical outcomes was adopted (Dindo et al., 2004).
- Calculation of the probability distribution of each patient outcome grade for the selected procedure in line with the central limit theorem.
- Scenario analysis to assess the importance of different IFs (i.e. their influence on the patient's expected outcome).

The simulation tool, which was implemented in Matlab®, consists of three main parts: i) initialisation of data, ii) quantitative evaluation of paths (iterative part) and iii) probability distribution calculation of the patient's outcome grade. The numerical inputs are i) the extremes of the NHU ranges, as assessed by expert surgeons; ii) the probability density functions (PDFs) of IFs, as defined by Onofrio and Trucco (2018); iii) the relative probabilities (α) of the identified EMs; iv) the IFs involved in each critical task according to the expert surgeons and v) the patient outcome grades associated with different EMs and the corresponding error recovery actions.

6. Pilot application in robotic surgery

Robotic surgery has been established as the best technique for surgical treatment of prostate cancer. As a result, RARP has recently increased in importance in the United States and Europe. It has been estimated that more than 75% of radical prostatectomies are performed using the DaVinci® system (Tanimoto et al., 2015). Many authors underline that robotic prostatectomy is a gold standard as the system enables a 3D view (as opposed to the 2D view provided by laparoscopy) and the instruments are highly precise, which minimises the possibility of surgical complications (Galfano et al., 2013).

The present study was done in collaboration with the Urology Department of a leading Italian hospital, ASST Grande Ospedale Metropolitano Niguarda, located in Milan. The surgeons involved in the study are members of a surgical team, one of whom is one of the authors of an RARP technique called Retzius-sparing RARP (Galfano et al., 2010). This technique enables a fully posterior approach without opening the Retzius and passing through the Douglas, not only for isolation of seminal vesicles (as in Montsouri's technique) but also for isolation of the whole prostate and anastomosis phase. The literature claimed that, although the technique initially appears complex, it allows one to obtain excellent results from both an oncological and functional point of view. Recently, Galfano et al. (2013) analysed the results of the first 200 patients treated with this procedure with a one-year minimum follow-up, concluding that the oncological results improved after a learning curve of 100 patients.

6.1. Hierarchical Task analysis of the prostatectomy procedure and selection of critical tasks

The hierarchical task analysis concerned the surgical procedure used at the Urology Department of the Italian Hospital "ASST Grande Ospedale Metropolitano Niguarda" in Milan. There are five major phases of the procedure: anaesthesia, room and robot preparation, patient preparation, surgery execution and closure. In the present study, we focused on only surgery execution. The time required to complete the surgery is, on average, an hour and a half (Galfano et al., 2013). Figure 3 shows the steps until direct validation of the hierarchical task analysis by surgeons. This approach aligns with previous experiences regarding healthcare task analysis documented in the literature (e.g. Joice et al., 1998; Tang et al., 2004a; Tang et al., 2004b; Lane at al. 2006; Tang et al., 2006; Huang and Gramopadhye, 2014).

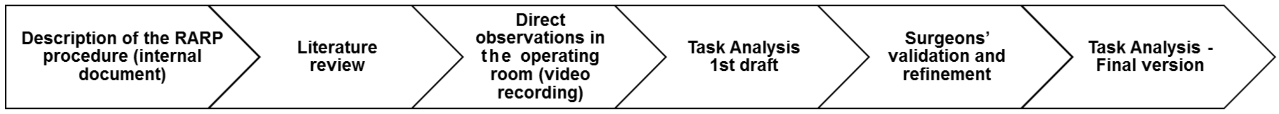

Figure 3. Steps to validate hierarchical task analysis

Investigation of surgical terms, practices and tools was needed for familiarisation with the context and collection of all relevant information about the specific surgical procedure. The description of the procedure in formal documents was very precise and informative in several parts, but not always. The distinction between tasks carried out at the operating table and those carried out at the robot console was not always clearly reported, and the beginning or end of elementary tasks was not always precisely identified. Consequently, a clear, complete and systematic description of the procedure and all relevant tasks was produced, with particular attention given to the sequence of tasks. Furthermore, research of the literature was performed to gather useful information on similar surgical procedures. Videotaped RARP procedures were watched on the institutional website for the DaVinci® system and other medical websites, and training manuals and technical protocols were consulted. During direct observation of the surgical procedure, we took notes and recorded two surgeries, placing cameras so we could obtain concurrent views of the operation at the table and at the console. Due to the availability of video recorded by an endoscopic camera integrated into the DaVinci® system, it was possible to produce a video from the perspective of both the endoscopic camera and the external cameras.

Once all the information was collected, a preliminary hierarchical task analysis was drafted for the surgery execution phase. In total, 11 elementary tasks were identified. Few corrections, mainly concerning terminology and proper synchronisation of parallel tasks, were made by the team of surgeons before validating the final version. Relying on medical evidence and the surgeons' judgment, the three most critical tasks were identified:

- Task 1 - Isolation of lateral peduncles and of posterior prostate surface.
- Task 2 - Santorini detachment from the anterior surface of the prostate.
- Task 3 - Anastomosis.

6.2 Deployment of the DET model of the surgical procedure

The first step was the identification of possible EMs in the execution of the three critical tasks and the corresponding recovery paths. It was based on surgeons' judgements; the EMs were obtained using a paper-based questionnaire. The relevant EMs and recovery paths for each critical task are (Teble 4):

Table 4. The relevant EMs and recovery paths (RPs) for each critical task

TASK1	TASK 2	TASK 3
EM 1.1: Incorrect identification of the surgical plan and subsequent prostate injury. EM 1.2: Poor control of bleeding. EM 1.3: Incorrect identification of the surgical plan and subsequent neurovascular bundle injury. EM 1.4: Rectum injury.	EM 2.1: Partial or complete opening of the Santorini with bleeding.	EM 3.1: Non-tight anastomosis. EM 3.2: Urethra injury. EM 3.3: Ligation of the catheter with the anastomosis points. EM 3.4: Suture of the posterior wall of the bladder with the anterior wall.
RP 1.1: Return to the beginning of the surgical step and identify the correct surgical plan. RP 1.2: Application of clips or stitches. RP 1.3: Return to the beginning of the surgical step and identify the correct surgical plan. RP 1.4: Identification of the lesion, suturing and repairing.	RP 2.1: Compression of the bleeding vessel with the vacuum cleaner by assistant surgeon (at patient table) and low pressure washing; increase in pneumoperitoneum; suture of bleeding vessels.	RP 3.1: Positioning of additional stitches and perform new anastomosis. RP 3.2: Positioning of additional stitches. RP 3.3: Sectioning the suture and perform new anastomosis. RP 3.4: Partial removal of anastomosis points and performing new anastomosis.

Figure 4 shows the schematic representation of sequences implemented into a DET taking into account these EMs and recovery paths for each critical task. Regarding the consequences associated with different failure modes, the team of surgeons was asked to grade the expected severity according to the Clavien-Dindo grading system of patient outcomes (Dindo et al., 2004), which is the most widely accredited grading system for surgery (Table 5).

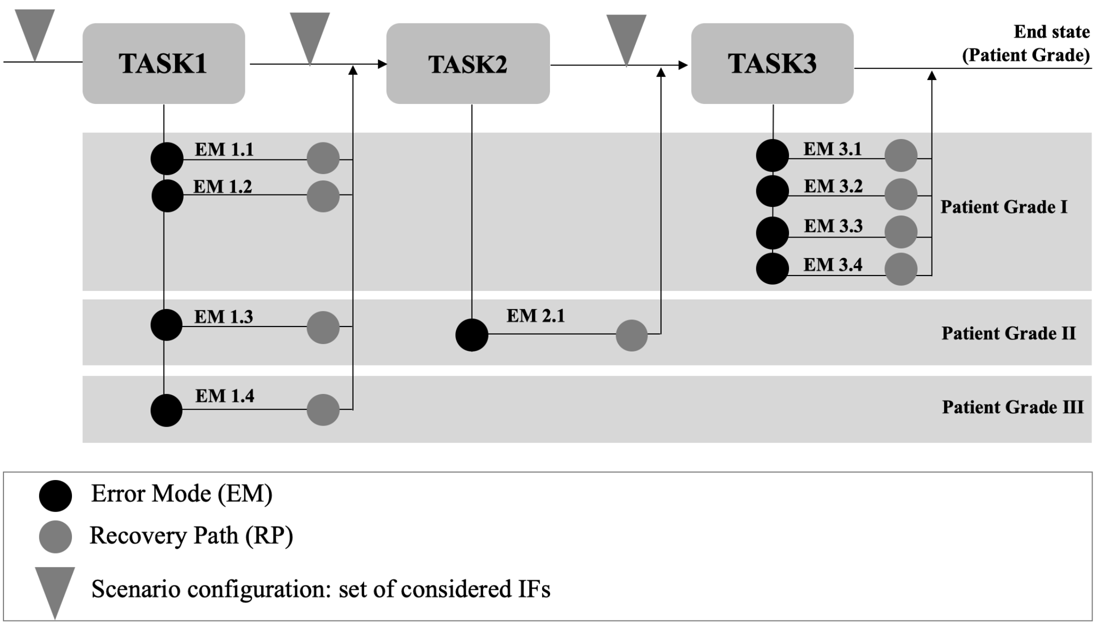

Figure 4. Schematic representation of the DET for the BA-RARP procedure.

Table 5. Clavien-Dindo grading system for patient outcomes (Source: Dindo et al., 2004).

Grade	Classification Criteria
Grade I	Any deviation from the normal postoperative course without the need for pharmacological treatment or surgical, endoscopic, and radiological interventions. Allowed therapeutic regimens are: drugs as antiemetics, antipyretics, analgesics, diuretics, electrolytes, and physiotherapy. This grade also includes wound infections opened at the bedside.
Grade II	Requiring pharmacological treatment with drugs other than such allowed for grade I complications. Blood transfusion and total parenteral nutrition are also included.
Grade III	Requiring surgical, endoscopic, or radiological intervention.
Grade IIIa	Intervention not under general anaesthesia.
Grade IIIb	Intervention during general anaesthesia.
Grade IV	Life-threatening complication (including CNS complications)* requiring IC/ICU management.
Grade IVa	Single organ dysfunction (including dialysis).
Grade IVb	Multiorgan dysfunction.
Grade V	Death of a patient.

*Brain haemorrhage, ischemic stroke, subarachnoid bleeding, but excluding transient ischemic attacks. CNS = central nervous system; IC = intermediate care; ICU = intensive care unit.

6. 3 Integration of HEART into the DET and Monte Carlo simulation

6.3.1 Selection of IFs

For each critical task, the team of surgeons was asked to select the IFs with the greatest impact on the surgeon's performance. The final set of IFs selected for the analysis are reported in Table 6.

Table 6. The final set of IFs selected for each critical task.

Critical task of the RARP procedure	*IFs selected for the modified HEART procedure*
1) Isolation of lateral peduncles and posterior prostate surface	- Noise and ambient talk (IF1) - Poor management errors (IF5) - Poor coordination (IF10)
2) Santorini detachment from the anterior surface of the prostate	- Noise and ambient talk (IF1) - Rude talk and disrespectful behaviour (IF7)
3) Anastomosis	- Noise and ambient talk (IF1) - Poor management errors (IF5) - Poor communication (IF9) - Poor coordination (IF10)

6.3.2 Simulation campaign

The probability distributions of patient grades associated with the execution of the three critical tasks was calculated using Monte Carlo simulation. Simulation runs were performed according to the following hypotheses and settings:

- Initialisation of the assessed PoA: since HEART adopts a linear and additive model, we assumed that there are no relationships between the randomly selected PoA values, and different random inputs were selected for each IF from the triangular distribution functions reported by Onofrio and Trucco (2018). We used fixed PoA values in a single run.

- Initialisation of the ANLU for all the critical tasks: rectangular distribution functions based on the 5–95th percentiles of HEART were set for random sampling.
- Calibration of the conditional probabilities of occurrence of the EMs: the probability of observing a specific EM, in case of surgeon's error in executing a critical task, was assumed as a single-point estimation directly elicited by surgeons (Table 7).
- Initialisation of patient outcome grades for each branch of the DET: based on surgeons' minimum and maximum estimates (Table 8), categorical distribution functions with k=2 and p_k=0.5 were used (due to unsufficient knowledge to calibrate different p values for the minimum and maximum grades). According to surgeons, the worst possible patient outcome due to complications in a RARP procedure is Grade 3 (i.e. 'requiring surgical, endoscopic or radiological intervention').

In order to achieve reliable results, the number of iterations in each simulation run (i.e. treated patients) was set at 20,000 after checking for good fit of the output distributions with a Gaussian distribution ($p<00.5$). Globally, we performed 12 simulations organised into the following scenarios:

- Scenario 1 - No IFs considered (1 simulation run).
- Scenario 2 - All IFs considered (1 simulation run).
- Scenarios 3 to 7 - Only one IF considered per simulation run (5 simulation runs).
- Scenarios 8 to 12 - Only one IF not considered per simulation run (5 simulation runs).

Table 7. Conditional probability of occurrence of EMs according to surgeons' judgements (X is the number of the critical task).

α factors	**EM X-1**		**EM X-2**		**EM X-3**		**EM X-4**	
	min	max	min	max	min	max	min	max
CT 1	0.39	0.425	0.1	0.6	0.01	0.05	0.425	0.58
CT 2	1	1	-	-	-	-	-	-
CT 3	0.28	0.5	0.1	0.5	0.2	0.6	0.3	0.57

Table 8. Grading of patient outcomes for different EMs according to surgeons (X is the critical task number associated with each EM).

Patient outcome grade*	**EM X-1**		**EM X-2**		**EM X-3**		**EM X-4**	
	min	max	min	max	min	max	min	max
CT 1	I	I	I	II	I	II	II	III
CT 2	I	II	-	-	-	-	-	--
CT 3	I	II	I	II	I	I	I	I

* Clavien-Dindo grading system (Table 5).

7. Discussion of results

The results are initially reported in the forms of frequency distributions of the patient outcome probability and quantiles of different patient outcomes (i.e. grades). Figure 5 shows the frequency distribution of grade probability when all the relevant IFs are active (simulation Scenario 2), whereas Figure 6 shows the probability of a Grade 0 outcome in the 95th percentile of patients, i.e. the probability that the 95% of patients experience a Grade 0 outcome ($p<0.05$). If 100% probability of success is expected in the absence of all relevant IFs, we observed that IF1 (noise and ambient talk) has the strongest impact, reducing the probability to 96.46%, which is close to the lowest probability of a successful surgical procedure when all the IFs are active (93.47%). Other critical IFs, in decreasing order of influence, are poor management of errors (IF5), poor coordination (IF10), poor communication (IF9) and rude talk and disrespectful behaviour (IF7).

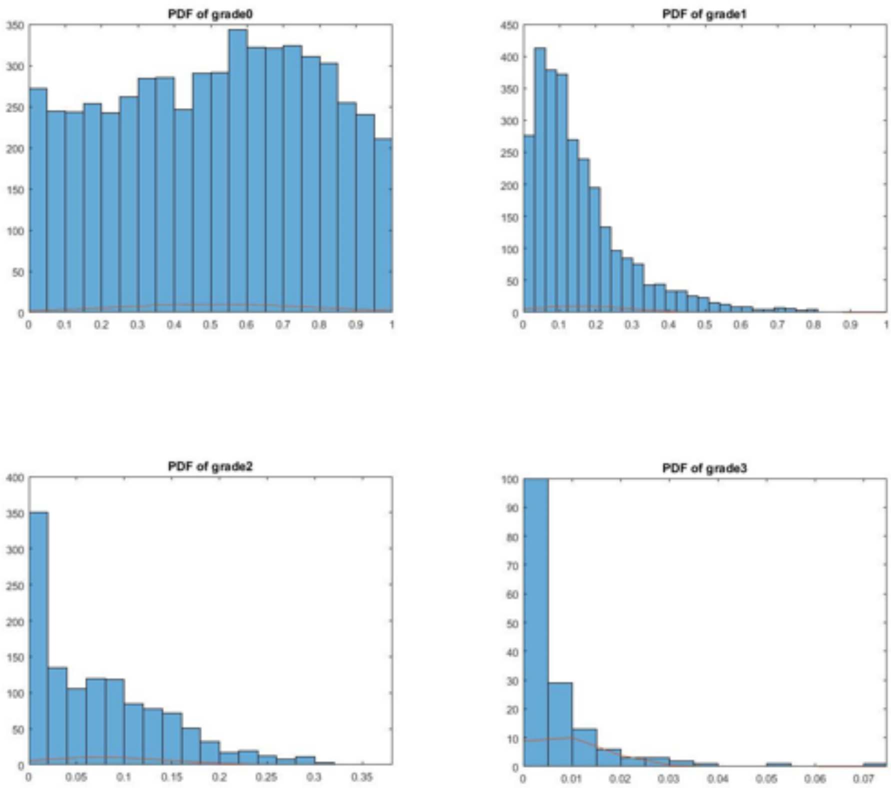

Figure 5. Frequency distribution of the patient outcome (grade) probability when all the selected IFs (5) are active (simulation scenario 2).

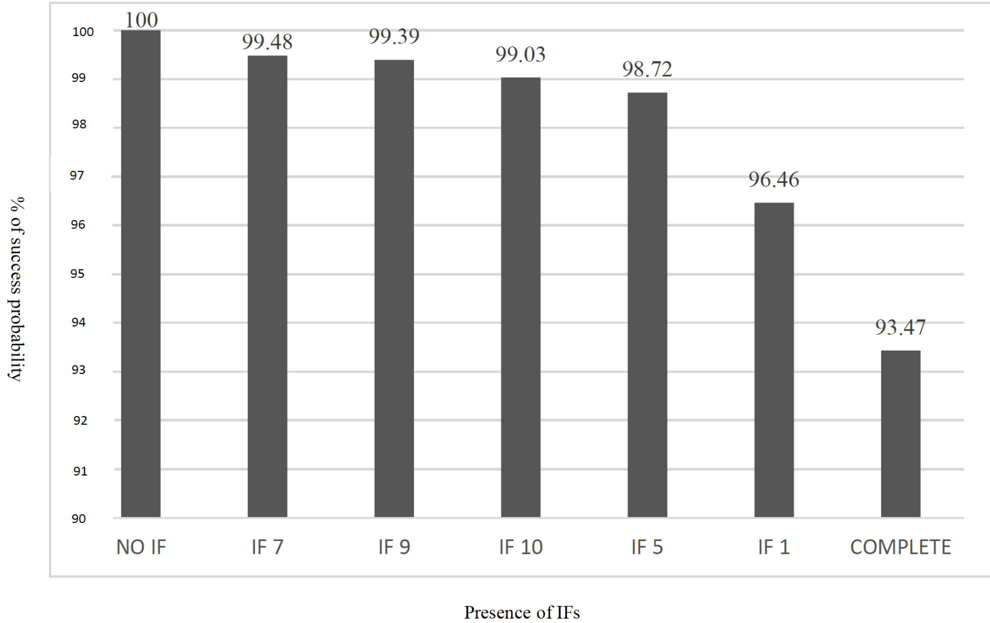

Figure 6. Probability of observing a Grade 0 patient outcome for the 95% of patients ($p<0.05$) under different combinations of IFs.

Figure 7 reports different probability values of a Grade 3 outcome for the 95th percentile of patients under different personal and team conditions. Since the surgeons estimated Grade 3 as a possible outcome only when an error occurred in the first critical task, the relevant IFs are the only ones involved in the Task 1, i.e. noise and ambient talk (IF1), poor management of errors (IF5) and poor coordination (IF10).

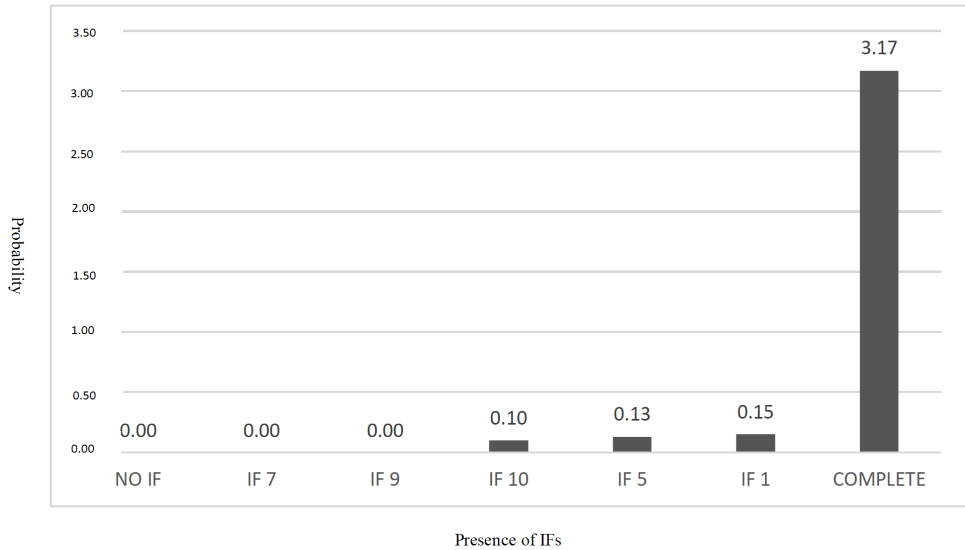

Figure 7. Probability of observing a Grade 3 patient outcome for the 95% of patients ($p<0.05$) under different combinations of IFs.

While the probability of a Grade 3 outcome is 0% under ideal conditions (i.e. no IFs), all the three relevant IFs show almost the same negligible impact, since the probability of a Grade 3 outcome increases of about 0.10–0.15% when they are taken separately. The probability of a Grade 3 outcome increases to 3.17% when the joint presence of the same subset of IFs is postulated. To achieve clearer and more robust data, we considered the probability of a Grade 3 outcome also for the 50th percentile of patients, which corresponds to the evaluation of the median of the sample (Figure 8). The results are consistent with those for the 95th percentile, but in this case, the impact of IF1 is almost double that of IF5 and IF10. Additionally, it is worth noting that the median of the PDF of a Grade 3 outcome is 0.27%, which is about one order of magnitude lower than that for the 95th percentile. Overall, these results clearly show a nonlinear cumulated effect on patient outcome degradation of influencing factors across different levels of the socio-technical system (Karsh & Brown, 2010).

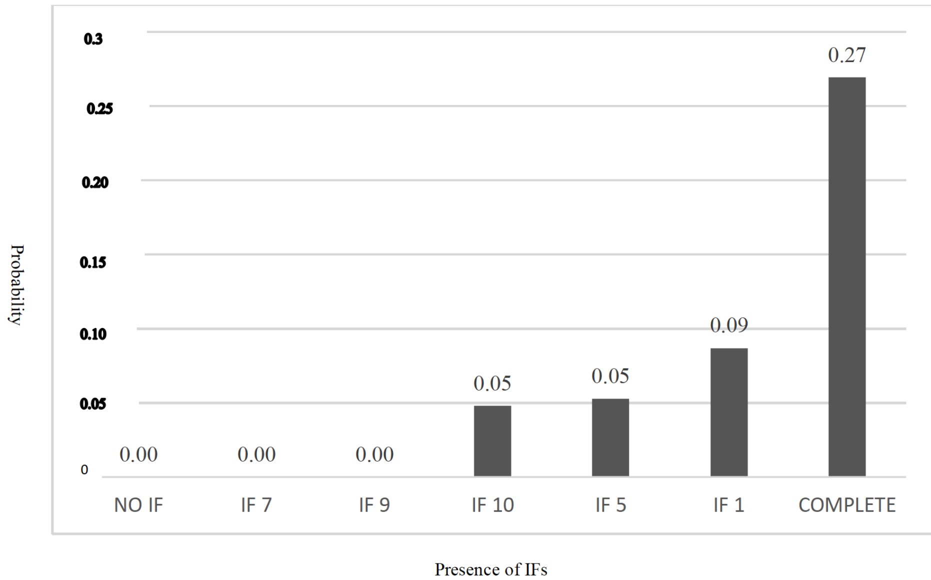

Figure 8. Probability of a Grade 3 patient outcome (50th percentile; $p<0.05$).

A second level of analysis was performed to investigate in greater detail the joint effect of different combinations of IFs. To this end, all the IFs selected for the study were clustered into three categories – personal, team-related, and organisational (Table 9) to generate three plausible scenarios. Results are reported in Table 10.

Table 9. IFs clustered into three categories (team-related, organisational and personal).

INFLUENCING FACTORS		**CATEGORY**
1	Noise and ambient talk	Team-related
5	Poor management of errors	Organisational
7	Rude talk and disrespectful behaviour	Team-related
9	Unclear communication	Team-related and personal
10	Poor coordination	Team-related and personal

It is apparent that the the worst scenario is the one where all the five IFs are active, resulting into the lowest probability (93.47%) of a Grade 0 oputcome for the 95% of the patient population and the highest probability of a Grade 3 outcome (3.17%). Notably, the cluster of IFs with the highest aggregate impact on surgical outcome is the team-related category (94.58 % of Grade 0 and 2.77 % of Grade 3 for the 95th percentile of patients), followed by the organisational category. The impact of personal factors alone (IF9 and IF10) on patient outcome seems marginal since the probability of a Grade 0 outcome is very close to the best scenario (99.55 %) and the probability of a Grade 3 outcome is negligible (0.09 %).

Table 10. Analysis of the influence of IFs by category (95th percentile probability; $p<0.05$).

PATIENT OUTCOME	COMPLETE	TEAM (IF1, 7, 9, 10)	ORGANISATIONAL (IF5)	PERSONAL (IF9, 10)
GRADE 0	93.47 %	94.58 %	98.72%	99.55 %
GRADE 3	3.17 %	2.77 %	0.13 %	0.09 %

The quantitative results of the present study shed a novel light on the role that team influencing factors play in complex technological contexts (Edmondson, 1996; Zohar et al., 2007; Zohar and Luria, 2005). The theoretical reference to Karsh's mesoergonomic framework (Karsh, 2006) allows to position team factors as the connecting elements between the individual and the organizational levels; according to present results, this positioning seems to attribute team factors an amplification role also, which drives the non linear accumulation of the effects attributable to personal and organisational factors.

Putting these results into a broader methodological perspective, present results contribute to strengthen the evidence in favour of mesoergonomic research (Karsh, 2014; Karsh and Brown, 2010; Karsh, 2006) which fosters the joint evaluation of variables explicitly measured at the organizational, unit and individual levels, so that cross-level relationships can be revealed and unfold.

8. Conclusions

This paper proposes a methodology for performing mesoergonomics analysis of surgical procedures, based on a modified version of HEART technique. The pilot application involved the analysis of the impact of personal, group and organizational factors on patient outcome of the Robot-Assisted Radical Prostatectomy procedure (Galfano et al., 2010). The study offers an original and relevant contribution to the development of domain-specific knowledge and, as such, fosters the application of HRA in healthcare as a reliable approach to system-level analysis of patient safety. More generally, the results support the relevance of mesoergonomics studies in healthcare where causal mechanisms between or among system levels are investigated (Karsh et al., 2014). The proposed methodology is suitable for further generalisation and application to different healthcare settings.

From a methodological stance, HEART was modified to incorporate the taxonomy of IFs that Onofrio and Trucco (2018) developed for the surgical context (both robotic and open surgery), taking into account uncertainties on their degree of influence on surgeons' technical performance. To the best of authors' knowledge, there are no published HRA studies addressing robotic surgery or healthcare-context-specific EPCs or IFs. The cross-matching method used to link method-specific EPCs with domain-specific IFs in this study can be generalised and adapted to other healthcare settings as well. In line with some previous applications (e.g. Chadwick and Fallon, 2012), the HEART technique was applied with a small team of experts, not a single assessor, to achieve a more insightful, comprehensive and reliable analysis. Additionally, DET modelling and the Montecarlo method were used to investigate the causal mechanisms within the surgical procedure (including error recovery loops), taking into account different sources of uncertainty in the form of stochastic parameters.

The study offers significant contributions to practice as it sheds light on the most relevant IFs that impair surgeons' performance and patient outcome. The results indicate that team-related factors have the highest impact on surgeons' performance, minimising the probability of a Grade 0 outcome and maximising the probability of a Grade 3 outcome (i.e. surgical complications). Even more significantly, the simulation approach, extended to the critical portion of the surgical procedure, highlighted that different IF categories can have a more-than-linear effect on surgeons' performance, even though the HEART algorithm adopts a linear model at the task level. Overall, the study successfully demonstrates how the HRA technique can be adopted and used by professionals for mesoergonomics studies in the healthcare context.

However, there are several limitations when it comes with implementing a mesoergonomics approach. At methodological level, the major issue comes from the limited available knowledge in the HRA domain on how to model and assess the interdependence between multiple influencing factors (De Ambroggi and Trucco, 2011). We argue that this aspect has a potentially relevant impact on the development of cross-level mesoergonomic analyses. Secondly, the study considered factors at organisational, group and individual levels only, without including the system level as it would be required for a complete mesoergonomic analysis (Karsh et al., 2014). Here it is worth mentioning that the interactions the system component has with the other elements of the mesoergonomic level cannot be adequately captured through the application of existing IFs or EPCs taxonomies used in HEART and in many other HRA quantitative techniques. Authors consider this issue as a relevant and urgent need of further research.

Acknowledgements

The authors wish to thank Dr. Antonio Galfano and his surgical group at the Urology Department of ASST Grande Ospedale Metropolitano Niguarda di Milano (Milan, Italy), to whom the authors are extremely grateful. The research presented in this paper originated from a development process that was carried out through a continuous interaction and collaboration between researchers and healthcare professionals. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

References

Al Naami, M., Anjum, M.N., Aldohayan, A., Al-Khayal, K., Alkharji, H. 2013. Robotic general surgery experience: a gradual progress from simple to more complex procedures. Int. J. Med. Robot. Comput. Assist. Surg. 9(4), 486–491.

Boring, R.L. 2007. Dynamic human reliability analysis: Benefits and challenges of simulating human performance. In T. Aven & J.E. Vinnem (Eds.), Risk, Reliability and Societal Safety, Volume 2: Thematic Topics. Proceedings of the European Safety and Reliability Conference (ESREL 2007) (pp. 1043-1049). London: Taylor & Francis.

Cagliano, A., Grimaldi, S., Rafele, C., 2011. A systemic methodology for risk management in healthcare sector. Saf. Sci. 49 (5), 695–708. Doi: 10.1016/j.ssci.2011. 01.006.

Castiglia, F., Giardina, M., Tomarchio, E., 2010. Risk analysis using fuzzy set theory of the accidental exposure of medical staff during brachytherapy procedures. J. Radiol. Prot. Off. J. Soc. Radiol. Prot. 30 (1), 49–62. Doi: 10.1088/0952-4746/30/1/004.

Chadwick, L., Fallon, E.F., 2012. Human reliability assessment of a critical nursing task in a radiotherapy treatment process. Appl. Ergon. 43 (1), 89–97. Doi: 10.1016/j.apergo.2011.03.011.

Cao, C., Rogers, G., 2006. Handbook of Human Factors and Ergonomics in Healthcare and Patient Safety, edited by Pascale Carayon. Lawrence Erlbaum.

Chang, Y.H.J., Mosleh, A., 2007. Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 2: IDAC performance influencing factors model. Reliab. Eng. Syst. Saf. 92 (8), 1014–1040.

Cuschieri, A., 2000. Human reliability assessment in surgery—a new approach for improving surgical performance and clinical outcome. Ann R Coll Surg Engl. 82(2): 83–87.

De Ambroggi, M., Trucco, P, 2011. Modelling and assessment of dependent performance shaping factors through Analytic Network Process. Reliability Engineering & System Safety, 96(7), 849-860.

Deeter, J., Rantanen, E., Human Reliability Analysis in Healthcare. 2012. Proceedings of Symposium on Human Factor and Ergonomics in Health Care, pp. 45-51.

Dindo D., Demartines N., Clavien P.A., 2004. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann. Surg. 240 (2), 205–213.

Edmondson, A., 1996. Learning from mistakes is easier said than done: group and organizational influences on the detection and correction of human error. Journal of Applied Behavioral Science 32, 5-28.

Galfano, A., Ascione, A., Grimaldi, S., Petralia, G., Strada, E., Bocciardi, A.M., 2010. A new anatomic approach for robot-assisted laparoscopic prostatectomy: a feasibility study for completely intrafascial surgery. Eur. Urol. Eur. Assoc. Urol. 58 (3), 457–461.

Galfano, A., Di Trapani, D., Sozzi, F., Strada, E., Petralia, G., Bramerio, M., Ascione, A., 2013. Beyond the learning curve of the Retzius-sparing approach for robot-assisted laparoscopic radical prostatectomy. Eur Urol. 64(6):974-80. Doi: 10.1016/j.eururo.2013.06.046.

Galfano, A., Ascione, A., Grimaldi, S., Petraglia, G., Strada, E., Bocciardi, A.M., 2010. A new anatomic approach for robot-assisted laparoscopic prostatectomy: a feasibility study for completely intrafascial surgery. Euro. Urol. 58 (3), 457–461.

Hamad, G.G., 2010. Minimally invasive surgery. Am. J. Surg. 199, 263–265.

Huang Y.-H and Gramopadhye A.K., 2014. Systematic Engineering Tools for Describing and Improving Medication Administration Processes at Rural Healthcare Facilities. Appl. Ergon, 45 (6), 1712-1724.

Inoue, K., Koizumi, A., 2004. Application of human reliability analysis to nursing errors in hospitals. Off. Publ. Soc. Risk Anal. 24 (6), 1459–1473. Doi: 10.1111/j.0272-4332.2004.00542.x.

Joice, P., Hanna, G.B., Cuschieri, A., 1998. Errors enacted during endoscopic surgery—a human reliability analysis. Appl. Ergon. 29 (6), 409–414.

Kim, J.W., Jung, W., Jang, S.C., Wang J.B. 2006. A case study for the selection of a railway human reliability analysis method. International Railway Safety Conference, October 22–27, 2006, Belfast.

Kohn, L.T., Corrigan, J.M., Donaldson, M.S., 2000. To Err is Human: Building a Safer Health System. Institute of Medicine (US), Committee on Quality of Health Care in America, National Academy Press, Washington, DC.

Karsh, B.T., Waterson, P., Holden, R.J., 2014. Crossing levels in systems ergonomics: A framework to support "mesoergonomic" inquiry. Appl. Ergon. 45 (1), 45–54. Doi: 10.1016/j.apergo.2013.04.021.

Karsh, B.-T., Brown, R., 2010. Macroergonomics and patient safety: the impact of levels on theory, measurement, analysis and intervention in patient safety research. Appl. Ergon. 41 (5), 674–681. Doi: 10.1016/j.apergo.2009.12.007.

Karsh, B.T., Holden, R.J., Alper, S.J., Or, C.K.L., 2006. A human factors engineering paradigm for patient safety: designing to support the performance of the healthcare professional. Qual. Saf. Health Care 15 (Suppl. 1), 59–65. Doi: 10.1136/qshc.2005.015974.

Karsh, B. 2006. Meso-Ergonomics: A New Paradigm for Macroergonomics Research. International Ergonomics Association 2006 Congress July 10-14, Maastricht.

Kim, J.W., Jung, W., 2003. A taxonomy of performance influencing factors for human reliability analysis of emergency tasks. J. Loss Prev. Process Ind. 16 (6), 479–495. Doi:10.1016/S0950-4230(03)00075 5.

Kirwan, B., Gibson, H., Kennedy, R., Edmunds, J., Cooksley, G., Umbers, I., 2005. Nuclear action reliability assessment (NARA): a data-based HRA tool. Saf. Reliab. 25 (2), 38–45.

Kirwan, B., Gibson, H. 2007. CARA: a human reliability assessment tool for air traffic safety management— technical basis and preliminary architecture, in: Redmill, F., Anderson, T. (Eds.), The Safety of Systems: Proceedings of the Fifteenth Safety-Critical Systems Symposium. Springer Verlag, London, pp. 197–214.

Kirwan, B., Gibson, H., Kennedy, R., Edmunds, J., Cooksley, G., Umbers, I., 2004. Nuclear action reliability assessment (NARA): a data-based HRA tool. Probabilistic safety assessment and management, pp 1206-1211. Springer, London. Doi: 10.1007/978-0-85729-410-4_195.

Kirwan, B., Kennedy, R., Taylor-Adams, S., & Lambert, B. 1997. The validation of three human reliability quantification techniques— THERP, HEART and JHEDI: Part II—Technique results and validation exercise. Applied Ergonomics, 28, 17–25. Doi: 10.1016/S0003-6870(96)00045-2.

Kirwan, B., 1994. A Guide to Practical Human Reliability Assessment. CRC Press, Taylor and Francis, Bristol.

Lai F., Entin E. 2005. Robotic surgery and the operating room team. Proc Hum Factors Ergon Soc Annu Meet 49:1070–1073. Doi: 10.1177/154193120504901115.
Lane, R., Stanton, N.A., Harrison, D., 2006. Applying hierarchical task analysis to medication administration errors. Appl. Ergon. 37, 669-679.
Lyons, M., 2009. Towards a framework to select techniques for error prediction: supporting novice users in the healthcare sector. Appl. Ergon. 40 (3), 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. Int. J. Risk Saf. Med. 16 (4), 223–237.
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. Ann. Surg. 257 (3), 476–482. Doi: 10.1097/SLA.0b013e318275b72a.
Onofrio, R., Trucco, P., 2018. Human reliability analysis (HRA) in surgery: identification and assessment of influencing factors. Safety Science 110: 110–123 Doi:10.1016/j.ssci.2018.08.004.
Pandya, D., Podofillini, L., Emert, F., Lomax, A. J., Dang, V. N., & Sansavini, G., 2020. Quantification of a human reliability analysis method for radiotherapy applications based on expert judgment aggregation. Reliability Engineering & System Safety, 194, 106489. Doi: 10.1016/j.ress.2019.05.001.
Pandya, D., Podofillini, L., Emert, F., Lomax, A.J., 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. Proceed. Inst. Mech. Eng. Part O: J. Risk Reliab. 232 (1), 3–37. Doi: 10.1177/1748006X17731903.
Randell, R., Honey, S., Alvarado, N., Pearman, A., Greenhalgh, J., Long, A., Gardner, P., Gill, A., Jayne, D., Dowding, D., 2016. Embedding robotic surgery into routine practice and impacts on communication and decision making: a review of the experience of surgical teams. Cogn. Technol. Work. 18, 423–437.
Rao, D., Kim, T., Dang, V.N., 2015. A dynamic event tree informed approach to probabilistic accident sequence modeling: Dynamics and variabilities in medium LOCA. Reliab. Eng. Syst. Saf. 142, 78–91. Doi: 10.1016/j.ress.2015.04.011.
Rao, K.D., Gopika V., Sanyasi Rao, V., V., S, Kushwaha, H., S., Verma, A., K., Srividya, A. 2009. Dynamic fault tree analysis using Monte Carlo simulation in probabilistic safety assessment. Reliability Engineering and System Safety 94 872–883.
Tang, B., Hanna, G.B., Bax, N.M., Cuschieri, A., 2004a. Analysis of technical surgical errors during initial experience of laparoscopic pyloromyotomy by a group of Dutch pediatric surgeons. Surg. Endosc. 18 (12), 1716–1720. Doi: 10.1007/s00464-004-8100-1.
Tang, B., Hanna, G.B., Carter, F., Adamson, G.D., Martindale, J.P., Cuschieri, A., 2006. Competence assessment of laparoscopic operative and cognitive skills: Objective structured clinical examination (OSCE) or observational clinical human reliability assessment (OCHRA). World J. Surg. 30 (4), 527–534. Doi: 10.1007/s00268-005-0157-z.
Tang, B., Hanna, G.B., Joice, P., Cuschieri, A., 2004b. Identification and categorization of technical errors by observational clinical human reliability assessment (OCHRA) during laparoscopic cholecystectomy. Arch. Surg. 139 (11), 1215–1220.
Tanimoto, R., Fashola, Y., Scotland K.B., Calvaresi A.E., Gomella L.G., Trabulsi E. J., Lallas C., D. 2015. Risk factors for biochemical recurrence after robotic assisted radical prostatectomy: a single surgeon experience. BMC Urol. 8, 15–27.
Trucco, P., Onofrio, R., Galfano, A., 2017. Human reliability analysis (HRA) for surgery: a modified HEART application to robotic surgery. Adv. Intell. Syst. Comput. 482, 27-37. Doi: 10.1007/978-3-319-41652-6_3.
Verbano, C., Turra, F., 2010. A human factors and reliability approach to clinical risk management: evidence from Italian cases. Saf. Sci. 48 (5), 625–639. Doi: 10.1016/j.ssci.2010.01.014.

Ward, J., Teng, Y.C., Horberry, T., Clarkson, P.J., 2013. Healthcare human reliability analysis – by HEART, in: Anderson, M. (Ed.), Contemporary Ergonomics and Human Factors. Taylor and Francis, Cambridge, UK, pp. 287–288. 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. Symposium on the Achievement of Reliability in Operating Plant, Safety and Reliability Society (SaRS), NEC, Birmingham.
Zohar, D., Livne, Y., Tenne-Gazit, O., Admi, H., Donchin, Y., 2007. Healthcare climate: a framework for measuring and improving patient safety. Critical Care Medicine 35 (5), 1312e1317.
Zohar, D., Luria, G., 2005. A multilevel model of safety climate: cross-level relationships between organization and group-level climates. Journal of Applied Psychology 90(4):616-28.